Fréchet means in Wasserstein space: theory and algorithms

THÈSE N^{O} 7601 (2017)

PRÉSENTÉE LE 28 AVRIL 2017

À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE DE STATISTIQUE MATHÉMATIQUE
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Yoav ZEMEL

acceptée sur proposition du jury:

Prof. T. Mountford, président du jury
Prof. V. Panaretos, directeur de thèse
Prof. W. Kendall, rapporteur
Prof. A. Munk, rapporteur
Prof. S. Morgenthaler, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2017

“If you can’t prove your theorem, keep shifting parts of the conclusion to the assumptions, until you can."

— Ennio de Giorgi

Eszterkémnek

Acknowledgements

I would first of all like to express my sincere gratitude to my advisor, Professor Victor Panaretos. His endless enthusiasm for research and quest for understanding the unfamiliar have been contagious. Given my tendency to look up for counter-examples and perhaps excessively focus on technicalities, Victor's inherent optimism and ability to see the big picture were hugely influential for me. He was always available when I was in need for his help, advice or feedback. No less helpful than his vast knowledge and intuition were his insights on the academic world and what one needs in order to be a successful researcher.

I would like to thank Professors Wilfrid Kendall, Stephan Morgenthaler, Thomas Mountford and Axel Munk for kindly agreeing to be part of my thesis committee, and for their constructive feedback and comments on this manuscript.

Anirvan Chakraborty, Pavol Guriçan, Eszter Major and Tomáš Rubín have generously read parts of the thesis, provided useful comments and found numerous typos and mistakes. I do hope that the remaining ones not hinder the readability of the text.

This research was supported by a European Research Council (ERC) Starting Grant Award to Victor Panaretos and I remain indebted to the ERC for this.

Part of this thesis grew out of work presented at the Mathematical Biosciences Institute (Ohio State University), during the *Statistics of Time Warping and Phase Variation* Workshop in November 2012. I wish to acknowledge the stimulating environment offered by the Institute.

I frequented the halls and corridors of the EPFL for a few years now, and have consequently enjoyed many formal and informal activities with friends and colleagues. I had numerous insightful conversations with Professors Bernard Dacorogna, Anthony Davison, Clément Hongler, Stephan Morgenthaler and Thomas Mountford. Discussions with Anirvan led to the construction in Section 3.4. Being the teaching assistant of Thomas Mountford for measure theory was not only a pleasure, but also provided me intuition that was crucial in countless occasions during this work.

I thank current and former members of SMAT, particularly Andrea, Anirvan, Guillaume, Kate, Marie, Matthieu, Mikael, Pavol, Shahin, Tomáš and Valentina for rendering working days at EPFL pleasant; the same thanks are due to colleagues from other groups: Alix, Hélène, Léo,

Acknowledgements

Linda, Peiman, Raphaël, Thomas, Yousra, Alfonso, Daria, Jacques and, last but not least, Rémy.

I was fortunate to share an office with Mikael, Marie(-Hélène) and, later, Pavol and Tomáš. Mikael was as keen as me to optimise our Swiss experience by means of skiing, hiking and wine-tasting, and has also been particularly helpful with numerous statistical and technical issues. Ever since she joined our office and the group, Marie has been a close friend and colleague, and I already miss our Monday-morning discussions on mathematics, delicacies of the French language, and other things. I very much enjoyed later conversations with Pavol and Tomáš, and wish I was less stressed and had more time to share activities with them.

Administrative help from Anna, Maroussia, Jocelyne and Nadia was also invaluable.

My family and friends, and in particular my parents, have been the source of constant support despite the geographical distance. I am not even sure I would have begun the endeavour of carrying out a Ph.D. without their encouragement. That, with hindsight, would have truly been regrettable.

My family-in-law has been the most welcoming since my arrival to Switzerland, substantially lessening the inevitable loneliness one experiences when moving to a foreign country.

During these years, Eszter's support was comforting and truly unconditional, culminating to levels I would have never asked for in the final stages of writing. My vocabulary and eloquence in English are not nearly sufficient to properly acknowledge her help and support, and I resort to another language: *köszönöm*. Finally, I am grateful to Aaron for unbeknowingly helping me to put things in perspective.

Lausanne, March 10, 2017 Yoav Zemel

Abstract

This work studies the problem of statistical inference for Fréchet means in the Wasserstein space of measures on Euclidean spaces, $\mathcal{W}_2(\mathbb{R}^d)$. This question arises naturally from the problem of separating amplitude and phase variation in point processes, analogous to a well-known problem in functional data analysis. We formulate the point process version of the problem, show that it is canonically equivalent to that of estimating Fréchet means in $\mathcal{W}_2(\mathbb{R}^d)$, and carry out estimation by means of M-estimation. This approach allows to achieve consistency in a genuinely nonparametric framework, even in a sparse sampling regime. For Cox processes on the real line, consistency is supplemented by convergence rates and, in the dense sampling regime, $\sqrt{n}$-consistency and a central limit theorem.

Computation of the Fréchet mean is challenging when the processes are multivariate, in which case our Fréchet mean estimator is only defined implicitly as the minimiser of an optimisation problem. To overcome this difficulty, we propose a steepest descent algorithm that approximates the minimiser, and show that it converges to a local minimum. Our techniques are specific to the Wasserstein space, because Hessian-type arguments that are commonly used for similar convergence proofs do not apply to that space. In addition, we discuss similarities with generalised Procrustes analysis. The key advantage of the algorithm is that it requires only the solution of pairwise transportation problems.

The results in the preceding paragraphs require properties of Fréchet means in $\mathcal{W}_2(\mathbb{R}^d)$ whose theory is developed, supplemented by some new results. We present the tangent bundle and exploit its relation to optimal maps in order to derive differentiability properties of the associated Fréchet functional, obtaining a characterisation of Karcher means. Additionally, we establish a new optimality criterion for local minima and prove a new stability result for the optimal maps that, enhanced with the established consistency of the Fréchet mean estimator, yields consistency of the optimal transportation maps.

Keywords: Fréchet mean, functional data analysis, geodesic variation, optimal transportation, phase variation, point process, random measure, registration, warping, Wasserstein distance.

Résumé

Dans cette thèse, nous étudions le problème d'inférence statistique des moyennes de Fréchet dans l'espace de Wasserstein de mesures sur les espaces euclidiens, $\mathcal{W}_2(\mathbb{R}^d)$. Cette question se pose naturellement lors de la séparation des variations d'amplitude et de phase dans les processus ponctuels, de manière analogue au problème bien connu en analyse des données fonctionnelles. Nous formulons ce problème pour les processus ponctuels, démontrons qu'il est canoniquement équivalent à l'estimation de moyennes de Fréchet dans $\mathcal{W}_2(\mathbb{R}^d)$ et effectuons cette estimation au moyen de l'estimation M. Cette approche permet d'obtenir de la consistance dans un cadre nonparamétrique, même sous un régime d'échantillonnage épars. Pour les processus Cox sur $\mathbb{R}$, la consistance est complémentée par les taux de convergences et, dans le régime d'échantillonnage dense, par la consistance-$\sqrt{n}$ et un théorème limite centrale.

Le calcul de la moyenne de Fréchet est difficile lorsque les processus sont multivariés. Dans ce cas, notre estimateur de la moyenne de Fréchet n'est défini que de manière implicite, en tant que solution d'un problème d'optimisation. Pour surmonter cette difficulté, nous proposons un algorithme de la plus forte pente qui approxime cette solution et démontrons qu'il converge à un minimum local. Nos techniques sont spécifiques à l'espace de Wasserstein, parce que des arguments de type hessienne, qui sont généralement utilisés pour des preuves similaires de convergence, ne s'appliquent pas à cet espace. De plus, nous examinons les similitudes avec l'analyse procrustéenne généralisée. L'avantage principal de l'algorithme est qu'il ne requiert que la solution des problèmes de transport entre des paires de mesures.

Les résultats des paragraphes précédents requièrent des propriétés des moyennes de Fréchet dans $\mathcal{W}_2(\mathbb{R}^d)$, dont la théorie qui est développée est complétée par de nouveaux résultats. Nous présentons l'espace tangent et exploitons sa relation avec les fonctions optimales, pour dériver des propriétés de différentiabilité de la fonctionnelle de Fréchet, obtenant une caractérisation de moyennes de Karcher. De plus, nous établissons un nouveau critère d'optimalité pour les minimums locaux et prouvons un nouveau résultat de stabilité pour les fonctions optimales qui, annexé à la consistance déjà établie de l'estimateur de la moyenne de Fréchet, apporte de la consistance aux fonctions de transport optimal.

Mots clefs: Analyse de données fonctionnelles, déformation, distance de Wasserstein, mesures aléatoires, moyenne de Fréchet, processus ponctuel, recalage, transport optimal, variation géodésique, variation de phase.

Contents

List of Figures

Notation

Measure theory

$\mathrm{Leb}(A)$	Lebesgue measure of a (measurable) subset $A \subseteq \mathbb{R}^d$
$P(\mathscr{X})$	the set of Borel probability measures on a space $\mathscr{X}$
$M_+(\mathscr{X})$	the set of Borel measures on a space $\mathscr{X}$
$M(\mathscr{X})$	the set of Borel signed measures on a space $\mathscr{X}$
δ_x or $\delta\{x\}$	Dirac measure at a point x; $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise
$\mu \otimes \nu$	the independence coupling measure defined by $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$
$T\#\mu$	push-forward measure defined as $[T\#\mu](A) = \mu(T^{-1}(A))$
$\mathrm{supp}\mu$	support of a measure μ, defined as the complement of the largest open set on which $\mu = 0$
$\mathscr{L}_p(\mu)$	the set of measurable functions $f : \mathscr{X} \to \mathscr{X}$ such that $x \mapsto \|f(x)\|_{\mathscr{X}} \in L_p(\mu)$ ($\mathscr{X}$ is a Banach space. If $\mathscr{X}$ is separable, then this is a Bochner space)

Set theory and topology

$A \setminus B$	the set difference $\{x \in A : x \notin B\}$
$\mathrm{int}A$	interior of a set A (largest open set included in A)
$\overline{A}$	closure of a set A (smallest closed set that contains A)
∂A	boundary of a set A, defined as the difference $\overline{A} \setminus \mathrm{int}A$
$\mathrm{conv}A$	convex-hull of a set A (smallest convex set containing A)
$\langle \cdot, \cdot \rangle$ ($\|\cdot\|$)	inner product (norm) of $\mathscr{X}$ when it is a Hilbert (Banach) space
$d(x, A)$	when $(\mathscr{X}, d)$ is a metric space and $\{x\} \cup A \subseteq \mathscr{X}$, this is $\inf_{a \in A} d(x, a)$
d_K	diameter of a nonempty subset K of a metric space, defined as $\sup_{x,y \in K} d(x, y)$
$C_b(\mathscr{X})$ or $C_b(\mathscr{X}, \mathbb{R})$	the set of continuous, bounded real-valued functions on $\mathscr{X}$
$C_b(U, K)$	the set of continuous functions $f : U \to K$ such that $\sup_{x \in U} \|f(x)\| < \infty$
$B_R(x_0)$, $\overline{B}_R(x_0)$	the sets $\{x \in \mathscr{X} : d(x, x_0) < R\}$ and $\{x : d(x, x_0) \leq R\}$ respectively. Here $\mathscr{X}$ is a metric space with metric d

Optimal transportation

$W_p(\mu,\nu)$	Wasserstein distance of order p between the measures μ and ν
$\mathscr{W}_p(\mathscr{X})$	Wasserstein space of order p on a space $\mathscr{X}$
$\mathscr{W}_p(K)$	for $K \subseteq \mathscr{X}$, this is the set $\{\mu \in \mathscr{W}_p(\mathscr{X}) : \mu(K) = 1\}$
$\mathbf{t}_\mu^\nu$	optimal transport map between μ and ν (when it exists and is unique)
$\log_\gamma$, $\exp_\gamma$, Tan_γ	log map, exponential map and tangent space at an absolutely continuous $\gamma \in \mathscr{W}_2$
F_μ and F_μ^{-1}	cumulative distribution function and quantile function of a probability measure $\mu \in P(\mathbb{R})$

Miscellaneous

$\mathbb{R}_+^d$	d-dimensional vectors with nonnegative coordinates
$\mathbf{i}$	the identity map
A^t	transpose of a matrix A
$\det A$	determinant of a matrix A
$\mathrm{tr} A$	trace of a matrix A
G^{den}	the set of Lebesgue points of a set $G \subseteq \mathbb{R}^d$
S_N	set of permutations: bijective functions from $\{1,\dots,N\}$ to itself
$\mathrm{dom} f$	the set of points at which $f : \mathscr{X} \to \mathbb{R} \cup \{\pm\infty\}$ is not $+\infty$
ϕ^*	Legendre transform of ϕ defined by $\phi^*(y) = \sup_{x\in\mathscr{X}} \langle x, y\rangle - \phi(x)$
$\partial\phi(x)$	subdifferential of ϕ at x
$X_n = O_{\mathbb{P}}(Y_n)$	the sequence (X_n / Y_n) is bounded in probability
$X_n = o_{\mathbb{P}}(Y_n)$	the sequence (X_n / Y_n) converges to 0 in probability
	(When X_n and Y_n are not random, we omit the subscript $\mathbb{P}$)

1 Introduction

In the early days of statistics, the data to be analysed typically came in the form of vectors in finite-dimensional Euclidean spaces. Though its roots date back to the years post World War II, the field of **functional data analysis** received considerable attention since the last quarter of the previous century. In this setting, instead of a finite sequence of numbers, the atoms are entire curves, lying in a function space of infinite dimensions. This formalism allows for modelling phenomena arising in an extremely rich variety of applications, such as growth curves, electricity consumption, weather, brain images, handwriting recognition, criminology and DNA dynamics. In some applications, however, the linear structure of function spaces is inappropriate, as the space in which the data lie has no obvious notion of addition. The ambient space of some medical data, for instance, is the quotient space of $\mathbb{R}^3$ over Euclidean similarities, which is a particular type of manifold called the shape space. The evolutionary history of a set of organisms is modelled by phylogenetic trees, elements of a stratified space. More recently, research on social networks led to statistical analysis on (possibly weighted) graphs, with the nodes being social units and edges representing affinity between them.

The type of datasets that motivated the work in this thesis arise in neuroscience, and arrive in the form of **random point patterns**, sometimes called **spike trains** and mathematically defined as **point processes**. In the simplest scenario, one observes for each individual a random set of points in the unit interval $K = [0,1]$ and the goal may be to define a sample mean, representing the “average" behaviour of the sample. Though the total number of observed points may be different for each observation, this is typically not the main source of variation of the sample. Rather, it is the way these points are distributed on the interval that differs across individuals. In view of that, it is convenient to normalise by the number of points and treat the point processes as discrete random probability measures on K.

Despite not being a linear space, the space of probability measures on K (denoted $P(K)$) is convex, and the linear sample average in $P(K)$ can be taken as a sample mean. There are at least two reasons why the linear mean is unsatisfactory as a representative of the sample. A first drawback is that the number of points it contains is much larger than each of the observations. A more fundamental problem, however, is that it does not properly take into

account the intrinsic time scales of each observations. Suppose as an example that each individual exhibits concentration of points in a small vicinity of a time spot, t_i, that differs between the observations. Then the linear average will contain multiple concentrations of points, one around each t_i. In contrast, a point pattern with many points around the average time $\bar{t}$ provides a better description of the dataset, having a shape that is similar to each individual.

Borrowing intuition and terminology from the functional case, we formulate this problem in terms of **amplitude and phase variation** in multivariate point processes. In the context of the above example, phase variation is the variation in the time spots t_i across individuals, whereas amplitude variation pertains to the fluctuations around a mean level that exist even without presence of phase variation. Models for phase variation involve random **deformations** of the observation window K, assumed in most applications to have mean identity and to be "increasing". We argue that the *canonical* way to view this problem relies upon a different geometric structure on the space $P(K)$, emanating from the **Wasserstein distance** between probability measures. The resulting space, conventionally referred to as the **Wasserstein space** and denoted $\mathcal{W}_2(K)$, is a metric space with a nonlinear geometry.

More specifically, a mean in a nonlinear space may be defined, in analogy to a well-known property of the arithmetic mean in Euclidean spaces, by the concept of **Fréchet mean**, the minimiser of a sum-of-squares functional on the space. We show that the classical assumptions on the deformations, being "increasing" and having mean identity (without which the model for phase variation is usually not even identifiable), lead one *inevitably* to the problem of estimating Fréchet means in $\mathcal{W}_2(K)$. This equivalence extends beyond the real line, and holds whenever K is a compact convex subset of a Euclidean space of arbitrary dimension.

A very fortunate property of the Wasserstein space, one that is the exception rather than the rule in most metric spaces, is that under weak regularity conditions Fréchet means exist and are unique. In practical applications, however, it is desirable to have a method of constructing them as a function of the data. With the notable exception of the real line and one-dimensional-type examples, explicit formulae for the Fréchet mean are not available and one needs to resort to numerical schemes. We propose an algorithm that reduces the problem of finding the Fréchet mean to pairwise problems involving only two measures at a time, for which efficient numerical methods exist. This algorithm can be elegantly interpreted as steepest descent in the Wasserstein space, and has connections to an algorithm used in the analysis of shapes, **generalised Procrustes analysis**.

The structure of the thesis follows.

The underlying geometry behind the Wasserstein space stems from the so-called **optimal transportation problem**, or **Monge–Kantorovich problem**, an optimisation problem with a long history and an immensely rich literature. Chapter 2 gives a short survey of the aspects of the problem that are relevant for the thesis. After introducing the problem, defining the terminology and notation, and discussing some basic results, we give in Section 2.2 a probabilistic

formulation for the optimal transportation problem. This formulation is sometimes more convenient, and is perhaps more natural to readers with a more probabilistic/statistical (rather than analytic) taste. Examples of cases where the description of the solutions is particularly simple are given in Sections 2.3 (the discrete case) and 2.5 (quadratic cost function); settings in which solutions are explicit include that of the real line (Section 2.6) and that of Gaussian distributions (Section 2.7). Like any convex optimisation problem, the optimal transportation one admits a dual problem, introduced in Section 2.4. The important yet technical issue of smoothness of solutions is briefly touched upon in Section 2.8.

One topic that will be covered in some detail is the stability of the solutions to perturbations (Section 2.9). Roughly speaking, we show that the optimal solution of the problem, given by a deformation of the space, is a continuous function of the parameters. This result will be important in deriving consistency results for the estimation of the deformations that bring about the phase variation of the data, and will also serve as a technical tool for the convergence proof of the steepest descent algorithm.

Chapter 3 is devoted to the Wasserstein space, or more precisely, the Wasserstein *spaces* $\mathcal{W}_p$, where $p \geq 1$ is an exponent. A fair amount of attention will be given to the relation between the topology of $\mathcal{W}_p$ and that of convergence of distribution, called **narrow topology** in this thesis (and weak topology in many other texts). In particular, this relation will be exploited in order to show existence results and to characterise compact sets in $\mathcal{W}_p$.

The Riemannian-type structure of the Wasserstein space is presented in Section 3.3, including a brief discussion on curvature. The tangent bundle will indeed turn out to be a crucial ingredient in the derivation of the gradient of the Fréchet functional that will be used in Chapter 5.

Preceded by the technical Section 3.4 that treats measurability issues, the longest and most important section in Chapter 3 is Section 3.5, where Fréchet means are introduced and discussed in some detail in the context of the Wasserstein space $\mathcal{W}_2$. We present existence and uniqueness results, as well as some characterisations and properties of Fréchet means. Assuming differentiability, a minimisation problem can be transformed to the problem of finding zeroes of a derivative. This leads to the notion of **Karcher means**, defined as local minima of the sum-of-squares functional, that are the centre of attention of Subsection 3.5.5. The last part of the section is concerned with the equivalence between the problem of finding Fréchet means and a multimarginal version of the optimal transportation problem involving more than two measures.

The main contributions of this thesis are in Chapters 4 and 5 and can be summarised as follows:

1. In Chapter 4, we formalise the problem of separation of amplitude and phase variation in multivariate point processes, and demonstrate that the canonical solution is intrinsically related to Fréchet means in the Wasserstein space $\mathcal{W}_2$. We show how the relevant objects

can be estimated consistently in a fully nonparametric fashion, supplemented with convergence rates and a central limit theorem in the case of the real line.

2. In Chapter 5, we propose an iterative algorithm for the computation of empirical multivariate Fréchet means. The motivation for the algorithm lies in the differentiability properties of the Wasserstein distance emanating from the tangent bundle, and it is elegantly interpretable as a steepest descent algorithm. We additionally provide a convergence analysis of the algorithm, and sketch an extension of it to the population level.

The presentation of the first contribution is based on the journal article Panaretos & Zemel [70], and that of the second is based on the preprint Zemel & Panaretos [94].

We next enumerate additional contributions of the thesis. In Chapter 2, the only new result is the stability of the optimal maps (Proposition 2.9.11). Other contributions in Chapter 3 include relating the concept of compatible measures to flatness of the Wasserstein space (3.3.2) and to common copulae, and making explicit the equivalence between the multimarginal problem and the Fréchet mean (Subsection 3.5.6). The differentiability properties of the Wasserstein distance were already known, but their application to Karcher and Fréchet means is made for the first time, and the extension to the population level is new. The optimality criterion for Karcher means (Theorem 3.5.18) is also new. Finally, the results in the measurability section 3.4 are most likely known, but the simplified construction that does not use abstract measurable selection theorems is probably new.

The text is meant to be readable from cover to cover, in case the reader is ambitious enough to do so. Concepts and results that require some digression from the main flow of the text (such as convex analysis or Bochner integrals) are defined en route when needed. Roughly speaking, the statistical core of the thesis is in Chapters 4 and 5, and each can be read more or less independently of the other. Both, but more so Chapter 5, require understanding of some parts of Chapter 3; nevertheless, someone with even superficial knowledge of Wasserstein spaces and of Fréchet means should not encounter major difficulties when reading the statements in Chapter 4. As for Chapter 2, it mainly serves as background for Chapter 3 and a hopefully gentle introduction to optimal transportation. A notable exception is the backbone stability result in Subsection 2.9.2 that will be used for showing convergence of optimal maps in Chapters 4 and 5, but is not required for grasping the main ideas.

2 Optimal transportation

In this chapter we introduce the problem of optimal transportation. General references on this field are the book by Rachev & Rüschendorf [73], the two books by Villani [88, 89], and the recent book by Santambrogio [83].

2.1 The Monge and the Kantorovich problems

In 1781 Monge [66] asked the following question: given a pile of sand and a pit, how can one optimally transport the sand into the pit? In modern mathematical terms, the problem can be formulated as follows. Given two measures μ and ν on some spaces $\mathscr{X}$ and $\mathscr{Y}$, and a cost function $c : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$, find a mass-preserving function $T : \mathscr{X} \to \mathscr{Y}$ that minimises the total transportation cost

$$C(T) = \int_{\mathscr{X}} c(x, T(x)) \, \mathrm{d}\mu(x).$$

By mass-preserving we mean that for any subset $B \subseteq \mathscr{Y}$ representing a part of the pit of size $\nu(B)$, exactly that same amount of sand must go into the pit. That is, we cannot shrink or expand the sand. The amount of sand allocated to B is $\{x \in \mathscr{X} : T(x) \in B\} = T^{-1}(B)$, so the mass preservation requirement is that $\mu(T^{-1}(B)) = \nu(B)$ for all $B \subseteq \mathscr{Y}$. This condition will be denoted by $T\#\mu = \nu$ and in words: ν is the push-forward of μ under T. To make the discussion mathematically rigorous, we must assume that c and T are measurable maps, and that $\mu(T^{-1}(B)) = \nu(B)$ for all measurable subsets of $\mathscr{Y}$. When the underlying measures are understood from the context, we call T a **transport map**. Specifying $B = \mathscr{Y}$, we see that no such T can exist unless $\mu(\mathscr{X}) = \nu(\mathscr{Y})$; we shall also assume unless explicitly specified otherwise that μ and ν are probability measures. In this setting, the Monge problem is to find the optimal transport map; that is, to solve

$$\inf_{T : T\#\mu = \nu} C(T).$$

We assume throughout this thesis that $\mathscr{X}$ and $\mathscr{Y}$ are complete and separable metric spaces. The space $\mathscr{X}$ has a topology induced from its metric, and it needs to be endowed with a σ-algebra in order to make it a measure space. These two structures can be made compatible via the standard choice of taking the **Borel σ-algebra** of $\mathscr{X}$; this is, by definition, the smallest σ-algebra containing the open sets of $\mathscr{X}$. Measures defined on the Borel σ-algebra of $\mathscr{X}$ are called **Borel measures**. Thus, if μ is a Borel measure on $\mathscr{X}$, then $\mu(A)$ is defined for any A that is open, or closed, or a countable union of closed sets, etc., and any continuous map on $\mathscr{X}$ is measurable. Similarly, we endow $\mathscr{Y}$ with its Borel σ-algebra. The product space $\mathscr{X} \times \mathscr{Y}$ is also complete and separable when endowed with its product topology; its Borel σ-algebra is generated by the product σ-algebra of those of $\mathscr{X}$ and $\mathscr{Y}$; thus, any continuous cost function $c : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ is measurable. It will be assumed without further notice that μ and ν are Borel measures on $\mathscr{X}$ and $\mathscr{Y}$ respectively, and that the cost function is continuous and nonnegative.

From Section 2.5 onwards, with the only exception being parts of Section 2.9, we will always impose further restrictions. Namely, we will assume that $\mathscr{Y} = \mathscr{X}$ is a complete and separable metric space with metric d. In that case, a natural cost function is a power of the distance between the source and the target, i.e.

$$c(x, y) = d^p(x, y), \qquad p \geq 0, \quad x, y \in \mathscr{X}. \tag{2.1}$$

In particular, c is continuous, hence measurable, if $p > 0$. The limit case $p = 0$ yields the discontinuous function $c(x, y) = \mathbf{1}\{x = y\}$, which nevertheless remains measurable because the diagonal $\{(x, x) : x \in \mathscr{X}\}$ is measurable in $\mathscr{X} \times \mathscr{X}$.

The problem introduced by Monge [66] is very difficult, mainly because the set of transport maps $\{T : T\#\mu = \nu\}$ is intractable. It may very well be empty: this will be the case if μ is a Dirac measure at some $x_0 \in \mathscr{X}$ (meaning that $\mu(A) = 1$ if $x_0 \in A$ and 0 otherwise) but ν is not. Indeed, in that case the set $B = \{T(x_0)\}$ satisfies $\mu(T^{-1}(B)) = 1 > \nu(B)$, so no such T can exist. This also shows that the problem is asymmetric in μ and ν: there always exists a map T such that $T\#\nu = \mu$ — the constant map $T(x) = x_0$ for all x is in fact the unique such map. A less extreme situation happens in the case of absolutely continuous measures. If μ and ν have densities f and g on $\mathbb{R}^d$ and T is continuously differentiable, then $T\#\mu = \nu$ if and only if for μ-almost all x

$$f(x) = g(T(x))|\det \nabla T(x)|.$$

This is a highly nonlinear equation in T, nowadays known as a particular case of a family of partial differential equations called **Monge–Ampère equations**. More than two centuries after the work of Monge, Caffarelli [23] cleverly used the theory of Monge–Ampère equations to deduce smoothness properties of transport maps (see Section 2.8).

As mentioned above, if $\mu = \delta\{x_0\}$ is a Dirac measure and ν is not, then no transport maps can exist, because the mass at x_0 must be sent to a unique point x_0. In 1942 Kantorovich [54] proposed a relaxation of Monge's problem in which mass can be split. In other words, for each point $x \in \mathscr{X}$ one constructs a probability measure μ_x that describes how the mass at x is split.

If μ_x is a Dirac measure at some y, then all the mass at x is sent to y. The formal mathematical object to represent this idea is a probability measure π on the product space $\mathscr{X} \times \mathscr{Y}$ (which is $\mathscr{X}^2$ in our particular setting). Here $\pi(A \times B)$ is the amount of sand that is being sent from the subset $A \subseteq \mathscr{X}$ into the part of the pit represented by $B \subseteq \mathscr{Y}$. The total mass sent from A is $\pi(A \times \mathscr{Y})$, and the total mass sent into B is $\pi(\mathscr{X} \times B)$. Thus, π is mesaure-preserving if and only if

$$\begin{aligned} &\pi(A \times \mathscr{Y}) = \mu(A), \qquad A \subseteq \mathscr{X} \quad \text{Borel;} \\ &\pi(\mathscr{X} \times B) = \nu(B), \qquad B \subseteq \mathscr{Y} \quad \text{Borel.} \end{aligned} \tag{2.2}$$

Probability measures satisfying (2.2) will be called **transference plans**, and the set of those will be denoted by $\Pi(\mu, \nu)$. We also say that π is a **coupling** of μ and ν, and that μ and ν are the first and second **marginal distributions**, or simply **marginals**, of π. The total cost associated with $\pi \in \Pi(\mu, \nu)$ is

$$C(\pi) = \int_{\mathscr{X} \times \mathscr{Y}} c(x, y) \, \mathrm{d}\pi(x, y).$$

In our setting of a complete separable metric space $\mathscr{X}$ one can in fact represent π as a collection of probability measures $\{\pi_x\}_{x \in \mathscr{X}}$ on $\mathscr{Y}$, in the sense that for all π-integrable functions

$$\int_{\mathscr{X} \times \mathscr{Y}} g(x, y) \, \mathrm{d}\pi(x, y) = \int_{\mathscr{X}} \left[\int_{\mathscr{Y}} g(x, y) \, \mathrm{d}\pi_x(y) \right] \mathrm{d}\mu(x).$$

The collection $\{\pi_x\}$ is that of the **conditional distributions**, and the iteration of integrals is called **disintegration**. For proofs of existence of conditional distributions, one can consult Dudley [31, Section 10.2] or Kallenberg [53, Chapter 5]. Conversely, the measure μ and the collection $\{\pi_x\}$ determine π uniquely by choosing g to be indicator functions. An interpretation of these notions in terms of random variables will be given in Section 2.2.

The Kantorovich problem is then to find the best transference plan, that is, to solve

$$\inf_{\pi \in \Pi(\mu, \nu)} C(\pi).$$

The Kantorovich problem is a relaxation of the Monge problem, because to each transport map T one can associate a transference plan $\pi = \pi_T$ of the same total cost. To see this, choose the conditional distribution π_x to be a Dirac at $T(x)$. Disintegration then yields

$$C(\pi) = \int_{\mathscr{X} \times \mathscr{Y}} c(x, y) \, \mathrm{d}\pi(x, y) = \int_{\mathscr{X}} \left[\int_{\mathscr{Y}} c(x, y) \, \mathrm{d}\pi_x(y) \right] \mathrm{d}\mu(x) = \int_{\mathscr{X}} c(x, T(x)) \, \mathrm{d}\mu(x) = C(T).$$

This choice of π satisfies (2.2) because $\pi(A \times B) = \mu(A \cap T^{-1}(B))$ and $\nu(B) = \mu(T^{-1}(B))$ for all Borel $A \subseteq \mathscr{X}$ and $B \subseteq \mathscr{Y}$.

Compared to the Monge problem, the relaxed problem has considerable advantages. Firstly, the set of transference plans is never empty: it always contains the product measure $\mu \otimes \nu$ de-

fined by $[\mu \otimes \nu](A) = \mu(A)\nu(B)$. Secondly, both the objective function $C(\pi)$ and the constraints (2.2) are linear in π, so the problem can be seen as infinite-dimensional linear programming. To be precise we need to endow the space of measures with a linear structure, and this is done in the standard way: define the space $M(\mathscr{X})$ of all finite signed Borel measures on $\mathscr{X}$. This is a vector space with $(\mu_1 + \alpha\mu_2)(A) = \mu_1(A) + \alpha\mu_2(A)$ for $\alpha \in \mathbb{R}$, $\mu_1, \mu_2 \in M(\mathscr{X})$ and $A \subseteq \mathscr{X}$ Borel. The set of probability measures on $\mathscr{X}$ is denoted by $P(\mathscr{X})$, and is a convex subset of $M(\mathscr{X})$. The set $\Pi(\mu,\nu)$ is then a convex subset of $P(\mathscr{X} \times Y)$, and as $C(\pi)$ is linear in π, the set of minimisers is a convex subset of $\Pi(\mu,\nu)$. Thirdly, there is a natural symmetry between $\Pi(\mu,\nu)$ and $\Pi(\nu,\mu)$. If π belongs to the former and we define $\tilde{\pi}(B \times A) = \pi(A \times B)$, then $\tilde{\pi} \in \Pi(\nu,\mu)$. If we set $\tilde{c}(y,x) = c(x,y)$, then

$$C(\pi) = \int_{\mathscr{X}\times\mathscr{Y}} c(x,y)\,\mathrm{d}\pi(x,y) = \int_{\mathscr{Y}\times\mathscr{X}} \tilde{c}(y,x)\,\mathrm{d}\tilde{\pi}(y,x) = \tilde{C}(\tilde{\pi}).$$

In particular, when $\mathscr{X} = \mathscr{Y}$ and $c = \tilde{c}$ is symmetric (as in (2.1)),

$$\inf_{\pi\in\Pi(\mu,\nu)} C(\pi) = \inf_{\tilde{\pi}\in\Pi(\nu,\mu)} \tilde{C}(\tilde{\pi}),$$

and $\pi \in \Pi(\mu,\nu)$ is optimal if and only if its natural counterpart $\tilde{\pi}$ is optimal in $\Pi(\nu,\mu)$. This symmetry will be fundamental in the definition of the Wasserstein distances in Chapter 3.

Perhaps most importantly, a minimiser for the Kantorovich problem exists under weak conditions. In order to show this we first recall some definitions. Let $C_b(\mathscr{X})$ be the space of real-valued, continuous bounded functions on $\mathscr{X}$. A sequence of probability measures $(\mu_n) \in M(\mathscr{X})$ is said to converge **narrowly** to $\mu \in M(\mathscr{X})$ if for all $f \in C_b(\mathscr{X})$, $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$. To avoid confusion with other types of convergence, we will usually write $\mu_n \to \mu$ narrowly; in the rare cases where a symbol is needed we shall use the notation $\mu_n \xrightarrow{n} \mu$. Of course, if $\mu_n \to \mu$ narrowly and $\mu_n \in P(\mathscr{X})$, then μ must be in $P(\mathscr{X})$ too (this is seen by taking $f \equiv 1$ and by observing that $\int f\,\mathrm{d}\mu \geq 0$ if $f \geq 0$).

Remark 1. *Many authors (Billingsley [17]; Villani [88, 89]) refer to this type of convergence as weak convergence. In terms of functional analysis, however, this should have been called weak-* convergence, since (at least when $\mathscr{X}$ is compact) $M(\mathscr{X})$ is the (topological) dual of $C_b(\mathscr{X})$, but the dual of $M(\mathscr{X})$ is larger than $C_b(\mathscr{X})$. We prefer to avoid this terminology and use the term narrow convergence like, for instance, Ambrosio, Gigli & Savaré [6].*

A collection of probability measures $\mathscr{K}$ is **tight** if for all $\epsilon > 0$ there exists a compact set K such that $\inf_{\mu\in\mathscr{K}} \mu(K) > 1-\epsilon$. If $\mathscr{K}$ is represented by a sequence (μ_n), then Prokhorov's theorem [17, Theorem 5.1] states that a subsequence of (μ_n) must converge narrowly to some probability measure μ.

We are now ready to show that the Kantorovich problem admits a solution when c is continuous and nonnegative and $\mathscr{X}$ and $\mathscr{Y}$ are complete separable metric spaces. Let (π_n) be a minimising sequence for C. Since μ and ν are Borel measures on the complete separable space $\mathscr{X}$, they must be tight [17, Theorem 1.3]. If K_1 and K_2 are compact with $\mu(K_1), \nu(K_2) > 1-\epsilon$, then $K_1 \times K_2$

is compact and for all $\pi \in \Pi(\mu,\nu)$, $\pi(K_1 \times K_2) > 1-2\epsilon$. It follows that the entire collection $\Pi(\mu,\nu)$ is tight, and by Prokhorov's theorem π_n has a limit π after extraction of a subsequence. For any integer K, $c_K(x,y) = \min(c(x,y),K)$ is a continuous bounded function, and

$$C(\pi_n) = \int c(x,y)\,\mathrm{d}\pi_n(x,y) \geq \int c_K(x,y)\,\mathrm{d}\pi_n(x,y) \to \int c_K(x,y)\,\mathrm{d}\pi(x,y), \qquad n \to \infty.$$

By the monotone convergence theorem the right-hand side converges to $C(\pi)$ as $K \to \infty$, and we conclude that

$$\liminf_{n\to\infty} C(\pi_n) \geq C(\pi) \qquad \text{if } \pi_n \to \pi \text{ narrowly.} \tag{2.3}$$

Since (π_n) was chosen as a minimising sequence for C, π must be a minimiser, and existence is established.

As we have seen, the Kantorovich problem is a relaxation of the Monge problem, in the sense that

$$\inf_{T:T\#\mu=\nu} C(T) = \inf_{\pi_T:T\#\mu=\nu} C(\pi) \geq \inf_{\pi\in\Pi(\mu,\nu)} C(\pi) = C(\pi^*),$$

for some optimal π^*. If $\pi^* = \pi_T$ for some transport map T, then we say that the solution is induced from a transport map. This will happen in two different and important cases that are discussed in Sections 2.3 and 2.5.

A remark about terminology is in order. Many authors talk about the **Monge–Kantorovich problem** or the **optimal transportation problem**. More often than not, they refer to what we call here the Kantorovich problem. Usually, however, one of the scenarios presented in Sections 2.3 and 2.5 is considered, in which case this does not result in ambiguity.

2.2 Probabilistic interpretation

The preceding section was an analytic presentation of the Monge and the Kantorovich problems. It is worth mentioning that the problem can be recast in probabilistic terms, and this is the topic of this section.

A **random element** on a complete separable metric space (in fact, any topological space) $\mathscr{X}$ is simply a measurable function X from some (generic) probability space $(\Omega,\mathscr{F},\mathbb{P})$ to $\mathscr{X}$ (with its Borel σ-algebra). The **probability law** (or **probability distribution**, **law** or **distribution**) is the probability measure $\mu_X = X\#\mathbb{P}$ defined on the space $\mathscr{X}$; this is the Borel measure satisfying $\mu_X(A) = \mathbb{P}(X \in A)$ for all Borel sets A.

Suppose that one is given two random elements X and Y taking values in $\mathscr{X}$ and $\mathscr{Y}$ respectively, and a cost function $c : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$. The Monge problem is to find a measurable function T

such that $T(X)$ has the same distribution as Y, and such that the expectation

$$C(T) = \int_{\mathscr{X}} c(x, T(x))\,\mathrm{d}\mu(x) = \int_{\Omega} c[X(\omega), T(X(\omega))]\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}c(X, T(X))$$

is minimised.

The Kantorovich problem is to find (a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and) a joint distribution (X, Y) with marginal distributions X and Y respectively, such that the probability law $\pi = (X, Y)\#\mathbb{P}$ minimises the expectation

$$C(\pi) = \int_{\mathscr{X}\times\mathscr{Y}} c(x, y)\,\mathrm{d}\pi(x, y) = \int_{\Omega} c[X(\omega), Y(\omega))]\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}_{\pi} c(X, Y).$$

Any such joint distribution is called a coupling of X and Y. Of course, $(X, T(X))$ is a coupling when $T(X)$ has the same distribution as Y. The measures π_x in the previous section are then interpreted as the conditional distribution of Y given $X = x$.

Consider now the important case where $\mathscr{X} = \mathscr{Y} = \mathbb{R}^d$, $c(x, y) = \|x - y\|^2$, and X and Y are square integrable random vectors ($\mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2 < \infty$). Let A and B be the covariance matrices of X and Y respectively, and notice that that of a coupling π must have the form $C = \begin{pmatrix} A & V \\ V^t & B \end{pmatrix}$ for a $d \times d$ matrix V. The covariance matrix of the difference $X - Y$ is

$$\begin{pmatrix} I_d & -I_d \end{pmatrix} \begin{pmatrix} A & V \\ V^t & B \end{pmatrix} \begin{pmatrix} I_d \\ -I_d \end{pmatrix} = A + B - V^t - V$$

so that

$$\mathbb{E}_{\pi} c(X, Y) = \mathbb{E}_{\pi}\|X - Y\|^2 = \|\mathbb{E}X - \mathbb{E}Y\|^2 + \mathrm{tr}_{\pi}[A + B - V^t - V].$$

Since only V depends on the coupling π, the problem is equivalent to that of maximising the trace of V, the covariance matrix between X and Y. This must be done subject to the constraint that a coupling π with covariance matrix C exists; in particular C has to be positive semidefinite.

2.3 The discrete case

There is a special case in which the Monge–Kantorovich problem reduces to a finite combinatorial problem. Although it may seem at first hand as an oversimplification of the original problem, it is of importance in practice because arbitrary measures can be approximated by discrete measures by means of the strong law of large numbers. Moreover, the discrete case is important in theory as well, as a motivating example for the Kantorovich duality (Section 2.4) and the property of cyclical monotonicity (Section 2.9).

Suppose that μ and ν are supported each on n distinct points and are uniform on these points:

$$\mu = \frac{1}{n}\left(\delta\{x_1\} + \cdots + \delta\{x_n\}\right), \qquad \nu = \frac{1}{n}\left(\delta\{y_1\} + \cdots + \delta\{y_n\}\right).$$

The only relevant costs are $c_{ij} = c(x_i, y_j)$, the collection of which can be represented by an $n \times n$ matrix. Transport maps T are associated with **permutations** in S_n, the set of all bijective functions from $\{1,\ldots,n\}$ to itself: given $\sigma \in S_n$, a transport map can be constructed by defining $T(x_i) = y_{\sigma(i)}$. If σ is not a permutation, then T will not be a transport map from μ to ν. Transference plans π are associated with $n \times n$ matrices: if M is such a matrix, then one can set $\pi(\{(x_i, y_j)\}) = M_{ij}$; this is the amount of mass sent from x_i to y_j. In order for π to a be a transference plan, it must be that $\sum_j M_{ij} = 1/n$ for all i and $\sum_i M_{ij} = 1/n$ for all j, and in addition M must be nonnegative. In other words, the matrix $M' = nM$ belongs to B_n, the set of bistochastic matrices of order n, defined as $n \times n$ matrices M' satisfying

$$M'_{ij} \geq 0, \quad i,j = 1,\ldots,n; \qquad \sum_{j=1}^{n} M'_{ij} = 1, \quad i = 1,\ldots,n; \qquad \sum_{i=1}^{n} M'_{ij} = 1, \quad j = 1,\ldots,n.$$

The Monge problem is therefore the combinatorial optimisation problem over permutations

$$\inf_{\sigma \in S_n} C(\sigma) = \frac{1}{n} \inf_{\sigma \in S_n} \sum_{i=1}^{n} c_{i,\sigma(i)},$$

and the Kantorovich problem is the linear program

$$\inf_{nM \in B_n} \sum_{i,j=1}^{n} c_{ij} M_{ij} = \inf_{M \in B_n/n} \sum_{i,j=1}^{n} c_{ij} M_{ij} = \inf_{M \in B_n/n} C(M).$$

If σ is a permutation, then one can define $M = M(\sigma)$ by $M_{ij} = 1/n$ if $j = \sigma(i)$ and 0 otherwise. Then $M \in B_n/n$ and $C(M) = C(\sigma)$. Such M (or, more precisely, nM) is called a **permutation matrix**.

The Kantorovich problem is a linear program with n^2 variables and $2n$ constraints. It must have a solution because B_n (hence B_n/n) is a compact (nonempty) set in $\mathbb{R}^{n^2}$ and the objective function is linear in the matrix elements, hence continuous. (This property is independent of the possibly infinite-dimensional spaces $\mathscr{X}$ and $\mathscr{Y}$ in which the points lie.) The Monge problem also admits a solution because S_n is a finite set. To see that the two problems are essentially the same we need to introduce the following notion. If B is a convex set, then $x \in B$ is an **extremal point** of B if it cannot be written as a convex combination $tz + (1-t)y$ for some distinct points $y, z \in B$. It is well known (Luenberger & Ye [63, Section 2.5]) that there exists an optimal solution that is extremal, so that it becomes relevant to identify the extremal points of B_n. It is fairly clear that each permutation matrix is extremal in B_n; the less obvious converse is known as Birkhoff's theorem, a proof of which can be found for instance at the end of the introduction in Villani [88] or (in a different terminology) in Luenberger & Ye [63, Section 6.5]. Thus, we have:

Proposition 2.3.1 (solution of discrete problem)**.** *There exists $\sigma \in S_n$ such that $M(\sigma)$ minimises $C(M)$ over B_n/n. Furthermore, if $\{\sigma_1, \dots, \sigma_k\}$ is the set of optimal permutations, then the set of optimal matrices is the convex hull of $\{M(\sigma_1), \dots, M(\sigma_k)\}$. In particular, if σ is the unique optimal permutation, then $M(\sigma)$ is the unique optimal matrix.*

We see that in this discrete case, the Monge and the Kantorovich problems coincide. One can of course use the simplex method [63, Chapter 3] to solve the linear program, but there are $n!$ vertices, and there is in principle no guarantee that the simplex method solves the problem efficiently. However, the constraints matrix has a very specific form (it contains only zeroes and ones and has a symmetric structure), so specialised algorithms for this problem exist. One of them is the Hungarian algorithm of Kuhn [60] or its variant of Munkres [67] that has a computational complexity of at most $O(n^4)$. Another alternative is the net flow algorithms described in [63, Chapter 6]. In particular, the algorithm of Edmonds & Karp [34] has a complexity of at most $O(n^3)$.

Interestingly, this special case is automatically symmetric, whatever the cost function is. Indeed, if σ is optimal from μ to ν, then its inverse σ^{-1} is optimal from ν to μ; and if M is optimal from μ to ν, then its transpose M^t is optimal from ν to μ.

It should be remarked that the special case described here could have been more precisely called “the discrete uniform case on the same number of points", as “the discrete case" could refer to any two finitely supported measures μ and ν. When the Monge problem is of interest and symmetry is desired, however, this turns out to be the only interesting case.

Indeed, suppose that μ is supported on n points and ν on m points. Then there cannot exist a transport map from μ to ν if $m > n$ and there cannot be a transport map from ν to μ if $n > m$. Consequently, if one is interested in solving both Monge problems, the only possible case is when $n = m$. If we now assume that the weights are ordered:

$$\mu = \sum_{i=1}^{n} a_i \delta\{x_i\}, \qquad \nu = \sum_{i=1}^{n} b_i \delta\{y_i\}, \qquad 0 \le a_1 \le \cdots \le a_n; \quad 0 \le b_1 \le \cdots \le b_n,$$

then transport maps exist if and only if $a_i = b_i$, $i = 1, \dots, n$. One can then split the problem into smaller uniform problems: suppose for example that $n = 7$ and $a_5 < a_6 = a_7$. Then x_7 and x_6 can only be sent to y_7 or y_6, and this creates a uniform discrete problem of size 2. Arguing inductively, we see that the only interesting discrete case for the Monge problem is the uniform one (with the same number of points). We will henceforth refer to this special case as “the discrete case".

2.4 Kantorovich duality

The discrete case of Section 2.3 is an example of a linear program and thus enjoys a rich duality theory (Luenberger & Ye [63, Chapter 4]). The goal of this section is to show that the

Kantorovich problem admits a dual problem and benefits from a similar theory.

2.4.1 Duality in the discrete case

We can represent any matrix M as a vector in $\mathbb{R}^{n^2}$, say $\vec{M}$, by enumeration of the elements row by row. If nM is bistochastic, i.e., $M \in B_n/n$, then the $2n$ constraints can be represented in a $(2n) \times n^2$ matrix A. For instance, if $n = 3$, then

$$A = \begin{pmatrix} 1 & 1 & 1 & & & & & & \\ & & & 1 & 1 & 1 & & & \\ & & & & & & 1 & 1 & 1 \\ 1 & & & 1 & & & 1 & & \\ & 1 & & & 1 & & & 1 & \\ & & 1 & & & 1 & & & 1 \end{pmatrix} \in \mathbb{R}^{6\times 9}$$

and the constraints read $A\vec{M} = n^{-1}(1,\dots,1) \in \mathbb{R}^{2n}$. In general, if I_n is the identity matrix, then A takes the form

$$A = \begin{pmatrix} I_n & & & \\ & I_n & & \\ & & \ddots & \\ & & & I_n \\ I_n & I_n & \dots & I_n \end{pmatrix}.$$

Thus the problem can be written

$$\min_M \vec{C}^t \vec{M} \qquad \text{subject to} \qquad A\vec{M} = \frac{1}{n}(1,\dots,1) \in \mathbb{R}^{2n}; \quad \vec{M} \geq 0.$$

The last constraint is to be interpreted coordinate-wise; all the elements of M must be non-negative. The **dual problem** is constructed by introducing one variable for each row of A, transposing the constraint matrix and interchanging the roles of the objective vector $\vec{C}$ and the constraints vector $b = n^{-1}(1,\dots,1)$. If we call the new variables $p_1,\dots,p_n$ and $q_1,\dots,q_n$, we see that each column of A contains exactly one p_i and one q_j and the n^2 columns exhaust all possibilities. Hence the dual problem is

$$\max_{p,q\in\mathbb{R}^n} b^t \begin{pmatrix} p \\ q \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n p_i + \frac{1}{n}\sum_{j=1}^n q_j \qquad \text{subject to} \quad p_i + q_j \leq c_{ij}, \quad i,j = 1,\dots,n. \tag{2.4}$$

An alternative approach to duality is via a minimax argument. Introduce dual variables p_i and q_j as above and $\lambda_{ij} \geq 0$ and define the Lagrangian $\mathscr{L} : \mathbb{R}^{n^2} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+^{n^2} \to \mathbb{R}$ by

$$\mathscr{L}(M,p,q,\lambda) = \sum_{i,j=1}^n c_{ij}M_{ij} + \sum_{i=1}^n p_i\left[\frac{1}{n} - \sum_{j=1}^n M_{ij}\right] + \sum_{j=1}^n q_j\left[\frac{1}{n} - \sum_{i=1}^n M_{ij}\right] - \sum_{i,j=1}^n \lambda_{ij}M_{ij}.$$

If M satisfies the constraints $M \in B_n$, then the coefficients of p_i and of q_j in $\mathscr{L}$ vanish and the coefficient of λ_{ij} is nonnegative. It follows that

$$\sup_{p,q\in\mathbb{R}^n;\lambda\in\mathbb{R}^{n^2}_+} \mathscr{L}(M,p,q,\lambda) = \sum_{i,j=1}^n c_{ij}M_{ij}.$$

If M is not in B_n, then the supremum is easily seen to be infinite. Thus the original minimisation problem on M can be written as

$$\inf_{M\in B_n/n} \sum_{i,j=1}^n c_{ij}M_{ij} = \inf_{M\in\mathbb{R}^{n^2}} \sup_{p,q\in\mathbb{R}^n;\lambda\in\mathbb{R}^{n^2}_+} \mathscr{L}(M,p,q,\lambda).$$

The dual problem can be obtained by interchanging the infimum and the supremum: define $g : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n^2}_+$ by

$$g(p,q,\lambda) = \inf_{M\in\mathbb{R}^{n^2}} \mathscr{L}(M,p,q,\lambda) = \frac{1}{n}\sum_{i=1}^n p_i + \frac{1}{n}\sum_{j=1}^n q_j + \inf_{M\in\mathbb{R}^{n^2}} \sum_{i,j=1}^n M_{ij}[c_{ij} - \lambda_{ij} - p_i - q_j],$$

and the dual problem as

$$\sup_{p,q,\lambda} g(p,q,\lambda).$$

The expression defining g can be simplified because one can evaluate the infimum. Indeed, it is trivially negative infinite if $c_{ij} \neq \lambda_{ij} - p_i - q_j$ for some (i,j). If $c_{ij} = \lambda_{ij} + p_i + q_j$ then the infimum in g vanishes. In that case λ_{ij} does not appear in the objective function directly, so it is convenient to write the constraints as $p_i + q_j = c_{ij} - \lambda_{ij}$. Now λ_{ij} is nonnegative but otherwise arbitrary, so this is equivalent to requiring $p_i + q_j \leq c_{ij}$ for all i and all j. The dual problem is therefore

$$\sup_{p,q\in\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n p_i + \frac{1}{n}\sum_{j=1}^n q_j \qquad \text{subject to} \quad p_i + q_j \leq c_{ij}, \quad i,j = 1,\dots,n,$$

which is (2.4).

In the context of duality, one uses the terminology **primal problem** for the original optimisation problem.

2.4.2 Duality in the general case

We now use this minimax approach in order to derive a dual problem to the Kantorovich problem. It will be more convenient to work with functional constraints rather than the set constraints (2.2) that define the set of transference plans $\Pi(\mu,\nu)$. This first step is carried out using the following lemma.

Lemma 2.4.1 (functional constraints for $\Pi(\mu,\nu)$)**.** *Let μ and ν be probability measures. Then*

$\pi \in \Pi(\mu,\nu)$ *if and only if for all integrable functions* $\varphi \in L_1(\mu)$, $\psi \in L_1(\nu)$,

$$\int_{\mathscr{X}\times\mathscr{Y}} [\varphi(x)+\psi(y)]\,\mathrm{d}\pi(x,y) = \int_{\mathscr{X}} \varphi(x)\,\mathrm{d}\mu(x) + \int_{\mathscr{Y}} \psi(y)\,\mathrm{d}\nu(y).$$

The proof follows from the fact that (2.2) yields the above equality when φ and ψ are indicator functions. One then uses linearity and approximations to deduce the result.

If we now define for integrable φ and ψ

$$A_\varphi^\psi(\pi) = \int_{\mathscr{X}\times\mathscr{Y}} [\varphi(x)+\psi(y)]\,\mathrm{d}\pi(x,y), \qquad b_\varphi^\psi = \int_{\mathscr{X}} \varphi(x)\,\mathrm{d}\mu(x) + \int_{\mathscr{Y}} \psi(y)\,\mathrm{d}\nu(y),$$

and recall that $M_+(\mathscr{X}\times\mathscr{Y})$ is the set of Borel measures on $\mathscr{X}\times\mathscr{Y}$, then the Kantorovich problem can be written as

$$\inf_{\pi\in M_+(\mathscr{X}\times\mathscr{Y})} C(\pi) = \int_{\mathscr{X}\times\mathscr{Y}} c(x,y)\,\mathrm{d}\pi(x,y) \qquad \text{subject to} \quad A_\varphi^\psi(\pi) = b_\varphi^\psi, \quad \varphi\in L_1(\mu); \psi \in L_1(\nu).$$

One can formally define the Lagrangian $\mathscr{L} : M_+(\mathscr{X}\times\mathscr{Y}) \times \mathbb{R}^{L_1(\mu)\times L_1(\nu)}$ as

$$\mathscr{L}(\pi,p) = C(\pi) + \sum_{\varphi,\psi} p_\varphi^\psi [b_\varphi^\psi - A_\varphi^\psi(\pi)],$$

but this is an uncountable sum that will not have a meaning for most values of π and p. A more fruitful approach is to view the functions φ and ψ themselves as dual variables, and define

$$\mathscr{L}(\pi,\varphi,\psi) = C(\pi) + b_\varphi^\psi - A_\varphi^\psi(\pi).$$

Let us now take a supremum over φ and ψ. If $\pi \notin \Pi(\mu,\nu)$, then $b_\varphi^\psi \neq A_\varphi^\psi(\pi)$ for some ψ and some φ. Choosing an arbitrary large negative or positive value for p_φ^ψ, we see that the supremum is infinite. On the other hand, if $\pi \in \Pi(\mu,\nu)$, the supremum is trivially $C(\pi)$. Thus

$$\inf_{\pi\in M_+(\mathscr{X}\times\mathscr{Y})} \sup_{(\varphi,\psi)\in L_1(\mu)\times L_1(\nu)} \mathscr{L}(\pi,\varphi,\psi) = \inf_{\pi\in\Pi(\mu,\nu)} C(\pi)$$

recovers the Kantorovich problem. Interchanging the supremum and the infimum and plugging in the definitions of $C(\pi)$ and $A_\varphi^\psi(\pi)$ yields the dual problem

$$\sup_{(\varphi,\psi)\in L_1(\mu)\times L_1(\nu)} b_\varphi^\psi + \inf_{\pi\in M_+(\mathscr{X}\times\mathscr{Y})} \int_{\mathscr{X}\times\mathscr{Y}} [c(x,y)-\varphi(x)-\psi(y)]\,\mathrm{d}\pi(x,y).$$

The infimum at the right-hand side is negative infinite if $\varphi(x_0)+\psi(y_0) > c(x_0,y_0)$ for some $x_0 \in \mathscr{X}$ and $y_0 \in \mathscr{Y}$, since we can take π to be a Dirac mass at (x_0,y_0) with arbitrarily large mass. If we define the set

$$\Phi_c = \big\{(\varphi,\psi) \in L_1(\mu)\times L_1(\nu) : \varphi(x)+\psi(y) \le c(x,y) \text{ for all } x,y\big\},$$

then the dual problem becomes

$$\sup_{(\varphi,\psi)\in L_1(\mu)\times L_1(\nu)} \int_{\mathscr{X}} \varphi(x)\,\mathrm{d}\mu(x) + \int_{\mathscr{Y}} \psi(y)\,\mathrm{d}\nu(y) \qquad \text{subject to} \quad (\varphi,\psi)\in\Phi_c.$$

Notice how this reduces to (2.4) in the discrete case.

If $\pi\in\Pi(\mu,\nu)$ and $(\varphi,\psi)\in\Phi_c$, then by Lemma 2.4.1

$$b_\varphi^\psi = \int_{\mathscr{X}\times\mathscr{Y}} [\varphi(x)+\psi(y)]\,\mathrm{d}\pi(x,y) \le C(\pi).$$

In particular, the supremum of b_φ^ψ is no larger than the infimum of $C(\pi)$. This result is known as **weak duality**, and it holds in full generality (provided that there exists some $\pi\in\Pi(\mu,\nu)$ for which $C(\pi)>-\infty$). More important and far more useful is **strong duality**:

Theorem 2.4.2 (Kantorovich duality)**.** *Let μ and ν be probability measures on complete separable metric spaces $\mathscr{X}$ and $\mathscr{Y}$ respectively and let $c:\mathscr{X}\times\mathscr{Y}$ be a nonnegative continuous function. Then*

$$\inf_{\pi\in\mathbb{P}(\mu,\nu)} \int_{\mathscr{X}\times\mathscr{Y}} c\,\mathrm{d}\pi = \sup_{(\varphi,\psi)\in\Phi_c} \int_{\mathscr{X}} \varphi\,\mathrm{d}\mu + \int_{\mathscr{Y}} \psi\,\mathrm{d}\nu.$$

We shall only use Theorem 2.4.2 in this form, but it holds in far more general circumstances (Villani [89, Theorem 5.10]; Rachev & Rüschendorf [73, Chapter 4]).

2.4.3 Relationship between the dual and primal problems

It is well-known (Luenberger & Ye [63, Section 4.4]) that the solutions to the primal and dual problems are related to each other via **complementary slackness**. In other words, solution of one problem provides a lot of information about the solution of the other problem. Here we show that this idea remains true for the Kantorovich primal and dual problems.

If one finds functions $(\varphi,\psi)\in\Phi_c$ and a transference plan $\pi\in\Pi(\mu,\nu)$ such that $C(\pi)=b_\varphi^\psi$, then by weak duality (φ,ψ) are optimal in Φ_c and π is optimal in $\pi\in\Pi(\mu,\nu)$. This is equivalent to

$$\int_{\mathscr{X}\times\mathscr{Y}} [c(x,y)-\varphi(x)-\psi(y)]\,\mathrm{d}\pi(x,y) = 0$$

which is in turn equivalent to

$$\varphi(x)+\psi(y)=c(x,y), \qquad \pi\text{-almost surely.}$$

It has already been established that there exists an optimal transference plan π^*. Let us assume that $C(\pi^*)<\infty$ (otherwise all transference plans are optimal). Then a pair $(\varphi,\psi)\in\Phi_c$

is optimal if and only if

$$\varphi(x) + \psi(y) = c(x,y), \qquad \pi^*\text{-almost surely.}$$

Conversely, if (φ^*, ψ^*) is an optimal pair, then π is optimal if and only if it is concentrated on the set

$$\{(x,y) : \varphi^*(x) + \psi^*(y) = c(x,y)\}.$$

In particular, if for a given x there exists a unique y such that $\varphi^*(x) + \psi^*(y) = c(x,y)$, then the mass at x must be sent entirely to y and not be split; if this is the case for μ-almost all x, then this relation defines y as a function of x and the resulting optimal π is in fact induced from a transport map. This idea provides a criterion for solvability of the Monge problem, see Villani [89, Theorem 5.30].

2.4.4 Unconstrained dual Kantorovich problem

It turns out that the dual Kantorovich problem can be recast as an unconstrained optimisation problem of only one function φ. The new formulation is not only conceptually simpler than the original one, but also sheds light on the properties of the optimal dual variables.

Since the dual objective function to be maximised

$$b_\varphi^\psi = \int_{\mathscr{X}} \varphi \, \mathrm{d}\mu + \int_{\mathscr{Y}} \psi \, \mathrm{d}\nu$$

is increasing in φ and ψ, one should seek functions that take values as large as possible subject to the constraint $\varphi(x) + \psi(y) \le c(x,y)$. Suppose that an oracle tells us that some $\varphi \in L_1(\mu)$ is a good candidate. Then the largest possible ψ satisfying $(\varphi, \psi) \in \Phi_c$ is defined as

$$\psi(y) = \inf_{x \in \mathscr{X}} c(x,y) - \varphi(x).$$

A function taking this form will be called c-**concave** [88, Chapter 2]; we say that ψ is the c-**transform** of φ and denote $\varphi^c = \psi$. It is not necessarily true that φ^c is integrable or even measurable, but if we neglect this difficulty, then it is obvious that

$$\sup_\psi b_\varphi^\psi = b_\varphi^{\varphi^c}.$$

The dual problem can thus be formulated as the unconstrained problem

$$\sup_{\varphi \in L_1(\mu)} \int_{\mathscr{X}} \varphi \, \mathrm{d}\mu + \int_{\mathscr{Y}} \varphi^c \, \mathrm{d}\nu.$$

One can apply this c-transform again and replace φ by

$$\varphi^{cc}(x) = (\varphi^c)^c(x) = \inf_{y\in\mathscr{Y}} c(x,y) - \varphi^c(y) \geq \varphi(x),$$

so that $b_{\varphi}^{\varphi^c} \leq b_{\varphi^{cc}}^{\varphi^c}$ but still $(\varphi^{cc},\varphi^c) \in \Phi_c$ (modulo measurability issues). An elementary calculation shows that in general $\varphi^{ccc} = \varphi^c$. Thus, for any function φ_1, the pair of functions $(\varphi,\psi) = (\varphi_1^{cc},\varphi_1^c)$ satisfies $\varphi^c = \psi$ and $\psi^c = \varphi$. We say that φ and ψ are **c-conjugate**.

Proposition 2.4.3 (existence of an optimal pair)**.** *Let μ and ν be probability measures on $\mathscr{X}$ and $\mathscr{Y}$ with optimal transference plan π^* such that $C(\pi^*)$ is finite. Then there exists an optimal pair (φ,ψ) for the dual Kantorovich problem. Furthermore, the pair can be chosen in a way that μ-almost surely, $\varphi = \psi^c$ and ν-almost surely, $\psi = \varphi^c$.*

This result is due to Ambrosio & Pratelli [7]. It is clear from the discussion above that once existence of an optimal pair (φ_1,ψ_1) is established, the pair $(\varphi,\psi) = (\varphi_1^{cc},\varphi_1^c)$ should be optimal; Ambrosio & Pratelli show that this pair can be modified up to null sets in order to be Borel measurable. Furthermore, $\int\varphi\,\mathrm{d}\mu + \int\psi\,\mathrm{d}\nu$ is always finite, since we require φ and ψ to be integrable. It follows that the condition $C(\pi^*) < \infty$ is necessary for their existence.

Whether $\varphi^c(y)$ is tractable to evaluate depends on the structure of c. Here is a concrete example. Assume that $\mathscr{X} = \mathscr{Y}$, denote their metric by d, and let $c(x,y) = d(x,y)$. If $\varphi = \psi^c$ is c-concave, then it is 1-Lipschitz. Indeed, by definition and the triangle inequality

$$\varphi(z) = \inf_y d(z,y) - \psi(y) \leq \inf_y d(x,y) + d(x,z) - \psi(y) = \varphi(x) + d(x,z).$$

Interchanging x and z yields $|\varphi(x) - \varphi(z)| \leq d(x,z)$.

Next, we claim that if φ is Lipschitz, then $\varphi^c(y) = -\varphi(y)$. Indeed, choosing $x = y$ in the infimum shows that $\varphi^c(y) \leq d(y,y) - \varphi(y) = -\varphi(y)$. But the Lipschitz condition on φ implies that for all x, $d(x,y) - \varphi(x) \geq -\varphi(y)$. In view of that, we can take in the dual problem φ to be Lipschitz and $\psi = -\varphi$, and the duality formula (Theorem 2.4.2) takes the form

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathscr{X}^2} d(x,y)\,\mathrm{d}\pi(x,y) = \sup_{\|\varphi\|_{Lip}\leq 1} \left| \int_{\mathscr{X}} \varphi\,\mathrm{d}\mu - \int_{\mathscr{X}} \varphi\,\mathrm{d}\nu \right|, \quad \|\varphi\|_{Lip} = \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)}. \tag{2.5}$$

This is known as the **Kantorovich–Rubinstein theorem** [88, Theorem 1.14]. (We have been a bit sloppy because φ may not be integrable. But if for some $x_0 \in \mathscr{X}$, $x \mapsto d(x,x_0)$ is in $L_1(\mu)$, then any Lipschitz function is μ-integrable. Otherwise one needs to restrict the supremum to bounded Lipschitz φ.)

Combining Proposition 2.4.3 with the preceding subsection, we see that if φ is optimal, then any optimal transference plan π^* must be concentrated on the set

$$\{(x,y) : \varphi(x) + \varphi^c(y) = c(x,y)\}.$$

If for μ-almost every x this equation defines y uniquely as a (measurable) function of x, then π^* is induced by a transport map. In the next section we present concrete examples as to when this happens.

2.5 The absolutely continuous case

2.5.1 Quadratic cost

Let us now consider the most important example of the Kantorovich problem. Suppose that $\mathscr{X} = \mathscr{Y}$ is a separable Hilbert space and the cost function is $c(x,y) = \|x-y\|^2/2$. Let φ be any function. Then

$$\varphi^c(y) = \inf_{x\in\mathscr{X}} \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2} - \langle x, y\rangle - \varphi(x) = \frac{\|y\|^2}{2} - \sup_{x\in\mathscr{X}} \langle x, y\rangle - \left(\frac{\|x\|^2}{2} - \varphi(x)\right).$$

Equivalently,

$$\frac{\|y\|^2}{2} - \varphi^c(y) = \sup_{x\in\mathscr{X}} \langle x, y\rangle - \left(\frac{\|x\|^2}{2} - \varphi(x)\right).$$

When viewed as a function of y, the right-hand side is the supremum of affine functions, hence enjoys some useful properties. We remind the reader that a function $f : \mathscr{X} \to \mathbb{R}\cup\{\infty\}$ is **convex** if $f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$ for all $x, y \in \mathscr{X}$ and $t \in [0,1]$. It is **lower semicontinuous** if for all $x \in \mathscr{X}$, $f(x) \le \liminf_{y\to x} f(y)$. Now affine functions are convex and lower semicontinuous, and it straightforward from the definitions that both convexity and lower semicontinuity are stable under the supremum operation. Thus the function $\|y\|^2/2 - \varphi^c(y)$ is convex and lower semicontinuous. In particular, it is measurable due to the following characterisation: f is lower semicontinuous if and only if $\{x : f(x) \le c\}$ is a closed set for all $c \in \mathbb{R}$. So in this particular case, there is no need to modify φ^c on a null set; it is already Borel measurable. From the preceding subsection, we now know that optimal dual functions φ and ψ must take the form of the difference between $\|\cdot\|^2/2$ and a convex function.

Given the vast wealth of knowledge on convex functions (Rockafellar [78]), it will be convenient to work with

$$\tilde{\varphi}(x) = \frac{\|x\|^2}{2} - \varphi(x), \qquad \text{and} \qquad \tilde{\varphi}^*(y) = \frac{\|y\|^2}{2} - \varphi^c(y) = \sup_{x\in\mathscr{X}} \langle x, y\rangle - \tilde{\varphi}(x),$$

so that

$$\varphi(x) + \psi(y) = c(x,y) \quad \Longleftrightarrow \quad \tilde{\varphi}(x) + \tilde{\varphi}^*(y) = \langle x, y\rangle.$$

The function $\tilde{\varphi}^*$ is known as the **Legendre transform** of $\tilde{\varphi}$ ([78, Chapter 26]; [88, Chapter 2]), and is of fundamental importance in convex analysis. It is, of course, convex and lower semicontinuous. Furthermore, we can assume that so is $\tilde{\varphi}$, because otherwise we may replace

it by $\tilde{\varphi}^{**}$.

What can we say about optimal transference plans π? If φ is optimal, then they must be concentrated on the set of (x, y) such that $\tilde{\varphi}(x) + \tilde{\varphi}^*(y) = \langle x, y \rangle$. By definition of the Legendre transform as a supremum, this happens if and only if the supremum is attained at x; equivalently

$$\tilde{\varphi}(z) - \tilde{\varphi}(x) \geq \langle z - x, y \rangle, \qquad z \in \mathscr{X}.$$

This condition is precisely the definition of y being a **subgradient** of $\tilde{\varphi}$ at x [78, Chapter 23].

Let us now further restrict the attention to a finite dimensional setting, where $\mathscr{X} = \mathscr{Y} = \mathbb{R}^d$ for some integer $d \geq 1$. It is then well-known [78, Theorem 25.1] that if $\tilde{\varphi}$ is differentiable, then the unique subdifferential at x is its gradient $\nabla\tilde{\varphi}(x)$. If we are fortunate and $\tilde{\varphi}$ is differentiable everywhere, or even μ-almost everywhere, then the optimal transference plan π is unique, and in fact induced from the transport map $\nabla\tilde{\varphi}$. The problem, of course, is that $\tilde{\varphi}$ may fail to be differentiable μ-almost surely. This is remedied by assuming some **regularity** on the source measure μ in order to make sure that *any* convex function be differentiable μ-almost surely, and is done via the following result, which is a simplified version of Theorem 25.5 in Rockafellar [78]. Another proof can be found in Alberti & Ambrosio [3, Chapter 2].

Theorem 2.5.1 (differentiability of convex functions)**.** *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be a convex function with domain $\operatorname{dom} f = \{x \in \mathbb{R}^d : f(x) < \infty\}$ and let $\mathscr{N}$ be the set of points at which f is not differentiable. Then $\mathscr{N} \cap \operatorname{int}(\operatorname{dom} f)$ has Lebesgue measure 0.*

Here $\operatorname{int} A$ means the interior of A, defined as the largest open set included in A. When A is convex and bounded, its boundary has Lebesgue measure zero: indeed, if $\operatorname{int} A$ is empty, then the closure of A lies in a lower dimensional subspace [78, Theorem 2.4]. Otherwise, without loss of generality $0 \in \operatorname{int} A$, and then by convexity of A, $\partial A \subseteq (1+\epsilon)A$ for all $\epsilon > 0$. When A is unbounded, write it as $\cup_n A \cap [-n, n]^d$. Since $\operatorname{dom} f$ is convex, it follows that in fact $\mathscr{N} \cap \operatorname{dom} f$ has Lebesgue measure zero.

Another issue that might arise is that optimal φ's might not exist. This is easily dealt with using Proposition 2.4.3. If we assume that μ and ν have finite second moments:

$$\int_{\mathbb{R}^d} \|x\|^2 \, \mathrm{d}\mu(x) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}^d} \|y\|^2 \, \mathrm{d}\nu(y) < \infty,$$

then any transference plan $\pi \in \Pi(\mu, \nu)$ has a finite cost, as is seen from integrating the elementary inequality $\|x - y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$ and using Lemma 2.4.1:

$$C(\pi) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} [\|x\|^2 + \|y\|^2] \, \mathrm{d}\pi(x, y) = \int_{\mathbb{R}^d} \|x\|^2 \, \mathrm{d}\mu(x) + \int_{\mathbb{R}^d} \|y\|^2 \, \mathrm{d}\nu(y) < \infty.$$

With these tools, we can now prove a fundamental existence and uniqueness result for the Monge–Kantorovich problem. It has been proven independently by several authors, including

Brenier [22], Cuesta-Albertos & Matrán [26], Knott & Smith [58] and Rachev & Rüschendorf [81].

Theorem 2.5.2 (quadratic cost in Euclidean spaces)**.** *Let μ and ν be probability measures on $\mathbb{R}^d$ with finite second moments, and suppose that μ is absolutely continuous with respect to Lebesgue measure. Then the solution to the Kantorovich problem is unique, and is induced from a transport map T that equals μ-almost surely the gradient of a convex function ϕ. Furthermore, the pair $(\|x\|^2/2 - \phi, \|y\|^2/2 - \phi^*)$ is optimal for the dual problem.*

Proof. The proof is now almost obvious. By Proposition 2.4.3 there exists an optimal dual pair (φ, ψ) such that $\phi(x) = \|x\|^2/2 - \varphi(x)$ is convex and lower semicontinuous, and by the discussion in Section 2.1, there exists an optimal π. Since ϕ is μ-integrable, it must be finite almost everywhere, i.e. $\mu(\mathrm{dom}\phi) = 1$. By Theorem 2.5.1, if we define $\mathscr{N}$ as the set of nondifferentiability points of ϕ, then $\mathrm{Leb}(\mathscr{N} \cap \mathrm{dom}\phi) = 0$; as μ is absolutely continuous, the same holds for μ. (Here Leb denotes Lebesgue measure.)

We conclude that $\mu(\mathrm{int}(\mathrm{dom}\phi) \setminus \mathscr{N}) = 1$. In other words, ϕ is differentiable μ-almost everywhere, and so for μ-almost any x, there exists a unique y such that $\phi(x) + \phi^*(y) = \langle x, y \rangle$, and $y = \nabla\phi(x)$. This shows that π is unique and induced from the transport map $\nabla\phi(x)$. Finally, $\nabla\phi$ is Borel measurable, since each of its coordinates can be written as $\limsup_{q \to 0, q \in \mathbb{Q}} q^{-1}(\phi(x + qv) - \phi(x))$ for some vector v (the canonical basis of $\mathbb{R}^d$), which is measurable because the limit superior is taken on countably many functions (and ϕ is measurable because it is lower semicontinuous). □

Theorem 2.5.2 gives a rather general situation (in terms of the measures μ and ν) in which the solution to the Kantorovich problem is given by a proper map. In the next subsection we show that it holds true for more general cost functions.

2.5.2 Strictly convex cost functions

When c is not the quadratic cost, we cannot open up the square and relate the Monge–Kantorovich problem to convexity. However, we can still apply the idea that $\varphi(x) + \varphi^c(y) = c(x, y)$ if and only if the infimum is attained at x. Indeed, recall that

$$\varphi^c(y) = \inf_{x \in \mathscr{X}} c(x, y) - \varphi(x),$$

so that $\varphi(x) + \varphi^c(y) = c(x, y)$ if and only if

$$\varphi(z) - \varphi(x) \leq c(z, y) - c(x, y), \qquad z \in \mathscr{X}.$$

Notice the similarity to the subgradient inequality in the previous subsection, with the sign being reversed. In analogy, we call the collection of y's satisfying the above in equality the

c-superdifferential of φ at x, and we denote it by $\partial^c\varphi(x)$. Of course, if $c(x,y) = \|x-y\|^2/2$, then $y \in \partial^c(x)$ if and only if y is a subgradient of $(\|\cdot\|^2/2 - \varphi)$ at x.

The idea is now to identify a class of functions c such that if φ is c-concave, then for most values of x, its c-superdifferential at x, $\partial^c\varphi(x)$, consists of a single point y. As before, if φ is dual optimal and $\varphi^c(x)$ contains a single point y for μ-almost all x, then the unique optimal $\pi \in \Pi(\mu,\nu)$ is again induced by the transport map $\partial^c\varphi$. As for the case of quadratic cost, we would like to find such a class for which $\partial^c\varphi(x)$ is unique *Lebesgue*-almost surely, and then it will automatically be the case μ-almost surely when μ is absolutely continuous.

It turns out that such a class is given by $c(x,y) = \|x-y\|^p/p$ for $p > 1$.

Theorem 2.5.3 (strictly convex costs in $\mathbb{R}^d$)**.** *Let $c(x,y) = \|x-y\|^p/p$ for some $p > 1$ and let μ and ν be probability measures on $\mathbb{R}^d$ with finite p-th moments such that μ is absolutely continuous with respect to Lebesgue measure. Then the solution to the Kantorovich problem with cost function $c(x,y) = \|x-y\|^p/p$ is unique and induced from a transport map T. Furthermore, there exists an optimal pair (ϕ,ϕ^c) of the dual problem, with ϕ c-concave. The solutions are related by*

$$T(x) = x - \nabla\phi(x)\|\nabla\varphi(x)\|^{1/(p-1)-1} \qquad (\mu\text{-almost surely}).$$

This result is due to Gangbo & McCann [37]. Let us show the easy part of the proof, which relates differentiability properties of ϕ to that of c. More precisely, define $h(v) = \|v\|^p/p$ so that $c(x,y) = h(\|x-y\|)$. Suppose now that $y \in \partial^c\phi(x)$ and let us assume that ϕ is subdifferentiable at x. That is, there exists a subgradient $u \in \mathbb{R}^d$ such that

$$\phi(z) - \phi(x) \geq \langle u, z-x\rangle + o(\|z-x\|).$$

Here and more generally, $o(\|z-x\|)$ denotes a function $r(z)$ (defined in a neighbourhood of x) such that $r(z)/\|z-x\| \to 0$ as $z \to x$. (If ϕ were convex then we could take $r \equiv 0$, so the definition for convex functions is equivalent, and then the inequality holds globally and not only locally.) But $y \in \partial^c\phi(x)$ means that

$$h(z-y) - h(x-y) = c(z,y) - c(x,y) \geq \phi(z) - \phi(x) \geq \langle u, z-x\rangle + o(\|z-x\|).$$

In other words, z is a subgradient of h at $x-y$. Now, since h is differentiable, it only has one subgradient $u = \nabla h(x-y)$. This means that u must be unique too. Now if ϕ were differentiable, then it must be that $u = \nabla\phi(x) = \nabla h(x-y)$. Since h is strictly convex, its gradient is invertible, so this equation defines y uniquely via

$$y = x - (\nabla h)^{-1}[\nabla\phi(x)],$$

which defines y as a function of x. So if, μ-almost everywhere, ϕ is differentiable and has a c-supergradient, then there is a unique transference plan induced by the transport map

$T(x) = x - (\nabla h)^{-1}[\nabla\phi(x)]$. Of course, we can assume that ϕ is c-concave; and this is exactly what Gangbo & McCann showed: if ϕ is c-concave, then it is differentiable Lebesgue-almost everywhere and further has at least (thus exactly) one c-supergradient.

It should be remarked that the result of Gangbo & McCann holds for much more general cost functions than $\|x-y\|^p/p$. Furthermore, if the cost function is sufficiently smooth, μ does not need to be absolutely continuous; it suffices that it not give positive measure to any set of Hausdorff dimension smaller or equal than $d-1$. If $d=1$ this means that Theorem 2.5.3 is still valid as long as μ has no atoms ($\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$), even if μ is not absolutely continuous.

In the same reference [37], Gangbo & McCann deal with strictly concave cost functions, where the situation is similar *if the supports of μ and ν are disjoint.* Analogous results in an infinite-dimensional setting can be found in Ambrosio, Gigli & Savaré [6, Theorem 6.2.10]. Although there is no obvious parallel for Lebesgue measure (i.e., translation invariant) on infinite-dimensional Banach spaces, one can still define absolute continuity via Gaussian measures. Indeed, $\mu \in P(\mathbb{R}^d)$ is absolutely continuous with respect to Lebesgue measure if and only if the following holds: if $\mathcal{N} \subset \mathbb{R}^d$ is such that $\nu(\mathcal{N}) = 0$ for any nondegenerate Gaussian measure ν, then $\mu(\mathcal{N}) = 0$. This definition can be extended to any separable Banach space $\mathcal{X}$ via projections, as follows. Let $\mathcal{X}^*$ be the (topological) dual of $\mathcal{X}$.

Definition 2.5.4 (Gaussian measures)**.** *A probability measure $\mu \in P(\mathcal{X})$ is a nondegenerate Gaussian measure if for any $\ell \in \mathcal{X}^* \setminus \{0\}$, $\ell\#\mu \in P(\mathbb{R})$ is a Gaussian measure with positive variance.*

Definition 2.5.5 (Gaussian null sets and absolutely continuous measures)**.** *A subset $\mathcal{N} \subset \mathcal{X}$ is a Gaussian null set if whenever ν is nondegenerate Gaussian measure, $\nu(\mathcal{N}) = 0$. A probability measure $\mu \in P(\mathcal{X})$ is absolutely continuous if μ vanishes on all Gaussian null sets.*

Clearly, if ν is a nondegenerate Gaussian measure, then it is absolutely continuous. In the sequel, $\mathcal{X}$ will usually be a Hilbert space, and then one can think of absolutely continuous measures as measures $\mu \in P(\mathcal{X})$ such that for any other $\nu \in P(\mathcal{X})$, there exists a unique optimal transference plan induced by a transport map T with respect to the cost $\|x-y\|^p/p$ (provided some moment conditions are satisfied). Therefore in the sense of optimal transportation, Definition 2.5.5 is an extension of the notion of absolute continuity of measures in $\mathbb{R}^d$ with respect to Lebesgue measure.

2.6 The one-dimensional case

When $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, the Monge–Kantorovich problem admits a unique solution in the situation described in the previous subsection. The main difference is that the solution has an explicit form in terms of the distribution functions of the measures. Specifically, let $\mu, \nu \in P(\mathbb{R})$ with distribution functions F and G respectively,

$$F(t) = \mu((-\infty, t]), \qquad G(t) = \nu((-\infty, t]), \qquad t \in \mathbb{R}.$$

Suppose that the cost function is $c(x,y) = |x-y|^2/2$ and let $x_1 \le x_2$, $y_1 \le y_2$. Since

$$c(y_2,x_1) + c(y_1,x_2) - c(y_1,x_1) - c(y_2,x_2) = (x_2-x_1)(y_2-y_1) \ge 0,$$

it seems natural to expect the optimal transport map to be monotonically increasing. In fact, the inequality holds whenever $c(x,y) = h(|x-y|)$ with $h : \mathbb{R}_+ \to \mathbb{R}$ convex, as elementary calculations show. It turns out that, on the real line, there is at most one such map: if T is increasing and $T\#\mu = \nu$, then for all $t \in \mathbb{R}$

$$G(t) = \nu((-\infty,t]) = \mu((-\infty, T^{-1}(t)]) = F(T^{-1}(t)).$$

If $t = T(x)$, then the above equation reduces to $T(x) = G^{-1}(F(x))$. This formula determines T uniquely, and has an interesting probabilistic interpretation: it is well-known that if X is a random variable with *continuous* distribution function F, then $F(X)$ follows a uniform distribution on $(0,1)$. Conversely, if U follows a uniform distribution, G is any distribution function, and

$$G^{-1}(u) = \inf G^{-1}([u,1]) = \inf\{x \in \mathbb{R} : G(x) \ge u\}, \qquad 0 < u < 1,$$

is the **quantile function** of X, then the random variable $G^{-1}(U)$ has distribution function G. We say that G is the **left-continuous inverse** of G. In terms of push-forward maps, we can write $F\#\mu = \mathrm{Leb}|_{[0,1]}$ and $G^{-1}\#\mathrm{Leb}|_{[0,1]} = \nu$, with Leb standing for Lebesgue measure, and it is restricted to the interval $[0,1]$. Consequently, we see that if F is continuous and G is arbitrary, then $T\#\mu = \nu$; we can view T as pushing μ forward to ν in two steps: firstly, μ is pushed forward to $\mathrm{Leb}|_{[0,1]}$ and secondly, $\mathrm{Leb}|_{[0,1]}$ is pushed forward to ν.

Using the change of variables formula, we see that the total cost of T is

$$C(T) = \int_{\mathbb{R}} c(G^{-1}(F(x)),x)\,\mathrm{d}\mu(x) = \int_0^1 c(G^{-1}(u),F^{-1}(u))\,\mathrm{d}u.$$

If F is discontinuous, then $F\#\mu$ is not Lebesgue measure, and T is not necessarily defined. But there will exist an optimal transference plan $\pi \in \Pi(\mu,\nu)$ which is monotone in the following sense: there exists a set $\Gamma \subset \mathbb{R}^2$ such that $\pi(\Gamma) = 1$ and whenever $(x_i,y_i) \in \Gamma$,

$$c(y_2,x_1) + c(y_1,x_2) - c(y_1,x_1) - c(y_2,x_2) \ge 0.$$

This is a particular case of the cyclical monotonicity that will be discussed in Section 2.9. Thus, if $x_1 < x_2$, then it must be that $y_1 \le y_2$. Since any distribution can be approximated by continuous distributions, in view of the above discussion, the following result from Villani [88, Theorem 2.18] should not be surprising.

Theorem 2.6.1 (optimal transportation in $\mathbb{R}$)**.** *Let $\mu,\nu \in P(\mathbb{R})$ with distribution functions F and*

G respectively and let the cost function be of the form $c(x,y) = h(|x-y|)$ with h convex. Then

$$\inf_{\pi\in\Pi(\mu,\nu)} C(\pi) = \int_0^1 h(G^{-1}(u) - F^{-1}(u))\,\mathrm{d}u.$$

Furthermore, if F is continuous, then the infimum is attained by the transport map $T = G^{-1}\circ F$, and if in addition $h(z) = \|z\|^p/p$ for some $p > 1$ and μ and ν have finite p-th moments, then the unique solution is the transference plan π induced by T.

This result allows in particular a direct evaluation of the Wasserstein distances for measures on the real line (see Chapter 3).

The only part of our formulation that is not explicitly proven in [88] is the last part about the uniqueness, in which case one may invoke Theorem 2.5.3 (if μ is not absolutely continuous, see the discussion after the sketch of the proof of that theorem).

When $p = 1$, the cost function is convex but not strictly, and solutions will not be unique. However, the total cost in Theorem 2.6.1 admits another representation that is often more convenient.

Proposition 2.6.2 (quantiles and distribution functions)**.** *If F and G are distribution functions, then*

$$\int_0^1 |G^{-1}(u) - F^{-1}(u)|\,\mathrm{d}u = \int_{\mathbb{R}} |G(x) - F(x)|\,\mathrm{d}x.$$

Proof. It is well known that $F^{-1}(u) \le x$ if and only if $u \le F(x)$. Let $A = \{u : G^{-1}(u) > F^{-1}(u)\} \subseteq (0,1)$ and notice that for $u \in A$, $F^{-1}(u) \le x < G^{-1}(u)$ if and only if $G(x) < u \le F(x)$. A similar equivalence holds when $u \in B = (0,1)\setminus A$. It follows from Fubini's theorem that

$$\int_A |G^{-1}(u) - F^{-1}(u)|\,\mathrm{d}u = \int_A \left(\int_{F^{-1}(u)}^{G^{-1}(u)} 1\,\mathrm{d}x\right)\mathrm{d}u = \int_{\mathbb{R}}\left(\int_{G(x)}^{F(x)} 1_A(u)\mathbf{1}\{F(x)\ge G(x)\}\,\mathrm{d}u\right)\mathrm{d}x;$$
$$\int_B |G^{-1}(u) - F^{-1}(u)|\,\mathrm{d}u = \int_B \left(\int_{G^{-1}(u)}^{F^{-1}(u)} 1\,\mathrm{d}x\right)\mathrm{d}u = \int_{\mathbb{R}}\left(\int_{F(x)}^{G(x)} 1_B(u)\mathbf{1}\{G(x)\ge F(x)\}\,\mathrm{d}u\right)\mathrm{d}x.$$

Since $1_A(u) + 1_B(u) = 1$, summing up these equalities yields the result. □

Corollary 2.6.3. *If $c(x,y) = |x-y|$ then under the conditions of Theorem 2.6.1*

$$\inf_{\pi\in\Pi(\mu,\nu)} C(\pi) = \int_{\mathbb{R}} |G(x) - F(x)|\,\mathrm{d}x.$$

This result, as well as the more general Theorem 2.6.1 do not assume that the total cost is finite, in which case both sides are infinite. Somewhat abusing the terminology, we will refer to $T = G^{-1}\circ F$ as *the* optimal map even in the rare cases where the total cost is infinite.

2.7 The Gaussian case with quadratic cost

Beside the one-dimensional case in the previous section, there is another special case in which not only uniqueness holds, but one also has an explicit solution to the Monge–Kantorovich problem.

Suppose that μ and ν are Gaussian measures on $\mathbb{R}^d$ with zero means and nonsingular covariance matrices A and B. By Theorem 2.5.2 we know that there exists a unique optimal map T such that $T\#\mu = \nu$. Since linear push-forwards of Gaussians are Gaussian, it seems natural to guess that T should be linear. This is indeed the case, as was shown independently by Dowson & Landau [29] and Olkin & Pukelsheim [69].

We present the argument of Bhatia [13, Exercise 1.2.13]. Since T is a linear map that should be the gradient of a convex function ϕ, it must be that ϕ is quadratic, i.e. $\phi(x) = \langle x, Ax\rangle$ for $x \in \mathbb{R}^d$ and some matrix A. The gradient of ϕ at x is $(A + A^t)x$ and the Hessian matrix is $A + A^t$. Thus $T = A + A^t$ and since ϕ is convex, the latter must be positive semidefinite.

Viewing T as a matrix leads to the *Ricatti equation* $TAT = B$ (since T is symmetric). This is a quadratic equation in T, and so we wish to take square roots in a way that would isolate T. This is done by multiplying the equation from both sides with $A^{1/2}$:

$$[A^{1/2}TA^{1/2}][A^{1/2}TA^{1/2}] = A^{1/2}TATA^{1/2} = A^{1/2}BA^{1/2} = [A^{1/2}B^{1/2}][B^{1/2}A^{1/2}].$$

Both sides are clearly positive semidefinite, and furthermore $A^{1/2}TA^{1/2}$ is positive semidefinite. By taking square roots and multiplying with $A^{-1/2}$ we finally find

$$T = A^{-1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{-1/2}.$$

A straightforward calculation shows that $TAT = B$ indeed, and T is positive definite, hence optimal. To calculate the transportation cost $C(T)$, observe that $(T - I)\#\mu$ is a centred Gaussian measure with covariance matrix

$$TAT - TA - AT + A = A + B - A^{1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{-1/2} - A^{-1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{1/2}.$$

If $Y \sim \mathcal{N}(0, C)$, then $\mathbb{E}\|Y\|^2$ equals the trace of C, denoted $\mathrm{tr}C$. Hence, by properties of the trace,

$$C(T) = \mathrm{tr}\left[A + B - 2(A^{1/2}BA^{1/2})^{1/2}\right]. \tag{2.6}$$

If $AB = BA$, the above formulae simplify to

$$T = B^{1/2}A^{-1/2}, \qquad C(T) = \mathrm{tr}\left[A + B - 2A^{1/2}B^{1/2}\right].$$

By continuity arguments, (2.6) is the total transportation cost between any two Gaussian distributions with zero means, even if A is singular.

If the means of μ and ν are m and n, one simply needs to translate the measures. The optimal map and the total cost are then

$$Tx = n - m + A^{-1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{-1/2}x; \qquad C(T) = \|n-m\|^2 + \mathrm{tr}[A+B-2(A^{1/2}BA^{1/2})^{1/2}].$$

From this we can deduce a lower bound on the total cost between *any* two measures in $\mathbb{R}^d$ in terms of their second order structure. This is worth mentioning, because such lower bounds are not very common. Once again, by continuity considerations this holds for arbitrary measures with possibly singular covariance matrices.

Proposition 2.7.1 (lower bound for quadratic cost)**.** *Let $\mu, \nu \in P(\mathbb{R}^d)$ be absolutely continuous measures with means m and n and covariance matrices A and B and let T be the optimal map. Then*

$$C(T) \geq \|n-m\|^2 + \mathrm{tr}[A+B-2(A^{1/2}BA^{1/2})^{1/2}].$$

Proof. It will be convenient here to use the probabilistic terminology of Section 2.2. Let X and Y be random variables with distributions μ and ν. Any coupling of X and Y will have covariance matrix of the form $C = \begin{pmatrix} A & V \\ V^t & B \end{pmatrix} \in \mathbb{R}^{2d\times 2d}$ for some matrix $V \in \mathbb{R}^{d\times d}$, constrained so that C is positive semidefinite [29]. This gives the lower bound

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}_\pi \|X-Y\|^2 = \|m-n\|^2 + \inf_{\pi\in\Pi(\mu,\nu)} \mathrm{tr}_\pi[A+B-2V] \geq \|m-n\|^2 + \inf_{V:C\geq 0} \mathrm{tr}[A+B-2V].$$

As we know from the Gaussian case, the last infimum is given by (2.6). □

2.8 Regularity of the transport maps

In the preceding two sections we have seen an explicit formula for the optimal transport map T between μ and ν. In the Gaussian case in $\mathbb{R}^d$, this map is linear, so it is of course very smooth (analytic). The densities of Gaussian measures are analytic too, so we see that T inherits the regularity of μ and ν. Using the formula for T, one can show that a similar phenomenon takes place in the one-dimensional case. Though we do not have a formula for T at our disposal when μ and ν are general absolutely continuous measures on $\mathbb{R}^d$, $d \geq 2$, it turns out that even in that case, T inherits the regularity of μ and ν if some convexity conditions are satisfied.

Let us first show a precise result in the case $d = 1$. Let F and G denote the distribution functions of μ and ν respectively. Suppose that G is continuously differentiable and that $G' > 0$ on some open interval (finite or not) I such that $\nu(I) = 1$. Then the inverse function theorem says that G^{-1} is also continuously differentiable. Recall that the **support** of a (Borel) probability measure μ (denoted $\mathrm{supp}\mu$) is the smallest closed set K such that $\mu(K) = 1$. Throughout this section, we will deal exclusively with the quadratic cost $c(x,y) = \|x-y\|^2/2$ on $\mathbb{R}^d$. Then, we have the following result:

Theorem 2.8.1 (regularity in $\mathbb{R}$)**.** *Let $\mu, \nu \in P(\mathbb{R})$ possess distribution functions F and G of class C^k, $k \geq 1$. Suppose further that* suppν *is an interval I (possibly unbounded) and that $G' > 0$ on (the interior of) I. Then the optimal map is of class C^k as well.*

Remark 2. *The result also holds if $k = 0$.*

Proof. The optimal map is $G^{-1} \circ F$ by Theorem 2.6.1, and the discussion in the preceding paragraph proves the result when $k = 1$, since we have a composition of C^1 functions. When $k = 2$, we let $H = G^{-1}$ and use the formula $H'(t) = 1/G'(H(t))$ for all $t \in (0,1)$. Then both G' and H are C^1, so that H' is C^1, and consequently H is C^2. By induction we see that if G is C^k, then so is H. If in addition F is C^k, then $T = G^{-1} \circ F$ is C^k.

For the case $k = 0$, observe that G is strictly increasing, because suppν is an interval. Since G is assumed continuous, so is $H = G^{-1}$, so that $T = H \circ F$ must be continuous too. □

The assumption on the support of ν is important: if μ is Lebesgue measure on $[0,1]$ and the support of ν is disconnected, then T cannot even be continuous, no matter how smooth ν is!

The argument above cannot be easily extended to measures on $\mathbb{R}^d$, $d \geq 2$, because there is no explicit formula available for the optimal maps. As before, we cannot expect the optimal map to be continuous if the support of ν is disconnected. It turns out that the right condition on the support of ν is not connectedness, but rather convexity. This was shown by Caffarelli, who was able to prove ([23] and the references within) the following regularity result.

Theorem 2.8.2 (regularity of transport maps)**.** *Fix open sets $\Omega_1, \Omega_2 \subseteq \mathbb{R}^d$ and absolutely continuous measures $\mu, \nu \in P(\mathbb{R}^d)$ with finite second moments and bounded densities f, g respectively, such that $\mu(\Omega_1) = 1 = \nu(\Omega_2)$. Suppose that Ω_2 is convex and that $f, g \in C^{k,\alpha}$ (their k-th derivatives are Hölder continuous of exponent $\alpha \in (0,1)$), $k \geq 0$. If either*

1. *both Ω_1 and Ω_2 are bounded and f, g are bounded below; or*
2. *both $\Omega_1 = \Omega_2 = \mathbb{R}^d$ and f and g are strictly positive,*

then the convex potential ϕ such that $\nabla\phi\#\mu = \nu$ satisfies $\phi \in C^{k+2,\alpha}$ on Ω_1.

If the first of these conditions hold then ϕ is in addition strictly convex.

One can find a statement of this result (without proof) in this version in Villani [88, Theorem 4.14]. Theorem 2.8.2 will be used in two ways in this thesis. Firstly, it is used to derive criteria for a Karcher mean to be the Fréchet mean (Theorem 3.5.18). Secondly, it allows one to obtain very smooth estimates for the transport maps. Indeed, any two measures μ and ν can be approximated by measures satisfying the second condition: one can approximate them by discrete measures using the law of large numbers and then employ a convolution with e.g. a Gaussian measure (see for instance Theorem 3.2.6). It is not at all obvious that the transport maps between the approximations converge to the transport maps between the original measures, and we will show this in the next section.

2.9 Stability of solutions under narrow convergence

In this section we discuss the behaviour of the solution to the Monge–Kantorovich problem when the measures μ and ν are replaced by approximations μ_n and ν_n. Since any measure can be approximated by discrete measures *or* by smooth measures, this allows us to benefit from both worlds. On one hand, approximating μ and ν with discrete measures leads to the finite discrete problem of Section 2.3 that can be solved exactly. On the other hand, approximating μ and ν with Gaussian convolutions thereof leads to very smooth measures (at least in $\mathbb{R}^d$), and so the regularity results of the previous section imply that the respective optimal maps will also be smooth. Finally, in applications, one would almost always observe the measures of interest μ and ν with a certain amount of noise, and it is therefore of interest to control the error introduced by the noise. In image analysis, μ can represent an image that has as undergone blurring, or some other perturbation (Amit, Grenander & Piccioni [8]). In other applications the noise could be due to sampling variation, where instead of μ one observes a discrete measure μ_N obtained from realisations $X_1, \dots, X_N$ of random elements with distribution μ as $\mu_N = N^{-1} \sum_{i=1}^{N} \delta\{X_i\}$ (see Chapter 4).

In Subsection 2.9.1 we show that the optimal transference plan π depends continuously on μ and ν. With this result under our belt, we then deduce an analogous property for the optimal map T from μ to ν given some regularity of μ, in Subsection 2.9.2.

We shall assume throughout this section that $\mu_n \to \mu$ and $\nu_n \to \nu$ narrowly, which, we recall, means that $\int_{\mathcal{X}} f \, \mathrm{d}\mu_n \to \int_{\mathcal{X}} f \, \mathrm{d}\mu$ for all continuous bounded $f : \mathcal{X} \to \mathbb{R}$. The collection of these functions is denoted by $C_b(\mathcal{X})$. The following equivalent conditions for narrow convergence will be used not only in this section, but in other parts of this work as well.

Lemma 2.9.1 (portmanteau)**.** *Let $\mathcal{X}$ be a complete separable metric space and let $\mu, \mu_n \in P(\mathcal{X})$. Then the following are equivalent:*

- *$\mu_n \to \mu$ narrowly;*
- *$F_n(x) \to F(x)$ for any continuity point x of F. Here $\mathcal{X} = \mathbb{R}^d$, F_n is the distribution function of μ_n and F is that of μ;*
- *for any open $G \subseteq \mathcal{X}$, $\liminf \mu_n(G) \geq \mu(G)$;*
- *for any closed $F \subseteq \mathcal{X}$, $\limsup \mu_n(F) \leq \mu(F)$;*
- *$\int h \, \mathrm{d}\mu_n \to \int h \, \mathrm{d}\mu$ for any bounded measurable h whose set of discontinuity points is a μ-null set.*

For a proof, see for instance Billingsley [17, Theorem 2.1]. The equivalence with the last condition can be found in Pollard [72, Section III.2].

2.9.1 Stability of transference plans and c-monotonicity

Here is a precise stability result, due to Schachermayer & Teichmann [84, Theorem 3]. As usual, we assume that $\mathscr{X}$ is a complete separable metric space.

Theorem 2.9.2 (narrow convergence and optimal plans)**.** *Let μ_n and ν_n converge narrowly to μ and ν respectively in $P(\mathscr{X})$ and let $c : \mathscr{X}^2 \to \mathbb{R}_+$ be continuous. If $\pi_n \in \Pi(\mu_n, \nu_n)$ are optimal transference plans and*

$$\limsup_{n\to\infty} \int_{\mathscr{X}^2} c(x,y)\,\mathrm{d}\pi_n(x,y) < \infty.$$

then (π_n) is a tight sequence and each of its narrow limits $\pi \in \Pi(\mu,\nu)$ is optimal.

One can even let c vary with n under some conditions, see Villani [89, Theorem 5.20].

A key idea in the proof of this result is to replace optimality of π with another property called **c-monotonicity**, which behaves nicely with respect to narrow convergence. To elucidate the importance of this property, we recall the discrete case of Section 2.3 where $\mu = N^{-1}\sum_{i=1}^N \delta\{x_i\}$ and $\nu = N^{-1}\sum_{i=1}^N \delta\{y_i\}$. There exists an optimal transference plan π induced from a permutation $\sigma_0 \in S_N$. Since the ordering of $\{x_i\}$ and $\{y_i\}$ is irrelevant in the representations of μ and ν, we may assume without loss of generality that σ_0 is the identity permutation. Then, by definition of optimality,

$$\sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{\sigma(i)}), \qquad \sigma \in S_N. \tag{2.7}$$

If σ is the identity except for a subset $i_1, \dots, i_n$, $n \le N$, then in particular

$$\sum_{k=1}^n c(x_{i_k}, y_{i_k}) \le \sum_{k=1}^n c(x_{i_k}, y_{i_{\sigma(k)}}), \qquad \sigma \in S_n,$$

and if we choose $\sigma(i_k) = i_{k-1}$ with $i_0 = i_n$, this writes

$$\sum_{k=1}^n c(x_{i_k}, y_{i_k}) \le \sum_{k=1}^n c(x_{i_k}, y_{i_{k-1}}). \tag{2.8}$$

By decomposing a permutation $\sigma \in S_N$ to disjoint cycles, one can verify that (2.8) implies (2.7). This will be useful since, as it turns out, a variant of (2.8) holds for arbitrary measures μ and ν for which there is no relevant finite N as in (2.7).

Definition 2.9.3 (c-monotone sets and measures)**.** *A set $\Gamma \subseteq \mathscr{X}^2$ is c-monotone if for any n and any $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$,*

$$\sum_{i=1}^n c(x_i, y_i) \le \sum_{i=1}^n c(x_i, y_{i-1}), \qquad (y_0 = y_n). \tag{2.9}$$

A probability measure π on $\mathscr{X}^2$ is c-monotone if there exists a c-monotone Borel set Γ such that

$\pi(\Gamma) = 1$.

The relevance of c-monotonicity becomes clear from the following observation. If μ and ν are discrete measures and σ is an optimal permutation for the Monge–Kantorovich problem, then the coupling $\pi = (1/N)\sum_{i=1}^{N} \delta\{(x_i, y_{\sigma(i)})\}$ is c-monotone. In fact, even if the optimal permutation is not unique, the set

$$\Gamma = \{(x_i, y_{\sigma(i)}) : i = 1, \dots, N, \sigma \in S_N \text{ optimal}\}$$

is c-monotone. Furthermore, $\pi \in \Pi(\mu, \nu)$ is optimal if and only if it is c-monotone, if and only if $\pi(S) = 1$. The following proposition extends the "only if" to arbitrary measures, when c is continuous. It is due to Gangbo and McCann [37, Theorem 2.3].

Proposition 2.9.4 (optimal plans are c-monotone)**.** *Let $\mu, \nu \in P(\mathcal{X})$ and suppose that the cost function c is nonnegative and continuous. Assume that the optimal $\pi \in \Pi(\mu, \nu)$ has a finite total cost. Then* suppπ *is c-monotone. In particular, π is c-monotone.*

The idea of the proof is that if for some $(x_1, y_1), \dots, (x_n, y_n)$ in the support of π,

$$\sum_{i=1}^{n} c(x_i, y_i) > \sum_{i=1}^{n} c(x_i, y_{i-1}),$$

then by continuity of c, the same inequality holds on some balls of positive measure. One can then replace π by a measure having (x_i, y_{i-1}) rather than (x_i, y_i) in its support, and this measure will incur a lower cost than π.

Thus, we see that optimal transference plans π solve infinitely many discrete Monge–Kantorovich problems emanating from their support. More precisely, for any finite collection (x_i, y_i), $i = 1, \dots, N$ and any permutation $\sigma \in S_N$, (2.7) is satisfied. Therefore the identity permutation is optimal between the measures $(1/N)\sum \delta\{x_i\}$ and $(1/N)\sum \delta\{y_j\}$.

It is not difficult to strengthen Proposition 2.9.4 and prove existence of a c-monotone set Γ that includes the support of *any* optimal transference plan π: take $\Gamma = \cup \text{supp}(\pi)$ for π optimal.

A major contribution of Schachermayer & Teichmann [84] was to prove the converse of Proposition 2.9.4.

Proposition 2.9.5 (c-monotone plans are optimal)**.** *Let $\mu, \nu \in P(\mathcal{X})$, $c : \mathcal{X}^2 \to \mathbb{R}_+$ continuous and $\pi \in \Pi(\mu, \nu)$ a c-monotone measure with $C(\pi)$ finite. Then π is optimal in $\Pi(\mu, \nu)$.*

Given these results, it is now instructive to prove Theorem 2.9.2.

Proof of Theorem 2.9.2. Since $\mu_n \to \mu$ narrowly, it is a tight sequence, and similarly for ν_n. Consequently, the entire set of plans $\cup_n \Pi(\mu_n, \nu_n)$ is tight too (see the discussion before deriving

(2.3)). Therefore, up to a subsequence, (π_n) has a narrow limit π. We need to show that π is c-monotone and that $C(\pi)$ is finite. The latter is easy, since

$$C(\pi) = \lim_{M\to\infty} \int_{\mathscr{X}^2} \min(c, M)\, \mathrm{d}\pi = \lim_{M\to\infty} \lim_{n\to\infty} \int_{\mathscr{X}^2} \min(c, M)\, \mathrm{d}\pi_n \leq \liminf_{n\to\infty} \int_{\mathscr{X}^2} c\, \mathrm{d}\pi_n < \infty.$$

To show that π is c-monotone, we fix $(x_1, y_1), \ldots, (x_N, y_N) \in \mathrm{supp}\pi$. Let us show that there exist $(x_k^n, y_k^n) \in \mathrm{supp}\pi_n$ that converge to (x_k, y_k). Once this is established, we conclude from the c-monotonicity of $\mathrm{supp}\pi_n$ and the continuity of c that

$$\sum_{k=1}^N c(x_k, y_k) = \lim_{n\to\infty} \sum_{k=1}^N c(x_k^n, y_k^n) \leq \lim_{n\to\infty} \sum_{k=1}^N c(x_k^n, y_{k-1}^n) = \sum_{k=1}^N c(x_k, y_{k-1}).$$

The existence proof for the sequence is standard. For all $\epsilon > 0$ let $B = B_\epsilon(x_k, y_k)$ be an open ball around (x_k, y_k). Then $\pi(B) > 0$ and by the portmanteau lemma 2.9.1, $\pi_n(B) > 0$ for sufficiently large n. It follows that there exist $(x_k^n, y_k^n) \in B \cap \mathrm{supp}\pi_n$. We can let $\epsilon = 1/m$, say, then for all $n \geq N_m$ we can find $(x_k^n, y_k^n) \in \mathrm{supp}\mu_n$ of distance $1/m$. We can choose $N_{m+1} > N_m$ without loss of generality in order to complete the proof. □

It should not come as a surprise that c-monotonicity takes a special form in the quadratic case. Indeed, when $\mathscr{X}$ is a separable Hilbert space and $c(x, y) = \|x - y\|^2 = (x - y)^2$, a c-monotone set is called **cyclically monotone**. Easy algebra shows that (2.9) is then equivalent to

$$\sum_{i=1}^n \langle y_i, x_{i+1} - x_i \rangle \leq 0, \qquad (x_{n+1} = x_1). \tag{2.10}$$

Recall that if $\phi : \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ is convex, then its subdifferential at x is the set of subgradients

$$\partial\phi(x) = \left\{ y \in \mathbb{R}^d : \phi(z) \geq \phi(x) + \langle y, z - x \rangle \text{ for any } z \in \mathscr{X} \right\}.$$

Suppose that $y_i \in \partial\phi(x_i)$ for all i. Summing up the subgradient inequalities at $z_i = x_{i+1}$ yields precisely (2.10). In other words, subdifferentials of convex functions are cyclically monotone. Rockafellar [77] showed that this is in fact a characterisation of cyclical monotonicity.

Theorem 2.9.6 (Rockafellar)**.** *A nonempty $\Gamma \subseteq \mathscr{X}^2$ is cyclically monotone if and only if it is included in the graph of the subdifferential of a lower semicontinuous convex function that is not identically infinite.*

The proof is constructive: given Γ, one fixes $(x_0, y_0) \in \Gamma$ and defines

$$\phi(x) = \sup \left\{ \langle y_0, x_1 - x_0 \rangle + \cdots + \langle y_{m-1}, x_m - x_{m-1} \rangle + \langle y_m, x - x_m \rangle : m \in \mathbb{N}, \quad (x_i, y_i) \in \Gamma \right\},$$

which is convex and lower semicontinuous (as a supremum of affine functions), and using the cyclical monotonicity equals 0 (hence not ∞) at x_0.

Importantly, the Kantorovich duality is not needed, and can in fact be *deduced* from these arguments. For instance, if π is optimal for (halved) quadratic cost, then its support is cyclically monotone. As such, it is included in the subgradient of a convex function ϕ. One can then verify that $(\|x\|^2/2 - \phi(x), \|y\|^2/2 - \phi^*(y))$ is optimal for the dual problem. These ideas can be extended to other cost functions: given a c-monotone set Γ, the potential can be constructed as (Rüschendorf [80])

$$\varphi(x) = \inf\left\{c(x_1, y_0) - c(x_0, y_0) + c(x_m, y_{m-1}) - c(x_{m-1}, y_{m-1}) + c(x, y_m) - c(x_m, y_m)\right\},$$

and then (φ, φ^c) is the solution to the dual problem.

2.9.2 Stability of transport maps

In this subsection, following Zemel & Panaretos [94, Section 7.5], we extend the narrow convergence of π_n to π of the previous subsection to convergence of optimal maps. Because of the applications we have in mind, we shall work exclusively in the Euclidean space $\mathscr{X} = \mathbb{R}^d$ with the quadratic cost function; our results can most likely be extended to more general situations.

In this setting, we know that optimal plans are supported on graphs of subdifferentials of convex functions. Suppose that π_n is induced by T_n and π is induced by T. Then in some sense, the narrow convergence of π_n to π yields convergence of the graphs of T_n to the graph of T. Our goal is to strengthen this to uniform convergence of T_n to T. Roughly speaking, we show the following: there exists a set A with $\mu(A) = 1$ and such that T_n converge uniformly to T on every compact subset of A. For the reader's convenience we give a user-friendly version here; a more general statement is given in Proposition 2.9.11 below.

Theorem 2.9.7 (uniform convergence of optimal maps)**.** *Let μ_n, μ be absolutely continuous measures with finite second moments on an open convex set $U \subseteq \mathbb{R}^d$ such that $\mu_n \to \mu$ narrowly, and let $\nu_n \to \nu$ narrowly with $\nu_n, \nu \in P(\mathbb{R}^d)$ with finite second moments. If T_n and T are continuous on U and $C(T_n)$ is bounded uniformly in n, then*

$$\sup_{x \in \Omega} \|T_n(x) - T(x)\| \to 0, \qquad n \to \infty,$$

for any compact $\Omega \subseteq U$.

A weaker result can be found in Villani [89, Corollary 5.23]: T_n converge to T in μ-measure. This result, however, assumes that $\mu_n = \mu$ and only ν_n is allowed to vary with n ([89, Remark 5.25]). On the flip side, the result in [89, Corollary 5.23] holds in a very general setting.

Since T_n and T are only defined up to Lebesgue null sets, it will be more convenient to work directly with the subgradients. That is, we view T_n and T as *set-valued* functions that to each $x \in \mathbb{R}^d$ assign a (possibly empty) subset of $\mathbb{R}^d$. In other words, T_n and T take values in the *power set* of $\mathbb{R}^d$, denoted by $2^{\mathbb{R}^d}$.

Let $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be convex, $y_1 \in \partial\phi(x_1)$ and $y_2 \in \partial\phi(x_2)$. Putting $n = 2$ in the definition of cyclical monotonicity (2.10) gives

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geq 0.$$

This property (which is weaker than cyclical monotonicity) is important enough to have its own name. Following the notation of Alberti & Ambrosio [3], we call a set-valued function (or multifunction) $u : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ **monotone** if whenever $y_i \in u(x_i)$, $i = 1, 2$,

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geq 0.$$

If $d = 1$, this simply means that u is a nondecreasing (set-valued) function. For example, one can define $u(x) = \{0\}$ for $x \in [0, 1)$, $u(1) = [0, 1]$ and $u(x) = \emptyset$ if $x \notin [0, 1]$. Next, u is said to be **maximally monotone** if no points can be added to its graph while preserving monotonicity:

$$\{\langle y' - y, x' - x \rangle \geq 0 \quad \text{whenever } y \in u(x)\} \quad \Longrightarrow \quad y' \in u(x').$$

It will be convenient to identify u with its graph; we will often write $(x, y) \in u$ to mean $y \in u(x)$. Note that $u(x)$ can be empty, even when u is maximally monotone. The previous example for u is not maximally monotone, but it will be if we modify $u(0)$ to be $(-\infty, 0]$ and $u(1)$ to be $[0, \infty)$.

Of course, if $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is convex, then $u = \partial\phi$ is monotone. It follows from Theorem 2.9.6 that u is maximally cyclical monotone (no points can be added to its graph while preserving cyclical monotonicity). It is not immediate, but not very difficult to show that u is actually maximally monotone; see [3, Section 7]. In what follows we will always work with subdifferentials of convex functions, so unless stated otherwise, u will always be assumed maximally monotone.

Maximally monotone functions enjoy the following very useful continuity property. It is proven in [3, Corollary 1.3] and will be used extensively below.

Proposition 2.9.8 (continuity at singletons)**.** *Let $x \in \mathbb{R}^d$ such that $u(x) = \{y\}$ is a singleton. Then u is nonempty on some neighbourhood of x and it is continuous at x: if $x_n \to x$ and $y_n \in u(x_n)$, then $y_n \to y$.*

Notice that this result implies that if a convex function ϕ is differentiable on some open set $E \subseteq \mathbb{R}^d$, then it is continuously differentiable there (Rockafellar [78, Corollary 25.5.1]).

If $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is any function, one can define its subgradient at x locally as

$$\partial f(x) = \{y : f(z) \geq f(x) + \langle y, z - x \rangle + o(\|z - x\|)\} = \left\{ y : \liminf_{z \to x} \frac{f(z) - f(x) + \langle y, z - x \rangle}{\|z - x\|} \geq 0 \right\}.$$

(See the discussion after Theorem 2.5.3.) When f is convex, one can remove the $o(\|z - x\|)$ term and the inequality holds for all z, i.e. globally and not locally. Since monotonicity is

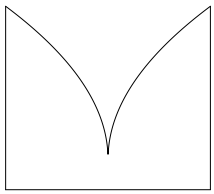


Figure 2.1: The set G in (2.11).

related to convexity, it should not be surprising that monotonicity is in some sense a local property. Suppose that $u(x_0) = \{y_0\}$ is a singleton and that for some $y^* \in \mathbb{R}^d$,

$$\langle y - y^*, x - x_0 \rangle \geq 0$$

for all $x \in \mathbb{R}^d$ and $y \in u(x)$. Then by maximality, y^* must equal y_0. By "local property" we mean that the conclusion $y^* = y_0$ holds if the above inequality holds for x in a small neighbourhood of x_0 (an open set that includes x_0). We will need a more general version of this result, replacing neighbourhoods by a weaker condition that can be related to Lebesgue points. The strengthening is somewhat technical; the reader can skip directly to Lemma 2.9.10 and assume that G is open without losing much intuition.

We remind the reader of the notation $B_r(x_0) = \{x : \|x - x_0\| < r\}$ for $r \geq 0$ and $x_0 \in \mathbb{R}^d$. The interior of a set $G \subseteq \mathbb{R}^d$ is denoted by int$G$ and the closure by $\overline{G}$. If G is measurable, then Leb$G$ denotes the Lebesgue measure of G. Finally, conv$G$ denotes the convex hull of G.

A point x_0 is a **Lebesgue point** (or of **Lebesgue density**) of a measurable set $G \subseteq \mathbb{R}^d$ if for any $\epsilon > 0$ there exists $t_\epsilon > 0$ such that

$$\frac{\text{Leb}(B_t(x_0) \cap G)}{\text{Leb}(B_t(x_0))} > 1 - \epsilon, \qquad 0 < t < t_\epsilon.$$

Here is an interesting example I learned from Tomáš Rubín. Define the set $G \subseteq \mathbb{R}^2$ by (see Figure 2.1)

$$G = \left\{(x, y) : |x| \leq 1,\ -0.2 \leq y \leq \sqrt{|x|}\right\}. \tag{2.11}$$

Then $(0,0)$ is a Lebesgue point of G (because the "slope" of the square root is infinite) but the fraction above is never one.

We denote the set of points of Lebesgue density of G by G^{den}. Here are some facts about G^{den}: clearly, int$G \subseteq G^{\text{den}} \subseteq \overline{G}$. Stein & Shakarchi [85, Chapter 3, Corollary 1.5] show that Leb$(G \setminus G^{\text{den}}) = 0$ (and Leb$(G^{\text{den}} \setminus G) = 0$, so G^{den} is very close to G). By the Hahn–Banach Theorem, $G^{\text{den}} \subseteq \text{int}(\text{conv}(G))$: indeed, if x is not in int(conv$G$) then there is a separating hyperplane between x and conv$G \supseteq G$, so the fraction above is at most $1/2$ for all $t > 0$.

The "density"-ness of Lebesgue points is materialised in the following classical result.

Lemma 2.9.9 (density points and distance)**.** *Let x_0 be a point of Lebesgue density of a measurable set $G \subseteq \mathbb{R}^d$. Then*

$$\delta(z) = \delta_G(z) = \inf_{x \in G} \|z - x\| = o(\|z - x_0\|), \qquad \text{as } z \to x_0.$$

Of course, this result holds for any $x_0 \in \overline{G}$ if the little o is replaced by big O, since δ is Lipschitz. When $x_0 \in \text{int}G$, this is trivial because δ vanishes on int$G$.

Proof. [85] give this as an exercise when $d = 1$; for completeness we provide a full proof.

For any $1 > \epsilon > 0$ there exists $0 < t_\epsilon$ such that for $t < t_\epsilon$,

$$\frac{\text{Leb}(B_t(x_0) \cap G)}{\text{Leb}(B_t(x_0))} > 1 - \epsilon^d.$$

Fix z such that $t = t(z) = \|z - x_0\| < t_\epsilon$. The intersection of $B_t(x_0)$ with $B_{2\epsilon t}(z)$ includes a ball of radius ϵt centred at $y = x_0 + (1-\epsilon)(z - x_0)$, so that

$$\frac{\text{Leb}(B_t(x_0) \cap B_{2\epsilon t}(z))}{\text{Leb}(B_t(x_0))} \geq \frac{\text{Leb}(B_{\epsilon t}(y))}{\text{Leb}(B_t(x_0))} = \epsilon^d.$$

It follows that $G \cap B_{2\epsilon t}(z)$ is nonempty. In other words: for any $\epsilon > 0$ there exists t_ϵ such that if $\|z - x_0\| < t_\epsilon$, then there exists $x \in G$ with $\|z - x\| \leq 2\epsilon\, t(z) = 2\epsilon \|z - x_0\|$. This means precisely that $\delta(z) = o(\|z - x_0\|)$ as $z \to x_0$. □

The important part here is the following corollary: for almost all $x \in G$, $\delta(z) = o(\|z - x\|)$ as $z \to x$. This can be seen in other ways: since δ is Lipschitz, it is differentiable almost everywhere. If $x \in \overline{G}$ and δ is differentiable at x, then $\nabla\delta(x)$ must be 0 (because δ is minimised there), and then $\delta(z) = o(\|z - x\|)$. We just showed that δ is differentiable with vanishing derivative at all Lebesgue points of x. The converse is not true: $G = \{\pm 1/n\}_{n=1}^{\infty}$ has no Lebesgue points, but $\delta(y) \leq 4y^2$ as $y \to 0$.

The locality of monotone functions can now be stated and proven as follows.

Lemma 2.9.10 (local monotonicity)**.** *Let $x_0 \in \mathbb{R}^d$ such that $u(x_0) = \{y_0\}$ and x_0 is a Lebesgue point of a set G satisfying*

$$\langle y - y^*, x - x_0 \rangle \geq 0 \qquad \forall x \in G\ \forall y \in u(x).$$

Then $y^* = y_0$. In particular, the result is true if the inequality holds on $G = O \setminus \mathcal{N}$ with $\emptyset \neq O$ open and $\mathcal{N}$ Lebesgue negligible.

Proof. Set $z_t = x_0 + t(y^* - y_0)$ for $t > 0$ small. It is possible that $z_t \notin G$; but Lemma 2.9.9

guarantees existence of $x_t \in G$ with $\|x_t - z_t\|/t \to 0$. By Proposition 2.9.8 $u(x_t)$ is nonempty for t small enough. For $y_t \in u(x_t)$,

$$\begin{aligned} 0 \le \langle y_t - y^*, x_t - x_0 \rangle &= \langle y_t - y^*, x_t - z_t \rangle + \langle y_t - y^*, z_t - x_0 \rangle \\ &= \langle y_t - y^*, x_t - z_t \rangle + t \langle y_t - y_0, y^* - y_0 \rangle - t\|y^* - y_0\|^2. \end{aligned}$$

It now follows from the Cauchy–Schwarz inequality that

$$\|y^* - y_0\|^2 \le \|y_t - y_0\| \|y^* - y_0\| + t^{-1}\|x_t - z_t\| \big(\|y_t - y_0\| + \|y^* - y_0\|\big).$$

As $t \searrow 0$ the right-hand side vanishes, since $y_t \to y_0$ (Proposition 2.9.8) and $\|x_t - z_t\|/t \to 0$. It follows that $y^* = y_0$. □

These continuity properties cannot be of much use unless $u(x)$ is a singleton for reasonably many values of x. Fortunately, this is indeed the case: the set of points x such that $u(x)$ contains more than one element has Lebesgue measure 0 (see Alberti & Ambrosio [3, Remark 2.3] for a stronger result). Another issue is that u may be empty, and convexity comes into play here again. Let $\mathrm{dom} u = \{x : u(x) \neq \emptyset\}$. Then there exists a convex closed set K such that

$$\mathrm{int} K \subseteq \mathrm{dom} u \subseteq K.$$

([3, Corollary 1.3(2)]). Although $\mathrm{dom} u$ itself may fail to be convex, it is almost convex in the above sense. By convexity, $K \setminus \mathrm{int} K$ has Lebesgue measure 0 (see the discussion after Theorem 2.5.1) and so the set of points in K where u is not a singleton,

$$\{x \in K : u(x) = \emptyset\} \cup \{x \in K : u(x) \text{ contains more than one point}\},$$

has Lebesgue measure 0, and $u(x)$ is empty for all $x \notin K$. (It is in fact not difficult to show that if $x \in \partial K$, then $u(x)$ cannot be a singleton, by the Hahn–Banach theorem.)

With this background on monotone functions at our disposal, we are now ready to state the stability result for the optimal maps. The following assumptions will be made unless stated otherwise.

Assumptions 1. *Let $\mu_n, \mu, \nu_n, \nu \in P(\mathbb{R}^d)$ with optimal couplings (with respect to quadratic cost) $\pi_n \in \Pi(\mu_n, \nu_n)$, $\pi \in \Pi(\mu, \nu)$ and convex potentials ϕ_n and ϕ respectively such that*

- *(*convergence*) $\mu_n \to \mu$ and $\nu_n \to \nu$ narrowly;*
- *(*finiteness*) the optimal couplings $\pi_n \in \Pi(\mu_n, \nu_n)$ satisfy*

$$\limsup_{n\to\infty} \int_{\mathcal{X}^2} \frac{1}{2}\|x - y\|^2 \, \mathrm{d}\pi_n(x, y) < \infty;$$

- *(*unique limit*) the optimal $\pi \in \Pi(\mu, \nu)$ is unique.*

We further denote the subgradients $\partial\phi_n$ and $\partial\phi$ by u_n and u respectively.

These assumptions imply that π has a finite total cost. This can be shown by the liminf argument in the proof of Theorem 2.9.2 but also from the uniqueness of π. As a corollary of the uniqueness of π, it follows that $\pi_n \to \pi$ narrowly; notice that this holds even if π_n is not unique for any n. We will now translate this narrow convergence to convergence of the maximal monotone maps u_n to u, in the following form.

Proposition 2.9.11 (uniform convergence of optimal maps)**.** *Let Assumptions 1 hold true and denote $E = \mathrm{supp}\mu$ and E^{den} the set of its Lebesgue points.*

Let Ω be a compact subset of E^{den} on which u is univalued (i.e. $u(x)$ is a singleton for all $x \in \Omega$). Then u_n converges to u uniformly on Ω: $u_n(x)$ is nonempty for all $x \in \Omega$ and all $n > N_\Omega$, and

$$\sup_{x\in\Omega} \sup_{y\in u_n(x)} \|y - u(x)\| \to 0, \qquad n \to \infty.$$

In particular, if u is univalued throughout $\mathrm{int}(E)$ *(so that $\phi \in C^1$ there), then uniform convergence holds for any compact $\Omega \subset \mathrm{int}(E)$.*

Corollary 2.9.12 (pointwise convergence μ-almost surely)**.** *If in addition μ is absolutely continuous then $u_n(x) \to u(x)$ μ-almost surely.*

Proof. We first claim that $E \subseteq \overline{\mathrm{dom}u}$. Indeed, for any $x \in E$ and any $\epsilon > 0$, the ball $B = B_\epsilon(x)$ has positive measure. Consequently, u cannot be empty on the entire ball, because otherwise $\mu(B) = \pi(B \times \mathbb{R}^d)$ would be 0. Since $\mathrm{dom}u$ is almost convex (see the discussion before Assumptions 1), this implies that actually $\mathrm{int}(\mathrm{conv}E) \subseteq \mathrm{dom}u$.

The rest is now easy: the set of points $x \in E$ for which $\Omega = \{x\}$ fails to satisfy the conditions of Proposition 2.9.11 is included in

$$(E \setminus E^{\mathrm{den}}) \cup \{x \in \mathrm{int}(\mathrm{conv}(E)) : u(x) \text{ contains more than one point}\},$$

which is μ-negligible because μ is absolutely continuous and both sets have Lebesgue measure 0. □

The remainder of this subsection is devoted to the proof of Proposition 2.9.11. This will be shown in two separate steps:

- if a sequence in the graph of u_n converges, then the limit is in the graph of u (Lemma 2.9.14);
- sequences in the graph of u_n are bounded if the domain is bounded (Proposition 2.9.16).

Each step will in turn be proven using an intermediate lemma.

Lemma 2.9.13 (points in the limit graph are limit points)**.** *Let $x_0 \in \mathrm{supp}\mu$ be such that $u(x_0) = \{y_0\}$ is a singleton. Then there exists a sequence $(x_n, y_n) \in u_n$ that converges to (x_0, y_0).*

Proof. This is essentially the same argument as used in the proof of Theorem 2.9.2. Invoking the continuity of u at x_0 (Proposition 2.9.8), for any k there exists $\delta = \delta_k > 0$ such that if $x \in B_\delta(x_0) = \{x : \|x - x_0\| < \delta\}$ then $u(x)$ is nonempty and if $y \in u(x)$, then $\|y - y_0\| < 1/k$. Assume without loss of generality that $\delta_k \to 0$, and set $B_k = B_{\delta_k}(x_0)$, $V_k = B_{1/k}(y_0)$. Then $u(B_k) \subseteq V_k$, so

$$\pi(B_k \times V_k) = \pi\{(x,y) : x \in B_k, y \in u(x) \cap V_k\} = \pi\{(x,y) : x \in B_k, y \in u(x)\} = \mu(B_k) > 0,$$

because B_k is a neighbourhood of $x_0 \in \mathrm{supp}(\mu)$. Since $B_k \times V_k$ is open, we have by the portmanteau lemma 2.9.1 that $\pi_n(B_k \times V_k) > 0$ for $n \geq N_k$. But π_n is concentrated on the graph of u_n, so when $n \geq N_k$ there exist $(x_n, y_n) \in u_n \cap [B_k \times V_k]$, so that $\|x_n - x_0\| < \delta_k$ and $\|y_n - y_0\| < 1/k$. This completes the proof. □

Lemma 2.9.14 (limit points are in the limit graph)**.** *Let x_0 be a Lebesgue point of $E = \mathrm{supp}\mu$ (for example $x_0 \in \mathrm{int}E$) such that $u(x_0) = \{y_0\}$ is a singleton. If a subsequence $(x_{n_k}, y_{n_k}) \in u_{n_k}$ converges to (x_0, y^*), then $y^* = y_0$.*

Proof. The set $\mathcal{N} \subseteq \mathbb{R}^d$ of points where u contains more than one element has Lebesgue measure zero. Moreover, there exists a neighbourhood V of x_0 on which u is nonempty (Proposition 2.9.8). It follows that x_0 is a Lebesgue point of $G = (E \cap V) \setminus \mathcal{N}$, and $u(x)$ has one and only element for every $x \in G$. Let us fix $(x, y) \in u$ with $x \in G$. Application of Lemma 2.9.13 to the sequence $\{u_{n_k}\}_{k=1}^\infty$ at x yields sequences $x'_{n_k} \to x$ and $y'_{n_k} \to y$ with $(x'_{n_k}, y'_{n_k}) \in u_{n_k}$. Consequently,

$$\left\langle y - y^*, x - x_0 \right\rangle = \lim_{l \to \infty} \left\langle y'_{n_k} - y_{n_k}, x'_{n_k} - x_{n_k} \right\rangle \geq 0, \qquad \forall x \in G \; \forall y \in u(x).$$

It now follows from Lemma 2.9.10 that $y^* = y_0$. □

We now know that if $u_n(x)$ converges and $u(x) = \{y\}$, then $u_n(x) \to y$. It therefore suffices to show that $u_n(x)$ remains in a bounded set. To this end we shall use another result about monotone functions: if x is in the convex hull of $x_1, \ldots, x_m$, $y_i \in u(x_i)$, and $y \in u(x)$, then $\|y\|$ can be bounded in terms of $\|y_i\|$ and the distance of x from the boundary of $\mathrm{conv}(x_1, \ldots, x_m)$. It will be convenient to introduce the ℓ_∞ balls $B_\epsilon^\infty(x_0) = \{x : \|x - x_0\|_\infty < \epsilon\}$ and their closures $\overline{B}_\epsilon^\infty(x_0)$, because unlike the ℓ_2 balls, ℓ_∞ balls are polytopes and equal the convex hull of their finitely many vertices. (For that purpose, we could have also chosen ℓ_1 balls.)

We will need the following easy result about ℓ_∞ balls: let $Z = \{z_1, \ldots, z_m\}$, $m = 2^d$ be a collection of vectors with the following property: for each collection $(e_1, \ldots, e_d) \in \{\pm 1\}^d$ there exists a vector $y \in Z$ such that $|y_j| > 1$ and $y_j e_j > 0$ for all $j = 1, \ldots, d$. Then $\mathrm{conv}Z \supseteq \overline{B}_1^\infty(0)$. In

geometric terms this means that if we have 2^d points that are "more extreme“ than the vertices of the unit ℓ_∞ ball around zero, then the convex hull of these points includes this ℓ_∞ ball.

The proof of this result is a straightforward consequence of the Hahn–Banach theorem. We show that $e = (e_1, \dots, e_d)$ cannot be separated from Z with a hyperplane for any $e_j \in \{\pm 1\}$. Indeed, let $x \in \mathbb{R}^d \setminus \{0\}$ be any vector and set $J = \{j : e_j x_j > 0\}$. Pick $w, y \in Z$ such that $w_j e_j > 0$ if and only if $y_j e_j < 0$ if and only if $j \in J$. Since $|w_j| > 1$ and $|y_j| > 1$ this gives $x_j y_j < x_j e_j < x_j w_j$ whenever $x_j \neq 0$ and since $x \neq 0$,

$$\langle x, y \rangle < \langle x, e \rangle < \langle x, w \rangle .$$

Lemma 2.9.15 (continuity of convex hulls)**.** *Let $Z = \{z_i\}_{i \in I} \subseteq \mathbb{R}^d$ be an arbitrary collection of points and let $\tilde{Z} = \{\tilde{z}_i\}_{i \in I}$ be another collection such that $\|\tilde{z}_i - z_i\|_\infty \leq \epsilon$ for all $i \in I$. If* $\mathrm{conv} Z \supseteq B_\rho^\infty(x_0)$*, then* $\mathrm{conv}\tilde{Z} \supseteq B_{\rho-\epsilon}^\infty(x_0)$.

Proof. Without loss of generality $\epsilon < \rho$. Fix $\epsilon < \rho' < \rho$. Each vertex of $\overline{B}_{\rho'}^\infty(x_0)$ takes the form

$$y = x_0 + \rho'(e_1, \dots, e_d), \qquad e_d \in \{\pm 1\},$$

and can be written as a (finite) convex combination $y = \sum a_i z_i$ with $z_i \in Z$. If we define $\tilde{y} = \sum a_i \tilde{z}_i \in \mathrm{conv}\tilde{Z}$, then $\|\tilde{y} - y\|_\infty \leq \epsilon$. It follows that $\tilde{y}$ is “more extreme" than the vertex

$$x = x_0 + (\rho' - \epsilon)(e_1, \dots, e_d)$$

of the ℓ_∞-ball $B_{\rho'-\epsilon}^\infty(x_0)$, in the sense that $y_j - x_0$ has a larger absolute value than $x_j - x_0$ but the same sign for all $j = 1, \dots, d$. For each of the 2^d vertices we can find a corresponding $\tilde{y}$, and consequently $\mathrm{conv}\tilde{Z} \supseteq B_{\rho'-\epsilon}^\infty(x_0)$ by the discussion before the lemma. Since $\rho' < \rho$ was arbitrary this completes the proof. □

Proposition 2.9.16 (boundedness)**.** *Let $\Omega \subseteq \mathrm{int}(\mathrm{conv}(\mathrm{supp}(\mu)))$ be compact. Then there exist $N(\Omega)$ and a constant $R(\Omega)$ such that for all $n > N(\Omega)$, $u_n(x)$ is nonempty for all $x \in \Omega$ and* $\sup_{x \in \Omega} \sup_{y \in u_n(x)} \|y\| \leq R(\Omega)$ *is bounded uniformly.*

Proof. If we set $E = \mathrm{supp}(\mu)$ and $F = \mathrm{conv}(E)$, then Ω is a compact subset of the open set $\mathrm{int}F$. Consequently, there exists $\delta = \delta(\Omega) > 0$ such that $\overline{B}_{3\delta}^\infty(\Omega) \subseteq \mathrm{int}F$. We may construct a finite collection $\{\omega_j\} \subseteq \Omega$ such that the union of $B_\delta^\infty(\omega_j)$ includes Ω. Since each vertex of $\cup_j \overline{B}_{3\delta}^\infty(\omega_j)$ is in F, it can be written as a convex combination of elements of E. Consequently, there exists a finite set $Z = \{z_1, \dots, z_m\} \subseteq E$ with $\mathrm{conv} Z \supseteq B_{3\delta}^\infty(\omega_j)$ for any j.

The ball $B_i = B_\delta^\infty(z_i)$ is an open neighbourhood of an element of $\mathrm{supp}\mu$ and therefore has positive measure, say $2\epsilon_i > 0$. By the portmanteau lemma 2.9.1 $\mu_n(B_i) > \epsilon_i$ for all n large and all $i = 1, \dots, m$. We can set $\epsilon = \min_i \epsilon_i > 0$ and invoke the tightness of $\{\nu_n\}$ to find a compact set K_ϵ with $\inf_n \nu_n(K_\epsilon) > 1 - \epsilon$. A simple calculation shows that this construction guarantees the

existence of $x_{ni} \in B_i$ and $y_{ni} \in u_n(x_{ni})$ such that $y_{ni} \in K_\epsilon$. Setting

$$\tilde{Z} = X_n = \{x_{n1}, \dots, x_{nm}\},$$

noticing that by definition $\|x_{ni} - z_i\|_\infty \le \delta$ and applying Lemma 2.9.15, we obtain

$$\mathrm{conv} X_n = \mathrm{conv}(\{x_{n1}, \dots, x_{nm}\}) \supseteq B^\infty_{3\delta-\delta}(\omega_j) = B^\infty_{2\delta}(\omega_j) \qquad \text{for all } j.$$

Recall that $B^\infty_\delta(\omega_j)$ cover Ω. From this it follows that $\mathrm{conv} X_n \supseteq B^\infty_\delta(\Omega) \supseteq B_\delta(\Omega)$ (since $\|x\| \ge \|x\|_\infty$, ℓ_2-balls are always included in ℓ_∞-balls of the same radius).

We are now in a position to employ the property of monotonicity mentioned above. From [3, Lemma 1.2(4)] we conclude that for any $\omega \in \Omega$ and any $y_0 \in u_n(\omega)$,

$$\|y_0\| \le \frac{[\sup_{x,z\in X_n} \|x - z\|][\max_{x\in X_n} \inf_{y\in u_n(x)} \|y\|]}{d(\omega, \mathbb{R}^d \setminus \mathrm{conv}(X_n))} \le \frac{1}{\delta} \left[\sup_{k,l} \|x_{nk} - x_{nl}\|\right] \left[\max_i \inf_{y\in u_n(x_{ni})} \|y\|\right].$$

To bound the infimum at the right-hand side, we can take y to be y_{ni}, which all lie in the compact set K_ϵ. To bound the supremum independently of n, we use the approximation $\|x_{nk} - z_k\| \le \sqrt{d}\|x_{nk} - z_k\|_\infty \le \sqrt{d}\delta$, so that $\|x_{nk} - x_{nl}\| \le 2\sqrt{d}\delta + \|z_k - z_l\|$. Hence

$$\forall n > N(\delta)\ \forall \omega \in \Omega \quad \forall y_0 \in u_n(\omega): \qquad \|y_0\| \le \frac{1}{\delta}\left(2\sqrt{d}\delta + \max_{k,l} \|z_k - z_l\|\right) \sup_{y\in K_\epsilon} \|y\|.$$

Recall that δ depends only on Ω, ϵ and Z only on δ, and K_ϵ only on ϵ, so $N(\delta) = N(\delta(\Omega))$ and the bound at the right-hand side does not depend on n.

Finally, the fact that u_n is not empty on Ω is a consequence of the almost convexity of $\mathrm{dom} u$ ([3, Corollary 1.3(2)]). □

Proof of Proposition 2.9.11. After all the hard work, the proof is now straightforward.

There exists N_Ω such that for all $n > N_\Omega$, $u_n(x)$ is nonempty and (Proposition 2.9.16)

$$\sup_{x\in\Omega} \sup_{y\in u_n(x)} \|y\| \le C_{\Omega,d} < \infty, \qquad n > N_\Omega,$$

where $C_{\Omega,d}$ is a constant that depends only on Ω (and the dimension d).

If uniform convergence did not hold, then one could find $\epsilon > 0$ and subsequences $(x_{n_k}, y_{n_k}) \in u_{n_k}$ with $x_{n_k} \in \Omega$ and

$$\|y_{n_k} - u(x_{n_k})\| > \epsilon, \qquad k = 1, 2, \dots.$$

Since the x_{n_k}'s are bounded (in Ω) and the y_{n_k}'s are bounded too, they have subsequences that converge to $x \in \Omega$ and some y, that must equal $u(x)$ by Lemma 2.9.14. Using again the

continuity of u at x (Proposition 2.9.8), we get (up to subsequences)

$$\epsilon < \|y_{n_k} - u(x_{n_k})\| \leq \|y_{n_k} - y\| + \|y - u(x)\| + \|u(x) - u(x_{n_k})\| \to 0, \qquad k \to \infty,$$

a contradiction. □

3 The Wasserstein space

The Kantorovich problem described in the previous chapter gives rise to a metric structure, the *Wasserstein distance*, in the space of probability measure $P(\mathscr{X})$ on a space $\mathscr{X}$. The resulting metric space, a subspace of $P(\mathscr{X})$, is commonly known as the *Wasserstein space* $\mathscr{W}$ (although, as Villani [89, bibliographical notes of Chapter 6] puts it, "this terminology is very questionable"; see also Bobkov & Ledoux [18, p. 4]). In the next chapter we shall see that this metric is in a sense canonical when dealing with warpings, that is, deformations of the space $\mathscr{X}$ (for example in Theorem 4.2.4). In this chapter we give the fundamental properties of the Wasserstein space. After some basic definitions, we describe the topological properties of that space in Section 3.2. It is then explained in Section 3.3 how $\mathscr{W}$ can be endowed with a sort of infinite-dimensional Riemannian structure. As we will consider random measures in this Wasserstein space, it will be necessary to deal with measurability issues; this is the purpose of the somewhat technical Section 3.4. Finally, the important concept of Fréchet mean is discussed in detail in Section 3.5 in the context of the Wasserstein space, both at the empirical and the population levels.

3.1 Definition, notation and basic properties

Let $\mathscr{X}$ be a separable Banach space. The p-**Wasserstein space** on $\mathscr{X}$ is defined and denoted by

$$\mathscr{W}_p(\mathscr{X}) = \left\{ \mu \in P(\mathscr{X}) : \int_{\mathscr{X}} \|x\|^p \, \mathrm{d}\mu(x) < \infty \right\}, \qquad p \geq 1.$$

We will sometimes abbreviate and write simply $\mathscr{W}_p$ instead of $\mathscr{W}_p(\mathscr{X})$.

Recall that if $\mu, \nu \in P(\mathscr{X})$, then $\Pi(\mu, \nu)$ is defined to be the set of measures $\pi \in P(\mathscr{X}^2)$ having μ and ν as marginals in the sense of (2.2). The p-**Wasserstein distance** between μ and ν is defined as the minimal total transportation cost between μ and ν in the Kantorovich problem

with respect to the cost function $c_p(x,y) = \|x-y\|^p$:

$$W_p(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} C_p(\pi)\right)^{1/p} = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathscr{X}\times\mathscr{X}} \|x_1-x_2\|^p \,\mathrm{d}\pi(x_1,x_2)\right)^{1/p}.$$

The Wasserstein distance between μ and ν is finite when both measures are in $\mathscr{W}_p(\mathscr{X})$, because

$$\|x_1-x_2\|^p \le 2^p\|x_1\|^p + 2^p\|x_2\|^p.$$

Thus W_p is finite on $[\mathscr{W}_p(\mathscr{X})]^2$; it is clearly nonnegative and symmetric and it is easy to see that $W_p(\mu,\nu) = 0$ if and only if $\mu = \nu$. A proof that W_p is a metric (satisfies the triangle inequality) on $\mathscr{W}_p$ can be found in Villani [88, Chapter 7].

The aforementioned setting is by no means the most general one can consider. Firstly, one can define W_p and $\mathscr{W}_p$ for $0 < p < 1$ by removing the power $1/p$ from the infimum and the limit case $p = 0$ yields the total variation distance. Another limit case can be defined as $W_\infty(\mu,\nu) = \lim_{p\to\infty} W_p(\mu,\nu)$. Moreover, W_p and $\mathscr{W}_p$ can be defined whenever $\mathscr{X}$ is a complete and separable metric space (or even only separable; see Clément and Desch [25]): one fixes some x_0 in $\mathscr{X}$ and replaces $\|x\|$ by $d(x,x_0)$. Although the topological properties below still hold at that level of generality (except when $p = 0$ or $p = \infty$), for the sake of simplifying the notation we restrict the discussion to Banach spaces. It will always be assumed without explicit mention that $1 \le p < \infty$.

The space $\mathscr{W}_p(\mathscr{X})$ is defined as the collection of measures μ such that $W_p(\mu,\delta_0) < \infty$ with δ_x being a Dirac measure at x. Of course, $W_p(\mu,\nu)$ can be finite even if $\mu,\nu \notin \mathscr{W}_p(\mathscr{X})$. But if $\mu \in \mathscr{W}_p(\mathscr{X})$ and $\nu \notin \mathscr{W}_p(\mathscr{X})$, then $W_p(\mu,\nu)$ is always infinite. This can be seen from the triangle inequality

$$\infty = W_p(\nu,\delta_0) \le W_p(\mu,\delta_0) + W_p(\mu,\nu).$$

In the sequel, we shall almost exclusively deal with measures in $\mathscr{W}_p(\mathscr{X})$.

The Wasserstein spaces are ordered in the sense that if $q \ge p$, then $\mathscr{W}_q(\mathscr{X}) \subseteq \mathscr{W}_p(\mathscr{X})$. This property extends to the distances, in the sense that

$$q \ge p \ge 1 \quad\Longrightarrow\quad W_q(\mu,\nu) \ge W_p(\mu,\nu). \tag{3.1}$$

To see this, let $\pi \in \Pi(\mu,\nu)$ be optimal with respect to q. Jensen's inequality for the convex function $z \mapsto z^{q/p}$ gives

$$W_q^q(\mu,\nu) = \int_{\mathscr{X}^2} \|x-y\|^q \,\mathrm{d}\pi(x,y) \ge \left(\int_{\mathscr{X}^2} \|x-y\|^p \,\mathrm{d}\pi(x,y)\right)^{q/p} \ge W_p^q(\mu,\nu).$$

The converse of (3.1) fails to hold in general, since it is possible that W_p is finite while W_q is

infinite. A converse can be established, however, if μ and ν are bounded:

$$q \geq p \geq 1, \quad \mu(K) = \nu(K) = 1 \quad \Longrightarrow \quad W_q(\mu,\nu) \leq W_p^{p/q}(\mu,\nu) \left(\sup_{x,y\in K} \|x-y\| \right)^{1-p/q}. \tag{3.2}$$

Indeed, if we denote the supremum by d_K and let π be now optimal with respect to p, then $\pi(K\times K) = 1$ and

$$W_q^q(\mu,\nu) \leq \int_{K^2} \|x-y\|^q \,\mathrm{d}\pi(x,y) \leq d_K^{q-p} \int_{K^2} \|x-y\|^p \,\mathrm{d}\pi(x,y) = d_K^{q-p} W_p^p(\mu,\nu).$$

Another useful property of the Wasserstein distance is the upper bound

$$\mathscr{W}_p(\mathbf{t}\#\mu, \mathbf{s}\#\mu) \leq \left(\int_{\mathscr{X}} \|\mathbf{t}(x) - \mathbf{s}(x)\|^p \,\mathrm{d}\mu(x) \right)^{1/p} = \| \, \|\mathbf{t}-\mathbf{s}\|_{\mathscr{X}} \, \|_{L_p(\mu)} \tag{3.3}$$

for any pair of measurable functions $\mathbf{t}, \mathbf{s} : \mathscr{X} \to \mathscr{X}$. Situations where this inequality holds as equality and $\mathbf{t}$ and $\mathbf{s}$ are optimal maps are related to **compatibility** of the measures μ, $\nu = \mathbf{t}\#\mu$ and $\rho = \mathbf{s}\#\mu$ (see Subsection 3.3.2) and will be of conceptual importance in the context of Fréchet means (see Section 3.5).

We also recall the notation $B_R(x_0) = \{x : \|x - x_0\| < R\}$ and $\overline{B}_R(x_0) = \{x : \|x - x_0\| \leq R\}$ for open and closed balls in $\mathscr{X}$.

3.2 Topological properties

3.2.1 Convergence, compact subsets

The topology of a space is determined by the collection of its closed sets. Since $\mathscr{W}_p(\mathscr{X})$ is a metric space, whether a set is closed or not depends on which sequences in $\mathscr{W}_p(\mathscr{X})$ converge. The following characterisation from Villani [88, Theorem 7.12] will be very useful.

Theorem 3.2.1 (convergence in Wasserstein space)**.** *Let $\mu, \mu_n \in \mathscr{W}_p(\mathscr{X})$. Then the following are equivalent:*

1. *$W_p(\mu_n, \mu) \to 0$ as $n \to \infty$;*
2. *$\mu_n \to \mu$ narrowly and $\int_{\mathscr{X}} \|x\|^p \,\mathrm{d}\mu_n(x) \to \int_{\mathscr{X}} \|x\|^p \,\mathrm{d}\mu(x)$;*
3. *$\mu_n \to \mu$ narrowly and*

$$\sup_n \int_{\{x:\|x\|>R\}} \|x\|^p \,\mathrm{d}\mu_n(x) \to 0, \qquad R \to \infty; \tag{3.4}$$

4. *for any $C > 0$ and any continuous $f : X \to \mathbb{R}$ such that $|f(x)| \leq C(1 + \|x\|^p)$ for all x,*

$$\int_{\mathscr{X}} f(x) \,\mathrm{d}\mu_n(x) \to \int_{\mathscr{X}} f(x) \,\mathrm{d}\mu(x).$$

Consequently, the Wasserstein topology is finer than the narrow topology induced on $\mathscr{W}_p(\mathscr{X})$ from $P(\mathscr{X})$. Indeed, let $\mathscr{A} \subseteq \mathscr{W}_p(\mathscr{X})$ be narrowly closed. If $\mu_n \in \mathscr{A}$ converge to μ in $\mathscr{W}_p(\mathscr{X})$, then $\mu_n \to \mu$ narrowly, so $\mu \in \mathscr{A}$. In other words, the Wasserstein topology has more closed sets than the induced narrow topology. Moreover, each $\mathscr{W}_p(\mathscr{X})$ is a narrowly closed subset of $P(\mathscr{X})$ by the same arguments that lead to (2.3). In view of Theorem 3.2.1, a common strategy to establish Wasserstein convergence is to first show tightness and obtain narrow convergence, hence a candidate limit; and then show that the stronger Wasserstein convergence actually holds. In some situations, the last part is automatic:

Corollary 3.2.2. *Let $K \subset \mathscr{X}$ be a bounded set and suppose that $\mu_n(K) = 1$ for all $n \geq 1$. Then $W_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \to \mu$ narrowly.*

Proof. This is immediate from (3.4). □

The fact that convergence in $\mathscr{W}_p$ is stronger than narrow convergence is exemplified in the following result. If $\mu_n \to \mu$ and $\nu_n \to \nu$ in $\mathscr{W}_p(\mathscr{X})$, then it obvious that $W_p(\mu_n, \nu_n) \to W_p(\mu, \nu)$. But the convergence is only narrow, then the Wasserstein distance is still lower semicontinuous:

$$\liminf_{n\to\infty} W_p(\mu_n, \nu_n) \geq W_p(\mu, \nu). \tag{3.5}$$

This follows from Theorem 2.9.2 and (2.3).

Before giving some examples it will be convenient to formulate Theorem 3.2.1 in probabilistic terms. Let X, X_n be random elements on $\mathscr{X}$ with laws $\mu, \mu_n \in \mathscr{W}_p(\mathscr{X})$. Assume without loss of generality that X, X_n are defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and write $W_p(X_n, X)$ to denote $W_p(\mu_n, \mu)$. Then $W_p(X_n, X) \to 0$ if and only if $X_n \to X$ narrowly and $\mathbb{E}\|X_n\|^p \to \mathbb{E}\|X\|^p$.

An early example of the use of Wasserstein metric in statistics is due to Bickel & Freedman [14]. Let X_n be independent and identically distributed random variables with mean zero and variance 1 and let Z be a standard normal random variable. Then $Z_n = \sum_{i=1}^n X_i/\sqrt{n}$ converge narrowly to Z by the central limit theorem. But $\mathbb{E}Z_n^2 = 1 = \mathbb{E}Z^2$, so $W_2(Z_n, Z) \to 0$. Let Z_n^* be a bootstrapped version of Z_n constructed by resampling the X_n's. If $W_2(Z_n^*, Z_n) \to 0$ then $W_2(Z_n^*, Z) \to 0$ and in particular Z_n^* has the same asymptotic distribution as Z_n.

Another consequence of Theorem 3.2.1 is that (in presence of narrow convergence) convergence of moments automatically yields convergence of smaller moments (there are, however, more elementary ways to see this). In the previous example, for instance, one can also conclude that $\mathbb{E}|Z_n|^p \to \mathbb{E}|Z|^p$ for any $p \leq 2$ by the last condition of the theorem. If in addition $\mathbb{E}X_1^4 < \infty$ then

$$\mathbb{E}Z_n^4 = 3 - \frac{3}{n} + \frac{\mathbb{E}X_1^4}{n} \to 3 = \mathbb{E}Z^4$$

(see Durrett [33, Theorem 2.3.5]) so $W_4(Z_n, Z) \to 0$ and all moments of order 4 or less converge.

Condition (3.4) is called **uniform integrability** of the function $x \mapsto \|x\|^p$ with respect to the collection (μ_n). Of course, it holds for a single measure $\mu \in \mathcal{W}_p(\mathcal{X})$ by the dominated convergence theorem. This condition allows us to characterise compact sets in the Wasserstein space. One should beware that when $\mathcal{X}$ is infinite dimensional, (3.4) alone is not sufficient in order to conclude that μ_n has a convergent subsequence: take μ_n to be Dirac measures at e_n with (e_n) an orthonormal basis of a Hilbert space $\mathcal{X}$ (or any sequence with $\|e_n\| = 1$ that has no convergent subsequence, if $\mathcal{X}$ is a Banach space). The uniform integrability (3.4) must be accompanied with narrow tightness, which is a consequence of (3.4) only when $\mathcal{X} = \mathbb{R}^d$.

Proposition 3.2.3 (compact sets in $\mathcal{W}_p$)**.** *A narrowly tight set $\mathcal{K} \subseteq \mathcal{W}_p$ is Wasserstein-tight (has a compact closure) if and only if*

$$\sup_{\mu \in \mathcal{K}} \int_{\{x : \|x\| > R\}} \|x\|^p \, \mathrm{d}\mu(x) \to 0, \qquad R \to \infty. \tag{3.6}$$

Proof. Suppose that (3.6) holds. If $\mu_n \in \mathcal{K}$, then there exists a measure μ_0 such that $\mu_n \to \mu_0$ narrowly (up to a subsequence), and as (3.4) holds for that subsequence, it converges in the Wasserstein space.

Conversely, if (3.6) does not hold, then we can find a sequence $\mu_n \in \mathcal{K}$ such that for some $\epsilon > 0$,

$$\int_{\{x : \|x\| > n\}} \|x\|^p \, \mathrm{d}\mu_n(x) > \epsilon, \qquad n = 1, 2, \ldots.$$

Obviously no subsequence of μ_n can converge in the Wasserstein space, in view of (3.4). Thus $\overline{\mathcal{K}}$ is not compact in $\mathcal{W}_p$. □

Corollary 3.2.4 (measures with common support)**.** *Let $K \subseteq \mathcal{X}$ be a compact set. Then*

$$\mathcal{K} = \mathcal{W}_p(K) = \{\mu \in P(\mathcal{X}) : \mu(K) = 1\} \subseteq \mathcal{W}_p(\mathcal{X})$$

is compact.

Proof. This is immediate, since $\mathcal{K}$ is narrowly tight and the supremum in (3.6) vanishes when R is larger than the finite quantity $\sup_{x \in K} \|x\|$. Finally, K is closed, so $\mathcal{K}$ is narrowly closed, hence Wasserstein closed, by the portmanteau lemma 2.9.1. □

For future reference we give another consequence of uniform integrability, called **uniform absolute continuity**

$$\forall \epsilon \; \exists \delta \; \forall n \; \forall A \subseteq \mathcal{X} \text{ Borel}: \qquad \mu_n(A) \le \delta \quad \Longrightarrow \quad \int_A \|x\|^p \, \mathrm{d}\mu_n(x) < \epsilon. \tag{3.7}$$

To show that (3.4) implies (3.7), let $\epsilon > 0$, choose $R = R_\epsilon > 0$ such that the supremum in (3.4) is smaller than $\epsilon/2$, and set $\delta = \epsilon/(2R^p)$. If $\mu_n(A) \le \delta$ then

$$\int_A \|x\|^p \, \mathrm{d}\mu_n(x) \le \int_{A \cap \overline{B}_R(0)} \|x\|^p \, \mathrm{d}\mu_n(x) + \int_{A \setminus \overline{B}_R(0)} \|x\|^p \, \mathrm{d}\mu_n(x) < \delta R^p + \epsilon/2 \le \epsilon.$$

3.2.2 Dense subsets and completeness

If we identify a measure $\mu \in \mathcal{W}_p(\mathcal{X})$ with a random variable X (having distribution μ), then X has a finite p-th moment in the sense that the real-valued random variable $\|X\|$ is in L_p. In view of that, it should not come as a surprise that $\mathcal{W}_p(\mathcal{X})$ enjoys topological properties similar to L_p spaces. In this subsection we give some examples of useful dense subsets of $\mathcal{W}_p(\mathcal{X})$ and then show that like $\mathcal{X}$, it is a complete separable metric space. In the next subsection we describe some of the negative properties that $\mathcal{W}_p(\mathcal{X})$ has, again in similarity with L_p spaces.

We first show that $\mathcal{W}_p(\mathcal{X})$ is separable. The core idea of the proof is the feasibility of approximating any measure with discrete measures as follows.

Let μ be a probability measure on $\mathcal{X}$, and let $X_1, X_2, \dots$ be a sequence of independent random elements in $\mathcal{X}$ with probability distribution μ. Then the **empirical measure** μ_n is defined as the random measure $(1/n)\sum_{i=1}^n \delta\{X_i\}$. The law of large numbers shows that for any (measurable) bounded or nonnegative $f : \mathcal{X} \to \mathbb{R}$, almost surely

$$\int_{\mathcal{X}} f(x) \, \mathrm{d}\mu_n(x) = \frac{1}{n}\sum_{i=1}^n f(X_i) \to \mathbb{E} f(X_1) = \int_{\mathcal{X}} f(x) \, \mathrm{d}\mu(x).$$

In particular when $f(x) = \|x\|^p$, we obtain convergence of moments of order p. Hence by Theorem 3.2.1, if $\mu \in \mathcal{W}_p(\mathcal{X})$ then $\mu_n \to \mu$ in $\mathcal{W}_p(\mathcal{X})$ if and only if $\mu_n \to \mu$ narrowly. We know that integrals of bounded functions converge with probability one, but the null set may depend on the chosen function and there are uncountably many such functions. When $\mathcal{X} = \mathbb{R}^d$, by the portmanteau lemma 2.9.1 we can replace the collection $C_b(\mathcal{X})$ by indicator functions of rectangles of the form $(-\infty, a_1] \times \cdots \times (-\infty, a_d]$ for $a = (a_1, \dots, a_d) \in \mathbb{R}^d$. It turns out that the countable collection provided by rational vectors a suffices (see the proof of Theorem 4.4.1 where this is done in a more complicated setting). For more general spaces $\mathcal{X}$, we need to find another countable collection $\{f_j\}$ such that convergence of the integrals of f_j for all j suffices for narrow convergence. Such a collection exists, by using bounded Lipschitz functions (Dudley, [31, Theorem 11.4.1]); an alternative construction can be found in Ambrosio, Gigli & Savaré [6, Section 5.1]. Thus, we have:

Proposition 3.2.5 (empirical measures in $\mathcal{W}_p$)**.** *For any $\mu \in P(\mathcal{X})$ and the corresponding sequence of empirical measures μ_n, $W_p(\mu_n, \mu) \to 0$ almost surely if and only if $\mu \in \mathcal{W}_p(\mathcal{X})$.*

Indeed, if $\mu \notin \mathcal{W}_p(\mathcal{X})$, then $W_p(\mu_n, \mu)$ is infinite for all n, since μ_n is compactly supported, hence in $\mathcal{W}_p(\mathcal{X})$.

Proposition 3.2.5 is the basis for constructing dense subsets of the Wasserstein space.

Theorem 3.2.6 (dense subsets of $\mathscr{W}_p$)**.** *The following collections of measures are dense in $\mathscr{W}_p(\mathscr{X})$:*

1. *finitely supported measures with rational weights;*
2. *compactly supported measures;*
3. *finitely supported measures with rational weights on a dense subset $A \subseteq \mathscr{X}$;*
4. *if $\mathscr{X} = \mathbb{R}^d$, the collection of absolutely continuous and compactly supported measures;*
5. *if $\mathscr{X} = \mathbb{R}^d$, the collection of absolutely continuous measures with strictly positive and analytic densities.*

In particular, $\mathscr{W}_p$ is separable (because $\mathscr{X}$ is separable and the third set is countable).

Proof. The first collection is dense by Proposition 3.2.5, and the second collection is larger than the first. Let $\mu = n^{-1}\sum_{i=1}^n \delta\{x_i\}$ be a finitely supported measure with rational weights (with x_i possibly not distinct) and $\epsilon > 0$. Pick $a_i \in A$ with $\|a_i - x_i\| < \epsilon$ and set $\nu = n^{-1}\sum_{i=1}^n \delta\{a_i\}$. Then $W_p(\mu,\nu) \le \epsilon$, and so the third set is also dense. Finally, for any $\sigma > 0$ define μ_σ as the convolution of μ with a uniform measure on a ball of size σ, i.e. with density

$$g(x) = \frac{\sigma^{-d}}{c_d}\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\|x - x_i\| \le \sigma\}, \qquad c_d = \mathrm{Leb}\left\{x \in \mathbb{R}^d : \|x\| \le 1\right\}.$$

Then μ_σ is absolutely continuous and compactly supported with

$$\mathscr{W}_p^p(\mu_\sigma, \mu) \le \frac{\sigma^{-d}}{c_d}\frac{1}{n}\sum_{i=1}^n \int_{B_\sigma(x_i)} \|x - x_i\|^p \,\mathrm{d}x \le \sigma^p.$$

It follows that the fourth collection is dense too. For the fifth, use the same convolution with a Gaussian measure instead of a uniform one. □

Proposition 3.2.7 (completeness)**.** *The Wasserstein space $\mathscr{W}_p(\mathscr{X})$ is complete.*

Proof. Two different proofs of this result can be found in Villani [89, Theorem 6.18] and in Ambrosio, Gigli & Savaré [6, Proposition 7.1.5]; we sketch an alternative argument here. Let (μ_n) be a Cauchy sequence in $\mathscr{W}_p(\mathscr{X})$. It follows from (3.1) that $W_1(\mu,\nu) \le W_p(\mu,\nu)$ for any $\mu, \nu \in P(\mathscr{X})$. Thus (μ_n) is a Cauchy sequence in $\mathscr{W}_1(\mathscr{X})$. In that space the Kantorovich–Rubinstein theorem (2.5) states that

$$W_1(\mu,\nu) = \sup_{\|\varphi\|_{Lip} \le 1} \left| \int_{\mathscr{X}} \varphi \,\mathrm{d}\mu - \int_{\mathscr{X}} \varphi \,\mathrm{d}\nu \right|, \qquad \|\varphi\|_{Lip} = \sup_{x \ne y} \frac{|\varphi(x) - \varphi(y)|}{\|x - y\|}.$$

In particular $W_1(\mu,\nu)$ is larger than the **bounded Lipschitz norm**

$$W_1(\mu,\nu) \geq \|\mu-\nu\|_{BL} = \sup_{\|\varphi\|_{BL}\leq 1} \left| \int_{\mathscr{X}} \varphi \,\mathrm{d}\mu - \int_{\mathscr{X}} \varphi \,\mathrm{d}\nu \right|, \qquad \|\varphi\|_{BL} = \|\varphi\|_{Lip} + \|\varphi\|_{\infty},$$

which metrises narrow convergence in $P(\mathscr{X})$ [31, Theorem 11.3.3]. Thus (μ_n) is a Cauchy sequence with $\|\cdot\|_{BL}$. Since $P(\mathscr{X})$ is complete with this norm [31, Corollary 11.5.5], (μ_n) converges narrowly to $\mu \in P(\mathscr{X})$. If we now fix N, then the lower semicontinuity of the Wasserstein distance (3.5) gives

$$W_p^p(\mu_N,\mu) \leq \liminf_{k\to\infty} W_p^p(\mu_N,\mu_k).$$

Since the sequence (μ_n) is Cauchy, the right-hand side vanishes as $N\to\infty$. Thus $W_p(\mu_N,\mu)\to 0$ and completeness is established. □

3.2.3 Negative topological properties

In the previous subsection we have shown that $\mathscr{W}_p(\mathscr{X})$ is separable and complete like L_p spaces. Just like them, however, the Wasserstein space is neither locally compact nor σ-compact. For this reason, existence proofs of Fréchet means in $\mathscr{W}_p(\mathscr{X})$ require tools that are more specific to this space, and do not rely upon local compactness (see Section 3.5).

Proposition 3.2.8 ($\mathscr{W}_p$ is not locally compact)**.** *Let $\mu\in\mathscr{W}_p(\mathscr{X})$ and let $\epsilon>0$. Then the Wasserstein ball*

$$\overline{B}_\epsilon(\mu) = \{\nu\in\mathscr{W}_p(\mathscr{X}) : W_p(\mu,\nu)\leq\epsilon\}$$

is not compact.

Proof. This is a generalisation of Remark 7.1.9 in Ambrosio, Gigli & Savaré [6] who prove it when μ is Dirac.

By Theorem 3.2.6 there exists a compactly supported measure ν with $W_p(\mu,\nu)<\epsilon/2$, so that $\overline{B}_{\epsilon/2}(\nu)\subseteq\overline{B}_\epsilon(\mu)$. We can consequently assume without loss of generality that there exists a compact $K\subset\mathscr{X}$ with $\mu(K)=1$.

Pick a sequence $x_n\in\mathscr{X}$ of elements that has no partial limits and that are of distance at least $\delta>0$ from K (i.e. such that $d(x_n,K)=\inf_{x\in K}\|x-x_n\|\geq\delta$ for all n, for instance $\|x_n\|\to\infty$), assume without loss of generality that $\epsilon<\delta$ and set

$$\mu_n = (1-\alpha_n)\mu + \alpha_n\delta\{x_n\}, \qquad \alpha_n = \epsilon^p / W_p^p(\mu,\delta\{x_n\}).$$

Then μ_n is a probability measure because

$$W_p^p(\mu,\delta\{x_n\}) = \int_K \|x-x_n\|^p \,\mathrm{d}\mu(x) \geq \delta^p \geq \epsilon^p,$$

so $\alpha_n \in [0,1]$ for all n. To bound $W_p(\mu_n,\mu)$ observe that we may leave the common $(1-\alpha_n)$ mass in place, so that

$$W_p^p(\mu_n,\mu) \le \alpha_n W_p^p(\mu,\delta\{x_n\}) = \epsilon^p \qquad \Longrightarrow \quad \mu_n \in \overline{B}_\epsilon(\mu).$$

We need to show that no subsequence of μ_n can converge in the Wasserstein space. By extracting a subsequence, we may assume that $\alpha_n \to \alpha \in [0,1]$. If (a subsequence of) μ_n converges in the Wasserstein space (or even narrowly), then the limit must be $(1-\alpha)\mu + \alpha\delta\{x\}$ with x a limit of (x_n). By the hypothesis on the sequence (x_n), this can only happen if $\alpha = 0$. To finish the proof we only need to show that $W_p(\mu_n,\mu)$ is bounded away from zero.

Clearly $W_p^p(\mu_n,\mu) \ge \alpha_n d^p(x_n,K)$; let us show that this is bounded below. Indeed, let $d_K = \sup_{x,y\in K}\|x-y\|$ be the diameter of K and observe that

$$W_p^p(\mu,\delta\{x_n\}) = \int_K \|x-x_n\|^p \,\mathrm{d}\mu(x) \le [d(x_n,K)+d_K]^p \le d^p(x_n,K)\left[1+\frac{d_K}{\delta}\right]^p,$$

so that

$$\alpha_n d^p(x_n,K) = \frac{\epsilon^p d^p(x_n,K)}{W_p^p(\mu,\delta\{x_n\})} \ge \frac{\epsilon^p\delta^p}{(\delta+d_K)^p} > 0.$$

Thus $\alpha = 0$ is impossible too and no subsequence of (μ_n) converges. □

From this we deduce:

Corollary 3.2.9. *The Wasserstein space $\mathcal{W}_p(\mathcal{X})$ is not σ-compact.*

Proof. If $\mathcal{K}$ is a compact set in $\mathcal{W}_p(\mathcal{X})$, then its interior is empty by Proposition 3.2.8. A countable union of compact sets has an empty interior (hence cannot equal the entire space $\mathcal{W}_p(\mathcal{X})$) by the Baire property, which holds on the complete metric space $\mathcal{W}_p(\mathcal{X})$ by the Baire category theorem (Dudley [31, Theorem 2.5.2]). □

3.3 The tangent bundle

Although the Wasserstein space $\mathcal{W}_p(\mathcal{X})$ is nonlinear in terms of measures, it *is* linear in terms of maps. Indeed, if $\mu \in \mathcal{W}_p(\mathcal{X})$ and $T_i : \mathcal{X} \to \mathcal{X}$ are such that $\|T_i\| \in L_p(\mu)$, then $(\alpha T_1 + \beta T_2)\#\mu \in \mathcal{W}_p(\mathcal{X})$ for all $\alpha,\beta \in \mathbb{R}$. Later, in Section 3.4, we shall see that $\mathcal{W}_p(\mathcal{X})$ is in fact homeomorphic to a subset of the space of such functions. The goal of this section is to exploit the linearity of the latter in order to define the tangent bundle of $\mathcal{W}_p$. This in particular will be used for deriving differentiability properties of the Wasserstein distance in Subsection 3.5.5. We assume here that $\mathcal{X}$ is a Hilbert space and, for simplicity only, that $p = 2$; the results below can be extended to any $p > 1$, see Ambrosio, Gigli & Savaré [6]. We recall that absolutely continuous measures

are assumed to be so with respect to Lebesgue measure if $\mathscr{X} = \mathbb{R}^d$ and otherwise refer to Definition 2.5.5.

3.3.1 Geodesics, the log map and the exponential map in $\mathscr{W}_2(\mathscr{X})$

Let $\gamma \in \mathscr{W}_2(\mathscr{X})$ be absolutely continuous and $\mu \in \mathscr{W}_2(\mathscr{X})$ arbitrary. From results in Section 2.5 we know that there exists a unique solution to the Monge–Kantorovich problem, and that solution is given by a transport map that we denote by $\mathbf{t}_\gamma^\mu$. Recalling that $\mathbf{i} : \mathscr{X} \to \mathscr{X}$ is the identity map, we can define a curve

$$\gamma_t = \left[\mathbf{i} + t(\mathbf{t}_\gamma^\mu - \mathbf{i})\right]\#\gamma, \qquad t \in [0,1].$$

This curve is known as McCann's interpolation (McCann [65, Equation 7]). As hinted in the introduction to this section, it is constructed via classical linear interpolation of the transport maps $\mathbf{t}_\gamma^\mu$ and the identity. Clearly $\gamma_0 = \gamma$, $\gamma_1 = \mu$ and from (3.3),

$$\begin{aligned} W_2(\gamma_t, \gamma) &\leq \sqrt{\int_{\mathscr{X}} \left[t(\mathbf{t}_\gamma^\mu - \mathbf{i})\right]^2 \mathrm{d}\gamma} &&= tW_2(\gamma,\mu); \\ W_2(\gamma_t, \mu) &\leq \sqrt{\int_{\mathscr{X}} \left[(1-t)(\mathbf{t}_\gamma^\mu - \mathbf{i})\right]^2 \mathrm{d}\gamma} &&= (1-t)W_2(\gamma,\mu). \end{aligned}$$

It follows from the triangle inequality in $\mathscr{W}_2$ that these inequalities must hold as equalities. Taking this one step further, we see that

$$W_2(\gamma_t, \gamma_s) = (t-s)W_2(\gamma,\mu), \qquad 0 \leq s \leq t \leq 1.$$

In other words, McCann's interpolation is a **constant-speed geodesic** in $\mathscr{W}_2(\mathscr{X})$.

In view of this, it seems reasonable to define the **tangent space** of $\mathscr{W}_2(\mathscr{X})$ at μ as (Ambrosio, Gigli & Savaré [6, Definition 8.5.1])

$$\mathrm{Tan}_\mu = \overline{\{t(\mathbf{t}-\mathbf{i}) : \mathbf{t} \text{ uniquely optimal between } \mu \text{ and } \mathbf{t}\#\mu;\, t>0\}}^{L_2(\mu)}.$$

Since $\mathbf{t}$ is uniquely optimal, $\mathbf{t}\#\mu \in \mathscr{W}_2(\mathscr{X})$ as well and $x \mapsto \|\mathbf{t}(x)\|$ is in $L_2(\mu)$, so $\mathrm{Tan}_\mu \subseteq L_2(\mu)$. (Strictly speaking, Tan_μ is a subset of the space of functions $f : \mathscr{X} \to \mathscr{X}$ such that $\|f\| \in L_2(\mu)$ rather than $L_2(\mu)$ itself, as in Definition 3.4.2, but we will write L_2 for simplicity.)

Since optimality of $\mathbf{t}$ is independent of μ, the only part of this definition that depends on μ is the closure operation. Although not obvious from the definition, this is a linear space.[1]

[1] There is an equivalent definition in terms of gradients, in which linearity is clear, see [6, Definition 8.4.1]: when $\mathscr{X} = \mathbb{R}^d$, it is $\mathrm{Tan}_\mu = \overline{\{\nabla f : f \in C_c^\infty(\mathbb{R}^d)\}}^{L_2(\mu)}$ (compactly supported C^∞ functions). When $\mathscr{X}$ is a separable Hilbert space, one takes C_c^∞ functions that depend on finitely many coordinates, called **cylindrical functions** [6, Definition 5.1.11]. The two definitions of the tangent space coincide by [6, Theorem 8.5.1].)

Since γ is absolutely continuous, the exponential map at γ

$$\exp_\gamma : \mathrm{Tan}_\gamma \to \mathscr{W}_2 \qquad \exp_\gamma(t(\mathbf{t}-\mathbf{i})) = \exp_\gamma([t\mathbf{t}+(1-t)\mathbf{i}]-\mathbf{i}) = [t\mathbf{t}+(1-t)\mathbf{i}]\#\gamma \quad (t\in\mathbb{R})$$

is surjective, as can be seen from its right inverse, the log map

$$\log_\gamma : \mathscr{W}_2 \to \mathrm{Tan}_\gamma \qquad \log_\gamma(\mu) = \mathbf{t}_\gamma^\mu - \mathbf{i},$$

defined throughout $\mathscr{W}_2$. Thus

$$\exp_\gamma(\log_\gamma(\mu)) = \mu, \quad \mu\in\mathscr{W}_2, \qquad \text{and} \qquad \log_\gamma(\exp_\gamma(t(\mathbf{t}-\mathbf{i}))) = t(\mathbf{t}-\mathbf{i}) \quad (t\in[0,1]),$$

because convex combinations of optimal maps are optimal maps as well. In particular, McCann's interpolant $\big[\mathbf{i}+t(\mathbf{t}_\gamma^\mu-\mathbf{i})\big]\#\gamma$ is mapped bijectively to the line segment $t(\mathbf{t}_\gamma^\mu-\mathbf{i})\in\mathrm{Tan}_\gamma$ through the log map.

3.3.2 Curvature and compatibility of measures

Let $\gamma,\mu,\nu\in\mathscr{W}_2(\mathscr{X})$ be absolutely continuous measures. Then by (3.3)

$$W_2^2(\mu,\nu) \le \int_{\mathscr{X}} \|\mathbf{t}_\gamma^\mu(x)-\mathbf{t}_\gamma^\nu(x)\|^2\,\mathrm{d}\gamma(x) = \|\log_\gamma(\mu)-\log_\gamma(\nu)\|^2.$$

In other words, the distance between μ and ν is smaller in $\mathscr{W}_2(\mathscr{X})$ then the distance between the corresponding vectors $\log_\gamma(\mu)$ and $\log_\gamma(\nu)$ in the tangent space Tan_γ. In the terminology of differential geometry, this means that the Wasserstein space has **nonnegative sectional curvature** at any absolutely continuous γ.

It is instructive to see when equality holds. Clearly $\mathbf{t}_\nu^\gamma = (\mathbf{t}_\gamma^\nu)^{-1}$, so a change of variables gives

$$W_2^2(\mu,\nu) \le \int_{\mathscr{X}} \|\mathbf{t}_\gamma^\mu(\mathbf{t}_\nu^\gamma(x))-x\|^2\,\mathrm{d}\nu(x).$$

Since the map $\mathbf{t}_\gamma^\mu\circ\mathbf{t}_\nu^\gamma$ pushes forward ν to μ, equality holds if and only if $\mathbf{t}_\gamma^\mu\circ\mathbf{t}_\nu^\gamma = \mathbf{t}_\nu^\mu$. This motivates the following definition.

Definition 3.3.1 (compatible measures)**.** *A collection of absolutely continuous measures $\mathscr{C}\subseteq\mathscr{W}_2(\mathscr{X})$ is **compatible** if for all $\gamma,\mu,\nu\in\mathscr{C}$, we have $\mathbf{t}_\gamma^\mu\circ\mathbf{t}_\nu^\gamma = \mathbf{t}_\nu^\mu$ (in $L_2(\nu)$).*

It appears that this notion was first introduced by Boissard, Le Gouic & Loubes [20] under the label of **admissible optimal maps** by defining families of gradients of convex functions (T_i) such that $T_j^{-1}\circ T_i$ is a gradient of a convex function for any i and j. For (any) fixed measure $\gamma\in\mathscr{C}$, compatibility of $\mathscr{C}$ is then equivalent to admissibility of the collection of maps $\{\mathbf{t}_\gamma^\mu\}_{\mu\in\mathscr{C}}$.

Remark 3. *The absolute continuity is not the important issue in the definition and was introduced in order to guarantee that $\mathbf{t}_\gamma^\mu$ exist and be unique for all $\gamma,\nu\in\mathscr{C}$. The definition of*

compatibility is valid as long as this is the case, which could very well happen if all the measures in $\mathscr{C}$ are uniform discrete measures on the same number of points.

A collection of two (absolutely continuous) measures is always compatible. More interestingly, if $\mathscr{X} = \mathbb{R}$, then the entire collection of absolutely continuous (or even just continuous) measures is compatible. This is because of the simple geometry of convex functions in $\mathbb{R}$: gradients of convex functions are nondecreasing, and this property is stable under composition. In a more probabilistic way of thinking, one can always push-forward μ to ν via the uniform distribution $\mathrm{Leb}|_{[0,1]}$ (see Section 2.6). Letting F_μ^{-1} and F_ν^{-1} denote the quantile functions, we have seen that

$$W_2(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L_2(0,1)}.$$

(As a matter of fact, in this specific case, the equality holds for all $p \geq 1$ and not only for $p = 2$.) In other words, $\mu \mapsto F_\mu^{-1}$ is an *isometry* from $\mathscr{W}_2(\mathbb{R})$ to the subset of $L_2(0,1)$ formed by (equivalence classes of) left-continuous nondecreasing functions on $(0,1)$. Since this is a convex subset of a Hilbert space, this property provides a very simple way to evaluate Fréchet means in $\mathscr{W}_2(\mathbb{R})$ (see Section 3.5). If $\gamma = \mathrm{Leb}|_{[0,1]}$, then $F_\mu^{-1} = \mathbf{t}_\gamma^\mu$ for all μ, so we can write the above equality as

$$W_2^2(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L_2(0,1)} = \|\log_\gamma(\mu) - \log_\gamma(\nu)\|^2,$$

so that if $\mathscr{X} = \mathbb{R}$, the Wasserstein space is essentially **flat** (has zero sectional curvature).

The importance of compatibility can be seen as mimicking the simple one-dimensional case in terms of a Hilbert space embedding. Let $\mathscr{C} \subseteq \mathscr{W}_2(\mathscr{X})$ be compatible and fix $\gamma \in \mathscr{C}$. Then for all $\mu, \nu \in \mathscr{C}$

$$W_2^2(\mu,\nu) = \int_{\mathscr{X}} \|\mathbf{t}_\gamma^\mu(x) - \mathbf{t}_\gamma^\nu(x)\|^2 \,\mathrm{d}\gamma(x) = \|\log_\gamma(\mu) - \log_\gamma(\nu)\|_{L_2(\gamma)}^2.$$

Consequently, once again, $\mu \mapsto \mathbf{t}_\gamma^\mu$ is an isometric embedding of $\mathscr{C}$ into $L_2(\gamma)$. Generalising the one-dimensional case, we shall see that this allows for easy calculations of Fréchet means by means of averaging transport maps (Theorem 3.5.21).

Example: Gaussian compatible measures. The Gaussian case presented in Section 2.7 is helpful in shedding light on the structure imposed by the compatibility condition. Let $\gamma \in \mathscr{W}_2(\mathbb{R}^d)$ be a standard Gaussian distribution with identity covariance matrix. Let Σ_μ denote the covariance matrix of a measure $\mu \in \mathscr{W}_2(\mathbb{R}^d)$. When μ and ν are centred nondegenerate Gaussian measures,

$$\mathbf{t}_\gamma^\mu = \Sigma_\mu^{1/2}; \qquad \mathbf{t}_\gamma^\nu = \Sigma_\nu^{1/2}; \qquad \mathbf{t}_\mu^\nu = \Sigma_\mu^{-1/2}[\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}]^{1/2}\Sigma_\mu^{-1/2},$$

so that γ, μ and ν are compatible if and only if

$$\mathbf{t}_\mu^\nu = \mathbf{t}_\gamma^\nu \circ \mathbf{t}_\mu^\gamma = \Sigma_\nu^{1/2}\Sigma_\mu^{-1/2}.$$

Since the matrix on the left-hand side must be symmetric, it must necessarily be that $\Sigma_\nu^{1/2}$ and $\Sigma_\mu^{-1/2}$ commute (if A and B are symmetric, then AB is symmetric if and only if $AB = BA$), or equivalently, if and only if Σ_ν and Σ_μ commute. We see that a collection $\mathscr{C}$ of Gaussian measures on $\mathbb{R}^d$ that includes the standard Gaussian distribution is compatible if and only if all the covariance matrices of the measures in $\mathscr{C}$ are *simultaneously diagonalisable*. In other words, there exists an orthogonal matrix U such that $D_\mu = U\Sigma_\mu U^t$ is diagonal for all $\mu \in \mathscr{C}$. In that case formula (2.6)

$$\mathscr{W}_2(\mu,\nu) = \operatorname{tr}[\Sigma_\mu + \Sigma_\nu - 2(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}] = \operatorname{tr}[\Sigma_\mu + \Sigma_\nu - 2\Sigma_\mu^{1/2}\Sigma_\nu^{1/2}]$$

simplifies to

$$\mathscr{W}_2(\mu,\nu) = \operatorname{tr}[D_\mu + D_\nu - 2D_\mu^{1/2}D_\nu^{1/2}] = \sum_{i=1}^{d}(\alpha_i - \beta_i)^2, \qquad \alpha_i = [D_\mu]_{ii}; \quad \beta_i = [D_\nu]_{ii},$$

and identifying the (nonnegative) number $a \in \mathbb{R}$ with the map $x \mapsto ax$ on $\mathbb{R}$, the optimal maps take the "orthogonal separable" form

$$\mathbf{t}_\mu^\nu = \Sigma_\nu^{1/2}\Sigma_\mu^{-1/2} = UD_\nu^{1/2}D_\mu^{-1/2}U^t = U \circ \left(\sqrt{\beta_1/\alpha_1}, \ldots, \sqrt{\beta_d/\alpha_d}\right) \circ U^t.$$

In other words, up to an orthogonal change of coordinates, the optimal maps take the form of d nondecreasing real-valued functions. This is yet another crystallisation the one-dimension-like structure of compatible measures.

With the intuition of the Gaussian case at our disposal, we can discuss a more general case. Suppose that the optimal maps are continuously differentiable. Then differentiating the equation $\mathbf{t}_\mu^\nu = \mathbf{t}_\gamma^\nu \circ \mathbf{t}_\mu^\gamma$ gives

$$\nabla\mathbf{t}_\mu^\nu(x) = \nabla\mathbf{t}_\gamma^\nu(\mathbf{t}_\mu^\gamma(x))\nabla\mathbf{t}_\mu^\gamma(x).$$

Since optimal maps are gradients of convex functions, their derivatives must be symmetric and positive semidefinite matrices. A product of such matrices stays symmetric if and only if they commute, so in this differentiable setting, compatibility is equivalent to commutativity of the matrices $\nabla\mathbf{t}_\gamma^\nu(\mathbf{t}_\mu^\gamma(x))$ and $\nabla\mathbf{t}_\mu^\gamma(x)$ for μ-almost all x. In the Gaussian case, the optimal maps are linear functions, so, of course, have constant derivatives and x does not appear in the matrices.

Boissard, Le Gouic & Loubes [20] give some examples of compatible measures in terms of the optimal maps. Let $\gamma \in \mathscr{W}_2(\mathbb{R}^d)$ be a fixed measure and define $\mathscr{C} = \mathbf{t}\#\gamma$ with $\mathbf{t}$ belonging to one of the following families. The first imposes the one-dimensional structure by varying only the behaviour of the norm of x, while the second allows for separation of variables that splits the d-dimensional problem into d one-dimensional ones.

Radial transformations. Consider the collection of functions $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ of the form $\mathbf{t}(x) =$

$xG(\|x\|)$ with $G:\mathbb{R}_+\to\mathbb{R}$ differentiable. Then a straightforward calculation shows that

$$\nabla\mathbf{t}(x) = G(\|x\|)I + [G'(\|x\|)/\|x\|]\; xx^t.$$

Since both I and xx^t are positive semidefinite, the above matrix is so if both G and G' are nonnegative. If $\mathbf{s}(x) = xH(\|x\|)$ is a function of the same form, then $\mathbf{s}(\mathbf{t}(x)) = xG(\|x\|)H(\|x\|G(\|x\|))$ which belongs to that family of functions (since G is nonnegative). Clearly

$$\nabla\mathbf{s}(\mathbf{t}(x)) = H\big[\|x\|G(\|x\|)\big]I + \Big[G(\|x\|)H'(\|x\|G(\|x\|))/\|x\|\Big]\; xx^t$$

commutes with $\nabla\mathbf{t}(x)$, since both matrices are of the form $aI + bxx^t$ with a,b scalars (that depend on x). In order to be able to change the base measure γ we need to check that the inverses belong to the family. But if $y = \mathbf{t}(x)$, then $x = ay$ for some scalar a that solves the equation

$$aG(a\|y\|) = 1.$$

Such a is guaranteed to be unique if $a\mapsto aG(a)$ is strictly increasing and it will exist (for y in the range of $\mathbf{t}$) if it is continuous. As a matter of fact, since the eigenvalues of $\nabla\mathbf{t}(x)$ are $G(a)$ and

$$G(a) + G'(a)a = (aG(a))', \qquad a = \|x\|,$$

the condition that $a\mapsto aG(a)$ is strictly increasing is sufficient (this is weaker than G itself increasing). Finally, differentiability of G is not required, so it is enough if G is continuous and $aG(a)$ is strictly increasing.

Separable variables. Consider the collection of functions $\mathbf{t}:\mathbb{R}^d\to\mathbb{R}^d$ of the form

$$\mathbf{t}(x_1,\dots,x_d) = (T_1(x_1),\dots,T_d(x_d)), \qquad T_i:\mathbb{R}\to\mathbb{R}, \tag{3.8}$$

with T_i continuous and strictly increasing. This is a generalisation of the compatible Gaussian case discussed above in which all the T_i's were linear. Here it is obvious that elements in this family are optimal maps and that the family is closed under inverses and composition, and compatibility follows.

As observed by Zemel & Panaretos [94], this family is characterised by measures having a *common dependence structure*. More precisely, we say that $C:[0,1]^d\to[0,1]$ is a **copula** if C is (the restriction of) a distribution function of a random vector having uniform margins. In other words, if there is a random vector $V = (V_1,\dots,V_d)$ with $\mathbb{P}(V_i\le a) = a$ for all $a\in[0,1]$ and all $j = 1,\dots,d$, and

$$\mathbb{P}(V_1\le v_1,\dots,V_d\le v_d) = C(v_1,\dots,v_d), \qquad u_i\in[0,1].$$

Nelsen [68] provides an overview on copulae. To any d-dimensional probability measure μ

one can assign a copula $C = C_\mu$ in terms of the distribution function G of μ and its marginals G_j as

$$G(a_1,\dots,a_d) = \mu((-\infty,a_1]\times\cdots\times(-\infty,a_d]) = C(G_1(a_1),\dots,G_d(a_d)).$$

If each G_j is surjective on $(0,1)$, which is equivalent to it being continuous, then this equation defines C uniquely on $(0,1)^d$, and consequently on $[0,1]^d$. If some marginal G_j is not continuous, then uniqueness is lost, but C still exists [68, Chapter 2]. The connection of copulae to compatibility becomes clear in the following lemma:

Lemma 3.3.2 (compatibility and copulae)**.** *The copulae associated with absolutely continuous measures $\mu,\nu\in\mathcal{W}_2(\mathbb{R}^d)$ are equal if and only if $\mathbf{t}_\mu^\nu$ takes the separable form* (3.8)*.*

Proof. Since G_j is continuous, classical arguments on quantile functions yield $G_j(G_j^{-1}(v)) = v$ for all $v\in(0,1)$, and the same holds for F_j. If μ and ν have the same copula then

$$G(G_1^{-1}(v_1),\dots,G_d^{-1}(v_d)) = C(v_1,\dots,v_d) = F(F_1^{-1}(v_1),\dots,F_d^{-1}(v_d)).$$

If we now change variables and set $v_j = F(x_j)$, then $F(x_1,\dots,x_d) = G(G_1^{-1}(F_1(x_1)),\dots,G_d^{-1}(F_d((x_d)))$ for all x_j in the range of F_j^{-1}. Defining now $T_j = G_j^{-1}\circ F_j$, it follows that $\nu = (T_1,\dots,T_d)\#\mu$, and this map is optimal, hence equals $\mathbf{t}_\mu^\nu$, because the T_j's are nondecreasing.

Conversely, $\mathbf{t}_\mu^\nu$ of the form (3.8) ensures that T_j is nondecreasing, since optimality will be violated otherwise. The push forward constraint of $\mathbf{t}_\mu^\nu$ means that T_j must push the j-th marginal of μ to that of ν; as we have seen in Section 2.6, this entails $T_j = G_j^{-1}\circ F_j$. Consequently for all $v_j\in(0,1)$,

$$C_\nu(v_1,\dots,v_d) = G(G_1^{-1}(v_1),\dots,G_d^{-1}(v_d)) = F(F_1^{-1}(v_1),\dots,F_d^{-1}(v_d)) = C_\mu(v_1,\dots,v_d).$$

□

Composition with linear functions. If $\phi:\mathbb{R}^d\to\mathbb{R}$ is convex with gradient $\mathbf{t}$ and A is a $d\times d$ matrix, then the gradient of the convex function $x\mapsto\phi(Ax)$ at x is $\mathbf{t}_A = A^t\mathbf{t}(Ax)$. Suppose ψ is another convex function with gradient $\mathbf{s}$ and that compatibility holds, i.e. $\nabla\mathbf{s}(\mathbf{t}(x))$ commutes with $\nabla\mathbf{t}(x)$ for all x. Then in order for

$$\nabla\mathbf{s}_A(\mathbf{t}_A(x)) = A^t\nabla\mathbf{s}(AA^t\mathbf{t}(Ax))A \qquad\text{and}\qquad \nabla\mathbf{t}_A(x) = A^t\nabla\mathbf{t}(Ax)A$$

to commute, it suffices that $AA^t = I$, i.e., that A be orthogonal. Consequently, if $\{\mathbf{t}\#\mu\}_{\mathbf{t}\in\mathbf{T}}$ are compatible, then so are $\{\mathbf{t}_U\#\mu\}_{\mathbf{t}\in\mathbf{T}}$ for any orthogonal matrix U.

3.4 Random measures in the Wasserstein space

Let μ be a fixed absolutely continuous probability measure in $\mathcal{W}_2(\mathcal{X})$. If $\Lambda \in \mathcal{W}_2(\mathcal{X})$ is another probability measure, then the transport map $\mathbf{t}_\mu^\Lambda$ as well the convex potential are functions of Λ. If Λ is now random, then we would like to be able to make probability statements about them. To this end it needs to be shown that $\mathbf{t}_\mu^\Lambda$ and the convex potential are *measurable* functions of Λ. The goal of this section is to develop a rigorous mathematical framework that justifies such probability statements. We show that all the relevant quantities are indeed measurable, and in particular establish the Fubini-type results in Propositions 3.4.8 and 3.4.13. The less-technically inclined reader may consider skipping this section at first reading.

Here is an example of a measurability result (Villani [89, Corollary 5.22]). Recall that $P(\mathcal{X})$ is the space of Borel probability measures on $\mathcal{X}$, endowed with the topology of narrow convergence that makes it a metric space. Let $\mathcal{X}$ be a complete separable metric space and $c : \mathcal{X}^2 \to \mathbb{R}_+$ a continuous cost function. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\Lambda, \kappa : \Omega \to P(\mathcal{X})$ be measurable maps. Then there exists a **measurable selection** of optimal transference plans. That is, a measurable $\pi : \Omega \to P(\mathcal{X}^2)$ such that $\pi(\omega) \in \Pi(\Lambda(\omega), \kappa(\omega))$ is optimal for all $\omega \in \Omega$.

Although this result is very general, it only provides information about π. If π is induced from a map T, it is not obvious how to construct T from π in a measurable way; we will therefore follow a different path. In order to have a (almost) self-contained exposition, we work in a somewhat simplified setting that nevertheless suffices for the sequel. For instance, rather than the narrow topology, we shall assume that the random measures are measurable with respect to the Wasserstein topology. Since the latter is finer (has more closed sets), this assumption is more restrictive. At least in the Euclidean case $\mathcal{X} = \mathbb{R}^d$, more general measurability results in the flavour of this section can be found in Fontbona, Guérin & Méléard [35]. On the other hand, we will not need to appeal to abstract measurable selection theorems as in [35, 89].

3.4.1 Measurability of measures and of optimal maps

Let $\mathcal{X}$ be a separable Banach space. (Most of the results below hold for any complete separable metric space but we will avoid this generality for brevity and simpler notation). The Wasserstein space $\mathcal{W}_p(\mathcal{X})$ is a metric space for any $p \geq 1$. We can thus define:

Definition 3.4.1 (random measure)**.** *A random measure Λ is any measurable map from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathcal{W}_p(\mathcal{X})$, endowed with its Borel σ-algebra.*

In what follows, whenever we call something random, we mean that it is measurable as a map from some generic unspecified probability space.

Optimal maps are functions from $\mathcal{X}$ to itself. In order to define random optimal maps, we need to define a topology and a σ-algebra on the space of such functions.

Definition 3.4.2 (the space $\mathcal{L}_p(\mu)$)**.** *Let $\mathcal{X}$ be a Banach space and μ a Borel measure on $\mathcal{X}$.*

Then the space $\mathscr{L}_p(\mu)$ is the space of measurable functions $f : \mathscr{X} \to \mathscr{X}$ such that

$$\|f\|_{\mathscr{L}_p(\mu)} = \left(\int_{\mathscr{X}} \|f(x)\|_{\mathscr{X}}^p \, \mathrm{d}\mu(x) \right)^{1/p} < \infty.$$

When $\mathscr{X}$ is separable, $\mathscr{L}_p(\mu)$ is an example of a **Bochner space**, but we will not use this terminology.

It follows from the definition that $\|f\|_{\mathscr{L}_p(\mu)}$ is the L_p norm of the map $x \mapsto \|f(x)\|_{\mathscr{X}}$ from $\mathscr{X}$ to $\mathbb{R}$:

$$\|f\|_{\mathscr{L}_p(\mu)} = \| \ \|f\|_{\mathscr{X}} \ \|_{L_p(\mu)}.$$

As usual we identify functions that equal almost everywhere. Clearly, $\mathscr{L}_p(\mu)$ is a normed vector space. It enjoys another property shared by L_p spaces — completeness:

Theorem 3.4.3 (Riesz–Fischer)**.** *The space $\mathscr{L}_p(\mu)$ is a Banach space.*

Proof. We repeat the proof of the Riesz–Fischer theorem of completeness of L_p spaces. Let f_n be a Cauchy sequence in $\mathscr{L}_p(\mu)$. For each k let n_k be such that $\|f_n - f_m\|_{\mathscr{L}_p(\mu)} < 1/k^2$ if $m, n \geq n_k$. Define $f : \mathscr{X} \to \mathscr{X}$ and $g : \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ by

$$f = f_1 + \sum_{k=1}^{\infty} f_{n_{k+1}} - f_{n_k}, \qquad g(x) = \|f_1(x)\|_{\mathscr{X}} + \sum_{k=1}^{\infty} \|f_{n_{k+1}}(x) - f_{n_k}(x)\|_{\mathscr{X}}.$$

Then $\|f(x)\|_{\mathscr{X}} \leq g(x)$ for all $x \in \mathscr{X}$ and

$$\|f\|_{\mathscr{L}_p(\mu)} \leq \|g\|_{L_p(\mu)} \leq \|f_1\|_{\mathscr{L}_p(\mu)} + \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

This means that for μ-almost every x, $g(x) < \infty$. Since $\mathscr{X}$ is complete, at each such point $f(x)$ is defined and belongs to $\mathscr{X}$. Clearly $\|f(x) - f_{n_k}(x)\|_{\mathscr{X}} \leq g(x)$ and $f_{n_k}(x) \to f(x)$ as $k \to \infty$, μ-almost surely. By the dominated convergence theorem $f_{n_k} \to f$ in $\mathscr{L}_p(\mu)$, and since $\{f_n\}$ is Cauchy it follows that $f_n \to f$. □

Random maps lead naturally to random measures:

Lemma 3.4.4 (push-forward with random maps)**.** *Let $\mu \in \mathscr{W}_p(\mathscr{X})$ and let $\mathbf{t}$ be a random map in $\mathscr{L}_p(\mu)$. Then $\Lambda = \mathbf{t}\#\mu$ is a continuous mapping from $\mathscr{L}_p(\mu)$ to $\mathscr{W}_p(\mathscr{X})$, hence a random measure.*

Proof. That Λ takes values in $\mathscr{W}_p$ follows from a change of variables

$$\int_{\mathscr{X}} \|x\|^p \, \mathrm{d}\Lambda(x) = \int_{\mathscr{X}} \|\mathbf{t}(x)\|^p \, \mathrm{d}\mu(x) = \|\mathbf{t}\|_{\mathscr{L}_p(\mu)} < \infty.$$

Since $W_p(\mathbf{t}\#\mu, \mathbf{s}\#\mu) \le \| \|\mathbf{t}-\mathbf{s}\|_{\mathscr{X}} \|_{L_p(\mu)} = \|\mathbf{t}-\mathbf{s}\|_{\mathscr{L}_p(\mu)}$ (see (3.3)), Λ is a continuous (in fact, 1-Lipschitz) function of $\mathbf{t}$. □

Conversely, $\mathbf{t}$ is a continuous function of Λ:

Lemma 3.4.5 (measurability of transport maps)**.** *Let Λ be a random measure in $\mathscr{W}_p(\mathscr{X})$ and let $\mu \in \mathscr{W}_p(\mathscr{X})$ such that $(\mathbf{i}, \mathbf{t}_\mu^\Lambda)\#\mu$ is the unique optimal coupling of μ and Λ. Then $\Lambda \mapsto \mathbf{t}_\mu^\Lambda$ is a continuous mapping from $\mathscr{W}_p(\mathscr{X})$ to $\mathscr{L}_p(\mu)$, so $\mathbf{t}_\mu^\Lambda$ is a random element in $\mathscr{L}_p(\mu)$. In particular, the result holds if $\mathscr{X}$ is a separable Hilbert space, $p > 1$, and μ is absolutely continuous.*

Proof. This result is more subtle than Lemma 3.4.4, since $\Lambda \mapsto \mathbf{t}_\mu^\Lambda$ is not necessarily Lipschitz.

Suppose that $\Lambda_n \to \Lambda$ in $\mathscr{W}_p(\mathscr{X})$ and fix $\epsilon > 0$. Define the sets

$$B_n = \{x : \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^\Lambda\| \ge \epsilon\},$$

so that

$$\|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^\Lambda\|_{\mathscr{L}_p(\mu)}^p = \int_{\mathscr{X}} \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^\Lambda\|^p \, \mathrm{d}\mu \le \epsilon^p + \int_{B_n} \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^\Lambda\|^p \, \mathrm{d}\mu.$$

Since $\|a-b\|^p \le 2^p\|a\|^p + 2^p\|b\|^p$, the last integral is no larger than

$$2^p \int_{B_n} \|\mathbf{t}_\mu^{\Lambda_n}\|^p \, \mathrm{d}\mu + 2^p \int_{B_n} \|\mathbf{t}_\mu^\Lambda\|^p \, \mathrm{d}\mu = 2^p \int_{(\mathbf{t}_\mu^{\Lambda_n})^{-1}(B_n)} \|x\|^p \, \mathrm{d}\Lambda_n(x) + 2^p \int_{(\mathbf{t}_\mu^\Lambda)^{-1}(B_n)} \|x\|^p \, \mathrm{d}\Lambda(x).$$

Since (Λ_n) and Λ are tight in the Wasserstein space, they must satisfy the absolute uniform continuity (3.7). Let $\delta = \delta_\epsilon$ as in (3.7). Invoking Corollary 5.23 in Villani [89], we see that $\mu(B_n) < \delta$ for all $n > N = N_\epsilon$. By the measure preserving property of the optimal maps, the last two integrals are taken on sets of measures at most δ. Consequently, for all $n > N_\epsilon$,

$$\|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^\Lambda\|_{\mathscr{L}_p(\mu)} \le \epsilon^p + 2^{p+1}\epsilon,$$

and this completes the proof upon letting $\epsilon \to 0$. □

Remark 4. *If $\mathscr{X} = \mathbb{R}^d$, $p = 2$ and μ is absolutely continuous, we can replace B_n above by a compact set S with μ-measure at least $1-\delta$, bound the integral on the complement of S as above (without needing to appeal to [89, Corollary 5.23]), and then use Proposition 2.9.11 to bound the integral on S.*

In Proposition 5.3.6 we show under some conditions that $\|\mathbf{t}_\mu^\Lambda\|_{\mathscr{L}_2(\mu)}$ is a continuous function *of the pair* (μ, Λ).

3.4.2 Random optimal maps and Fubini's theorem

From now on we assume that $\mathscr{X}$ is a separable Hilbert space and that $p = 2$. The results can most likely be generalised to all $p > 1$ (see Ambrosio, Gigli & Savaré [6, Section 10.2]), but we shall not need to do this here.

In some cases (Theorem 3.5.15) we would like to apply Fubini's theorem in the form

$$\mathbb{E}\int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \mathbb{E}\left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \left\langle \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0.$$

In order for this to even make sense, we need to have a meaning for "expectation" in the spaces $\mathscr{L}_2(\theta_0)$ and $L_2(\theta_0)$, both of which are Banach spaces. There are several nonequivalent definitions for integrals in such spaces (Hildebrant [47]); the one which will be the most convenient to us is the Bochner integral.

Definition 3.4.6 (Bochner integral)**.** *Let B be a Banach space and let f be a simple random element taking values in B:*

$$f(\omega) = \sum_{j=1}^{n} f_j \mathbf{1}\{\omega \in \Omega_j\}, \qquad \Omega_j \in \mathscr{F}, \qquad f_j \in B.$$

Then the Bochner integral (or expectation) of f is defined by

$$\mathbb{E}f = \sum_{j=1}^{n} \mathbb{P}(\Omega_j) f_j \in B.$$

If f is measurable and there exists a sequence f_n of simple random elements such that $\|f_n - f\| \to 0$ almost surely and $\mathbb{E}\|f_n - f\| \to 0$, then the Bochner integral of f is defined as the limit

$$\mathbb{E}f = \lim_{n\to\infty} \mathbb{E}f_n.$$

The space of functions for which the Bochner integral is defined is the **Bochner space** $L_1(\Omega; B)$, but we will use neither this terminology nor the notation. It is not difficult to see that Bochner integrals are well-defined: the expectations do not depend on the representation of the simple functions nor on the approximating sequence, and the limit exists in B (because it is complete). More on Bochner integrals can be found in Hsing & Eubank [49, Section 2.6] or Dunford, Schwartz, Bade & Bartle [32, Chapter III.6]. It turns out that separability is quite important in this setting:

Lemma 3.4.7 (approximation of separable functions)**.** *Let $f : \Omega \to B$ be measurable. Then there exists a sequence of simple functions f_n such that $\|f_n(\omega) - f(\omega)\| \to 0$ for almost all ω if and only if $f(\Omega \setminus \mathscr{N})$ is separable for some $\mathscr{N} \subseteq \Omega$ of probability zero. In that case, f_n can be chosen so that $\|f_n(\omega)\| \leq 2\|f(\omega)\|$ for all $\omega \in \Omega$.*

Functions satisfying this approximation condition are sometimes called **strongly measurable**

or **Bochner measurable**. In view of the lemma, we will call them **separately valued**, since this is the condition that will need to be checked in order to define their integrals.

Proof. If f is a limit of simple functions f_n, and $\mathcal{N}$ is the set on which $f_n(\omega)$ does not converge to f, then $f(\Omega \setminus \mathcal{N})$ is included in the closure of the union of $f_n(\Omega \setminus \mathcal{N})$. This is a countable union of finite sets; hence $f(\Omega \setminus \mathcal{N})$ is separable.

Conversely, let (b_j) be dense in $f(\Omega \setminus \mathcal{N})$. For each n, $f(\Omega \setminus \mathcal{N})$ is included in the countable union $\cup_k B_{1/n}(b_k)$. By the monotone convergence theorem, there exists a finite $M = M(n)$ such that the probability that f is in the first M balls is at least $1 - 1/n$. If we make these balls disjoint ($C_1 = B_{1/n}(b_1); C_{k+1} = B_{1/n}(b_{k+1}) \setminus \cup_{j=1}^{k} B_{1/n}(b_j)$) and let

$$f_n(\omega) = \sum_{k=1}^{M(n)} b_k \mathbf{1}\{f(\omega) \in C_k\},$$

then f_n is a simple function and $\mathbb{P}(\|f_n - f\| \geq 1/n) < 1/n$, so that $\|f_n - f\| \to 0$ in probability. Consequently, there exists a subsequence f_{n_k} that converges to f almost surely on $\Omega \setminus \mathcal{N}$. Finally, define $g_n(\omega) = f_n(\omega)\mathbf{1}\{\|f_n(\omega)\| \leq 2\|f(\omega)\|\}$. The sequence (g_{n_k}) satisfies the desired properties. □

Two remarks are in order. Firstly, if B itself is separable, then $f(\Omega)$ will obviously be separable. Secondly, the set $\mathcal{N}' \subset \Omega \setminus \mathcal{N}$ on which (g_{n_k}) does not converge to f may fail to be measurable, but must have outer probability zero (it is included in a measurable set of measure zero) [32, Lemma III.6.9]. This can be remedied by assuming that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete. It will not, however, be necessary to do so, since this measurability issue will not alter the Bochner expectation of f.

Proposition 3.4.8 (Fubini for optimal maps)**.** *Let Λ be a random measure in $\mathcal{W}_2(\mathcal{X})$ such that $\mathbb{E}W_2(\delta_0, \Lambda) < \infty$ and let $\theta_0, \theta \in \mathcal{W}_2(\mathcal{X})$ such that $\mathbf{t}_{\theta_0}^{\Lambda}$ and $\mathbf{t}_{\theta_0}^{\theta}$ exist (and are unique) with probability one. (For example if θ_0 is absolutely continuous.) Then*

$$\mathbb{E}\int_{\mathcal{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathcal{X}} \mathbb{E}\left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathcal{X}} \left\langle \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0. \tag{3.9}$$

Proof. First we remark that projections are continuous: if $\mathbf{t}$ and $\mathbf{s}$ are random functions in $\mathscr{L}_2(\theta_0)$, then $\langle \mathbf{t}, \mathbf{s} \rangle$ is a random function in $L_2(\theta_0)$. Thus, the integral on the left-hand side of (3.9) is a random variable, and so the expectation is taken on $\mathbb{R}$. In the middle integral, the expectation is the Bochner expectation of the random element $\left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle$ in $L_2(\theta_0)$. The expectation on the right-hand side of (3.9) is the Bochner expectation of the random element $\mathbf{t}_{\theta_0}^{\Lambda}$ in $\mathscr{L}_2(\theta_0)$.

Suppose initially that Λ is a simple function, that is

$$\Lambda(\omega) = \sum_{j=1}^{n} \lambda_j \mathbf{1}\{\omega \in \Omega_j\}, \qquad \lambda_j \in \mathscr{W}_2(\mathscr{X}); \qquad \Omega = \bigcup_{j=1}^{n} \Omega_j.$$

If we let $\alpha_j = \mathbb{P}\Omega_j$, then equation (3.9) states that

$$\sum_{j=1}^{n} \alpha_j \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \sum_{j=1}^{n} \alpha_j \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \left\langle \sum_{j=1}^{n} \alpha_j \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0,$$

which is true by linearity and by finiteness of each of the summands:

$$\int_{\mathscr{X}} \left| \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right| \mathrm{d}\theta_0 \le \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0} \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0} = W_2(\theta_0, \lambda_j) W_2(\theta_0, \theta) < \infty.$$

Now suppose that Λ is measurable and $\mathbb{E}W_2(\Lambda, \delta_0) < \infty$. Since $\mathscr{X}$ is separable, the Wasserstein space $\mathscr{W}_2(\mathscr{X})$ is separable too, so $\Lambda(\Omega)$ is separable. But it has been shown that $\Lambda \mapsto \mathbf{t}_{\theta_0}^{\Lambda}$ is continuous from $\mathscr{W}_2(\mathscr{X})$ to $\mathscr{L}_2(\theta_0)$ (Lemma 3.4.5). Consequently, $\mathbf{t}_{\theta_0}^{\Lambda}(\Omega)$ is separable, and by Lemma 3.4.7 there exists a sequence of simple functions $\mathbf{t}_n(\omega)$ that converge to $\mathbf{t}_{\theta_0}^{\Lambda}(\omega)$ for almost every ω and $\|\mathbf{t}_n\|_{\mathscr{L}_2(\theta_0)} \le 2\|\mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)}$. We may assume without loss of generality that $\mathbf{t}_n$ are optimal maps: indeed, define the simple random measures $\Lambda_n = \mathbf{t}_n \# \theta_0$. Then

$$\|\mathbf{t}_{\theta_0}^{\Lambda_n}\|_{\mathscr{L}_2(\theta_0)} = W_2(\Lambda_n, \delta_0) = \|\mathbf{t}_n\|_{\mathscr{L}_2(\theta_0)},$$

and $\Lambda_n(\omega) \to \Lambda(\omega)$ by Lemma 3.4.4, so $\mathbf{t}_{\theta_0}^{\Lambda_n(\omega)} \to \mathbf{t}_{\theta_0}^{\Lambda(\omega)}$ almost surely by Lemma 3.4.5. Thus, $\mathbf{t}_n$ can be replaced by $\mathbf{t}_{\theta_0}^{\Lambda_n}$.

As (3.9) has been established for Λ_n, it suffices to show that each expression of (3.9) equals the limit as $n \to \infty$ of the same expression with Λ replaced by Λ_n.

We begin with the right-hand side. Since for all $\omega \in \Omega$

$$\sup_n \|\mathbf{t}_{\theta_0}^{\Lambda_n(\omega)}\|_{\mathscr{L}_2(\theta_0)} \le 2\|\mathbf{t}_{\theta_0}^{\Lambda(\omega)}\|_{\mathscr{L}_2(\theta_0)} = 2W_2(\Lambda(\omega), \delta_0),$$

and the latter is integrable, it follows from the dominated convergence theorem that $\mathbb{E}\|\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)} \to 0$ and the definition of the Bochner integral implies that

$$\mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} \to \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} \qquad \text{in } \mathscr{L}_2(\theta_0).$$

Consequently

$$\int_{\mathscr{X}} \left| \left\langle \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right| \mathrm{d}\theta_0 \le \sqrt{\int_{\mathscr{X}} \left\| \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} \right\|^2 \mathrm{d}\theta_0} \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0}$$

vanishes as $n \to \infty$, since $\|\mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)} \to 0$.

Next we deal with the middle integral of (3.9). We have by continuity of the projections that almost surely

$$\left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \to \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \qquad \text{in } L_2(\theta_0), \qquad n \to \infty,$$

and as before

$$\sup_n \left\| \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right\|_{L_2(\theta)} \le W_2(\theta_0, \theta) \sup_n \left\| \|\mathbf{t}_{\theta_0}^{\Lambda_n}\|_{\mathscr{X}} \right\|_{L_2(\theta)} \le 2W_2(\theta_0, \theta) \left\| \mathbf{t}_{\theta_0}^{\Lambda} \right\|_{\mathscr{L}_2(\theta_0)}$$

so again the dominated convergence theorem gives

$$\mathbb{E} \left\| \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle - \left\langle \mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right\|_{L_2(\theta_0)} \to 0, \qquad n \to \infty.$$

Of course the same holds if we subtract the identity from $\mathbf{t}_{\theta_0}^{\Lambda_n}$ and $\mathbf{t}_{\theta_0}^{\Lambda}$. The definition of the Bochner integral means that

$$\mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \to \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \qquad \text{in } L_2(\theta),$$

which of course implies

$$\int_{\mathscr{X}} \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 \to \int_{\mathscr{X}} \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0, \qquad n \to \infty.$$

Lastly we treat the left-hand side of (3.9). Define the random variables

$$Y_n = \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0, \qquad Y = \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0.$$

Then again

$$\sup_n |Y_n| \le 2W_2(\theta_0, \theta) \left\| \mathbf{t}_{\theta_0}^{\Lambda} \right\|_{\mathscr{L}_2(\theta_0)}$$

and

$$|Y_n - Y| \le W_2(\theta_0, \theta) \|\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{t}_{\theta_0}^{\Lambda}\|_{L_2(\theta_0)} \to 0, \qquad n \to \infty,$$

so the dominated convergence theorem applies and $\mathbb{E}Y_n \to \mathbb{E}Y$. □

3.4.3 Measurability of the convex potentials in $\mathscr{W}_2$

Let λ be an absolutely continuous measure supported on a convex compact set $K \subset \mathbb{R}^d$ and let $\mathbf{t}$ be a random deformation of K, i.e. a random element in $\mathscr{L}_2(\lambda)$ taking values in K. Then $\Lambda = \mathbf{t}\#\lambda$ is a random measure on K by Lemma 3.4.4. The goal of this subsection is to establish

sufficient conditions for

$$\mathbb{E}\mathbf{t} = \mathbf{i} \qquad \Longrightarrow \qquad \{\lambda\} = \underset{\theta \in \mathscr{W}_2(\mathscr{X})}{\arg\min}\, \mathbb{E} W_2^2(\theta, \Lambda).$$

That is to say, we wish to find conditions that guarantee the implication that if the mean of $\mathbf{t}$ is the identity function, then λ is the Fréchet mean of the random measure Λ (see Section 3.5). This property will allow to "retract" statistical deformation models to the nonlinear Wasserstein space (see Subsection 4.2.2).

The idea is to use the Kantorovich duality and express the Wasserstein distance as the sum of two integrals involving the convex potential ϕ (of which $\mathbf{t}$ is the gradient). In order to do this we need to have ϕ as a measurable function of $\mathbf{t}$; this is the purpose of this subsection.

Without loss of generality, suppose that K is the minimal convex compact set with $\lambda(K) = 1$, i.e. $K = \overline{\text{conv}(\text{supp}\lambda)}$. If the random map $\mathbf{t}$ is optimal, then it is a subgradient of a convex function defined λ-almost everywhere, and that subgradient is nonempty for all $x \in U = \text{int}K$ (see Subsection 2.9.2). In fact, $\mathbf{t}$ is not only well-defined, but also continuous (in the set-valued sense) λ-almost surely. It is therefore not very restrictive to assume that $\mathbf{t}$ is in fact continuous on U. In other words, $\mathbf{t}$ belongs to the space

$$C_b(U, K) = \{f : U \to K; f \text{continuous}\},$$

which, since K is compact, is of course a subset of

$$C_b(U, \mathscr{X}) = \{f : U \to \mathscr{X}; f \text{ continuous and } \|f\| \text{ is bounded}\}.$$

(To simplify we will henceforth write "f is bounded" when $\|f\|$ is a bounded function.) We will therefore explore the properties of random elements $C_b(U, \mathscr{X})$. For ease of reference, we gather up the following assumptions that will be used in the sequel. The compactness can be replaced by weaker conditions (see Remark 6) but this generalisation will not be pursued in the thesis.

Assumptions 2. *Let $\mathscr{X}$ be a separable Hilbert space, and suppose that:*

- *$K \subset \mathscr{X}$ is nonempty, compact and convex;*
- *$U \subset \mathscr{X}$ is convex, contains 0, and has a compact closure;*
- *$\lambda \in \mathscr{W}_2(\mathscr{X})$ is a probability measure satisfying $\lambda(U) = 1$.*

If $\mathscr{X} = \mathbb{R}^d$ and λ is absolutely continuous, then we can take $K = \overline{\text{conv}(\text{supp}\lambda)}$ and $U = \text{int}K$. But if $\mathscr{X}$ is infinite-dimensional, then U cannot be open because open nonempty sets of $\mathscr{X}$ do not have compact closures. The assumption that U contains the origin is for convenience only, the general case being easily deduced via a translation of λ.

Continuous functions from U to K may in general fail to have limits at the boundary of U, so they cannot be extended to functions in $C_b(\overline{U},\mathscr{X})$. This creates some complications, because the space $C_b(U,\mathscr{X})$ is not separable, unlike $C_b(\overline{U},\mathscr{X})$. It is easy to see that $f \in C_b(U,\mathscr{X})$ can be extended to a function in $C_b(\overline{U},\mathscr{X})$ if and only if f is uniformly continuous. In any case, we still have:

Proposition 3.4.9. *The space $C_b(U,\mathscr{X})$ endowed with the supremum norm is a Banach space.*

Proof. Only completeness is not immediate from the definition, and is a straightforward extension of Theorems 2.4.8 and 2.4.9 in Dudley [31].

Let (f_n) be a Cauchy sequence. Then for all x, $(f_n(x))$ is a Cauchy sequence in $\mathscr{X}$ so has a limit $f(x)$. Let $\epsilon > 0$ and let N_ϵ such that $\|f_n - f_m\|_\infty < \epsilon$ for $n, m > N_\epsilon$. For each $x \in U$ choose $m = m(x,\epsilon) > N_\epsilon$ such that $\|f(x) - f_m(x)\|_{\mathscr{X}} < \epsilon$. Then

$$\|f(x) - f_n(x)\|_{\mathscr{X}} \le \|f(x) - f_m(x)\|_{\mathscr{X}} + \|f_n - f_m\|_\infty < 2\epsilon, \qquad n > N_\epsilon.$$

Thus $f_n \to f$ uniformly. Clearly f is bounded, because $\|f\|_\infty \le \|f_n - f\|_\infty + \|f_n\|_\infty$ for $n = N_\epsilon + 1$. To show that f is continuous at x, let $n = N_\epsilon + 1$ and choose $\delta > 0$ such that $\|x - y\|_{\mathscr{X}} < \delta \Longrightarrow \|f_n(x) - f_n(y)\|_{\mathscr{X}} < \epsilon$. Then

$$\|f(x) - f(y)\|_{\mathscr{X}} \le \|f(x) - f_n(x)\|_{\mathscr{X}} + \|f_n(x) - f_n(y)\|_{\mathscr{X}} + \|f_n(y) - f(y)\|_{\mathscr{X}} < 5\epsilon,$$

whenever $\|x - y\|_{\mathscr{X}} < \delta$. Thus $f \in C_b(U,\mathscr{X})$ and the proof is complete. □

Since $C_b(U,\mathscr{X})$ is a Banach space, we can define Bochner integrals on it. Furthermore, K is convex and closed, so $C_b(U,K)$ is a convex closed subset of $C_b(U,\mathscr{X})$. Thus, if a Bochner integrable random element $\mathbf{t} \in C_b(U,\mathscr{X})$ takes values (with probability one) in $C_b(U,K)$, then $\mathbb{E}\mathbf{t}$ is also in $C_b(U,K)$. Now $C_b(U,\mathscr{X})$ is more convenient than $\mathscr{L}_2(\lambda)$ because pointwise evaluations are well-defined. Since $C_b(U,\mathscr{X})$ has a stronger norm than $\mathscr{L}_2(\lambda)$, no measurability issues arise:

Lemma 3.4.10 (random elements in C_b)**.** *Let $\mathbf{t}$ be a random element in $C_b(U,\mathscr{X})$. Then $\mathbf{t}(x)$ is a random variable for any $x \in U$. In addition (the equivalence class of) $\mathbf{t}$ is a random element in $\mathscr{L}_2(\lambda)$ and $\mathbf{t}\#\lambda$ is a random measure in $\mathscr{W}_2(\mathscr{X})$.*

Proof. For all $x \in U$, the evaluation $e_x(f) = f(x)$ is continuous from $C_b(U,\mathscr{X})$ to $\mathscr{X}$. Furthermore

$$\|\mathbf{t} - \mathbf{s}\|^2_{\mathscr{L}_2(\lambda)} = \int_U \|\mathbf{t}(x) - \mathbf{s}(x)\|^2_{\mathscr{X}} \, \mathrm{d}\lambda(x) \le \int_U \|\mathbf{t} - \mathbf{s}\|^2_\infty \, \mathrm{d}\lambda(x) = \|\mathbf{t} - \mathbf{s}\|^2_\infty,$$

so the identity map from $C_b(U,\mathscr{X})$ to $\mathscr{L}_2(\lambda)$ is continuous (in fact, 1-Lipschitz). The first two statements follow as composition of measurable functions. The assertion on $\mathbf{t}\#\lambda$ is a corollary of Lemma 3.4.4. □

For $\mathbf{t} \in C_b(U, \mathscr{X})$ define its "potential" function $\phi : U \to \mathbb{R}$ by the line integral

$$\phi(x) = \phi_{\mathbf{t}}(x) = \int_0^1 \langle \mathbf{t}(sx), x \rangle \, \mathrm{d}s \qquad \text{(well-defined because } U \text{ is convex and contains 0)}$$

and the Legendre transform $\phi^* : K \to \mathbb{R}$ of ϕ by

$$\phi^*(y) = \sup_{x \in U} \langle x, y \rangle - \phi(x).$$

Remark 5. *If $0 \notin U$, we fix another point $x_0 \in U$ and define the potential as*

$$\phi(x) = \frac{1}{2}\|x_0\|^2 + \int_0^1 \langle \mathbf{t}(x_0 + s(x - x_0)), x - x_0 \rangle \, \mathrm{d}s,$$

and all the results still hold modulo this translation.

The following lemma collects some properties of $\phi_{\mathbf{t}}$ and $\phi^*_{\mathbf{t}}$.

Lemma 3.4.11 (convex potentials and measurability)**.** *Let $\mathbf{t} \in C_b(U, \mathscr{X})$. Then ϕ is bounded. If $\mathbf{t}$ is uniformly continuous, then so is ϕ. If $\mathbf{t}$ is the gradient in the Gâteaux sense of some function, then ϕ is Gâteaux differentiable and $\nabla\phi = \mathbf{t}$. Furthermore ϕ^* is uniformly continuous and bounded. Finally, when $\mathbf{t}$ is uniformly continuous, $\phi : C_b(U, \mathscr{X}) \to C_b(U) = C_b(U, \mathbb{R})$ is Lipschitz continuous as well as $\phi \mapsto \phi^*$ from $C_b(U)$ to $C_b(K)$. In particular, $\phi_{\mathbf{t}}$ and $\phi^*_{\mathbf{t}}$ are random elements in $C_b(U)$ and $C_b(K)$.*

Proof. Clearly ϕ is bounded by $\sup_{u \in U} \|\mathbf{t}(u)\|_{\mathscr{X}} \sup_{x \in U} \|x\|_{\mathscr{X}}$. If $\mathbf{t}$ is uniformly continuous, then for any $\epsilon > 0$ there exists $\delta > 0$ such that if $\|x - y\|_{\mathscr{X}} < \delta$ and $x, y \in U$, then $\|\mathbf{t}(x) - \mathbf{t}(y)\|_{\mathscr{X}} < \epsilon$. Since $s \in [0,1]$, $\|sx - sy\|_{\mathscr{X}} \le \|x - y\|_{\mathscr{X}}$. Write

$$\int_0^1 \langle \mathbf{t}(sx), x \rangle \, \mathrm{d}s - \int_0^1 \langle \mathbf{t}(sy), y \rangle \, \mathrm{d}s = \int_0^1 \langle \mathbf{t}(sx) - \mathbf{t}(sy), x \rangle \, \mathrm{d}s + \int_0^1 \langle \mathbf{t}(sy), x - y \rangle \, \mathrm{d}s$$

and notice that if $\|x - y\|_{\mathscr{X}} < \delta$, then the first integral on the right-hand side is bounded by $\epsilon \sup_{x \in U} \|x\|_{\mathscr{X}}$ and the second by $\delta \sup_{v \in U} \|\mathbf{t}(v)\|_{\mathscr{X}}$. Both bounds vanish as ϵ and $\delta \to 0$ and are independent of x, so ϕ is uniformly continuous.

If $\mathbf{t} = \nabla\psi$, then by the mean value theorem, $\psi(x) - \psi(0) = \phi(x)$, so $\nabla\phi = \nabla\psi = \mathbf{t}$.

By the Cauchy–Schwarz inequality,

$$|\phi_{\mathbf{t}}(x) - \phi_{\mathbf{r}}(x)| \le \|\mathbf{t} - \mathbf{r}\|_\infty \|x\|_{\mathscr{X}} \le \|\mathbf{t} - \mathbf{r}\|_\infty \sup_{x \in U} \|x\|_{\mathscr{X}} < \infty,$$

so $\phi : C_b(U, \mathscr{X}) \to C_b(U, \mathbb{R})$ is Lipschitz. It is also obvious that $\|\phi^*_{\mathbf{t}} - \phi^*_{\mathbf{r}}\|_\infty \le \|\phi_{\mathbf{t}} - \phi_{\mathbf{r}}\|_\infty$ and that ϕ^* is bounded on U because ϕ is bounded and both U and K are bounded. Uniform continuity can be verified directly as follows. Fix $\delta > 0$ and $y, z \in K$ with $\|z - y\| < \delta$. Then for

any $\epsilon > 0$ we can pick some $x \in U$ such that

$$\phi^*(z) \le \langle x, z\rangle - \phi(x) + \epsilon = \big\langle x, y\big\rangle - \phi(x) + \epsilon + \big\langle x, z-y\big\rangle \le \phi^*(y) + \epsilon + \delta \sup_{x\in U} \|x\|.$$

Letting $\epsilon \to 0$ and since the supremum is finite, we see that $\phi^*(z) - \phi^*(y) \le \sup_{x\in U}\|x\|\|y-z\|$; interchanging the roles of y and z above shows that in fact $|\phi^*(z) - \phi^*(y)| \le \sup_{x\in U}\|x\|\|y-z\|$, so ϕ^* is even Lipschitz. □

Lemma 3.4.12 (random dual integrals)**.** *Let* $\mathbf{t}$ *be a Bochner integrable random element in* $C_b(U,\mathscr{X})$ *with potential* $\phi_{\mathbf{t}}$ *and its Legendre transform* $\phi^*_{\mathbf{t}}$*, and let* $\Lambda = \mathbf{t}\#\lambda$*. Suppose that with probability one,* $\mathbf{t}$ *takes values in* $C_b(U,K)$*. Then*

$$\int_{\mathscr{X}} \phi^*_{\mathbf{t}}(x)\,\mathrm{d}\Lambda(x) = \int_K \phi^*_{\mathbf{t}}(x)\,\mathrm{d}\Lambda(x)$$

is an integrable random variable.

Proof. Clearly $\Lambda(K) = 1$ almost surely, hence the equality of the integrals. As it has been established that $(\phi^*_{\mathbf{t}}, \Lambda)$ is measurable in the product space (Lemmas 3.4.10 and 3.4.11), it is sufficient to show that the integral above is a continuous function of this pair. By Lemma 3.4.11 $\phi^*_{\mathbf{t}}$ is a random element in $C_b(K)$. Now if $f_n \to f$ in $C_b(K)$ and $\mu_n \to \mu$ narrowly (a fortiori if $\mu_n \to \mu$ in the Wasserstein space), then

$$\int_K f_n\,\mathrm{d}\mu_n - \int_K f\,\mathrm{d}\mu = \int_K (f_n - f)\,\mathrm{d}\mu_n + \int_K f\,\mathrm{d}(\mu_n - \mu) \to 0,$$

because $f_n \to f$ uniformly on K and f is continuous and bounded.

Finally, we have $|\phi(x)| \le \|x\| \sup_{z\in U}\|\mathbf{t}(z)\|$ so $\|\phi\|_\infty \le \|\mathbf{t}\|_\infty \sup_{x\in K}\|x\|$. Thus $\|\phi^*\|_\infty$ is, by the Cauchy–Schwarz inequality, bounded by $\sup_{x\in K}\|x\| \sup_{x\in U}\|x\| + \|\mathbf{t}\|_\infty \sup_{x\in K}\|x\|$. As Λ is a probability measure, the integral is bounded by the same quantity, which has a finite expectation because K and U are bounded and $\mathbb{E}\|\mathbf{t}\|_\infty < \infty$. □

We are now ready to state the next Fubini result:

Proposition 3.4.13 (Fubini for convex potentials)**.** *Let* $\mathbf{t}$ *be a random map in* $C_b(U,K)$ *that is uniformly continuous and let* $\phi = \phi_{\mathbf{t}}$ *the corresponding potential. Then*

$$\mathbb{E}\phi(x) = \int_0^1 \langle \mathbb{E}\mathbf{t}(sx), x\rangle\,\mathrm{d}s, \qquad x \in U, \qquad \textit{and} \quad \mathbb{E}\int_K \phi\,\mathrm{d}\theta = \int_K \mathbb{E}\phi\,\mathrm{d}\theta \qquad \forall \theta \in P(K).$$

Proof. Both equalities hold when $\mathbf{t}$ is a simple function. If $\mathbf{t}$ is separately valued, then there exists a sequence $\mathbf{t}_n$ that converge uniformly to $\mathbf{t}$ with probability one, and $\|\mathbf{t}_n\|_\infty \le 2\|\mathbf{t}\|_\infty$, the expectation of which is smaller than $\sup_{x\in K}\|x\| < \infty$. Hence $\mathbb{E}\mathbf{t}_n \to \mathbb{E}\mathbf{t}$ in the Bochner sense,

which means that $\|\mathbb{E}\mathbf{t}_n \to \mathbb{E}\mathbf{t}\|_\infty \to 0$ and so

$$\int_0^1 \langle \mathbb{E}\mathbf{t}_n(sx), x\rangle \, \mathrm{d}s \to \int_0^1 \langle \mathbb{E}\mathbf{t}(sx), x\rangle \, \mathrm{d}s, \qquad n \to \infty.$$

Furthermore $\|\phi_n - \phi\|_\infty \le \|\mathbf{t}_n - \mathbf{t}\|_\infty$ and $\|\phi_n\|_\infty \le \|\mathbf{t}_n\|_\infty \sup_{x\in U}\|x\|$, which is integrable. It follows that $\mathbb{E}\phi_n(x) \to \mathbb{E}\phi(x)$ by the dominated convergence theorem and the first equality is proven. For the second, we see that $\mathbb{E}\phi_n \to \mathbb{E}\phi$, that $\int \phi_n \, \mathrm{d}\theta \to \int \phi \, \mathrm{d}\theta$ and the integrals with respect to ϕ_n are bounded by the integrable quantity $\|\mathbf{t}\|_\infty \sup_{x\in K}\|x\|$ as above. In the next lemma we show that $\mathbf{t}$ is separately valued. □

Lemma 3.4.14. *The set* $\{\mathbf{t} \in C_b(U,\mathscr{X}) : \mathbf{t}$ *is uniformly continuous*$\}$ *is separable.*

Proof. If $\mathbf{t}$ is uniformly continuous on U, then $\mathbf{t}$ can be extended to a continuous function on the compact set $\overline{U}$. We can therefore assume without loss of generality that U is compact.

Step 1: reduction to real-valued functions. Suppose momentarily that there exists a countable dense subset D of $C(U) = \{f : U \to \mathbb{R} \text{ continuous}\}$. Let $(e_1, \dots,)$ be an orthonormal basis of $\mathscr{X}$. We claim that the countable set

$$\tilde{D} = \bigcup_{n=1}^{\infty} \left\{ x \mapsto \sum_{i=1}^{n} g_i(x) e_i : g_i \in D \right\}$$

is dense in $C_b(U,\mathscr{X})$. Indeed, fix $f \in C_b(U,\mathscr{X})$ and let $\epsilon > 0$. Then $f(U)$ is a compact subset of $\mathscr{X}$. Consequently, there exists n such that $f(U)$ is covered by an ϵ-neighbourhood of the subspace $S = \mathrm{span}(e_1, \dots, e_n)$. (This follows from total boundedness of $f(U)$ and Parseval's equality.) Then $u = \mathrm{Proj}_S \circ f$ is continuous and satisfies $\sup_{x\in U}\|u(x) - f(x)\|_{\mathscr{X}} \le \epsilon$. Let $u_i(x) = \langle u(x), e_i\rangle$, so that $u(x) = \sum_{i=1}^n u_i(x)e_i$. For each i there exists $g_i \in D$ such that $\sup_{x\in U}|u_i(x) - g_i(x)|_{\mathscr{X}} < \epsilon/n$. Then $\tilde{g}(x) = \sum g_i(x)e_i$ is such that $\tilde{g} \in \tilde{D}$ and $\sup_x \|\tilde{g}(x) - u(x)\|_{\mathscr{X}} < \epsilon$, so $\|\tilde{g} - f\|_\infty < 2\epsilon$. Density of $\tilde{D}$ is thus established.

Step 2: existence of D. Let $\{x_k\}$ be a countable dense set of U. Define $f_k(x) = \|x - x_k\|$ for $k \ge 1$ and $f_0(x) \equiv 1$. Then the algebra generated by $\{f_k\}_{k=0}^\infty$ separates points in U and contains the constant functions, so is dense in $C(U)$ by the Stone–Weierstrass theorem. Notice that this result holds whenever U is a compact metric space. □

Remark 6 (beyond compactness)**.** *In view of the application we have in mind (Theorem 4.2.4),* $\overline{U}$ *and* K *were assumed compact. This can certainly be relaxed. For instance, one can endow the space* $C(K,\mathscr{X})$ *with the metric* $d(f,g) = \min(1, \|f - g\|_\infty)$ *that induces the same topology as the infinity norm, for* K *possibly unbounded. Further conditions will then need to be imposed, however, in order to guarantee integrability.*

3.5 Fréchet means in $\mathscr{W}_2$

3.5.1 The Fréchet functional

If H is a Hilbert space (or a closed convex subspace thereof) and $x_1, \dots, x_N \in H$, then the empirical mean $\overline{x}_N = N^{-1} \sum x_i$ minimises the sum of squared distances from the x_i's. That is, if we define

$$F(\theta) = \sum_{i=1}^{N} \|\theta - x_i\|^2, \qquad \theta \in H,$$

then $\theta = \overline{x}_N$ is the unique minimiser of F. This is easily seen by "opening the squares" and writing

$$F(\theta) = F(\overline{x}_N) + N\|\theta - \overline{x}_N\|^2.$$

This property allows one to generalise the notion of mean to non-Hilbertian spaces, such as the Wasserstein space. The generalisation is attributed to Fréchet [36], whence the terminology.

Definition 3.5.1 (empirical Fréchet functional and mean)**.** *The Fréchet functional associated with measures $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathscr{X})$ is*

$$F : \mathscr{W}_2(\mathscr{X}) \to \mathbb{R} \qquad F(\gamma) = \frac{1}{2N} \sum_{i=1}^{N} W_2^2(\gamma, \mu^i), \qquad \gamma \in \mathscr{W}_2(\mathscr{X}). \tag{3.10}$$

A Fréchet mean of $(\mu^1, \dots, \mu^N)$ is a minimiser of F in $\mathscr{W}_2(\mathscr{X})$ (if it exists).

In analysis, a Fréchet mean is often called a **barycentre**. We stick to "Fréchet mean" that is more popular in statistics.

The factor $1/(2N)$ is of course irrelevant for the definition of Fréchet mean. It is introduced in order to have simpler expressions for the derivatives (Theorems 3.5.13 and 3.5.15) and to be compatible with a population version:

Definition 3.5.2 (population Fréchet mean)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$. The Fréchet mean of Λ is the minimiser (if it exists and is unique) of the Fréchet functional*

$$F(\gamma) = \frac{1}{2}\mathbb{E}W_2^2(\gamma, \Lambda), \qquad \gamma \in \mathscr{W}_2(\mathscr{X}). \tag{3.11}$$

The first reference that deals with empirical Fréchet means in $\mathscr{W}_2(\mathbb{R}^d)$ is the seminal paper of Agueh & Carlier [2]. They treat the seemingly more general weighted Fréchet functional

$$F(\gamma) = \frac{1}{2} \sum_{i=1}^{N} w_i W_2^2(\gamma, \mu^i), \qquad 0 \le w_i, \quad \sum_{i=1}^{N} w_i = 1,$$

but, at least conceptually, this generality is superfluous. Indeed, if all the w_i's are rational, then the weighted functional can be encompassed in (3.10) by taking some of the μ^i's to be the same. The case of irrational w_i's is then treated with continuity arguments. Moreover, (3.11) encapsulates (3.10) as well as the weighted version when Λ can take finitely many values. For these reasons, we will only discuss (3.10) and its population counterpart (3.11) in the sequel.

3.5.2 The one-dimensional case

When $\mathscr{X} = \mathbb{R}$, there is a simple expression for the Fréchet mean because $\mathscr{W}_2(\mathbb{R})$ can be imbedded in a Hilbert space. Indeed, recall that

$$W_2(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L_2(0,1)}$$

(see Subsection 3.3.2 or Section 2.6). In view of that, $\mathscr{W}_2(\mathbb{R})$ can be seen as the convex closed subset of $L_2(0,1)$ formed by equivalence classes of left-continuous nondecreasing functions on $(0,1)$: any quantile function is left-continuous and nondecreasing, and any such function G can be seen to be the inverse function of the distribution function, the **right-continuous inverse** of G

$$F(x) = \inf\{t \in (0,1) : G(t) > x\} = \sup\{t \in (0,1) : G(t) \le x\}.$$

If Λ is a random measure in $\mathscr{W}_2(\mathbb{R})$, then F_Λ^{-1} can be viewed as a random element in the Hilbert space $L_2(0,1)$. Let us assume that $\mathbb{E}\|F_\Lambda^{-1}\|^2$ is finite. Then the unique Fréchet mean of F_Λ^{-1} is $\mathbb{E}F_\Lambda^{-1}$ (defined as a Bochner integral). If we can show that $\mathbb{E}F_\Lambda^{-1}$ is a quantile function, it will follow that the Fréchet mean of Λ in $\mathscr{W}_2(\mathbb{R})$ is the measure λ having $\mathbb{E}F_\Lambda^{-1}$ as a quantile function. This is fairly obvious intuitively, and holds trivially in the empirical case (3.10); the reader not interested in the technical details can safely skip to Theorem 3.5.3 below.

We will always take F_Λ^{-1} as the unique left-continuous nondecreasing in the equivalence class. It needs to be shown that $\mathbb{E}F_\Lambda^{-1}$ is left-continuous and nondecreasing. Although F_Λ^{-1} is an element of L_2 where pointwise evaluations are undefined, the left-continuity allows us to define them without resorting to the construction of Section 3.4: for any $t \in (0,1)$, the quantity

$$\lim_{m\to\infty} m\int_{t-1/m}^{t} F_\Lambda^{-1}(u)\,\mathrm{d}u = \lim_{m\to\infty} m\left\langle F_\Lambda^{-1}, 1[t-1/m, t]\right\rangle$$

is a random variable (measurable from $(\Omega,\mathscr{F},\mathbb{P})$ to $\mathbb{R}$), and the limit exists and equals $F_\Lambda^{-1}(t)$ by left-continuity. Furthermore for all m,

$$|m\left\langle F_\Lambda^{-1}, 1[t-1/m, t]\right\rangle| \le \|F_\Lambda^{-1}\|$$

and this is integrable, so the dominated convergence theorem gives

$$\mathbb{E}F_\Lambda^{-1}(t) = \lim_{m\to\infty} m\mathbb{E}\left\langle F_\Lambda^{-1}, 1[t-1/m, t]\right\rangle = \lim_{m\to\infty} m\left\langle \mathbb{E}F_\Lambda^{-1}, 1[t-1/m, t]\right\rangle.$$

(The last inequality is a consequence of Fubini's theorem in the Bochner sense, like in Proposition 3.4.8). Using this, one can easily deduce the desired properties. If $s < t$, then $F_\Lambda^{-1}(s) \le F_\Lambda^{-1}(t)$, so $\mathbb{E}F_\Lambda^{-1}(s) \le \mathbb{E}F_\Lambda^{-1}(t)$. This implies that the sequence is nondecreasing in m. To prove left-continuity, fix $t \in (0,1)$ and $\epsilon > 0$. Pick m such that

$$m\int_{t-1/m}^{t} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u = m\left\langle \mathbb{E}F_\Lambda^{-1}, 1[t-1/m,t]\right\rangle \ge \mathbb{E}F_\Lambda^{-1}(t) - \epsilon/2,$$

and then $\delta < 1/m$ such that

$$m\int_{t-1/m}^{t-\delta} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge m\int_{t-1/m}^{t} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u - \epsilon/2 \ge \mathbb{E}F_\Lambda^{-1}(t) - \epsilon,$$

which exists because the integral converges. If $s \in (t-\delta, t)$ then using the monotonicity of $\mathbb{E}F_\Lambda^{-1}$ again and noticing that $s - 1/k > t - \delta$ for k large, we obtain

$$\mathbb{E}F_\Lambda^{-1}(s) = \lim_{k\to\infty} k\int_{s-1/k}^{s} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge \mathbb{E}F_\Lambda^{-1}(t-\delta) \ge m\int_{t-1/m}^{t-\delta} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge \mathbb{E}F_\Lambda^{-1}(t) - \epsilon,$$

This proves that $\mathbb{E}F_\Lambda^{-1}$ can be viewed as a quantile function, and we can finally conclude:

Theorem 3.5.3 (Fréchet means in $\mathcal{W}_2(\mathbb{R})$)**.** *Let Λ be a random measure in $\mathcal{W}_2(\mathbb{R})$ such that $F(\delta_0) < \infty$. Then the Fréchet mean of Λ is the unique measure λ whose quantile function F_λ^{-1} equals $\mathbb{E}F_\Lambda^{-1}$.*

Interestingly, no regularity is needed in order for the Fréchet mean to be unique. This is not the case for higher dimensions, see Proposition 3.5.8 below. If there is some regularity, then one can state Theorem 3.5.3 in terms of optimal maps, because F_Λ^{-1} is the optimal map from $\mathrm{Leb}|_{[0,1]}$ to Λ. If $\gamma \in \mathcal{W}_2(\mathbb{R})$ is any absolutely continuous (or even just continuous) measure, then Theorem 3.5.3 can be stated as follows: the Fréchet mean of Λ is the measure $[\mathbb{E}\mathbf{t}_\gamma^\Lambda]\#\gamma$. A generalisation of this result to compatible measures (see Definition 3.3.1) is given in Theorem 3.5.21.

3.5.3 Existence and uniqueness

Fréchet means on a general metric space M may fail to exist and even if they do, they are not necessarily unique. Usually, existence proofs are easier: for example, since the Fréchet functional F is continuous on M (as we show below), one often invokes local compactness of M in order to establish existence of a minimiser. Unfortunately, a different strategy is needed when $M = \mathcal{W}_2(\mathcal{X})$, because the Wasserstein space is not locally compact (Proposition 3.2.8).

We will only consider the space $\mathcal{W}_2(\mathcal{X})$; results concerning $\mathcal{W}_p(\mathcal{X})$ for other values of p can be found in Le Gouic & Loubes [42]. The first thing to notice is that F is indeed continuous (this is clear for the empirical version). This property has nothing to do with the Wasserstein geometry and is valid when $\mathcal{W}_2(\mathcal{X})$ is replaced by any metric space. Assume that F is finite at

γ. If θ is any other measure in $\mathscr{W}_2(\mathscr{X})$, write

$$2F(\gamma) - 2F(\theta) = \mathbb{E}[W_2(\gamma,\Lambda) - W_2(\theta,\Lambda)]\,[W_2(\gamma,\Lambda) + W_2(\theta,\Lambda)]$$

so that by the triangle inequality in $\mathscr{W}_2$,

$$2|F(\gamma) - F(\theta)| \le W_2(\gamma,\theta)[2\mathbb{E}W_2(\gamma,\Lambda) + W_2(\theta,\gamma)] \le W_2(\gamma,\theta)[2\mathbb{E}W_2^2(\gamma,\Lambda) + 2 + W_2(\theta,\gamma)].$$

Since $F(\gamma) < \infty$, the right-hand side vanishes as $\theta \to \gamma$ in $\mathscr{W}_2(\mathscr{X})$. Now that we know that F is continuous, the same upper bound shows that it is in fact locally Lipschitz.

Using the lower semicontinuity (3.5), one can prove existence on $\mathbb{R}^d$ rather easily.

Proposition 3.5.4 (existence of Fréchet means)**.** *The Fréchet functional associated with any random measure Λ in $\mathscr{W}_2(\mathbb{R}^d)$ admits a minimiser.*

Proof. The assertion is clear if F is identically infinite. Otherwise, let (γ_n) be a minimising sequence. We wish to show that the sequence is tight. Define $L = \sup_n F(\gamma_n) < \infty$ and observe that since $x \le 1 + x^2$ for all $x \in \mathbb{R}$,

$$\mathbb{E}W_2(\gamma_n,\Lambda) \le 1 + \mathbb{E}W_2^2(\gamma_n,\Lambda) \le 2L + 1, \qquad n = 1,2,\ldots.$$

By the triangle inequality

$$L' = \mathbb{E}W_2(\delta_0,\Lambda) \le W_2(\delta_0,\gamma_1) + \mathbb{E}W_2(\gamma_1,\Lambda) \le W_2(\delta_0,\gamma_1) + 2L + 1$$

so that for all n

$$\left(\int_{\mathbb{R}^d} \|x\|^2 \,\mathrm{d}\gamma_n(x)\right)^{1/2} = W_2(\gamma_n,\delta_0) \le \mathbb{E}W_2(\gamma_n,\Lambda) + \mathbb{E}W_2(\Lambda,\delta_0) \le 2L + 1 + L' < \infty.$$

Since closed and bounded sets in $\mathbb{R}^d$ are compact, it follows that (γ_n) is a tight sequence. We may assume that $\gamma_n \to \gamma$ narrowly, then use (3.5) and Fatou's lemma to obtain

$$2F(\gamma) = \mathbb{E}W_2^2(\gamma,\Lambda) \le \mathbb{E}\liminf_{n\to\infty} W_2^2(\gamma_n,\Lambda) \le \liminf_{n\to\infty} \mathbb{E}W_2^2(\gamma_n,\Lambda) = 2\inf F.$$

Thus γ is a minimiser of F, and existence is established. □

If $\mathscr{X}$ is an infinite dimensional Hilbert space, existence still holds under a compactness assumption. We first prove a result about the support of the Fréchet mean. On $\mathbb{R}^d$ at the empirical level, one can say more about the support (see Corollary 5.5.1).

Proposition 3.5.5 (support of Fréchet mean)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ and let $K \subseteq \mathscr{X}$ be a convex closed set such that $\mathbb{P}[\Lambda(K) = 1] = 1$. If γ minimises F, then $\gamma(K) = 1$.*

Remark 7. *For any closed $K \subseteq \mathscr{X}$ and any $\alpha \in [0,1]$, the set $\{\Lambda \in \mathscr{W}_p(\mathscr{X}) : \Lambda(K) \ge \alpha\}$ is closed*

in $\mathscr{W}_p(\mathscr{X})$ because $\{\Lambda \in P(\mathscr{X}) : \Lambda(K) \geq \alpha\}$ is narrowly closed by the portmanteau lemma (Lemma 2.9.1).

Proof. Let $\mathrm{proj}_K : \mathscr{X} \to K$ denote the projection onto the set K, which is well-defined since K is closed and convex, and of course satisfies

$$\|x - y\| \geq \|x - \mathrm{proj}_K(y)\|, \qquad x \in K, \quad y \in \mathscr{X},$$

with equality if and only if $y \in K$. Let $\pi \in \Pi(\Lambda, \gamma)$ be optimal. By the hypothesis $\Lambda(K) = 1$, so that the above inequality holds for Λ-almost every x and all y, hence π-almost surely. Define the projection $\gamma_K = \mathrm{proj}_K \# \gamma$ of γ onto K, and recall that $\mathbf{i}$ denotes the identity mapping on $\mathscr{X}$. Then $(\mathbf{i} \times \mathrm{proj}_K)\#\pi \in \Pi(\Lambda, \gamma_K)$ and

$$W_2^2(\Lambda, \gamma) = \int_{K \times \mathscr{X}} \|x - y\|^2 \, \mathrm{d}\pi(x, y) \geq \int_{K \times \mathscr{X}} \|x - \mathrm{proj}_K(y)\|^2 \, \mathrm{d}\pi(x, y) \geq W_2^2(\Lambda, \gamma_K).$$

Taking expectations gives $F(\gamma) \geq F(\mathrm{proj}_K \# \gamma)$. Again equality holds if and only if $\gamma(K) = 1$, in which case $\mathrm{proj}_K \# \gamma = \gamma$, which completes the proof. □

Corollary 3.5.6. *If there exists a compact convex K satisfying the hypothesis of Proposition 3.5.5, then the Fréchet functional admits a minimiser supported on K.*

Proof. Proposition 3.5.5 allows us to restrict the domain of F to $\mathscr{W}_2(K)$, the collection of probability measures supported on K. Since this set is compact in $\mathscr{W}_2(\mathscr{X})$ (Corollary 3.2.4), the result follows from continuity of F. □

Next, we turn to uniqueness of Fréchet means. The first step is to exclude infinite values of F. (Of course, this is not needed in the empirical version.) This is yet again a general property that holds for any metric space, not only for $\mathscr{W}_2(\mathscr{X})$.

Lemma 3.5.7 (finiteness of the Fréchet functional)**.** *Suppose that $F(\gamma_0) < \infty$ for some $\gamma_0 \in \mathscr{W}_2(\mathscr{X})$. Then F is finite everywhere on $\mathscr{W}_2(\mathscr{X})$.*

Proof. It follows from the triangle inequality in $\mathscr{W}_2$ that for all γ

$$2F(\gamma) \leq W_2^2(\gamma_0, \gamma) + 2W_2(\gamma, \gamma_0)\mathbb{E}W_2(\gamma_0, \Lambda) + \mathbb{E}W_2^2(\gamma_0, \Lambda) < \infty,$$

because both expectations are finite. □

Example: let (a_n) be a sequence of positive numbers that sum up to one. Let $x_n = 1/a_n$ and suppose that Λ equals $\delta\{x_n\}$ with probability a_n. Then

$$\mathbb{E}W_2^2(\Lambda, \delta_0) = \sum_{n=1}^{\infty} a_n x_n^2 = \sum_{n=1}^{\infty} 1/a_n = \infty,$$

and by Lemma 3.5.7 F is identically infinite. Henceforth, we say that F is finite when the condition in Lemma 3.5.7 holds.

A general situation in which Fréchet means are unique is when the Fréchet functional is strictly convex. Although this is not always the case in the Wasserstein space, at least weak convexity holds: if $\gamma_1, \gamma_2 \in \mathscr{W}_2(\mathscr{X})$, $\pi_i \in \Pi(\gamma_i, \Lambda)$, then the linear interpolant $t\pi_1 + (1-t)\pi_2 \in \Pi(t\gamma_1 + (1-t)\gamma_2, \Lambda)$ for all $t \in [0,1]$; taking π_i to be optimal, we obtain

$$W_2^2(t\gamma_1 + (1-t)\gamma_2, \Lambda) \leq \int_{\mathscr{X}^2} \|x-y\|^2 \, \mathrm{d}[t\pi_1 + (1-t)\pi_2](x,y) = tW_2^2(\gamma_1, \Lambda) + (1-t)W_2^2(\gamma_2, \Lambda). \tag{3.12}$$

Remark 8. *The Wasserstein distance is not convex along geodesics. That is, if we replace the linear interpolant $t\gamma_1 + (1-t)\gamma_2$ by McCann's interpolant, then $t \mapsto W_2^2(\gamma_t, \Lambda)$ is not necessarily convex (see Example 9.1.5 in [6]).*

If we can upgrade (3.12) to strict convexity, then uniqueness of minimisers will be guaranteed. This is the case if Λ is regular enough with positive probability:

Proposition 3.5.8 (uniqueness of Fréchet means)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ with finite Fréchet functional. If Λ is absolutely continuous with positive (inner) probability, then the Fréchet mean of Λ is unique (if it exists).*

This is a particular case of results of Álvarez-Esteban, del Barrio, Cuesta-Albertos & Matrán [4] (see Corollary 2.9 there).

Remark 9. *It is not obvious that the set of absolutely continuous measures is measurable in $\mathscr{W}_2$. We assume that there exists a Borel set $A \subset \mathscr{W}_2$ such that $\mathbb{P}(\Lambda \in A) > 0$ and all measures in A are absolutely continuous.*

Proof of Proposition 3.5.8. By taking expectations in (3.12), one sees that F is convex on $\mathscr{W}_2(\mathscr{X})$ with respect to linear interpolants. Let Λ be absolutely continuous, and let γ_i arbitrary. Then equality in (3.12) holds if and only if $\pi_t = t\pi_1 + (1-t)\pi_2 = (\mathbf{t}_\Lambda^{t\gamma_1 + (1-t)\gamma_2} \times \mathbf{i})\#\Lambda$. But π_t is supported on the graphs of two functions: $\mathbf{t}_\Lambda^{\gamma_1}$ and $\mathbf{t}_\Lambda^{\gamma_2}$. Consequently, equality can hold only if these two maps equal Λ-almost surely, or, equivalently, if $\gamma_1 = \gamma_2$. We can thus conclude that

$$\Lambda \text{ absolutely continuous} \quad \Longrightarrow \quad \gamma \mapsto \frac{1}{2} W_2^2(\gamma, \Lambda) \text{ strictly convex.}$$

As F was already shown to be weakly convex in any case, it follows that

$$\mathbb{P}(\Lambda \text{ absolutely continuous}) > 0 \quad \Longrightarrow \quad F \text{ strictly convex.}$$

Since strictly convex functionals have at most one minimiser, this completes the proof. □

3.5.4 The Agueh–Carlier characterisation

Agueh & Carlier [2] provide a useful sufficient condition for γ being the Fréchet mean. When $\mathscr{X} = \mathbb{R}^d$, this condition is also necessary [2, Proposition 3.8], hence a characterisation of Fréchet means in $\mathbb{R}^d$. It will allow to easily deduce some equivariance results for Fréchet means with respect to independence (Lemma 3.5.10) and rotations (3.5.11). More importantly, it provides a sufficient condition under which a local minimum of F is a global minimum (Theorem 3.5.18) and the same idea can be used to relate the population Fréchet mean to the expected value of the optimal maps (Theorem 4.2.4).

Proposition 3.5.9 (Fréchet means and potentials)**.** *Let $\mu^1, \dots, \mu^N \in \mathscr{W}_2(\mathscr{X})$ be absolutely continuous, let $\gamma \in \mathscr{W}_2(\mathscr{X})$ and denote by ϕ_i^* the convex potentials of $\mathbf{t}_{\mu^i}^{\gamma}$. If $\phi_i = \phi_i^{**}$ are such that*

$$\frac{1}{N}\sum_{i=1}^{N} \phi_i(x) \leq \frac{1}{2}\|x\|^2, \qquad \forall\, x \in \mathscr{X}, \qquad \text{with equality } \gamma\text{-almost surely,}$$

then γ is the unique Fréchet mean of $\mu^1, \dots, \mu^N$.

Proof. Uniqueness follows from Proposition 3.5.8. If $\theta \in \mathscr{W}_2(\mathscr{X})$ is any measure, then the Kantorovich duality yields

$$\begin{aligned} W_2^2(\gamma, \mu^i) &= \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \phi_i(x)\right) \mathrm{d}\gamma(x) + \int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi_i^*(y)\right) \mathrm{d}\mu^i(y); \\ W_2^2(\theta, \mu^i) &\geq \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \phi_i(x)\right) \mathrm{d}\theta(x) + \int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi_i^*(y)\right) \mathrm{d}\mu^i(y). \end{aligned}$$

Summation over i gives the result. □

A population version of this result, based on similar calculations, is shown in Theorem 4.2.4. (The compactness assumption imposed there can be relaxed.)

The next two results are formulated in $\mathbb{R}^d$ because then the converse of Proposition 3.5.9 is proven to be true. If one could extend [2, Proposition 3.8] to any separable Hilbert $\mathscr{X}$, then the two lemmas below will hold with $\mathbb{R}^d$ replaced by $\mathscr{X}$.

Lemma 3.5.10 (independent Fréchet means)**.** *Let $\mu^1, \dots, \mu^N$ and $\nu^1, \dots, \nu^N$ be absolutely continuous measures in $\mathscr{W}_2(\mathbb{R}^{d_1})$ and $\mathscr{W}_2(\mathbb{R}^{d_2})$ with Fréchet means μ and ν respectively. Then the independent coupling $\mu \otimes \nu$ is the Fréchet mean of $\mu^1 \otimes \nu^1, \dots, \mu^N \otimes \nu^N$.*

By induction (or a straightforward modification of the proof), one can show that the Fréchet mean of $(\mu^i \otimes \nu^i \otimes \rho^i)$ is $\mu \otimes \nu \otimes \rho$, and so on.

Proof. Agueh & Carlier [2, Proposition 3.8] show that there exist convex lower semicontinuous potentials ψ_i^* on $\mathbb{R}^{d_1}$ and φ_i^* on $\mathbb{R}^{d_2}$ whose gradients push μ forward to μ^i and ν to ν^i

respectively, and such that

$$\frac{1}{N}\sum_{i=1}^{N}\psi_i(x) \le \frac{1}{2}\|x\|^2, \quad x \in \mathbb{R}^{d_1}; \qquad \frac{1}{N}\sum_{i=1}^{N}\varphi_i(y) \le \frac{1}{2}\|y\|^2, \quad y \in \mathbb{R}^{d_2},$$

with equality μ- and ν-almost surely respectively. Define the convex function $\phi_i : \mathbb{R}^{d_1+d_2} \to \mathbb{R} \cup \{\infty\}$ by $\phi_i(x,y) = \psi_i(x) + \varphi_i(y)$. Then the gradient of

$$\phi_i^*(x,y) = \sup_{u,v} \langle x,u\rangle + \langle y,v\rangle - \psi_i(u) - \varphi_i(v) = \psi_i^*(x) + \varphi_i^*(y)$$

pushes $\mu^i \otimes \nu^i$ forward to $\mu \otimes \nu$ and

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i(x,y) \le \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 = \frac{1}{2}\|(x,y)\|^2, \quad (x,y) \in \mathbb{R}^{d_1+d_2},$$

with equality $\mu\otimes\nu$-almost surely. By Proposition 3.5.9, $\mu\otimes\nu$ is the Fréchet mean of $(\mu^i \otimes \nu^i)$. □

Lemma 3.5.11 (rotated Fréchet means)**.** *If μ is the Fréchet mean of the absolutely continuous measures $\mu^1,\dots,\mu^N$ and U is orthogonal, then $U\#\mu$ is the Fréchet mean of $U\#\mu^1,\dots,U\#\mu^N$.*

Bonneel, Rabin, Peyré & Pfister sketch a proof of this statement in [21, Proposition 1], and it also appears implicitly in Boissard, Le Gouic & Loubes [20, Proposition 4.1]; we give an alternative argument here.

Proof. If $x \mapsto \phi(x)$ is convex, then $x \mapsto \phi(U^{-1}x)$ is convex with gradient $U\nabla\phi(U^{-1}x)$ at (almost all) x and conjugate $x \mapsto \phi^*(U^{-1}x)$. If ϕ_i are convex potentials with $\nabla\phi_i^*\#\mu = \mu^i$, then $\nabla[(\phi_i \circ U^{-1}]^*) = \nabla(\phi_i^* \circ U^{-1})$ pushes $U\#\mu$ forward to $U\#\mu^i$ and by [2, Proposition 3.8]

$$\frac{1}{N}\sum_{i=1}^{N}(\varphi_i \circ U^{-1})(Ux) = \frac{1}{N}\sum_{i=1}^{N}\phi_i(x) \le \frac{1}{2}\|x\|^2 = \frac{1}{2}\|Ux\|^2$$

with equality for μ-almost any x. A change of variables $y = Ux$ shows that the set of points y such that $\sum(\varphi_i \circ U^{-1})(y) < N\|y\|^2/2$ is $(U\#\mu)$-negligible, completing the proof. □

3.5.5 Differentiability of the Fréchet functional and Karcher means

Since we seek to minimise the Fréchet functional F, it would be helpful if F were differentiable, because we could then find at least local minima by solving the equation $F' = 0$. This observation of Karcher [55] leads to the notion of **Karcher mean**.

Definition 3.5.12 (Karcher mean)**.** *Let F be a Fréchet functional associated with some random measure Λ in $\mathscr{W}_2(\mathscr{X})$. Then γ is a Karcher mean for Λ if F is differentiable at γ and $F'(\gamma) = 0$.*

Of course, if γ is a Fréchet mean for Λ and F is differentiable at γ, then $F'(\gamma)$ must vanish. In

this subsection we build upon the work of Ambrosio, Gigli & Savaré [6] and determine the derivative of the Fréchet functional. This will not only allow for a simple characterisation of Karcher means in terms of the optimal maps $\mathbf{t}_\gamma^\Lambda$ (Proposition 3.5.16), but will also be the cornerstone of the construction of a steepest descent algorithm for empirical calculation of Fréchet means (see Section 5.1).

It turns out that the tangent bundle structure described in Section 3.3 gives rise to a differentiable structure in the Wasserstein space. Fix $\mu^0 \in \mathcal{W}_2(\mathcal{X})$ and consider the function

$$F_0 : \mathcal{W}_2(\mathcal{X}) \to \mathbb{R}, \qquad F_0(\gamma) = \frac{1}{2} W_2^2(\gamma, \mu^0).$$

Ambrosio, Gigli & Savaré [6, Corollary 10.2.7] show that when γ is absolutely continuous,

$$\lim_{W_2(\nu,\gamma)\to 0} \frac{F_0(\nu) - F_0(\gamma) + \int_{\mathcal{X}} \left\langle \mathbf{t}_\gamma^{\mu^0}(x) - x, \mathbf{t}_\gamma^\nu(x) - x \right\rangle \mathrm{d}\gamma(x)}{W_2(\nu,\gamma)} = 0.$$

Parts of the proof of this result (the limit superior above is ≤ 0; the limit inferior is bounded below) are reproduced in Proposition 3.5.14. The integral above can be seen as the inner product

$$\left\langle \mathbf{t}_\gamma^{\mu^0} - \mathbf{i}, \mathbf{t}_\gamma^\nu - \mathbf{i} \right\rangle$$

in the space $L_2(\gamma)$ that includes as a (closed) subspace the tangent space Tan_γ. In terms of this inner product and the log map, we can write

$$F_0(\nu) - F_0(\gamma) = -\left\langle \log_\gamma(\mu^0), \log_\gamma(\nu) \right\rangle + o(W_2(\nu,\gamma)), \qquad \nu \to \gamma \quad \text{in } \mathcal{W}_2,$$

so that F_0 is Fréchet-differentiable[2] at γ with derivative

$$F_0'(\gamma) = -\log_\gamma(\mu^0) = -\left(\mathbf{t}_\gamma^{\mu^0} - \mathbf{i}\right) \in \mathrm{Tan}_\gamma.$$

By linearity, one immediately obtains:

Theorem 3.5.13 (gradient of the Fréchet functional)**.** *Fix a collection of measures $\mu^1, \ldots, \mu^N \in \mathcal{W}_2(\mathcal{X})$. When $\gamma \in \mathcal{W}_2(\mathcal{X})$ is absolutely continuous, the Fréchet functional*

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^N W_2^2(\gamma, \mu^i), \qquad \gamma \in \mathcal{W}_2(\mathcal{X})$$

is Fréchet-differentiable and

$$F'(\gamma) = -\frac{1}{N} \sum_{i=1}^N \log_\gamma(\mu^i) = -\frac{1}{N} \sum_{i=1}^N \left(\mathbf{t}_\gamma^{\mu_i} - \mathbf{i}\right).$$

[2] The notion of Fréchet derivative is also named after Fréchet, but is not directly related to Fréchet means.

We now wish to extend this result to the population version (3.11). This will follow immediately if we can interchange the expectation and the derivative in the form

$$F'(\gamma) = \frac{1}{2}(\mathbb{E}W_2^2)'(\gamma,\Lambda) = \mathbb{E}\left(\frac{1}{2}W_2^2\right)'(\gamma,\Lambda) = -\mathbb{E}(\mathbf{t}_\gamma^\Lambda - \mathbf{i}).$$

In order to do this we will use dominated convergence in conjunction with uniform bounds on the slopes

$$u(\theta,\Lambda) = \frac{0.5W_2^2(\theta,\Lambda) - 0.5W_2^2(\theta_0,\Lambda) + \int_{\mathscr{X}}\langle \mathbf{t}_{\theta_0}^\Lambda - i, \mathbf{t}_{\theta_0}^\theta - i\rangle \,\mathrm{d}\theta_0}{W_2(\theta,\theta_0)}, \qquad u(\theta_0,\Lambda) = 0. \tag{3.13}$$

Proposition 3.5.14 (slope bounds)**.** *Let θ_0, Λ and θ be probability measures with θ_0 absolutely continuous, and set $\delta = W_2(\theta,\theta_0)$. Then*

$$\frac{1}{2}\delta - W_2(\theta_0,\Lambda) - \sqrt{2W_2^2(\theta_0,\delta_0) + 2W_2^2(\Lambda,\delta_0)} \le u(\theta,\Lambda) \le \frac{1}{2}\delta,$$

where u is defined by (3.13)*. If the measures are compatible in the sense of Definition 3.3.1 then in fact $u(\theta,\Lambda) = \delta/2$.*

Proof. We repeat the calculations of Ambrosio, Gigli & Savaré (Theorem 10.2.2 and Proposition 10.2.6) for the particular case $p = 2$. Define a three-coupling $\boldsymbol{\mu} = (\mathbf{i}, \mathbf{t}_{\theta_0}^\Lambda, \mathbf{t}_{\theta_0}^\theta)\#\theta_0 \in P(\mathscr{X}^3)$ and notice that its relevant projections are optimal couplings of (θ_0,Λ) and (θ_0,θ) but not necessarily of (Λ,θ). By definition

$$\int_{\mathscr{X}}\langle \mathbf{t}_{\theta_0}^\Lambda - \mathbf{i}, \mathbf{t}_{\theta_0}^\theta - \mathbf{i}\rangle \,\mathrm{d}\theta_0 = \int_{\mathscr{X}^3}\langle x_2 - x_1, x_3 - x_1\rangle \,\mathrm{d}\boldsymbol{\mu}; \quad W_2^2(\theta_0,\Lambda) = \int_{\mathscr{X}^3}\|x_2 - x_1\|^2 \,\mathrm{d}\boldsymbol{\mu};$$
$$\delta^2 = W_2^2(\theta_0,\theta) = \int_{\mathscr{X}^3}\|x_1 - x_3\|^2 \,\mathrm{d}\boldsymbol{\mu}; \quad W_2^2(\theta\ ,\Lambda) \le \int_{\mathscr{X}^3}\|x_2 - x_3\|^2 \,\mathrm{d}\boldsymbol{\mu}.$$

Integrating the equality

$$\frac{1}{2}\|x_2 - x_3\|^2 - \frac{1}{2}\|x_2 - x_1\|^2 + \langle x_2 - x_1, x_3 - x_1\rangle = \frac{1}{2}\|x_1 - x_3\|^2 \tag{3.14}$$

with respect to $\boldsymbol{\mu}$ yields the second inequality of the proposition. If the measures are compatible then the relevant marginal of $\boldsymbol{\mu}$ is optimal for (Λ,θ), and the inequality holds as equality.

For the other inequality, let $\boldsymbol{\beta}$ be another three-coupling that optimally couples (θ_0,θ) and (Λ,θ). Then

$$W_2^2(\theta,\Lambda) = \int_{\mathscr{X}^3}\|x_2 - x_3\|^2 \,\mathrm{d}\boldsymbol{\beta} \quad \text{and} \quad W_2^2(\theta_0,\Lambda) \le \int_{\mathscr{X}^3}\|x_1 - x_2\|^2 \,\mathrm{d}\boldsymbol{\beta}.$$

Integration of (3.14) with respect to $\boldsymbol{\beta}$ yields

$$\frac{1}{2}W_2^2(\theta,\Lambda) - \frac{1}{2}W_2^2(\theta_0,\Lambda) \ge \frac{1}{2}\delta^2 - \int_{\mathscr{X}^3}\langle x_2 - x_1, x_3 - x_1\rangle \,\mathrm{d}\boldsymbol{\beta}.$$

All that remains is to bound the last displayed integral by a constant times δ, when the integral is taken with respect to either $\boldsymbol{\beta}$ or $\boldsymbol{\mu}$. To this end, we apply the Cauchy–Schwarz inequality

$$\left|\int_{\mathscr{X}^3}\langle x_2-x_1,x_3-x_1\rangle\,\mathrm{d}\boldsymbol{\mu}\right| \le \sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\boldsymbol{\mu}}\sqrt{\int_{\mathscr{X}^3}\|x_3-x_1\|^2\,\mathrm{d}\boldsymbol{\mu}} = \delta W_2(\theta_0,\Lambda),$$
$$\left|\int_{\mathscr{X}^3}\langle x_2-x_1,x_3-x_1\rangle\,\mathrm{d}\boldsymbol{\beta}\right| \le \sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\boldsymbol{\beta}}\sqrt{\int_{\mathscr{X}^3}\|x_3-x_1\|^2\,\mathrm{d}\boldsymbol{\beta}}$$

where the last displayed square root again equals δ, and

$$\sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\boldsymbol{\beta}} \le \sqrt{\int_{\mathscr{X}^3}2\|x_1\|^2\,\mathrm{d}\boldsymbol{\beta}+\int_{\mathscr{X}^3}2\|x_2\|^2\,\mathrm{d}\boldsymbol{\beta}} = \sqrt{2W_2^2(\theta_0,\delta_0)+2W_2^2(\Lambda,\delta_0)}.$$

This completes the proof. □

Theorem 3.5.15 (population Fréchet gradient)**.** *Let Λ be a random measure with finite Fréchet functional F. Then F is Fréchet-differentiable at any absolutely continuous θ_0 in the Wasserstein space, and $F'(\theta_0) = \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i} \in L_2(\theta_0)$. More precisely,*

$$\frac{F(\theta)-F(\theta_0)+\int_{\mathscr{X}}\langle \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}-\mathbf{i},\mathbf{t}_{\theta_0}^{\theta}-\mathbf{i}\rangle\,\mathrm{d}\theta_0}{W_2(\theta,\theta_0)} \to 0, \qquad \theta\to\theta_0 \quad \textit{in } \mathscr{W}_2.$$

Thus, the Fréchet derivative of F can be identified with the map $-(\mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}-\mathbf{i})$ in the tangent space at θ_0, a subspace of $L_2(\theta_0)$.

In particular, the conclusion of the theorem holds if $\Lambda(K)=1$ almost surely for some non random bounded set K.

Proof. Introduce the slopes $u(\theta,\Lambda)$ defined by (3.13). Then for all Λ, $u(\theta,\Lambda)\to 0$ as $W_2(\theta,\theta_0)\to 0$, by the differentiability properties established above. Let us show that $\mathbb{E}u(\theta,\Lambda)\to 0$ as well. By Proposition 3.5.14, the expectation of u is bounded above by a constant that does not depend on Λ, and below by the negative of

$$\mathbb{E}W_2(\theta_0,\Lambda)+\mathbb{E}\sqrt{2W_2^2(\theta_0,\delta_0)+2W_2^2(\Lambda,\delta_0)} \le \sqrt{2}W_2(\theta_0,\delta_0)+\mathbb{E}W_2(\theta_0,\Lambda)+\sqrt{2}\mathbb{E}W_2(\Lambda,\delta_0).$$

Both expectations are finite by the hypothesis on Λ because the Fréchet functional is finite. The dominated convergence theorem yields

$$\mathbb{E}u(\theta,\Lambda) = \frac{F(\theta)-F(\theta_0)+\mathbb{E}\int_{\mathscr{X}}\langle \mathbf{t}_{\theta_0}^{\Lambda}-i,\mathbf{t}_{\theta_0}^{\theta}-i\rangle\,\mathrm{d}\theta_0}{W_2(\theta,\theta_0)} \to 0, \qquad W_2(\theta_0,\theta)\to 0.$$

The measurability of the integral and the result then follow from Fubini's theorem (see Proposition 3.4.8). □

Proposition 3.5.16. *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ with finite Fréchet functional F, and*

let γ be absolutely continuous in $\mathscr{W}_2(\mathscr{X})$. Then γ is a Karcher mean of Λ if and only if $\mathbb{E}\mathbf{t}_\gamma^\Lambda - \mathbf{i} = 0$ in $L_2(\gamma)$. Furthermore, if γ is a Fréchet mean of Λ, then it is also a Karcher mean.

Proof. The characterisation of Karcher means follows immediately from Theorem 3.5.15. Suppose that $F'(\gamma) \neq 0$ and define $\mathbf{t} = \mathbb{E}\mathbf{t}_\gamma^\Lambda$ and $W = \mathbf{t} - \mathbf{i}$. If we show that actually $\mathbf{t} = \mathbf{t}_\gamma^{\#\gamma}$, i.e. that $\mathbf{t}$ is optimal, then the result will follow immediately. Indeed, if we set $\nu_s = [\mathbf{i} + s(W - \mathbf{i})]\#\gamma$, then $W_2(\nu_s, \gamma) = s\|W\|_{\mathscr{L}_2(\gamma)}$ for $s \in [0,1]$ and by Theorem 3.5.15,

$$0 = \lim_{s\to 0^+} \frac{F(\nu_s) - F(\gamma) + \int_{\mathscr{X}} \langle W(x), sW(x)\rangle \, \mathrm{d}\gamma(x)}{s\|W\|_{\mathscr{L}_2(\gamma)}} = \lim_{s\to 0^+} \frac{F(\nu_s) - F(\gamma)}{s\|W\|_{\mathscr{L}_2(\gamma)}} + \|W\|_{\mathscr{L}_2(\gamma)}.$$

Since $\|W\|_{\mathscr{L}_2(\gamma)} > 0$, this means that $F(\nu_s) - F(\gamma)$ is negative when s is small, and therefore γ cannot be the Fréchet mean.

Let us now show the optimality of $\mathbf{t}$. If Λ is a simple random measure, then the result follows immediately. Otherwise, there exists a sequence of simple optimal maps $\mathbf{t}_n$ that converge to $\mathbf{t}$ in $\mathscr{L}_2(\gamma)$ (see the proof of Proposition 3.4.8). Let us show that $\mathbf{t}$ is monotone. There exists a set B with $\gamma(B) = 1$ such that

$$\big\langle \mathbf{t}_n(y) - \mathbf{t}_n(x), y - x\big\rangle \geq 0, \qquad x, y \in B \quad n = 1, 2, \ldots.$$

Fix an integer k, let $R = R_k$ such that $\gamma[B_R(0)] \geq 1 - 1/k$ and define $D_k \subseteq \mathscr{X}^2$ by

$$D_k = \big\{(x,y) : x, y \in B \cap B_R(0), \quad \big\langle \mathbf{t}(y) - \mathbf{t}(x), y - x\big\rangle < -2/k\big\}.$$

If $(x,y) \in D_k$ then

$$\|\mathbf{t}_n(x) - \mathbf{t}(x)\| \geq \frac{1/k}{\|x - y\|} \geq \frac{1/k}{2R}$$

or that same lower bound holds for $\|\mathbf{t}_n(y) - \mathbf{t}(y)\|$. By Markov's inequality and since $\|\mathbf{t}_n - \mathbf{t}\|_{\mathscr{L}_2(\gamma)} \to 0$, when $n \geq N_k$ is large enough this happens with γ measure at most $1/k$. Define

$$B_k = B \cap B_R(0) \cap \{x : \|\mathbf{t}_n(x) - \mathbf{t}(x)\| \leq 1/(2Rk)\}, \qquad n = N_k.$$

Then $\gamma(B_k) \geq 1 - 2/k$ and

$$\big\langle \mathbf{t}(y) - \mathbf{t}(x), y - x\big\rangle \geq -\frac{2}{k}, \qquad x, y \in B_k.$$

If we now set

$$B' = \cap_{j=1}^\infty \cup_{k=j}^\infty B_k,$$

then $\gamma(B') = 1$ and $\big\langle \mathbf{t}(y) - \mathbf{t}(x), y - x\big\rangle \geq 0$ for all $x, y \in B'$. Similarly one shows that $\mathbf{t}$ is cyclically monotone, that is, the measure $\pi = (\mathbf{i}, \mathbf{t})\#\gamma$ is cyclically monotone, hence optimal by

Proposition 2.9.5. □

These results will be of more use if we can show that the Fréchet mean is absolutely continuous, because only then F is provably differentiable. This is provided by the following proposition of Agueh & Carlier [2, Proposition 5.1], at least for $\mathbb{R}^d$.

Proposition 3.5.17 (L_∞-regularity of Fréchet means)**.** *Let $\mu^1,\dots,\mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ and suppose that μ^1 is absolutely continuous with density bounded by M. Then the Fréchet mean of $\{\mu^i\}$ is absolutely continuous with density bounded by $N^d M$ and is consequently a Karcher mean.*

We prove a population version of Proposition 3.5.17 in Theorem 5.6.2.

It may happen that a collection $\mu^1,\dots,\mu^N$ of absolutely continuous measures have a Karcher mean that is not a Fréchet mean; an example in $\mathbb{R}^2$ is given in Álvarez-Esteban, del Barrio, Cuesta-Albertos & Matrán [5]. But a Karcher mean γ is "almost" a Fréchet mean in the following sense. By Proposition 3.5.16, $N^{-1}\sum \mathbf{t}_\gamma^{\mu^i}(x) = x$ for γ-almost all x. If, on the other hand, the equality holds *for all* $x \in \mathscr{X}$, then γ is the Fréchet mean by taking integrals and applying Proposition 3.5.9. One can hope that under regularity conditions, the γ-almost sure equality can be upgraded to equality everywhere. Indeed, this is the case:

Theorem 3.5.18 (optimality criterion for Karcher means)**.** *Let $U \subseteq \mathbb{R}^d$ be an open convex set and let $\mu^1,\dots,\mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ be probability measures on U with bounded strictly positive densities $g^1,\dots,g^N$. Suppose that an absolutely continuous Karcher mean γ is supported on U with bounded strictly positive density f there. Then γ is the Fréchet mean of $\mu^1,\dots,\mu^N$ if one of the following holds:*

1. *$U = \mathbb{R}^d$ and the densities $f, g^1,\dots,g^N$ are of class C^1 (or C^α for some $\alpha > 0$);*
2. *U is bounded and the densities $f, g^1,\dots,g^N$ are bounded below on U.*

Proof. The result exploits Caffarelli's regularity theory for Monge–Ampère equations in the form of Theorem 2.8.2. In the first case, there exist C^1 (in fact, $C^{2,\alpha}$) convex potentials φ_i on $\mathbb{R}^d$ with $\mathbf{t}_\gamma^{\mu^i} = \nabla\varphi_i$, so that $\mathbf{t}_\gamma^{\mu^i}(x)$ is a singleton for all $x \in \mathbb{R}^d$. The set $\{x \in \mathbb{R}^d : \sum \mathbf{t}_\gamma^{\mu^i}(x)/N \neq x\}$ is γ-negligible (and hence Lebesgue-negligible) and open by continuity. It is therefore empty, so $F'(\gamma) = 0$ everywhere, and γ is the Fréchet mean (see the discussion before the theorem).

In the second case, by the same argument we have $\sum \mathbf{t}_\gamma^{\mu^i}(x)/N = x$ for all $x \in U$. Since U is convex, there must exist a constant C such that $\sum \varphi_i(x) = C + N\|x\|^2/2$ for all $x \in U$, and we may assume without loss of generality that $C = 0$. If one repeats the proof of Proposition 3.5.9, then $F(\gamma) \leq F(\theta)$ for all $\theta \in P(U)$. By continuity considerations the inequality holds for all $\theta \in P(\overline{U})$ (Theorem 3.2.6) and since $\overline{U}$ is closed and convex, γ is the Fréchet mean by Proposition 3.5.5. □

3.5.6 Relation to multimarginal formulation and the compatible case

In [38] Gangbo & Święch consider a *multimarginal* Monge–Kantorovich problem in the following sense. Let $\mu^1,\dots,\mu^N$ be N measures in $\mathscr{W}_2(\mathscr{X})$ and let $\Pi(\mu^1,\dots,\mu^N)$ be the set of probability measures in $\mathscr{X}^N$ having μ^i as marginals. The problem is to minimise

$$G(\pi)=\frac{1}{2N^2}\int_{\mathscr{X}^N}\sum_{i<j}\|x_i-x_j\|^2\,\mathrm{d}\pi(x_1,\dots,x_N),\qquad\text{over}\quad\pi\in\Pi(\mu^1,\dots,\mu^N).$$

The factor $1/(2N^2)$ is of course irrelevant for the minimisation and its purpose will be clarified shortly. If $N=2$ we obtain the Kantorovich problem with quadratic cost. The probabilistic interpretation (as in Section 2.2) is that one is given random variables $X_1,\dots,X_N$ with probability laws $\mu^1,\dots,\mu^N$ and one seeks a joint distribution, say Z, on $\mathscr{X}^N$ minimising

$$\frac{1}{2N^2}\mathbb{E}_Z\sum_{i<j}\|X_i-X_j\|^2.$$

We refer to elements of $\Pi(\mu^1,\dots,\mu^N)$ as **multicouplings** (of $\mu^1,\dots,\mu^N$). Just like in the Kantorovich problem, there always exists an optimal multicoupling π. Given the results of Section 2.5, it should not come as a surprise that if $\mu^1,\dots,\mu^N$ are all absolutely continuous in $\mathbb{R}^d$, then π is unique, and takes the form

$$\pi=(\mathbf{i},\mathbf{s}_2,\dots,\mathbf{s}_N)\#\mu^1,$$

for some functions $\mathbf{s}_2,\dots,\mathbf{s}_N:\mathbb{R}^d\to\mathbb{R}^d$. In probabilistic terms, the optimal coupling Z is the vector $(X_1,\mathbf{s}_2(X_2),\dots,\mathbf{s}_N(X_N))$. The functions $\mathbf{s}_j$ are not gradients of convex functions, but are rather of the form (provided some extra regularity holds) $\mathbf{t}_j^{-1}\circ\mathbf{t}_1$ with $\mathbf{t}_j$ gradients of convex functions. In other words, there exists a measure $\rho=\mathbf{t}_1\#\mu^1$ such that the optimal π is

$$\pi=\left(\mathbf{t}_\rho^{\mu^1},\dots,\mathbf{t}_\rho^{\mu^N}\right)\#\rho.$$

It is tempting to conjecture that ρ is the Fréchet mean of $\mu^1,\dots,\mu^N$, but this need not be the case. As the one-dimensional case shows, ρ is not unique: one can take any absolutely continuous measure θ and notice that the compatibility of all measures in $\mathscr{W}_2(\mathbb{R})$ gives

$$\left(\mathbf{t}_\rho^{\mu^1},\dots,\mathbf{t}_\rho^{\mu^N}\right)\#\rho=\left(\mathbf{t}_\rho^{\mu^1}\circ\mathbf{t}_\theta^{\rho},\dots,\mathbf{t}_\rho^{\mu^N}\circ\mathbf{t}_\theta^{\rho}\right)\#\theta=\left(\mathbf{t}_\theta^{\mu^1},\dots,\mathbf{t}_\theta^{\mu^N}\right)\#\theta.$$

As we show below, the measure ρ *can* be the Fréchet mean, and any other θ must be such that

$$\mathbf{t}_\rho^{\mu^j}\circ\mathbf{t}_\theta^{\rho}=\mathbf{t}_\theta^{\mu^j},\qquad j=1,\dots,N.$$

We see that compatibility of measures is strongly related to the multimarginal problem. Indeed, it turns out that if (μ^i) are compatible, then the $\mathbf{s}_j$'s *are* the optimal maps from μ^1 to μ^j, that is $\mathbf{s}_j=\mathbf{t}_{\mu^1}^{\mu^N}$.

Let us now show how the multimarginal problem is equivalent to the problem of finding the Fréchet mean of $\mu^1, \dots, \mu^N$. The first thing to observe is that the objective function can be written as

$$G(\pi) = \int_{\mathscr{X}^N} \frac{1}{2N} \sum_{i=1}^N \|x_i - M(x)\|^2 \, \mathrm{d}\pi(x), \qquad M(x) = M(x_1, \dots, x_n) = \frac{1}{N} \sum_{i=1}^N x_i.$$

As Agueh & Carlier [2, Proposition 4.2] show (in $\mathbb{R}^d$ but their proof extends as is to any Hilbert space), solving the multimarginal problem gives the Fréchet mean:

Proposition 3.5.19 (Fréchet means via multicouplings)**.** *Let π be a solution to the multimarginal problem with marginals $\mu^1, \dots, \mu^N$. Then $\gamma = M\#\pi$ is a Fréchet mean of $\mu^1, \dots, \mu^N$ and*

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^N W_2^2(\gamma, \mu^i) = \int_{\mathscr{X}^N} \frac{1}{2N} \sum_{i=1}^N \|x_i - M(x_1, \dots, x_N)\|^2 \, \mathrm{d}\pi(x_1, \dots, x_N) = G(\pi).$$

Proof. Let $\pi \in \Pi(\mu^1, \dots, \mu^N)$ and define $\gamma = M\#\pi$. Then for all i

$$\int_{\mathscr{X}^N} \|x_i - M(x_1, \dots, x_N)\|^2 \, \mathrm{d}\pi(x_1, \dots, x_N) \geq W_2^2(\mu^i, \gamma),$$

because the relevant projection of π to $\mathscr{X}^2$ is a coupling of μ^i and γ. Summation over i yields $F(\gamma) \leq G(\pi)$.

Now let γ be any absolutely continuous measure, so that $\mathbf{t}_i = \mathbf{t}_\gamma^{\mu^i}$ is well-defined for all i, and set $\pi = (\mathbf{t}_1, \dots, \mathbf{t}_N)\#\gamma$. Let $\bar{\mathbf{t}} = N^{-1} \sum \mathbf{t}_i$, so that

$$\sum_{i=1}^N \|\mathbf{t}_i(x) - \bar{\mathbf{t}}(x)\|^2 \leq \sum_{i=1}^N \|\mathbf{t}_i(x) - x\|^2, \qquad x \in \mathscr{X}.$$

Integration with respect to γ and a change of variables yield

$$G(\pi) = \int_{\mathscr{X}} \frac{1}{2N} \sum_{i=1}^N \|\mathbf{t}_i(x) - \bar{\mathbf{t}}(x)\|^2 \, \mathrm{d}\gamma(x) \leq \int_{\mathscr{X}} \frac{1}{2N} \sum_{i=1}^N \|\mathbf{t}_i(x) - x\|^2 \, \mathrm{d}\gamma(x) = F(\gamma).$$

From the two established inequalities we see that

$$\inf_{\gamma \text{ absolutely continuous}} F(\gamma) \geq \inf_\pi G(\pi) \geq \inf_\gamma F(\gamma).$$

But the two infima over γ are equal, since F is continuous and absolutely continuous measures constitute a dense set in $\mathscr{W}_2(\mathscr{X})$[3]. □

[3] For $\mathscr{X} = \mathbb{R}^d$ this was shown in Theorem 3.2.6, but the idea of convolving with Gaussian measures works for $\mathscr{X}$ separable Hilbert space. Agueh & Carlier use a more direct approach without resorting to approximations, where they invoke the gluing lemma.

Conversely, the Fréchet mean leads to an optimal multicoupling (Zemel & Panaretos [94, Theorem 2]):

Theorem 3.5.20 (multicoupling via Fréchet means)**.** *Let $\mu^1,\dots,\mu^N$ be probability measures in $\mathscr{W}_2(\mathscr{X})$ with absolutely continuous Fréchet mean γ. (For example, when $\mathscr{X}=\mathbb{R}^d$ and μ^1 has a bounded density.) Then*

$$\pi = \left(\mathbf{t}_\gamma^{\mu^1},\dots,\mathbf{t}_\gamma^{\mu^N}\right)\#\gamma$$

is an optimal multicoupling of $\mu^1,\dots,\mu^N$.

Proof of Theorem 3.5.20. By Proposition 3.5.19 it suffices to show that $F(\gamma)=G(\pi)$. Since γ is a Karcher mean (Proposition 3.5.16), $M(x)=x$ π-almost surely, so that

$$G(\pi) = \frac{1}{2N}\int_{\mathscr{X}}\sum_{i=1}^N \|\mathbf{t}_\gamma^{\mu^i}-\mathbf{i}\|^2\,\mathrm{d}\gamma = \frac{1}{2N}\sum_{i=1}^N W_2^2(\gamma,\mu^i) = F(\gamma),$$

proving optimality of π. □

It is natural to ask whether such an equivalence still holds for the population Fréchet mean. However, defining the multimarginal problem in full generality is not obvious because unless $\Lambda(\Omega)$ is countable (i.e. Λ is a discrete random measure in $\mathscr{W}_2(\mathscr{X})$), the elements π should be taken as probability measures in an uncountable product of $\mathscr{X}$. If, however, there is more structure in Λ, then the problem can be defined and solved in terms of stochastic processes; see Pass [71]. In this work, when dealing with population Fréchet means, we will not consider the multimarginal formulation.

Boissard, Le Gouic & Loubes [20] noticed that compatibility of $\mu^1,\dots,\mu^N$ according to Definition 3.3.1 allows for a simple solution to the problem of finding their Fréchet mean. As showed in the beginning of this subsection, this is equivalent to solving the multimarginal problem. Returning to the original form of G, we see that for any $\pi\in\Pi(\mu^1,\dots,\mu^N)$ we have

$$G(\pi) = \frac{1}{2N^2}\int_{\mathscr{X}^N}\sum_{i<j}\|x_i-x_j\|^2\,\mathrm{d}\pi(x_1,\dots,x_N) \geq \frac{1}{2N^2}\sum_{i<j}W_2^2(\mu^i,\mu^j),$$

because the (i,j)-th marginal of π is a coupling of μ^i and μ^j. Thus, if equality above holds for π, then π is optimal and $M\#\pi$ is the Fréchet mean by Proposition 3.5.19. This is indeed the case for $\pi=(\mathbf{i},\mathbf{t}_{\mu^1}^{\mu^2},\dots,\mathbf{t}_{\mu^1}^{\mu^N})\#\mu^1$ because the compatibility gives:

$$\int_{\mathscr{X}^N}\|x_i-x_j\|^2\,\mathrm{d}\pi(x_1,\dots,x_N) = \int_{\mathscr{X}}\left\|\mathbf{t}_{\mu^1}^{\mu^i}-\mathbf{t}_{\mu^1}^{\mu^j}\right\|^2\mathrm{d}\mu^1 = \int_{\mathscr{X}}\left\|\mathbf{t}_{\mu^1}^{\mu^i}\circ\mathbf{t}_{\mu^j}^{\mu^i}-\mathbf{i}\right\|\mathrm{d}\mu^j = W_2^2(\mu^i,\mu^j).$$

We may thus conclude, in a slightly more general form (γ was μ^1 above):

Theorem 3.5.21 (Fréchet mean of compatible measures)**.** *Suppose that* $\{\gamma, \mu^1, \ldots, \mu^N\}$ *are compatible measures. Then*

$$\left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{t}_{\gamma}^{\mu^i}\right] \# \gamma$$

is the Fréchet mean of $(\mu^1, \ldots, \mu^N)$.

A population version is given in Theorem 5.6.3.

4 Phase variation and Fréchet means

4.1 Amplitude and phase variation

Following Panaretos & Zemel [70], we describe the problem of separation of amplitude and phase variation in point processes. To build the intuition we discuss the functional case first.

As the functional case will only serve as a motivation to the point process case discussed next, our treatment will mostly be heuristic and superficial. Rigorous proofs and more precise details can be found in Horváth & Kokoszka [48] or Hsing & Eubank [49]. The notion of amplitude and phase variation is discussed in the more applied books by Ramsay & Silverman [75, 76]. One can also consult the review by Wang, Chiou & Müller [90], where amplitude and phase variation are discussed in Section 5.2.

4.1.1 The functional case

Let K denote the unit cube $[0,1]^d \subset \mathbb{R}^d$. A real random function $Y = (Y(x) : x \in K)$ can, broadly speaking, have two types of variation. The first, *amplitude variation*, results from $Y(x)$ being a random variable for every x and describes its fluctuations around the mean level $m(x) = \mathbb{E}Y(x)$, usually encoded by the variance $\operatorname{var} Y(x)$. For this reason, it can be referred to as "variation in the y-axis". More generally, for any finite set $x_1, \dots, x_n$, the $n \times n$ covariance matrix with entries $\kappa(x_i, x_j) = \operatorname{cov}[Y(x_i), Y(x_j)]$ encapsulates (up to second order) the stochastic deviations of the random vector $(Y(x_1), \dots, Y(x_n))$ from its mean, in analogy with the multivariate case. Heuristically, one then views amplitude variation as the collection $\kappa(x, y)$ for x, y in a sense we discuss next.

One typically views Y as a random element in the separable Hilbert space $L_2(K)$, assumed to have $\mathbb{E}\|Y\|^2 < \infty$ and continuous sample paths, so that in particular $Y(x)$ is a random variable for all $x \in K$. Then the **mean function**

$$m(x) = \mathbb{E}Y(x), \quad x \in K$$

and the **covariance kernel**

$$\kappa(x,y) = \operatorname{cov}[Y(x), Y(y)], \quad x, y \in K$$

are well-defined and finite; we shall assume that they are continuous, which is equivalent to Y being **mean-square continuous**:

$$\mathbb{E}[Y(y) - Y(x)]^2 \to 0, \qquad y \to x.$$

The covariance kernel κ gives rise to the **covariance operator** $\mathscr{R} : L_2(K) \to L_2(K)$, defined by

$$(\mathscr{R}f)(y) = \int_K \kappa(x,y) f(x)\,\mathrm{d}x,$$

a self-adjoint positive semidefinite Hilbert-Schmidt operator on $L_2(K)$. The justification to this terminology is the observation that when $m = 0$, for all bounded $f, g \in L_2(K)$,

$$\mathbb{E}\langle Y, f\rangle\langle Y, g\rangle = \mathbb{E}\left[\int_{K^2} Y(x) f(x) Y(y) g(y)\,\mathrm{d}(x,y)\right] = \int_K g(y)(\mathscr{R}f)(y)\,\mathrm{d}y,$$

and so, without the restriction to $m = 0$,

$$\operatorname{cov}\left[\langle Y, f\rangle, \langle Y, g\rangle\right] = \int_K g(y)(\mathscr{R}f)(y)\,\mathrm{d}y = \langle g, \mathscr{R}f\rangle.$$

The covariance operator admits an eigendecomposition $(r_k, \phi_k)_{k=1}^{\infty}$ such that $r_k \searrow 0$, $\mathscr{R}\phi_k = r_k\phi_k$ and (ϕ_k) is an orthonormal basis of $L_2(K)$. One then has the celebrated **Karhunen–Loève expansion**

$$Y(x) = m(x) + \sum_{k=1}^{\infty} \langle Y - m, \phi_k\rangle \phi_k(x) = m(x) + \sum_{k=1}^{\infty} \xi_k \phi_k(x).$$

A major feature in this expansion is the separation of the functional part from the stochastic part: the functions $\phi_k(x)$ are deterministic; the random variables ξ_k are scalars. This separation actually holds for any orthonormal basis; the role of choosing the eigenbasis of $\mathscr{R}$ is making ξ_k *uncorrelated*:

$$\operatorname{cov}(\xi_k, \xi_l) = \operatorname{cov}\left[\langle Y, \phi_k\rangle, \langle Y, \phi_l\rangle\right] = \langle \phi_l, \mathscr{R}\phi_k\rangle$$

vanishes when $k \neq l$ and equals r_k otherwise. For this reason, it is not surprising that using as ϕ_k the eigenfunctions yields the optimal representation of Y. Here optimality is with respect to truncations: for any other basis (ψ_k) and any M,

$$\mathbb{E}\left\| Y - m - \sum_{k=1}^{M} \langle Y - m, \psi_k\rangle \psi_k \right\|^2 \geq \mathbb{E}\left\| Y - m - \sum_{k=1}^{M} \langle Y - m, \phi_k\rangle \phi_k \right\|^2$$

so that (ϕ_k) provides the best finite-dimensional approximation to Y. The approximation

error on the right-hand side equals

$$\mathbb{E}\left\| \sum_{k=M+1}^{\infty} \xi_k \phi_k \right\|^2 = \sum_{k=M+1}^{\infty} r_k$$

and depends on how quickly the eigenvalues of $\mathscr{R}$ decay.

One carries out inference for m and κ on the basis of a sample $Y_1, \ldots, Y_n$ by

$$\widehat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x), \qquad x \in K$$

and

$$\widehat{\kappa}(x,y) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x) Y_i(y) - \widehat{m}(x)\widehat{m}(y),$$

from which one proceeds to estimate $\mathscr{R}$ and its eigendecomposition.

We have seen that amplitude variation in the sense described above is linear and dealt with using linear operations. There is another, qualitatively different type of variation, *phase variation*, that is nonlinear and does not have an obvious finite-dimensional analogue. It arises when in addition to the randomness in the values $Y(x)$ itself, an extra layer of stochasticity is present in its domain of definition. In mathematical terms, there is a random invertible **warp function** (sometimes called **deformation** or **warping**) $T : K \to K$ and instead of $Y(x)$, one observes realisations from the model

$$\tilde{Y}(x) = Y(T^{-1}(x)), \qquad x \in K.$$

For this reason, phase variation can be viewed as "variation in the x-axis". When $d = 1$, the set K is usually interpreted as a time interval, and then the model stipulates that each individual has its own time scale. Typically, the warp function is assumed to be a homeomorphism of K independent of Y and often some additional smoothness is imposed, say $T \in C^2$. One of the classical examples is growth curves of children, of which a dataset from the Berkeley growth study (Jones & Bayley [51]) is shown in Figure 4.1. The curves are the derivatives of the height of a sample of children as a function of time, from birth until the age of 18. One clearly notices the presence of the two types of variation in the figure. The initial velocity for all children is the highest immediately or shortly after birth, and in most cases decreases sharply during the first two years. Then follows a period of acceleration for another year or so, and so on. Despite presenting qualitatively similar behaviour, the curves differ substantially not only in the magnitude of the peaks but also in their location. For instance, one green curve has a local minimum at the age of three, while a red one has maximum at that same time point. It is apparent that if one tries to estimate the mean function by averaging the curves at each time x, the shape of the resulting estimate would look very different from each of the curves. Thus, this pointwise averaging (known as the *cross-sectional mean*) fails to represent the typical

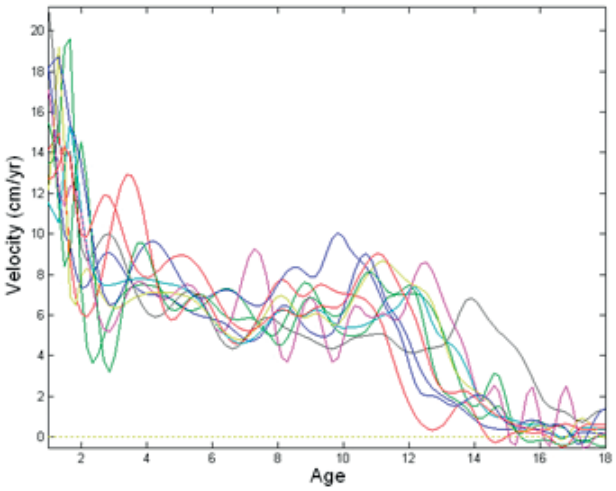


Figure 4.1: Derivatives of growth curves from the Berkeley dataset.

behaviour. This phenomenon is seen more explicitly in the next example.

The terminology of amplitude and phase comes from trigonometric functions, from which we derive an artificial example that illustrates the difficulties of estimation in the presence of phase variation. Let A and B be symmetric random variables and consider the random function

$$\tilde{Y}(x) = A\sin[8\pi(x+B)]. \tag{4.1}$$

(Strictly speaking, $x \mapsto x+B$ is not from $[0,1]$ to itself; for illustration purposes, we assume in this example that $K = \mathbb{R}$.) The random variable A generates the amplitude variation, while B represents the phase variation. In Figure 4.2 we plot four realisations and the resulting empirical means for the two extreme scenarios where $B = 0$ (no phase variation) or $A = 1$ (no amplitude variation). In the left panel of the figure, we see that the sample mean (in thick blue) lies between the observations and has a similar form, so can be viewed as the curve representing the typical realisation of the random curve. This is in contrast to the right panel, where the mean is qualitatively different from all curves in the sample: though periodicity is still present, the peaks and troughs have been flattened, and the sample mean is much more diffuse than any of the observations.

The phenomenon illustrated in Figure 4.2 is hardly surprising, since as mentioned earlier amplitude variation is linear while phase variation is not, and taking sample means is a linear operation. Let us see in formulae how this phenomenon occurs. When $A = 1$ we have

$$\mathbb{E}\tilde{Y}(x) = \sin 8\pi x \mathbb{E}\cos 8\pi B + \cos 8\pi x \mathbb{E}\sin 8\pi B.$$

Since B is symmetric the second term vanishes, and unless B is trivial the expectation of the

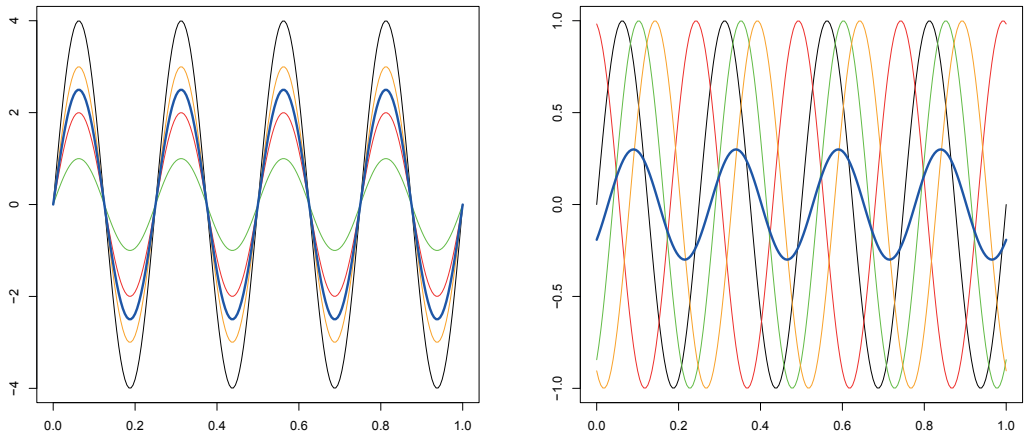


Figure 4.2: Four realisations of (4.1) with means in thick blue. Left: amplitude variation ($B = 0$); right: phase variation ($A = 1$).

cosine is smaller than one in absolute value. Consequently, the expectation of $\tilde{Y}(x)$ is the original function $\sin 8\pi x$ multiplied by a constant of magnitude strictly less than one, resulting in peaks of smaller magnitude.

In the general case, where $\tilde{Y}(x) = Y(T^{-1}(x))$ and Y and T are independent, we have

$$\mathbb{E}\tilde{Y}(x) = \mathbb{E}m(T^{-1}(x))$$

and

$$\operatorname{cov}[\tilde{Y}(x), \tilde{Y}(y)] = \mathbb{E}\kappa(T^{-1}(x), T^{-1}(y)) + \operatorname{cov}(m(T^{-1}(x), m(T^{-1}(y))).$$

From this several conclusions can be drawn. Let $\tilde{\mu} = \mu(T^{-1}(x))$ be the conditional mean function given T. Then the value mean function itself, $\mathbb{E}\tilde{\mu}$, at x_0 is determined not by a single point, say x, but rather by all the values of m at the possible outcomes of $T^{-1}(x)$. In particular, if x_0 was a local maximum for m, the value of $\mathbb{E}\tilde{\mu}(x_0)$ will be strictly smaller than $m(x_0)$; the phase variation results in smearing m.

At this point an important remark should be made. Whether or not phase variation is problematic depends on the specific application. If one is interested indeed in the mean and covariance functions of $\tilde{Y}$, then the standard empirical estimators will be consistent, since $\tilde{Y}$ itself is a random function. But if it is rather m, the mean of Y, that is of interest, then the confounding of the amplitude and phase variation will lead to inconsistency. This can also be seen from the formula

$$\tilde{Y}(x) = m(T^{-1}(x)) + \sum_{k=1}^{\infty} \xi_k \phi_k(T^{-1}(x)).$$

The above series is *not* the Karhunen–Loève expansion of $\tilde{Y}$; the simplest way to notice this is the observation that $\phi_k(T^{-1}(x))$ includes both the functional component ϕ_k and the random component $T^{-1}(x)$. The true Karhunen–Loève expansion of $\tilde{Y}$ will in general be qualitatively very different from that of Y, not only in terms of the mean function but also in terms of the covariance operator and, consequently, its eigenfunctions and eigevalues. As illustrated in the trigonometric example, the typical situation is that the mean $\mathbb{E}\tilde{Y}$ is more diffuse than m, and the decay of the eigenvalues $\tilde{r}_k$ of the covariance operator is slower that that of r_k; as a result, one needs to truncate the sum at high threshold in order to capture a substantial enough part of the variability. In the example model (4.1), the Karhunen–Loève expansion has a single term besides the mean if $B = 0$, while having two terms if $A = 1$.

When one is indeed interested in the mean m and the covariance κ, the random function T pertaining to the phase variation is nuisance parameter. Given a sample $\tilde{Y}_i = Y_i \circ T_i^{-1}$, $i = 1,\dots,n$, there is no point in taking pointwise means of $\tilde{Y}_i$, because the curves are *misaligned*; $\tilde{Y}_1(x) = Y_1(T_1^{-1}(x))$ should not be compared with $\tilde{Y}_2(x)$, but rather with $Y_2(T_1^{-1}(x)) = \tilde{Y}_2(T_1^{-1}(T_2(x))$. To overcome this difficulty, one seeks estimators $\widehat{T_i}$ such that

$$\widehat{Y}_i(x) = \tilde{Y}_i(\widehat{T_i}(x)) = Y_i(T_i^{-1}(\widehat{T_i}(x)))$$

is approximately $Y_i(x)$. In other words, one tries to align the curves in the sample to have a common time scale. Such a procedure is called **curve registration**. Once registration has been carried out, one proceeds the analysis on $\widehat{Y}_i(x)$ assuming only amplitude variation is now present: estimate the mean m by

$$\widehat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{Y}_i(x)$$

and the covariance κ by its analogous counterpart. Put differently, registering the curves amounts to *separating the two types of variation*. This step is crucial regardless of whether the warp functions are considered as nuisance or an analysis of the warp functions is of interest in the particular application.

There is an obvious identifiability problem in the model $\tilde{Y} = Y \circ T^{-1}$. If S is any (deterministic) function, then the model with (Y, T) is statistically indistinguishable from the model with $(Y \circ S, T \circ S)$. It is therefore often assumed that $\mathbb{E}T = \mathbf{i}$ is the identity and in addition, in nearly all application, that T is monotonically increasing (if $d = 1$).

Discretely observed data. One cannot measure the height of person at every single instant of her life. In other words, it is rare in practice that one has access to the entire curve. A far more common situation is that one observes the curves *discretely*, i.e., at a finite number of points. The conceptually simplest setting is that one has access to a grid $x_1,\dots,x_J \in K$, and the data come in the form

$$\tilde{y}_{ij} = \tilde{Y}_i(t_j),$$

with possibly additional additive measurement error. The problem is to find, given $\tilde{y}_{ij}$, consistent estimators of T_i and of the original, aligned functions Y_i. We briefly discuss some methods for carrying out this separation of amplitude and phase variation. In the next subsection we formulate the analogous problem with functions replaced by point processes.

One of the first registration techniques employes dynamic programming (Wang & Gasser [91]) and dates back to Sakoe & Chiba [82]. Landmark registration consists of identifying salient features for each curve, called **landmarks**, and aligning them (Gasser & Kneip [39]; Gervini & Gasser [40]). In pairwise synchronisation (Tang & Müller [86]) one aligns each pair of curves and then derives an estimator of the warp functions by linear averaging of the pairwise registration maps. Another class of methods involves a template curve, to which each observation is registered, minimising a discrepancy criterion; the template is then iteratively updated (Wang & Gasser [92]; Ramsay & Li [74]). James [50] defines a “feature function" for each curve and uses the moments of the feature function to guarantee identifiability. Elastic registration employs the Fisher–Rao metric that is invariant to warpings and calculates averages in the resulting quotient space (Tucker, Wu & Srivastava [87]). Other techniques include semiparametric modelling (Rønn [79]; Gervini & Gasser [41]) and principal components registration (Kneip & Ramsay [57]). More details can be found in the review article by Marron, Ramsay, Sangalli & Srivastava [64].

It is fair to say that no single registration method arises as the canonical solution to the functional registration problem. Indeed, most need to make additional structural and/or smoothness assumptions on the warp maps, further to the basic identifiability conditions requiring that T be increasing and that $\mathbb{E}T$ equal the identity. We now argue that the point process case, in contrast, admits a canonical framework, without needing additional assumptions.

4.1.2 The point process case

A point process is the mathematical object that represents the intuitive notion of a random collection of points in a space $\mathscr{X}$. It is formally defined as a measurable map Π from a generic probability space into the space of (possibly infinite) Borel integer-valued measures of $\mathscr{X}$ in such a way that $\Pi(B)$ is a measurable real-valued random variable for all Borel subsets B of $\mathscr{X}$. The quantity $\Pi(B)$ represents the random number of points observed in the set B. Among the plethora of books on point processes, let us mention Daley & Vere-Jones [28] and Karr [56]. Kallenberg [52] treats more general objects, *random measures*, of which point processes are a peculiar special case. We will assume for convenience that Π is a measure on a compact subset $K \subset \mathbb{R}^d$.

Amplitude variation of Π can be understood in analogy with the functional case. One defines the mean measure

$$\lambda(A) = \mathbb{E}\Pi(A), \qquad A \subset K \text{ Borel}$$

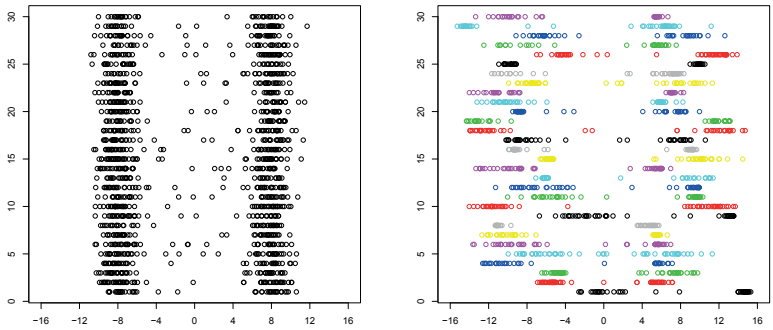


Figure 4.3: Unwarped (left) and warped Poisson point processes.

and, provided that $\mathbb{E}[\Pi(K)]^2 < \infty$, the covariance measure

$$\kappa(A,B) = \operatorname{cov}[\Pi(A),\Pi(B)] = \mathbb{E}\Pi(A)\Pi(B) - \lambda(A)\lambda(B),$$

the latter being a finite signed Borel measure on K. Just like in the functional case, these two objects encapsulate the second-order stochastic properties of the law of Π. Given a sample $\Pi_1,\dots,\Pi_n$ of independent point processes distributed as Π, the natural estimators

$$\widehat{\lambda}(A) = \frac{1}{n}\sum_{i=1}^{n}\Pi_i(A); \qquad \widehat{\kappa}(A,B) = \frac{1}{n}\sum_{i=1}^{n}\Pi_i(A)\Pi_i(B) - \widehat{\lambda}(A)\widehat{\lambda}(B),$$

are consistent and the former asymptotically normal [56, Proposition 4.8].

Phase variation then pertains to a random warp function $T : K \to K$ (independent of Π) that deforms Π: if we denote the points of Π by $x_1,\dots,x_K$ (with K random), then instead of (x_i), one observes $T(x_1),\dots,T(x_K)$. In symbols, what this means is that the data arise as $\widetilde{\Pi} = T\#\Pi$. We refer to Π as the *original point processes*, and $\widetilde{\Pi}$ as the *warped point processes*. An example of 30 warped and unwarped point processes is shown in Figure 4.3. In both panels of the figure, the point patterns present qualitatively similar structure: there are two peaks of high concentration of points, while in between the peaks there are relatively few of them. The difference between the two panels is in the position and concentration of those peaks. In the left panel, only amplitude variation is present, and the location/concentration of the peaks is the same across all observations. In contrast, phase variation results in shifting the peaks to different places for each of the observations, while also smearing or sharpening them. Clearly, estimation of the mean measure of a subset A by averaging the number of observed points in A would not be satisfactory as an estimator of λ when carried out with the warped data. As in the functional case, it will only be consistent for the measure $\tilde{\lambda}$ defined by

$$\tilde{\lambda}(A) = \mathbb{E}\lambda(T^{-1}(A)), \qquad A \subseteq \mathscr{X},$$

and $\tilde{\lambda} = \mathbb{E}[T\#\lambda]$ misses most (or at least a significant part) of the bimodal structure of λ and is far more diffuse.

Since Π and T are independent, the conditional expectation of $\widetilde{\Pi}$ given T is

$$\mathbb{E}\widetilde{\Pi}(A)|T = \mathbb{E}\Pi(T^{-1}(A))|T = \lambda(T^{-1}(A)) = [T\#\lambda](A).$$

Consequently, we define the **conditional mean measure** $\Lambda = T\#\lambda$. The problem of separation of amplitude and phase variation can now be stated as follows. On the basis of a sample $\widetilde{\Pi}_1, \ldots, \widetilde{\Pi}_n$, find estimators of (T_i) and (Π_i). Registering the point processes amounts to constructing estimators, *the registration maps* $\widehat{T_i^{-1}}$ such that the aligned points

$$\widehat{\Pi_i} = \widehat{T_i^{-1}}\#\widetilde{\Pi}_i = [\widehat{T_i^{-1}} \circ T_i]\#\Pi_i$$

are close to the original points Π_i.

Poisson processes. A special but important case is that of a Poisson process. Gaussian processes probably yield the most elegant and rich theory in functional data analysis, and so do Poisson processes when it comes to point processes. We say that Π is a **Poisson process** when the following two conditions hold. (1) For any disjoint collection $(A_1, \ldots, A_n)$ of sets, the random variables $\Pi(A_1), \ldots, \Pi(A_n)$ are independent; and (2) for every Borel $A \subset \mathscr{X}$, $\Pi(A)$ follows a Poisson distribution with mean $\lambda(A)$:

$$\mathbb{P}(\Pi(A) = k) = e^{-\lambda(A)} \frac{[\lambda(A)]^k}{k!}.$$

Conditional on T, the random variables $\widetilde{\Pi}(A_k) = \Pi(T^{-1}(A_k))$, $k = 1, \ldots, n$ are independent as the sets $(T^{-1}(A_k))$ are disjoint; and $\widetilde{\Pi}(A)$ follows a Poisson distribution with mean $\lambda(T^{-1}(A)) = \Lambda(A)$. This is precisely the definition of a **Cox process**: conditional on the **driving measure** Λ, $\widetilde{\Pi}$ is a Poisson process with mean measure λ. For this reason, it also called a **doubly stochastic process**; in our context, the phase variation is associated with the stochasticity of Λ while the amplitude one is associated with the Poisson variation conditional on Λ.

As in the functional case there are problems with identifiability: the model (Π, T) cannot be distinguished from the model $(S\#\Pi, T \circ S^{-1})$ for any invertible $S : K \to K$. It is thus natural to assume that $\mathbb{E}T$ is the identity map (otherwise set $S = \mathbb{E}T$, i.e., replace Π by $[\mathbb{E}T]\#\Pi$ and T by $T \circ [\mathbb{E}T]^{-1}$).

Constraining T to have mean identity is nevertheless not sufficient for the model $\widetilde{\Pi} = T\#\Pi$ to be identifiable. The reason is that given the two point sets $\widetilde{\Pi}$ and Π, there are many functions that push forward the latter to the former. This ambiguity can be dealt with by assuming some sort of *regularity* or *parsimony* for T. For example, when $K = [a, b]$ is a subset of the real line, imposing T to be monotonically increasing guarantees its uniqueness. In multiple dimensions there is no obvious analogue for increasing functions. One possible definition is the monotonicity described in Subsection 2.9.2 (p. 35):

$$\big\langle T(y) - T(x), y - x \big\rangle \geq 0, \qquad x, y \in K.$$

This property is rather weak in a sense we describe now. Let $K \subseteq \mathbb{R}^2$ and write $y \geq x$ if and only if $y_i \geq x_i$ for $i = 1,2$. It is natural to expect the deformations to maintain the *lexicographic order* in $\mathbb{R}^2$:

$$y \geq x \quad \Longrightarrow \quad T(y) \geq T(x).$$

If we require in addition that the ordering must be preserved for all quadrants: for $z = T(x)$ and $w = T(y)$

$$\{y_1 \geq x_1, y_2 \leq x_2\} \qquad \Longrightarrow \{w_1 \geq z_1, w_2 \leq z_2\},$$

then monotonicity is automatically satisfied. In that sense it is arguably not very restrictive.

Monotonicity is weaker than cyclical monotonicity (see (2.10), p. 34, with $y_i = T(x_i)$), which is itself equivalent to the property of being the subgradient of a convex function. But if extra smoothness is present and T is a gradient of some function $\phi : K \to \mathbb{R}$, then ϕ must be convex and T is then cyclically monotone. Consequently, we will make the following assumptions:

- the expected value of T is the identity;
- T is a gradient of a convex function.

In the functional case, at least on the real line, these two conditions are imposed on the warp functions in virtually all applications, often accompanied with additional assumptions about smoothness of T, its structural properties, or its distance from the identity (see p. 95). In the next section, we show how these two conditions alone lead to the Wasserstein geometry and open the door to consistent, fully nonparametric separation of the amplitude and phase variation.

4.2 Wasserstein geometry and phase variation

4.2.1 Equivariance properties of the Wasserstein distance

A first hint to the relevance of Wasserstein metrics in $\mathscr{W}_p(\mathscr{X})$ for deformations of the space $\mathscr{X}$ is that for all $p \geq 1$ and all $x, y \in \mathscr{X}$,

$$W_p(\delta_x, \delta_y) = \|x - y\|,$$

where δ_x is as usual the Dirac measure at $x \in \mathscr{X}$. This is in contrast to metrics such as the bounded Lipschitz distance (that metrises narrow convergence) or the total variation distance on $P(\mathscr{X})$. Recall that these are defined by

$$\|\mu - \nu\|_{BL} = \sup_{\|\varphi\|_{BL} \leq 1} \left| \int_{\mathscr{X}} \varphi \, \mathrm{d}\mu - \int_{\mathscr{X}} \varphi \, \mathrm{d}\nu \right|; \qquad \|\mu - \nu\|_{TV} = \sup_A |\mu(A) - \nu(A)|,$$

so that

$$\|\delta_x - \delta_y\|_{BL} = \min(1, \|x-y\|); \qquad \|\delta_x - \delta_y\|_{TV} = \begin{cases} 1 & x \neq y \\ 0 & x = y. \end{cases}$$

In words, the total variation metric "does not see the geometry" of the space $\mathscr{X}$. This is less so for the bounded Lipschitz distance, that does take small distances into account but not large ones.

Another property (shared by BL and TV) that holds for the specific case $p = 2$ is equivariance with respect to translations. It is more convenient to state it using the probabilistic formalism of Section 2.2 (p. 11). Let X and Y be random elements in $\mathscr{X}$, and suppose that the optimal coupling between them is attained by a map T, the gradient of a convex function ϕ. Then

$$W_2(X,Y) = \mathbb{E}\|T(X) - X\|^2.$$

If a is a fixed point in $\mathscr{X}$, $X' = X + a$ and $Y' = Y + a$, then $T'(x) = a + T(x-a)$ pushes forward X' to Y' and does so optimally, as the gradient of the convex function $x \mapsto \langle a, x\rangle + \phi(x-a)$. This leads to the rather obvious fact that

$$W_2(X + a, Y + a) = W_2(X,Y).$$

The same holds even if the optimal coupling is not given by a proper map. In terms of measures, the result states the following. Let $\mu * \delta_a$ denote the convolution of μ with the Dirac mass at a. Then

$$W_2(\mu * \delta_a, \nu * \delta_a) = W_2(\mu, \nu).$$

This carries over to Fréchet means in an obvious way.

Lemma 4.2.1 (Fréchet means and translations)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ with finite Fréchet functional and $a \in \mathscr{X}$. Then γ is a Fréchet mean of Λ if and only if $\gamma * \delta_a$ is the Fréchet mean of $\Lambda * \delta_a$.*

One can say more. Denote the first moment (mean) of $\mu \in \mathscr{W}_1(\mathscr{X})$ by

$$m : \mathscr{W}_1(\mathscr{X}) \to \mathscr{X} \qquad m(\mu) = \int_{\mathscr{X}} x \, \mathrm{d}\mu(x).$$

If only μ is translated, then

$$W_2(\mu * \delta_a, \nu) = W_2(\mu,\nu) + (a - [m(\mu) - m(\nu)])^2 - [m(\mu) - m(\nu)]^2,$$

which is minimised at $a = m(\mu) - m(\nu)$. This leads to the following conclusion:

Proposition 4.2.2 (first moment of Fréchet mean)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathscr{X})$ with*

finite Fréchet functional with Fréchet mean γ. Then

$$\int_{\mathscr{X}} x \, \mathrm{d}\gamma(x) = \mathbb{E} \int_{\mathscr{X}} x \, \mathrm{d}\Lambda(x).$$

4.2.2 Canonicity of Wasserstein distance in measuring phase variation

The purpose of this subsection is show that the standard functional data analysis assumptions on the warp function T, having mean identity and being increasing, are equivalent to purely geometric conditions on T and the conditional mean measure $\Lambda = T\#\lambda$. Put differently, if one is willing to assume that $\mathbb{E}T = \mathbf{i}$ and that T is increasing, then one is led *unequivocally* to the problem of estimation of Fréchet means in the Wasserstein space $\mathscr{W}_2(\mathscr{X})$. When $\mathscr{X} \neq \mathbb{R}$, "increasing" is interpreted as being the gradient of a convex function, as explained at the end of Subsection 4.1.2.

We begin with the one-dimensional case, slightly generalising Proposition 1 in Panaretos & Zemel [70]. The explicit formulae available when $\mathscr{X} = \mathbb{R}$ allow for a more transparent argument, and for simplicity we will assume some regularity.

Let $K \subseteq \mathbb{R}$ be a nonempty closed convex set, that is, a possibly unbounded interval, and let T be a real-valued continuous and strictly increasing function. Typically one assumes that K is compact and T is a homeomorphism on K, which will happen if and only if $T(K) = K$, but this will not always be necessary. We will therefore assume the following:

Assumptions 3. *The continuous and injective random map $T : K \to \mathbb{R}$ (a random element in $C_b(K)$) satisfies the following two conditions:*

(A1) ***Unbiasedness:*** *$\mathbb{E}[T(x)] = x$ for all $x \in K$.*

(A2) ***Regularity:*** *T is monotone increasing.*

The relevance of the Wasserstein geometry to phase variation becomes clear in the following Proposition, that shows that Assumptions 3 are equivalent to geometric assumptions on the Wasserstein space $\mathscr{W}_2(\mathbb{R})$. We say that a measure λ is **diffuse** or **nonatomic** if $\lambda(\{x\}) = 0$ for all $x \in \mathbb{R}$; equivalently, λ has a continuous distribution function F_λ.

Proposition 4.2.3 (mean identity warp functions and Fréchet means in $\mathscr{W}_2(\mathbb{R})$)**.** *Let $\lambda \in \mathscr{W}_2(\mathbb{R})$ be a diffuse probability with support K and let T be a continuous injective random map such that $\mathbb{E}W_2^2(T\#\lambda, \lambda) < \infty$. Then T satisfies Assumptions 3 if and only if it satisfies:*

(B1) ***Unbiasedness:*** *for any $\gamma \in \mathscr{W}_2(\mathbb{R})$*

$$\mathbb{E}W_2^2(T\#\lambda, \lambda) \leq \mathbb{E}W_2^2(T\#\lambda, \gamma).$$

(B2) ***Regularity:*** *if $Q : K \to \mathbb{R}$ is such that $T\#\lambda = Q\#\lambda$, then with probability one*

$$\int_K \left| T(x) - x \right|^2 \mathrm{d}\lambda(x) \leq \int_K \left| Q(x) - x \right|^2 \mathrm{d}\lambda(x), \qquad \textit{almost surely.}$$

These assumptions have a clear interpretation: (B1) stipulates that λ is the Fréchet mean of the random measure $\Lambda = T\#\lambda$, while (B2) states that T must be the optimal map from λ to Λ, that is, $T = \mathbf{t}_\lambda^\Lambda$.

Proof of Proposition 4.2.3. If T satisfies (B2) then, as an optimal map, it must be nondecreasing λ-almost surely. Conversely, if T is nondecreasing, then it is optimal. Hence (A2) and (B2) are equivalent.

Assuming (A2), we now show that (A1) and (B1) are equivalent. Condition (B1) is equivalent to

$$\mathbb{E}\|F_{T\#\lambda}^{-1} - F_\lambda^{-1}\|_{L_2(0,1)}^2 = \mathbb{E}W_2^2(T\#\lambda, \lambda) \leq \mathbb{E}W_2^2(T\#\lambda, \gamma) = \mathbb{E}\|F_{T\#\lambda}^{-1} - F_\gamma^{-1}\|_{L_2(0,1)}^2, \quad \gamma \in \mathscr{W}_2(\mathbb{R}),$$

which is in turn equivalent to $\mathbb{E}F_\Lambda^{-1} = F_\lambda^{-1}$ (see Subsection 3.5.2). Condition (A2) and the assumptions on T imply that $F_\Lambda(x) = F_\lambda(T^{-1}(x))$. Now F_λ is continuous and strictly increasing on K (since $\mathrm{supp}\lambda = K$), and $T^{-1}(x) \in K$ for all x, so that $F_\Lambda^{-1}(u) = T(F_\lambda^{-1}(u))$. Thus (B1) is equivalent to $\mathbb{E}T(x) = x$ for all x in the range of F_λ^{-1}, which is K (or at least the interior of K) by the hypothesis on λ. □

The situation in more than one dimension is similar but the proof is less transparent. Taking $\mathscr{X} = \mathbb{R}$ below, we see that the warp functions do not have to be strictly increasing and λ does not have to be diffuse in order for the implication (A1-A2)⇒(B1-B2) to hold true, at least when K is bounded (but boundedness can most likely be relaxed, see Remark 6, p. 71). When $\mathscr{X} = \mathbb{R}^d$, one can take any compact convex $K \subset \mathscr{X}$ and choose $U = \mathrm{int}K$ to be its interior, leading to a somewhat cleaner formulation. See Bigot & Klein [16, Theorem 5.1] for a similar result, albeit in a parametric setting.

Theorem 4.2.4 (mean identity warp functions and Fréchet means)**.** *Fix a convex subset U of a separable Hilbert space $\mathscr{X}$ with compact closure K and a probability measure $\lambda \in P(U)$. Let $\mathbf{t} \in C_b(U, \mathscr{X})$ be a random map such that with probability one $\mathbf{t}$ is uniformly continuous, takes values in K, and equals the gradient of its convex potential $\phi_\mathbf{t}$. If $\mathbb{E}\mathbf{t}(x) = x$ for all x in a dense subset of U and $\Lambda = \mathbf{t}\#\lambda$, then*

$$\mathbb{E}W_2^2(\lambda, \Lambda) \leq \mathbb{E}W_2^2(\theta, \Lambda) \qquad \forall \theta \in \mathscr{W}_2(\mathscr{X}).$$

Proof. Since $\mathbb{P}[\Lambda(K) = 1] = 1$ and K is compact and convex, we may assume that $\theta \in P(K)$ by Corollary 3.5.6. Moreover, $K = \overline{U}$, so any measure in $P(K)$ can be approximated in $\mathscr{W}_2$ by measures in $P(U)$ (Theorem 3.2.6), and the functional $\theta \mapsto \mathbb{E}W_2^2(\theta, \Lambda)$ is continuous on $P(K) = \mathscr{W}_2(K)$, so we may further assume that $\theta \in P(U)$.

By Theorem 2.5.2, $\mathbf{t} = \mathbf{t}_\lambda^\Lambda$ is optimal and the pair $(\|x\|^2/2 - \phi, \|y\|^2/2 - \phi^*)$ is dual optimal. Invoking strong duality for λ and weak duality for θ, we find

$$W_2^2(\lambda,\Lambda) = \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \phi(x)\right) \mathrm{d}\lambda(x) + \int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi^*(y)\right) \mathrm{d}\Lambda(y);$$
$$W_2^2(\theta,\Lambda) \geq \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \phi(x)\right) \mathrm{d}\theta(x) + \int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi^*(y)\right) \mathrm{d}\Lambda(y).$$

Since $\mathbf{t}$ is separately valued by Lemma 3.4.14, we may invoke Lemma 3.4.12 and Proposition 3.4.13 to write

$$\mathbb{E}W_2^2(\lambda,\Lambda) = \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \mathbb{E}\phi(x)\right) \mathrm{d}\lambda(x) + \mathbb{E}\int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi^*(y)\right) \mathrm{d}\Lambda(y);$$
$$\mathbb{E}W_2^2(\theta,\Lambda) \geq \int_{\mathscr{X}} \left(\frac{1}{2}\|x\|^2 - \mathbb{E}\phi(x)\right) \mathrm{d}\theta(x) + \mathbb{E}\int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi^*(y)\right) \mathrm{d}\Lambda(y).$$

But $\mathbb{E}\mathbf{t}$ is continuous (by the bounded convergence theorem and boundedness of K), so equals the identity for all $x \in U$. Again by Proposition 3.4.13, it follows that $\mathbb{E}\phi(x) = \|x\|^2/2$ for all $x \in U$, perhaps up to an additive constant. Since $\lambda(U) = 1 = \theta(U)$, the integrals with respect to λ and θ vanish, and this completes the proof. □

4.3 Estimation of Fréchet means

4.3.1 Oracle case

In view of the canonicity of the Wasserstein geometry in Subsection 4.2.2, separation of amplitude and phase variation of the point processes $\widetilde{\Pi}_i$ essentially requires computing Fréchet means in the 2-Wasserstein space. It is both conceptually important and technically convenient to introduce the case where an oracle reveals the conditional mean measures $\Lambda = T\#\lambda$ entirely. Thus, assuming that $\lambda \in \mathscr{W}_2(\mathscr{X})$ is the unique Fréchet mean of a random measure Λ, the goal is to estimate the structural mean λ on the basis of independent and identically distributed realisations $\Lambda_1, \ldots, \Lambda_n$ of λ.

Given that λ is defined as the minimiser of the Fréchet functional

$$F(\gamma) = \frac{1}{2}\mathbb{E}W_2^2(\Lambda,\gamma), \qquad \gamma \in \mathscr{W}_2(\mathscr{X}),$$

it is natural to estimate λ by a minimiser, say λ_n, of the empirical Fréchet functional

$$F_n(\gamma) = \frac{1}{2n}\sum_{i=1}^n W_2^2(\Lambda_i,\gamma), \qquad \gamma \in \mathscr{W}_2(\mathscr{X}).$$

In subsection 3.5.3 it is shown that λ_n exists if $\mathscr{X} = \mathbb{R}^d$ or if the measures Λ_i have a compact support.

When $\mathscr{X} = \mathbb{R}$, λ_n can be seen to be an **unbiased** estimator of λ in a generalised sense of

Lehmann [62] (see Subsection 4.3.5).

4.3.2 Discretely observed measures

In practice, one does not have the fortune of fully observing the inherently infinite-dimensional objects $\Lambda_1, \ldots, \Lambda_n$. A far more realistic scenario is that one only has access to a discrete version of Λ_i, say $\widetilde{\Lambda}_i$. The simplest situation is when $\widetilde{\Lambda}_i$ arises as an empirical measure of the form $\tau^{-1}\sum_{i=1}^{\tau}\delta\{Y_j\}$, where Y_j are independent with distribution Λ_i. More generally, $\widetilde{\Lambda}_i$ can be a normalised point process $\widetilde{\Pi}_i$ with mean measure $\tau\Lambda_i$, i.e.

$$\widetilde{\Lambda}_i = \frac{1}{\widetilde{\Pi}_i(\mathscr{X})}\widetilde{\Pi}_i \quad \text{with} \quad \mathbb{E}\widetilde{\Pi}_i(A)|\Lambda_i = \tau\Lambda_i(A), \qquad A \subseteq \mathscr{X} \text{ Borel.}$$

This encapsulates the case of empirical measure when τ is an integer and $\widetilde{\Pi}_i$ is a **binomial point process**. The parameter τ is the expected number of observed points over the entire space $\mathscr{X}$; clearly, the larger τ is, the more information $\widetilde{\Pi}_i$ gives on Λ_i.

Except if $\widetilde{\Lambda}_i$ is an empirical measure, there is one difficulty in the above setting that needs to be addressed. Unless $\widetilde{\Pi}_i$ is binomial, there is a positive probability that $\widetilde{\Pi}_i(\mathscr{X}) = 0$ and no points pertaining to Λ_i are observed. In the asymptotic setup below, conditions will be imposed to ensure that this probability becomes negligible as $n \to \infty$. For concreteness we define $\widetilde{\Lambda}_i = \lambda^{(0)}$ for some fixed measure $\lambda^{(0)}$ that will be of minor importance. This can be a Dirac measure at 0, a certain fixed Gaussian measure, or (normalised) Lebesgue measure on some bounded set in case $\mathscr{X} = \mathbb{R}^d$. We can now replace the estimator λ_n by $\widetilde{\lambda}_n$, defined as any minimiser of

$$\widetilde{F}_n(\gamma) = \frac{1}{2n}\sum_{i=1}^{n} W_2^2(\widetilde{\Lambda}_i, \gamma), \qquad \gamma \in \mathscr{W}_2(\mathscr{X}).$$

Once again results in Subsection 3.5.3 guarantee the existence of $\widetilde{\lambda}_n$. Indeed, each $\widetilde{\Lambda}_i$ has finite, hence compact support, with the possible exception of the case $\widetilde{\Lambda}_i = \lambda^{(0)}$. Thus, if the latter is compactly supported, then $\widetilde{\lambda}_n$ must exist; in any case there is no problem whatsoever if $\mathscr{X} = \mathbb{R}^d$. With the notable exception of the case $\mathscr{X} = \mathbb{R}$, $\widetilde{\lambda}_n$ is not guaranteed to be unique.

As a generalisation of the discrete case discussed in Section 2.3, the Fréchet mean of discrete measures can be computed exactly. Suppose that $N_i = \widetilde{\Pi}_i(\mathscr{X})$ is nonzero for all i. Then each $\widetilde{\Lambda}_i$ is a discrete measure supported on N_i points. One can then recast the multimarginal formulation (see Subsection 3.5.6) as a finite linear program, solve it and "average" the solution as in Proposition 3.5.19 in order to obtain $\widetilde{\lambda}_n$ (an alternative linear programming formulation for finding a Fréchet mean is given by Anderes, Borgwardt & Miller [9]). Thus $\widetilde{\lambda}_n$ can be computed in finite time, even when $\mathscr{X}$ is infinite-dimensional.

Finally, a remark about measurability is in order. Point processes can be viewed as random elements in $M_+(\mathscr{X})$ endowed with the **vague topology** induced from convergence of integrals

of continuous functions with compact support. If μ_n converge to μ vaguely, and a_n are numbers that converge to a, then $a_n\mu_n \to a\mu$ vaguely. Thus $\tilde{\Lambda}_i$ is a continuous function of the pair $(\widetilde{\Pi}_i, \widetilde{\Pi}_i(\mathscr{X}))$ and can be viewed as a random measure with respect to the vague topology. When it is known a-priori that the mean measures Λ_i are always supported on a fixed compact set $K \subset \mathscr{X}$, the vague topology is equivalent to the weak topology, which is in turn equivalent to the Wasserstein topology and $\tilde{\Lambda}_i$, like Λ_i itself, can be viewed as measurable mappings into $\mathscr{W}_2(K)$.

4.3.3 Smoothing

Even when the computational complexity involved in calculating $\widetilde{\lambda}_n$ is tractable, there is another reason not to use it as an estimator for λ. If one has a-priori knowledge that λ is smooth, it is often desirable to estimate it by a smooth measure. One way to achieve this would be to apply some smoothing technique to $\widetilde{\lambda}_n$ using, e.g., kernel density estimation. However, unless the number of observed points from each measure is the same $N_1 = \cdots = N_n = N$, $\widetilde{\lambda}_n$ will usually be concentrated on many points, essentially $N_1 + \cdots + N_n$ of them. In other words, the Fréchet mean is concentrated on many more points than each of the measures $\widetilde{\Lambda}_i$, thus potentially hindering its usefulness as a mean because it will not be a representative of the sample.

This is most easily seen when $\mathscr{X} = \mathbb{R}$, in which case each $\widetilde{\Lambda}_i$ is a discrete uniform measure on points $x_1^i < x_2^i < \cdots < x_{N_i}^i$, where we assume for simplicity that the points are not repeated (that is, that Λ_i is diffuse). If we now set G_i to be the distribution function of $\widetilde{\Lambda}_i$, then the quantile function G_i^{-1} is piecewise constant on each interval $(k, k+1]/N_i$ with jumps at

$$G_i^{-1}(k/N_i) = x_k^i, \qquad k = 1, 2, \ldots, N_i.$$

The Fréchet mean has quantile function $G^{-1}(u) = n^{-1}\sum G_i^{-1}(u)$ and will have jumps at every point of the form k/N_i for $k \leq N_i$ and $i = 1, \ldots, n$. In the worst case scenario, when no pair from N_i has a common divisor, there will be

$$\left(\sum_{i=1}^{n} N_i - 1\right) + 1 = \sum_{i=1}^{n} N_i - n + 1$$

jumps for G^{-1}, which is the number of points on which the Fréchet mean will be supported. (All the G_i^{-1}'s have a jump at one which thus needs to be counted once rather than n times.)

By counting the number of redundancies in the constraints matrix of the linear program, one can show that this is in general an upper bound on the number of support points of the Fréchet mean.

An alternative approach is to first smooth each observation $\widetilde{\lambda}_n$ and then calculate the Fréchet mean. Since it is easy to bound the Wasserstein distances when dealing with convolutions, we will employ kernel density estimation, although other smoothing approaches could be used

as well.

To simplify the exposition we provide the technical details only when $\mathscr{X} = \mathbb{R}^d$, but a similar construction will work when the dimension of $\mathscr{X}$ is infinite. Let $\psi : \mathbb{R}^d \to (0,\infty)$ be a continuous, bounded, strictly positive isotropic density function with unit variance: $\psi(x) = \psi_1(\|x\|)$ with ψ_1 nondecreasing and

$$\int_{\mathbb{R}^d} \|x\|^2 \psi(x)\,\mathrm{d}x = 1 = \int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x.$$

(Besides the boundedness all these properties can be relaxed, and if $\mathscr{X} = \mathbb{R}$ even boundedness is not necessary.) A classical example for ψ is the standard Gaussian density in $\mathbb{R}^d$. Define the rescaled version $\psi_\sigma(x) = \sigma^{-d}\psi(x/\sigma)$ for all $\sigma > 0$. We can then replace $\widetilde{\Lambda}_i$ by a smooth proxy $\widetilde{\Lambda}_i * \psi_\sigma$. If $\widetilde{\Lambda}_i$ is a sum of Dirac masses at $x_1, \dots, x_{N_i}$, then

$$\widetilde{\Lambda}_i * \psi_\sigma \quad \text{has density} \quad g(x) = \frac{1}{N_i} \sum_{j=1}^{N_j} \psi_\sigma(x - x_i).$$

If $N_i = 0$ one can either use $\lambda^{(0)}$ or $\lambda^{(0)} * \psi_\sigma$; this event will have negligible probability anyway.

For the purpose of approximating $\widetilde{\Lambda}_i$, this convolution is an acceptable estimator, because as was seen in the proof of Theorem 3.2.6,

$$W_2^2(\widetilde{\Lambda}_i, \widetilde{\Lambda}_i * \psi_\sigma) \le \sigma^2.$$

But the measure $\widetilde{\Lambda}_i$ has a strictly positive density throughout $\mathbb{R}^d$. If we know that Λ is supported on a convex compact $K \subset \mathbb{R}^d$, it is desirable to construct an estimator that has the same support K. The first idea that comes to mind is to project $\widetilde{\Lambda}_i * \psi_\sigma$ to K (see Proposition 3.5.5), as this will further decrease the Wasserstein distance; but the resulting measure will then have positive mass on the boundary of K, and will not be absolutely continuous. We will therefore use a different strategy: eliminate all the mass outside K and redistribute it on K. The simplest way to do this is to restrict $\widetilde{\Lambda}_i * \psi_\sigma$ to K and renormalise the restriction to be a probability measure. For technical reasons, it will be more convenient to bound the Wasserstein distance when the restriction and renormalisation is done separately on each point of $\widetilde{\Lambda}_i$. This yields the measure

$$\widehat{\Lambda_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{\delta\{x_j\} * \psi_\sigma}{[\delta\{x_j\} * \psi_\sigma](K)}\bigg|_K, \tag{4.2}$$

It is most likely true that $W_2^2(\widetilde{\Lambda}_i, \widehat{\Lambda_i}) \le \sigma^2$ still holds; we show in Lemma 4.4.2 below that this inequality holds up to a constant. It is apparent that $\widehat{\Lambda_i}$ is a continuous function of $\widetilde{\Lambda}_i$ and σ; in any case this is not particularly important because σ will vanish, so $\widehat{\Lambda_i} = \widetilde{\Lambda}_i$ asymptotically.

Finally, the estimator $\widehat{\lambda}_n$ for λ is defined as the minimiser

$$\widehat{F}_n(\gamma) = \frac{1}{2n}\sum_{i=1}^{n} W_2^2(\widehat{\Lambda}_i, \gamma), \qquad \gamma \in \mathcal{W}_2(\mathcal{X}).$$

Since the measures $\widehat{\Lambda_i}$ are absolutely continuous, $\widehat{\lambda}_n$ is unique. We refer to $\widehat{\lambda}_n$ as the **regularised Fréchet–Wasserstein estimator**.

In the case $\mathcal{X} = \mathbb{R}$, $\widehat{\lambda}_n$ can be constructed via averaging of quantile functions. Let $\widehat{G}_i$ be the distribution function of $\widehat{\Lambda}_i$. Then $\widehat{\lambda}_n$ is the measure with quantile function

$$F^{-1}_{\widehat{\lambda}_n}(u) = \frac{1}{n}\sum_{i=1}^{n} \widehat{G}_i^{-1}(u), \qquad u \in (0,1),$$

and disribution function

$$F_{\widehat{\lambda}_n}(x) = [F^{-1}_{\widehat{\lambda}_n}]^{-1}(x).$$

By construction, the $\widehat{G}_i$ are continuous and strictly increasing, so the inverses are proper inverses and one does not to use the right-continuous inverse as in Subsection 3.5.2 (p. 73).

If $\mathcal{X} = \mathbb{R}^d$ and $d \geq 2$, then there is no explicit expression for $\widehat{\lambda}_n$, although it exists and is unique. In Section 5.1 we present a steepest descent algorithm that approximately constructs $\widehat{\lambda}_n$ by taking advantage of the differentiability properties of the Fréchet functional $\widehat{F}_n$ in Subsection 3.5.5.

4.3.4 Estimation of warpings and registration maps

Once estimators $\widehat{\Lambda_i}$, $i = 1, \ldots, n$ and $\widehat{\lambda}_n$ are constructed, it is natural to estimate the map $T_i = \mathbf{t}^{\Lambda_i}_{\lambda}$ and its inverse $T_i^{-1} = \mathbf{t}^{\lambda}_{\Lambda_i}$ (when Λ_i are absolutely continuous; see the discussion after Assumptions 4 below) by the plug-in estimators

$$\widehat{T_i} = \mathbf{t}^{\widehat{\Lambda}_i}_{\widehat{\lambda}_n}, \qquad \widehat{T_i^{-1}} = (\widehat{T_i})^{-1} = \mathbf{t}^{\widehat{\lambda}_n}_{\widehat{\Lambda}_i}.$$

The latter, the registration maps, can then be used in order to register the points Π_i via

$$\widehat{\Pi_i^{(n)}} = \widehat{T_i^{-1}} \# \widetilde{\Pi}_i^{(n)} = \left[\widehat{T_i^{-1}} \circ T_i\right] \# \Pi_i^{(n)}.$$

It is thus reasonable to expect that if $\widehat{T_i^{-1}}$ is a good estimator, then its composition with T_i should be close to the identity and $\widehat{\Pi_i}$ should be close to Π_i.

4.3.5 Unbiased estimation when $\mathscr{X} = \mathbb{R}$

In the same way Fréchet means extend the notion of mean to non-Hilbertian spaces, they also extend the definition of unbiased estimators. Let H be a separable Hilbert space (or a convex subset thereof) and suppose that $\widehat{\theta}$ is a random element in H whose distribution μ_θ depends on a parameter $\theta \in H$. Then $\widehat{\theta}$ is **unbiased** for θ if for all $\theta \in H$

$$\mathbb{E}_\theta \widehat{\theta} = \int_H x \, \mathrm{d}\mu_\theta(x) = \theta.$$

(We use the standard notation $\mathbb{E}_\theta g(\widehat{\theta}) = \int g(x) \, \mathrm{d}\mu_\theta(x)$ in the sequel.) This is equivalent to

$$\mathbb{E}_\theta \|\theta - \widehat{\theta}\|^2 \leq \mathbb{E}_\theta \|\gamma - \widehat{\theta}\|^2, \qquad \forall \theta, \gamma \in H.$$

In view of that, one can define unbiased estimators of $\lambda \in \mathscr{W}_2$ as measurable functions $\delta = \delta(\Lambda_1, \ldots, \Lambda_n)$ for which

$$\mathbb{E}_\lambda W_2^2(\lambda, \delta) \leq \mathbb{E}_\lambda W_2^2(\gamma, \delta), \qquad \forall \gamma, \theta \in \mathscr{W}_2.$$

This definition was introduced by Lehmann [62].

Unbiased estimators allow us to avoid the problem of over-registering (the so-called “pinching effect"; Kneip & Ramsay [57, Section 2.4]; Marron et al. [64, p. 476]). An extreme example of over-registration is if one “aligns" all the observed patterns into a single fixed point x_0. The registration will then seem “successful" in the sense of having no residual phase variation, but the estimation is clearly biased because the points are not registered to the correct reference measure. Thus, requiring the estimator to be unbiased is an alternative to penalising the registration maps.

Due to the Hilbert space embedding of $\mathscr{W}_2(\mathbb{R})$, it is possible to characterise unbiased estimators in terms of a simple condition on their quantile functions. As a corollary, the Fréchet mean of $\{\Lambda_1, \ldots, \Lambda_n\}$ (λ_n) is unbiased. Our regularised Fréchet–Wasserstein estimator $\hat{\lambda}_n$ can then be interpreted as *approximately unbiased*, since it approximates the unobservable λ_n.

Proposition 4.3.1 (unbiased estimators in $\mathscr{W}_2(\mathbb{R})$)**.** *Let Λ be a random measure in $\mathscr{W}_2(\mathbb{R})$ with finite Fréchet functional and let λ be the unique Fréchet mean of Λ (Theorem 3.5.3). An estimator δ constructed as a function of a sample $(\Lambda_1, \ldots, \Lambda_n)$ is unbiased for λ if and only if the left-continuous representatives (in $L_2(0,1)$) satisfy $\mathbb{E}F_\delta^{-1}(x) = F_\lambda^{-1}(x)$ for all $x \in (0,1)$.*

Proof of Proposition 4.3.1. The proof is straightforward from the definition: δ is unbiased if and only if for all λ and all γ,

$$\mathbb{E}_\lambda \|F_\lambda^{-1} - F_\delta^{-1}\|_{L_2}^2 \leq \mathbb{E}_\lambda \|F_\gamma^{-1} - F_\delta^{-1}\|_{L_2}^2,$$

which is equivalent to $\mathbb{E}_\lambda F_\delta^{-1} = F_\lambda^{-1}$. In other words, these two functions must equal almost everywhere on $(0,1)$, and their left-continuous representatives must equal everywhere (the

fact that $\mathbb{E}_\lambda F_\delta^{-1}$ has such a representative was established in Subsection 3.5.2).

To show that $\delta = \lambda_n$ is unbiased, we simply invoke Theorem 3.5.3 twice to see that

$$\mathbb{E}F_\delta^{-1} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} F_{\Lambda_i}^{-1} = \mathbb{E}F_\Lambda^{-1} = F_\lambda^{-1},$$

which proves unbiasedness of δ. □

4.4 Consistency

In functional data analysis, one typically assumes that the number of curves n as well as the number of observed points m both diverge to infinity. An analogous framework for point processes would similarly require the number of point processes n as well as the expected number of points τ per processes to diverge. A technical complication arises, however, because the mean measures do not suffice to characterise the distribution of the processes. Indeed, if one is given a point processes Π with mean measure λ (not necessarily a probability measure), and τ is an integer, there is no unique way to define a process $\Pi^{(\tau)}$ with mean measure $\tau\lambda$. One can define $\Pi^{(\tau)} = \tau\Pi$, so that every point in Π will be counted τ times. Such a construction, however, can never yield a consistent estimator of λ, even when $\tau \to \infty$.

Another way to generate a point process with mean measure $\tau\lambda$ is to take a superposition of τ independent copies of Π. In symbols, this means

$$\Pi^{(\tau)} = \Pi_1 + \cdots + \Pi_\tau,$$

with (Π_i) independent, each having the same distribution as Π. This superposition is the analogue of an “iid" scheme that gives the possibility to use the law of large numbers. If τ is not an integer, then this construction is not well-defined but can be made so by assuming that the distribution of Π is **infinitely divisible**. The reader willing to assume that τ is always an integer can safely skip to Subsection 4.4.1; all the main ideas are developed first for integer values of τ and then extended to the general case.

A point process Π is infinitely divisible if for every integer m there exists a collection of m independent and identically distributed $\Pi_i^{(1/m)}$ such that

$$\Pi = \Pi_1^{(1/m)} + \cdots + \Pi_m^{(1/m)} \qquad \text{in distribution.}$$

If Π is infinitely divisible and $\tau = k/m$ is rational, then can define $\pi^{(\tau)}$ using km independent copies of $\Pi^{(1/m)}$:

$$\Pi^{(\tau)} = \sum_{i=1}^{km} \Pi_i^{(1/m)}.$$

One then deals with the case of irrational τ via duality and continuity arguments, as follows.

Define the **Laplace functional** of Π by

$$L_\Pi(f) = \mathbb{E}\left[e^{-\Pi f}\right] = \mathbb{E}\left[\exp -\int_{\mathscr{X}} f \,\mathrm{d}\Pi\right], \qquad f : \mathscr{X} \to \mathbb{R}_+ \quad \text{Borel measurable.}$$

The Laplace functional characterises the distribution of the point process, generalising the notion of Laplace transform of a random variable or vector (Karr [56, Theorem 1.12]). The expectation is of course finite, because f is a nonnegative function and Π is a nonnegative measure. By definition, it translates convolutions into products. When $\Pi = \Pi^{(1)}$ is infinitely divisible, the Laplace functional L_1 of Π takes the form (Kallenberg [52, Chapter 6]; Karr [56, Theorem 1.43])

$$L_1(f) = \mathbb{E}\left[e^{-\Pi^{(1)} f}\right] = \exp\left[-\int_{M_+(\mathscr{X})} (1 - e^{-\mu f}) \,\mathrm{d}\boldsymbol{\rho}(\mu)\right] \qquad \text{for some } \boldsymbol{\rho} \in M_+(M_+(\mathscr{X})).$$

The Laplace functional of $\Pi^{(\tau)}$ is $L_\tau(f) = [L_1(f)]^\tau$ for any rational τ, which simply amounts to multiplying the measure $\boldsymbol{\rho}$ by the scalar τ. One can then do the same for an irrational τ, and the resulting Laplace functional determines the distribution of $\Pi^{(\tau)}$ for all $\tau > 0$.

4.4.1 Consistent estimation of Fréchet means

We are now ready to define our asymptotic setup. The following assumptions will be made. Notice that the Wasserstein geometry does not appear explicitly in these assumptions, but is rather *derived* from them in view of Theorem 4.2.4.

Assumptions 4. *Let $K \subset \mathbb{R}^d$ be a compact convex nonempty set, λ an absolutely continuous probability measure on K and τ_n a sequence of positive numbers. Let Π be a point processes on K with mean measure λ. Finally, define $U = \mathrm{int} K$.*

- *For every n, let $\{\Pi_1^{(n)}, \dots, \Pi_n^{(n)}\}_{n=1}^\infty$ be independent point processes, each having the same distribution as a superposition of τ_n copies of Π.*
- *Let T be a random injective function on U (viewed as a random element in $C_b(U, U)$) with nonsingular derivative $\nabla T(x) \in \mathbb{R}^{d\times d}$ for almost all $x \in U$, that is uniformly continuous and is a gradient of a convex function. Let $\{T_1, \dots, T_n\}$ be independent and identically distributed as T.*
- *For every $x \in U$, assume that $\mathbb{E}T(x) = x$.*
- *Assume that the collections $\{T_n\}_{n=1}^\infty$ and $\{\Pi_i^{(n)}\}_{i\le n, n=1}^\infty$ are independent.*
- *Let $\widetilde{\Pi}_i^{(n)} = T_i \# \Pi_i^{(n)}$ be the warped point processes, having **conditional mean measures** $\Lambda_i = T_i \# \lambda = \tau_n^{-1} \mathbb{E}\left\{\widetilde{\Pi}_i^{(n)} \middle| T_i\right\}$.*
- *Define $\widehat{\Lambda}_i$ by the smoothing procedure (4.2), using bandwidth $\sigma_i^{(n)} \in [0, 1]$ (possibly random).*

The dependence of the estimators on n will sometimes be tacit. But Λ_i does not depend on n.

By virtue of Theorem 4.2.4, λ is a Fréchet mean of the random measure $\Lambda = T\#\lambda$. Uniqueness of this Fréchet mean will follow from Proposition 3.5.8 if we show that Λ is absolutely continuous with positive probability. This is indeed the case, since T is injective and has a nonsingular Jacobian matrix; see Lemma 5.5.3 in Ambrosio, Gigli & Savaré [6]. The Jacobian assumption can be relaxed when $\mathscr{X} = \mathbb{R}$, because Fréchet means are always unique in this case by Theorem 3.5.3.

Notice that there is no assumption about the dependence between rows. Assumptions 4 thus cover, in particular, two different scenarios:

- *Full independence*: here the point processes are independent across rows, that is, $\Pi_i^{(n)}$ and $\Pi_i^{(n+1)}$ are also independent.
- *Nested observations*: here $\Pi_i^{(n+1)}$ includes the same points as $\Pi_i^{(n)}$ and additional points, that is, $\Pi_i^{(n+1)}$ is a superposition of $\Pi_i^{(n)}$ and another point process distributed as $(\tau_{n+1} - \tau_n)\Pi$.

The full independence scenario is more difficult, because extra stochasticity is present; this is analogous to the following fact: the strong law of large numbers holds as soon as $\mathbb{E}|X| < \infty$, but if instead of a sequence we have a triangular array, then the requirement becomes $\mathbb{E}X^2 < \infty$ (this is the Hsu–Robbins–Erdös theorem; see Gut [45, Theorem 11.2]).

Needless to say, Assumptions 4 also encompass binomial processes when τ_n are integers, as well as Poisson processes or, more generally, Poisson cluster processes.

We now state and prove the consistency result for the estimators of the conditional mean measures Λ_i and the structural mean measure λ. This is a stronger version of Theorem 1 in Panaretos & Zemel [70] where it was assumed that τ_n must diverge to infinity faster than $\log n$.

Theorem 4.4.1 (consistency)**.** *If Assumptions 4 hold, $\sigma_n = n^{-1}\sum_{i=1}^n \sigma_i^{(n)} \to 0$ almost surely and $\tau_n \to \infty$ as $n \to \infty$, then:*

1. *The estimators $\widehat{\Lambda}_i$ defined by* (4.2)*, constructed with bandwidth $\sigma = \sigma_i^{(n)}$, are Wasserstein-consistent for the conditional mean measures: for all i such that $\sigma_i^{(n)} \xrightarrow{p} 0$*

$$W_2\left(\widehat{\Lambda_i}, \Lambda_i\right) \xrightarrow{p} 0, \qquad \text{as } n \to \infty;$$

2. *The regularised Fréchet–Wasserstein estimator of the structural mean measure (as described in Section 4.3) is strongly Wasserstein-consistent,*

$$W_2(\widehat{\lambda}_n, \lambda) \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty.$$

Convergence in 1. holds almost surely under the additional conditions that $\sum_{n=1}^{\infty} \tau_n^{-2} < \infty$ *and* $\mathbb{E}\left[\Pi(\mathbb{R}^d)\right]^4 < \infty$. *If* $\sigma_n \to 0$ *only in probability, then convergence in 2. still holds in probability.*

Theorem 4.4.1 still holds without smoothing ($\sigma_n = 0$). In that case, $\widehat{\lambda}_n = \tilde{\lambda}_n$ is possibly not unique, and the theorem should be interpreted in a set-valued sense (as in Proposition 2.9.8): almost surely, *any* choice of minimisers $\tilde{\lambda}_n$ converges to λ as $n \to \infty$.

The preceding paragraph notwithstanding, we will usually assume that some smoothing *is* present, in which case $\widehat{\lambda}_n$ is unique and absolutely continuous by Proposition 3.5.17. The uniform Lipschitz bounds for the objective function show that if we restrict the relevant measures to be absolutely continuous, then $\widehat{\lambda}_n$ is a continuous function of $(\widehat{\Lambda_1}, \dots, \widehat{\Lambda_n})$ and hence $\widehat{\lambda}_n : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathscr{W}_2(K)$ is measurable; this is again a minor issue because many arguments in the proof hold for each $\omega \in \Omega$ separately. Thus, even if $\widehat{\lambda}_n$ is not measurable, the proof shows that the convergence holds outer almost surely or in outer probability.

The first step in proving consistency is to show that the Wasserstein distance between the unsmoothed and the smoothed estimators of Λ_i vanishes with the smoothing parameter. The exact rate of decay will be important to later establish the rate of convergence of $\widehat{\lambda}_n$ to λ and is determined next.

Lemma 4.4.2 (smoothing error)**.** *There exists a finite constant* $C_{\psi,K}$ *depending only on* ψ *and on* K *such that*

$$W_2^2\left(\widehat{\Lambda}_i, \widetilde{\Lambda}_i\right) \le C_{\psi,K}\sigma^2 \quad \text{if } \sigma \le 1. \tag{4.3}$$

The constant $C_{\psi,K}$ is explicit. When $\mathscr{X} = \mathbb{R}$, a more refined construction allows to improve this constant in some situations, see [70, Lemma 1].

Proof. The idea is that (4.2) is a sum of measures with mass $1/N_i$ that can be all sent to the relevant point x_j. This would have not been the case if the normalisation was carried out for all the points simultaneously.

Denote the total number of points by $N_i = \widetilde{\Pi}_i(\mathbb{R}^d)$, suppose that it is nonzero and let $\Psi(A) = \int_A \psi(x)\,\mathrm{d}x$ be the probability measure corresponding to the density ψ. For every $y \in K$ define $\tilde{\mu}_y = \delta\{y\} * \psi_\sigma$ and its restricted renormalised version $\mu_y = (1/\tilde{\mu}_y(K))\tilde{\mu}_y|_K$. Then $\widehat{\Lambda}_i = (1/N_i)\sum_{j=1}^{N_i} \mu_{x_j}$ with $N_i \ge 1$ and $x_j \in K$ (because $\Lambda_i(K) = 1$).

A coupling (certainly not optimal, unless $N_i = 1$) of $\widehat{\Lambda}_i$ and $\widetilde{\Lambda}_i = \widetilde{\Pi}_i/N_i$ can be constructed by sending the $1/N_i$ mass of μ_{x_j} to x_j. This gives

$$W_2^2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) \le \frac{1}{N_i}\sum_{j=1}^{N_i} W_2^2(\mu_{x_j}, \delta\{x_j\}) = \frac{1}{N_i}\sum_{j=1}^{N_i} \frac{1}{\tilde{\mu}_{x_j}(K)} \int_K \|x - x_j\|^2 \psi_\sigma(x - x_j)\,\mathrm{d}x.$$

A change of variables shows that each of the last displayed integrals is bounded by σ^2, since ψ

was assumed to have unit variance and so ψ_σ has variance σ^2. The proof will be complete if we can find a lower bound for $\tilde{\mu}_y(K)$ that is uniform in σ and in $y \in K$. Clearly

$$\tilde{\mu}_y(K) = \int_K \psi_\sigma(x-y)\,\mathrm{d}x = \int_{(K-y)/\sigma} \psi(x)\,\mathrm{d}x = \Psi\left(\frac{K-y}{\sigma}\right).$$

Let us first eliminate σ. The set $K_y = K - y$ is a convex set that includes the origin; it follows that $K_y \subseteq (1+\epsilon)K_y$ for all $\epsilon > 0$. Consequently $K_y/\sigma \supseteq K_y$ as long as $\sigma \le 1$. Since the smoothing parameter will anyway vanish, this restriction to small values of σ is not binding. Recalling that $\psi(x) = \psi_1(\|x\|)$ with ψ_1 nonincreasing and strictly positive, we find

$$\Psi\left(\frac{K-y}{\sigma}\right) \ge \Psi(K-y) = \int_{K-y} \psi(x)\,\mathrm{d}x \ge \int_{K-y} \psi_1(d_K)\,\mathrm{d}x = \psi_1(d_K)\mathrm{Leb}K > 0.$$

We have again used the notation $d_K = \sup\{\|x-y\| : x, y \in K\}$ for the finite diameter of the compact set K.

If we now define $C_{\psi,K} = [\psi_1(d_K)\mathrm{Leb}K]^{-1} < \infty$, then putting everything together gives

$$W_2^2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) \le \frac{1}{N_i}\sum_{j=1}^{N_i} W_2^2(\mu_{x_j}, \delta\{x_j\}) \le C_{\psi,K}\sigma^2 \qquad \text{if } \sigma \le 1.$$

Finally, if $N_i = 0$, then by construction $W_2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) = 0$. □

Here is an example: suppose that $K = [0,\infty)^2$ and $y = 0$. Then $\tilde{\mu}_y(K) = 1/4$ for all $\sigma > 0$. But actually $W_2^2(\mu_y, \delta_y) = W_2^2(\tilde{\mu}_y, \delta_y) = \sigma^2$ by the isotropy of ψ: this can be seen by "folding" each quadrant onto the positive quadrant in $\mathbb{R}^2$. If now K is $[0,1]^2$, then after this folding there is still mass in the positive quadrant outside of K, so in fact $W_2^2(\mu_y, \delta_y) < \sigma^2$.

Proof of Theorem 4.4.1. Let us first show the convergence in probability of $\widehat{\Lambda_i}$ to Λ_i. Let $N_i^{(n)} = \Pi_i^{(n)}(\mathscr{X}) = \widetilde{\Pi}_i^{(n)}(\mathscr{X})$ denote the total number of observed points. We may assume without loss of generality that τ_n are integers: otherwise, we replace τ_n by $\lfloor\tau_n\rfloor$ and Λ_i by $(\tau_n/\lfloor\tau_n\rfloor)\Lambda_i$ which converges to Λ_i because the fraction $\tau_n/\lfloor\tau_n\rfloor \to 1$.

Then, $\Pi_i^{(n)}$ has the same distribution as the superposition of τ_n independent copies of Π, say $\{P_j^{(n)}\}$, and $\widetilde{\Pi}_i^{(n)}$ has the same distribution as a superposition of $\{\widetilde{P}_j^{(n)}\}$, independent copies of $T_i\#\Pi$ which have mean measure Λ_i. Consequently, (e.g., Karr [56, Proposition 4.8])

$$\frac{1}{\tau_n}\widetilde{\Pi}_i^{(n)} \overset{d}{=} \frac{1}{\tau_n}\sum_{j=1}^{\tau_n} \widetilde{P}_j^{(n)} \overset{n}{\to} \Lambda_i, \qquad \text{in probability,}$$

with '$\overset{n}{\to}$' denoting narrow convergence of measures. (The convergence will be almost surely if $\widetilde{P}_j^{(n+1)} = \widetilde{P}_j^{(n)}$; but otherwise the convergence is only in probability unless further conditions are imposed).

The proof of this result is just a conditional version of the empirical measure setting in Proposition 3.2.5 with n replaced by τ_n: for any continuous bounded $f : K \to \mathbb{R}$,

$$\int_K f \, \mathrm{d}\frac{1}{\tau_n}\widetilde{\Pi}_i^{(n)} \to \int_K f \, \mathrm{d}\Lambda_i, \qquad \text{in probability,}$$

and one then finds a countable collection (f_j) that suffices to conclude the narrow convergence. In particular when $f \equiv 1$ we obtain $N_i^{(n)}/\tau_n \xrightarrow{p} 1$ and conclude from Slutsky's theorem that

$$\widetilde{\Pi}_i^{(n)}/N_i^{(n)} \xrightarrow{n} \Lambda_i \qquad \text{in probability.} \tag{4.4}$$

The narrow convergence is equivalent to Wasserstein convergence, since K is compact (Corollary 3.2.2). Finally, by Lemma 4.4.2 and the triangle inequality

$$W_2(\widehat{\Lambda}_i, \Lambda_i) \le W_2\left(\Lambda_i, \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}\right) + W_2\left(\frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}, \widehat{\Lambda}_i\right) \le W_2\left(\Lambda_i, \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}\right) + \sqrt{C_{\psi,K}}\sigma_i^{(n)} \to 0,$$

because $\sigma_i^{(n)} \to 0$ as $n \to \infty$. This proves claim (1) in probability.

Let us now prove claim (2). Recall the definitions of the following functionals, defined on $\mathscr{W}_2(K)$,

$$\begin{aligned}
F(\gamma) &= \frac{1}{2}\mathbb{E}W_2^2(\Lambda, \gamma);\\
F_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\Lambda_i, \gamma);\\
\tilde{F}_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\widetilde{\Lambda}_i, \gamma), \qquad \widetilde{\Lambda}_i = \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}} \qquad \text{or } \lambda^{(0)} \text{ if } N_i^{(n)} = 0;\\
\widehat{F}_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\widehat{\Lambda}_i, \gamma), \qquad \widehat{\Lambda}_i = \lambda^{(0)} \text{ if } N_i^{(n)} = 0.
\end{aligned}$$

Assumptions 4 imply that λ is the unique minimiser of F, and we wish to show that any sequences of minimisers $\widehat{\lambda}_n$ of $\widehat{F}_n$ must converge to λ. To this end we shall bound the differences between any two consecutive functionals uniformly in γ. This is possible because all the relevant measures lie in a bounded set of the Wasserstein space $\mathscr{W}_2(\mathbb{R}^d)$. Indeed, if μ, ν and ρ are probability measures on K, then

$$W_2(\mu,\nu) \le \sqrt{\sup_{\pi \in P(K^2)} \int_{K^2} \|x-y\|^2 \, \mathrm{d}\pi(x,y)} \le \sqrt{\sup_{x,y\in K} \|x-y\|^2} = d_K < \infty; \tag{4.5}$$

$$|W_2^2(\mu,\rho) - W_2^2(\nu,\rho)| = |W_2(\mu,\rho) + W_2(\nu,\rho)||W_2(\mu,\rho) - W_2(\nu,\rho)| \le 2d_K W_2(\mu,\nu), \tag{4.6}$$

so that

$$\sup_{\gamma\in\mathscr{W}_2(K)} |\widehat{F}_n(\gamma) - \tilde{F}_n(\gamma)| \le \frac{d_K}{n}\sum_{i=1}^n W_2\left(\widehat{\Lambda}_i, \widetilde{\Lambda}_i\right) \le d_K\sqrt{C_{\psi,K}}\frac{1}{n}\sum_{i=1}^n \sigma_i^{(n)}$$

by Lemma 4.4.2. The right-hand side vanishes by our assumptions.

Similarly,

$$\sup_{\gamma\in\mathscr{W}_2(K)} |\tilde{F}_n(\gamma) - F_n(\gamma)| \leq \frac{1}{n}\sum_{i=1}^{n} W_2\left(\Lambda_i, \widetilde{\Lambda}_i\right) = \frac{1}{n}\sum_{i=1}^{n} X_{ni} = \overline{X}_n.$$

Now X_{ni} is a function of T_i and $\Pi_i^{(n)}$, so by construction $(X_{ni})_{i=1}^n$ are independent and identically distributed. Therefore $\mathbb{E}\overline{X}_n = \mathbb{E}X_{n1}$. Since $X_{ni} \in [0, d_K]$ by (4.5) and $X_{ni} \to 0$ in probability by (4.4), we have $\mathbb{E}\overline{X}_n \to 0$ by the bounded convergence theorem. In general L_1 convergence does not imply almost sure convergence, but here we deal with averages so the latter can be established. The centred versions $Y_{ni} = X_{ni} - \mathbb{E}X_{ni}$ are again bounded, and repeating the proof of the fourth moment law of large numbers (Durrett [33, Theorem 2.3.5]), we have

$$\mathbb{P}\left(\left(\overline{X}_n - \mathbb{E}\overline{X}_n\right)^4 > \epsilon\right) = \mathbb{P}(\overline{Y_n}^4 > \epsilon) \leq \frac{n\mathbb{E}\left[Y_{n1}^4\right] + 3n(n-1)\mathbb{E}\left[Y_{n1}^2\right]}{\epsilon^4 n^4} \leq \frac{3\max(d_K^4, d_K^2)}{\epsilon^4 n^2}.$$

Put $\epsilon = n^{-1/5}$ and apply the Borel–Cantelli lemma while observing that $\mathbb{E}\overline{X}_n \to 0$ to conclude $|\overline{X}_n| \leq |\overline{X}_n - \mathbb{E}\overline{X}_n| + |\mathbb{E}\overline{X}_n| \to 0$ almost surely.

Uniform convergence of F_n to F comes from a combination of the uniform Lipschitz bound (4.6), the strong law of large numbers and compactness of $\mathscr{W}_2(K)$ (Corollary 3.2.4). For each $\gamma \in \mathscr{W}_2$,

$$F_n(\gamma) \xrightarrow{a.s.} F(\gamma),$$

Fix $\epsilon > 0$, invoke the total boundedness of $\mathscr{W}_2(K)$ to find a finite ϵ-cover $\gamma_1, \ldots, \gamma_m$, $m = m(\epsilon)$. By virtue of (4.6), F_n and F are uniformly d_K-Lipschitz. For any $\gamma \in \mathscr{W}_2(K)$ choose j such that $W_2(\gamma, \gamma_j) < \epsilon$. Then

$$\begin{aligned}
|F_n(\gamma) - F(\gamma)| &\leq |F_n(\gamma) - F_n(\gamma_j)| + |F_n(\gamma_j) - F(\gamma_j)| + |F(\gamma_j) - F(\gamma)| \\
&\leq d_K W_2(\gamma, \gamma_j) + |F_n(\gamma_j) - F(\gamma_j)| + d_K W_2(\gamma, \gamma_j) \\
&\leq 2d_K\epsilon + |F_n(\gamma_j) - F(\gamma_j)|.
\end{aligned}$$

Thus almost surely

$$\limsup_{n\to\infty} \sup_{\gamma\in\mathscr{W}_2(K)} |F_n(\gamma) - F(\gamma)| \leq 2d_K\epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude that

$$\sup_{\gamma\in\mathscr{W}_2(K)} |\widehat{F}_n(\gamma) - F(\gamma)| \to 0, \qquad \text{almost surely.}$$

Convergence of minimisers is now standard. If a subsequence of $\widehat{\lambda}_n$ converges to μ, then the uniform convergence of $\widehat{F}_n$ to F and the continuity of F imply that $\widehat{F}_{n_k}(\widehat{\lambda}_{n_k}) \to F(\mu)$. The

definition of $\widehat{\lambda}_n$ gives $\widehat{F}_{n_k}(\widehat{\lambda}_{n_k}) \le \widehat{F}_{n_k}(\lambda) \to F(\lambda)$. Consequently, $F(\mu) \le F(\lambda)$ and it must be that $\mu = \lambda$ because λ is the unique minimiser of F. Since $\widehat{\lambda}_n$ is a sequence in the compact set $\mathscr{W}_2(K)$, this means that $W_2(\widehat{\lambda}_n, \lambda) \to 0$ almost surely.

Lastly, we prove convergence almost surely in (1) under the more stringent assumptions on τ_n and on Π, mentioned in the end of the theorem's statement. Let us begin by showing that for all $a = (a_1, \dots, a_d) \in \mathbb{R}^d$,

$$\mathbb{P}\left(\frac{\widetilde{\Pi}_i^{(n)}((-\infty, a])}{\tau_n} - \Lambda_i((-\infty, a]) \to 0 \right) = 1.$$

To simplify we shall write a instead of $(-\infty, a]$ henceforth. Recall that $\widetilde{P}_j^{(n)}$ are generic point processes, distributed as $T_i \# \Pi$ and independent across j. We may assume that they are constructed as $T_i \# P_j^{(n)}$ with $P_j^{(n)}$ distributed as Π.

Define the random variables

$$X_{nj} = \widetilde{P}_j^{(n)}(a) - \Lambda_i(a), \quad j = 1, \dots, \tau_n; \qquad S_n = \sum_{j=1}^{\tau_n} X_{nj}.$$

The idea is now to use the fourth-moment law of large numbers (Durrett [33, Theorem 2.3.5]) conditional on Λ_i. Here is an informal argument. Since $\Lambda_i = T_i \# \lambda$, conditioning on Λ_i is equivalent to conditioning on T_i. The random variables X_{nj} have conditional mean zero by construction; and since T_i and $\{\Pi_i^{(n)}\}$ are independent, X_{nj} are also conditionally independent across j. It follows that

$$\mathbb{E}[S_n^4 | T_i] = \sum_{j=1}^{\tau_n} \mathbb{E}[X_{nj}^4 | T_i] + \sum_{j<l} \mathbb{E}[X_{nj}^2 X_{nl}^2 | T_i] = \tau_n \mathbb{E}[X_{11}^4 | T_i] + 3\tau_n(\tau_n - 1)\mathbb{E}[X_{11} X_{12} | T_i].$$

To see this formally, we set $k = \tau_n$ and define $\Phi : (M(U))^k \times C_b(U, K) \to \mathbb{R}_+$ by

$$\Phi(p_1, \dots, p_k, f) = \left[\sum_{j=1}^{k} f \# p_j(a) - f \# \lambda(a) \right]^4, \quad f \in C_b(U, K); \quad p_j \in M(U).$$

(Recall that $M(U)$ is the collection of finite Borel measures on U endowed with the topology of narrow convergence.) Then $S_n^4 = \Phi(P_1^{(n)}, \dots, P_k^{(n)}, T_i)$ and we claim that Φ is continuous (hence measurable). Indeed, if V_n are random vectors that converge narrowly to V and f_n are continuous functions that converge uniformly to f, then $f_n(V_n) \to f(V)$ narrowly by the continuous mapping theorem and Slutsky's theorem. Another application of Slutsky's theorem then shows that Φ is continuous. Finally, Φ is integrable because $0 \le f \# \lambda(a) \le 1$ and $\mathbb{E}[T_i \# P_j^{(n)}(a)]^4 \le \mathbb{E}[\Pi(\mathbb{R}^d)]^4 < \infty$ by the hypothesis.

Since $\{P_j^{(n)}\}$ and T_i are independent, one can evaluate the conditional expectation $\mathbb{E}[S_n^4 | T_i]$ by

taking the expectation with respect to P. That is, if we define $g : C_b(U,K) \to \mathbb{R}_+$ by

$$g(f) = \mathbb{E}_P\left[\Phi(P_1^{(n)},\dots,P_k^{(n)},f)\right] = \int_{[M(U)]^k} \Phi(p_1,\dots,p_k,f)\,\mathrm{d}(p_1,\dots,p_k), \quad f \in C_b(U,K),$$

then Lemma 6.2.1 in Durrett [33] gives $\mathbb{E}[S_n^4|T_i] = g(T_i)$.

The same idea shows that for each j,

$$\mathbb{E}[X_{nj}|T_i] = \int_{M(U)} T_i\#p_j(a)\,\mathrm{d}p_j - T_i\#\lambda(a) = \lambda(T_i^{-1}(a)) - \lambda(T_i^{-1}(a)) = 0.$$

This provides the formal justification for the expression for $\mathbb{E}[S_n^4|T_i]$. If we now take the expectation with respect to T_i and apply Markov's inequality, we obtain

$$P\left[\left(\frac{S_n}{\tau_n}\right)^4 > \epsilon\right] \le \frac{\mathbb{E}[S_n^4]}{\epsilon^4\tau_n^4} = \frac{\tau_n\mathbb{E}[X_{11}^4] + 3\tau_n(\tau_n-1)\mathbb{E}[X_{11}^2X_{12}^2]}{\epsilon^4\tau_n^4}.$$

Since the expectations are finite, the right-hand side is bounded by a constant times τ_n^{-2}, which is a convergent sum by the hypothesis. The result now follows from the Borel–Cantelli lemma.

Now that we have convergence for a fixed $a \in \mathbb{R}^d$, we use a standard approximation by rationals to obtain the convergence for *all* $a \in \mathbb{R}^d$. Indeed, we have

$$\mathbb{P}\left(\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \to 0 \text{ for any } a \in \mathbb{Q}^d\right) = 1.$$

If $a \in \mathbb{R}^d$ is arbitrary, then we can find rational sequences $a^k \nearrow a \swarrow b^k$ that converge monotonically coordinatewise to a. We can then use the approximations

$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \le \frac{\widetilde{\Pi}_i^{(n)}(b^k)}{\tau_n} - \Lambda_i(b^k) + \Lambda_i(b^k) - \Lambda_i(a);$$
$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \ge \frac{\widetilde{\Pi}_i^{(n)}(a^k)}{\tau_n} - \Lambda_i(a^k) + \Lambda_i(a^k) - \Lambda_i(a).$$

The resulting errors

$$\Lambda_i(b^k) - \Lambda_i(a) = \Lambda_i((-\infty,b^k] \setminus (-\infty,a]) \quad \text{and} \quad \Lambda_i(a^k) - \Lambda_i(a) = -\Lambda_i((-\infty,a] \setminus (-\infty,a^k])$$

both vanish as $k \to \infty$: the first set converges monotonically to the empty set; the second one does not converge to empty set but rather to $(-\infty,a] \setminus (-\infty,a)$, which is a union of d rays of dimension $d-1$. When a is a continuity point of Λ_i, this is still a Λ_i-null set. We may therefore conclude that with probability one

$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \to 0, \qquad \text{for all } a \in \mathbb{R}^d \text{ continuity point of } \Lambda_i.$$

Taking $a = \infty$, we see that $\tau_n / N_i^{(n)} \to 1$ almost surely, so that

$$\frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}} \to \Lambda_i \text{ narrowly.}$$

Since all these measures are concentrated on the compact set $K \subset \mathbb{R}^d$, the convergence holds in Wasserstein distance too. Finally,

$$W_2(\widehat{\Lambda}_i, \Lambda_i) \leq W_2(\widehat{\Lambda}_i, \widetilde{\Pi}_i^{(n)} / N_i^{(n)}) + W_2(\widetilde{\Pi}_i^{(n)} / N_i^{(N)}, \Lambda_i) \to 0, \qquad n \to \infty,$$

by Lemma 4.4.2 if $\sigma_i^{(n)} \to 0$. □

4.4.2 Consistency of warp functions and inverses

We next discuss the consistency of the warp and registration function estimators. These are key elements in order to align the observed point patterns $\widetilde{\Pi}_i$. Recall that we have consistent estimators $\widehat{\Lambda_i}$ for Λ_i and $\widehat{\lambda}_n$ for λ. Then $T_i = \mathbf{t}_\lambda^{\Lambda_i}$ is estimated by $\mathbf{t}_{\widehat{\lambda}_n}^{\widehat{\Lambda_i}}$ and T_i^{-1} is estimated by $\mathbf{t}_{\widehat{\Lambda_i}}^{\widehat{\lambda}_n}$. We will make the following extra assumptions that make the statements more transparent (otherwise one needs to replace K with the set of Lebesgue points of the supports of λ and Λ_i).

Assumptions 5 (strictly positive measures)**.** *In addition to Assumptions 4 suppose that:*

1. *λ has a positive density on K (equivalently,* $\mathrm{supp}\lambda = K$*);*
2. *T is almost surely surjective on $U = \mathrm{int}K$ (thus a homeomorhpism of U).*

As a consequence $\mathrm{supp}\Lambda = \mathrm{supp}(T\#\lambda) = \overline{T(\mathrm{supp}\lambda)} = K$ almost surely.

Theorem 4.4.3 (consistency of optimal maps)**.** *Let Assumptions 5 be satisfied in addition to the hypotheses of Theorem 4.4.1. Then for any i such that $\sigma_i^{(n)} \xrightarrow{p} 0$ and any compact set $S \subseteq \mathrm{int}K$,*

$$\sup_{x \in S} \| \widehat{T_i^{-1}}(x) - T_i^{-1}(x) \| \xrightarrow{\mathrm{p}} 0, \qquad \sup_{x \in S} \| \widehat{T_i}(x) - T_i(x) \| \xrightarrow{\mathrm{p}} 0.$$

Almost sure convergence can be obtained under the same provisions made at the end of the statement of Theorem 4.4.1.

A few technical remarks are in order. First and foremost, it is not clear that the two suprema are measurable. Even though T_i and T_i^{-1} are random elements in $C_b(U, \mathbb{R}^d)$, their estimators are only defined in an L_2 sense. The proof of Theorem 4.4.3 is done ω-wise. That is, for any ω in the probability space such that Theorem 4.4.1 holds, the two suprema vanish as $n \to \infty$. In other words, the convergence holds in outer probability or outer almost surely.

Secondly, assuming positive smoothing, the random measures $\widehat{\Lambda_i}$ are smooth with densities bounded below on K, so $\widehat{T_i^{-1}}$ are defined on the whole of U (possibly as set-valued functions

on a Λ_i-null set). But the only known regularity result for $\widehat{\lambda}_n$ is an upper bound on its density (Proposition 3.5.17), so it is unclear what is its support and consequently what is the domain of definition of $\widehat{T_i}$.

Lastly, when the smoothing parameter σ is zero, $\widehat{T_i}$ and $\widehat{T_i^{-1}}$ are not defined. Nevertheless, theorem 4.4.3 still holds in the set-valued formulation of Proposition 2.9.11, of which it is a rather simple corollary:

Proof of Theorem 4.4.3. The proof simply amounts to setting up the scene in order to apply Proposition 2.9.11 of stability of optimal maps. We define

$$\mu_n = \widehat{\Lambda}_i; \qquad \nu_n = \widehat{\lambda}_n; \qquad \mu = \Lambda_i; \qquad \nu = \lambda; \qquad u_n = \widehat{T}_i^{-1}; \qquad u = T_i^{-1},$$

and verify the conditions of the proposition. The narrow convergence of μ_n to μ and ν_n to ν is the conclusion of Theorem 4.4.1; the finiteness is apparent because K is compact and the uniqueness follows from the assumed absolute continuity of Λ_i. Since in addition T_i^{-1} is uniquely defined on $U = \mathrm{int}K$ which is an open convex set, the restrictions on Ω in Proposition 2.9.11 are redundant. Uniform convergence of $\widehat{T_i}$ to T_i is proven in the same way. □

Corollary 4.4.4 (consistency of point pattern registration)**.** *For any i such that $\sigma_i^{(n)} \xrightarrow{p} 0$,*

$$W_2\left(\frac{\widehat{\Pi_i^{(n)}}}{N_i^{(n)}}, \frac{\Pi_i^{(n)}}{N_i^{(n)}}\right) \xrightarrow{\mathrm{p}} 0.$$

The division by the number of observed points ensures that the resulting measures are probability measures; the relevant information is contained in the point patterns themselves, which is invariant under this normalisation.

Proof. Since $\widehat{\Pi_i^{(n)}} = \widehat{T_i^{-1}} \circ T_i \# \Pi_i^{(n)}$, we have the upper bound on the squared Wasserstein distance:

$$\int_K \|\widehat{T_i^{-1}}(T_i(x)) - x\|^2 \,\mathrm{d}\frac{\Pi_i^{(n)}}{N_i^{(n)}},$$

and this is well-defined (that is, $N_i^{(n)} > 0$) almost surely for n large enough by Lemma 4.6.1. Fix a compact $\Omega \subseteq \mathrm{int}K$ and split the integral to Ω and its complement. Then

$$\int_{K\setminus\Omega} \|\widehat{T_i^{-1}}(T_i(x)) - x\|^2 \,\mathrm{d}\frac{\Pi_i^{(n)}}{N_i^{(n)}} \leq d_K^2 \frac{\Pi_i^{(n)}(K\setminus\Omega)}{\tau_n}\frac{\tau_n}{N_i^{(n)}} \xrightarrow{\mathrm{as}} d_K^2 \lambda(K\setminus\Omega),$$

by the law of large numbers. By writing $\mathrm{int}K$ as a countable union of compact sets (and since λ is absolutely continuous), this can be made arbitrarily small by choice of Ω.

We can easily bound the integral on Ω itself by

$$\int_\Omega \|\widehat{T_i^{-1}}(T_i(x)) - x\|^2 \,\mathrm{d}\frac{\Pi_i^{(n)}}{N_i^{(n)}} \le \sup_{x\in\Omega} \|\widehat{T_i^{-1}}(T_i(x)) - x\|^2 = \sup_{y\in T_i(\Omega)} \|\widehat{T_i^{-1}}(y) - T_i^{-1}(y)\|^2.$$

But $T_i(\Omega)$ is a compact subset of $U = \mathrm{int}K$, because $T_i \in C_b(U,U)$. The right-hand side therefore vanishes as $n \to \infty$ by Theorem 4.4.3, and this completes the proof. □

We conclude with a discussion on possible extensions pertaining to the boundary of K.

Indeed, stronger statements can be made when we can control the behaviour at the boundary of K. For example, when $\mathscr{X} = \mathbb{R}$, $K = [a,b]$ and the construction guarantees that $\widehat{T_i^{-1}}(a) = a$ and $\widehat{T_i^{-1}(b)} = b$, because in the one-dimensional case we *do* know that the Fréchet mean $\widehat{\lambda}_n$ is strictly positive on K. Consequently, the convergence in Theorem 4.4.3 actually holds on the whole of K. This can also be seen in elementary ways by properties of nondecreasing functions on the real line [70].

The interpretation of this property when $d = 1$ in terms of the set-valued framework is more propitious for extensions to multivariate setups. Let u be the set-valued function represented by T_i^{-1}. If $x = b \in \partial K$, then $u(x)$ is a subset of the ray $[b,\infty)$ (because u is nondecreasing and $u(z) \to b$ as $z \nearrow b$). In other words, there is a unique $y \in K$ that can be an element of $u(x)$, namely $y = b$. The same thing happens at $x = a$, which is the only other point of the boundary of K.

Now suppose that $\mathscr{X} = \mathbb{R}^d$ and u is as above. Assume that for each $x \in \partial K$, $u(x) \cap K$ contains exactly one element y. Let x_n be a sequence in U that converges to $x \in \partial K$. If $y_n \in u(x_n)$ and $y_n \to y$, then it is not difficult to see that $y \in u(x)$ (this property is called upper semicontinuity of set-valued functions and proven in Alberti & Ambrosio [3, Corollary 1.3]). Since y_n must be in K, it follows that they must converge to y. The same convergence holds when $y_n \in u_n(x_n)$, where u_n is represented by $\widehat{T_i^{-1}}$. In other words, we have extended the uniform convergence on compact subsets of U to uniform convergence on U itself.

Finally, for Corollary 4.4.4 we have assumed that $T_i(x) \in U$ for all $x \in U$. Let us see two sufficient conditions for this to be a consequence rather than an assumption: one in terms of T_i, the other in terms of the geometry of K. What we do know is that $T_i(x) \in K$ for all $x \in U$ and it is of interest to see whether this property suffices. Suppose that $y = T_i(x) \in \partial K$ for some $x \in \mathrm{int}K$. By the Hahn–Banach theorem there exists $\alpha \in \mathbb{R}^d \setminus \{0\}$ with $\langle y, \alpha \rangle \ge \sup \langle K, \alpha \rangle$. Let $x' = x + t\alpha$ for $t > 0$ small enough such that $x' \in U$. Then $y' = T_i(x') \in K$, so that

$$0 \le \langle y' - y, x' - x \rangle = t \langle y' - y, \alpha \rangle.$$

One way to obtain a contradiction is to assume that T_i is strictly monotone on U; and this happens when the convex potential of T_i is strictly convex on U.

Another way is to assume that α separates y from K strictly, in the sense that

$$\langle y, \alpha \rangle > \langle y', \alpha \rangle, \qquad y' \in K \setminus \{y\}.$$

When such a strict separator exists (and $y \in K$), we say that y is an **exposed** point of K. When this is the case, the inequality $0 \leq t\langle y' - y, \alpha \rangle$ entails $y' = y$, because $t > 0$. This is a contradiction to the injectivity of T_i. Hence when *any* boundary point of K is exposed, T_i must map U into U. Examples for such K include the unit ball or any ellipsoid in $\mathbb{R}^d$ and more generally, when it can be written as $\partial K = \{x : \varphi_K(x) = 0\}$, for some strictly convex function φ_K. Indeed, if α creates a supporting hyperplane to K at y and $\langle \alpha, y \rangle = \langle \alpha, y' \rangle$ for $y \neq y'$, then as φ_K is strictly convex on the line segment $[y, y']$, it is impossible that $y' \in K$ without the hyperplane intersecting the interior of K. Although this condition excludes some interesting cases, perhaps most prominently polyhedral sets such as $K = [0,1]^d$, such sets can be approximated by convex sets that do satisfy it (Krantz [59, Proposition 1.12]).

4.5 Illustrative examples

In this section we illustrate the estimation framework put forth in this chapter by considering an example of a structural mean λ with a bimodal density on the real line. The unwarped point patterns Π originate from Poisson processes with mean measure λ and, consequently, the warped points $\widetilde{\Pi}$ are Cox processes (see Subsection 4.1.2). Another scenario involving triangular densities can be found in Panaretos & Zemel [70].

4.5.1 Explicit classes of warp maps

As a first step we introduce a class of random warp maps satisfying Assumptions 3, that is, increasing maps that have as mean the identity function. The construction is a mixture version of similar maps considered by Wang & Gasser in [91, 92].

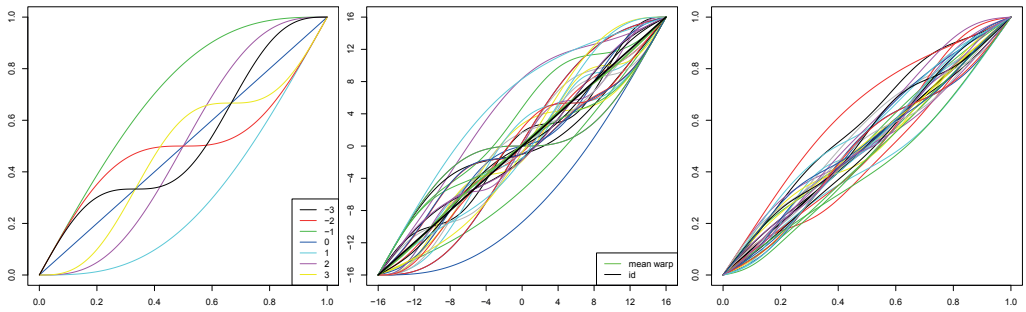
For any integer k define $\zeta_k : [0,1] \to [0,1]$ by

$$\zeta_0(x) = x, \qquad \zeta_k(x) = x - \frac{\sin(\pi k x)}{|k|\pi}, \quad k \in \mathbb{Z} \setminus \{0\}. \tag{4.7}$$

Clearly $\zeta_k(0) = 0$, $\zeta_k(1) = 1$ and ζ_k is smooth and strictly increasing for all k. Figure 4.4(a) plots ζ_k for $k = -3, \ldots, 3$. To make ζ_k a random function we let k be an integer-valued random variable. If the latter is symmetric, then we have

$$\mathbb{E}[\zeta_k(x)] = x, \qquad x \in [0,1].$$

By means of mixtures, we replace this discrete family by a continuous one: let $J > 1$ be an integer and $V = (V_1, \ldots, V_J)$ be a random vector following the flat Dirichlet distribution (uniform on the set of nonnegative vectors with $v_1 + \cdots + v_J = 1$). Take independently k_j



(a) (b) (c)

Figure 4.4: (a) The functions $\{\zeta_{-3},\ldots,\zeta_3\}$; (b) Realisations of T defined by (4.8) with $J=2$ and k_j symmetrisations of Poisson random variables with mean 3; (c) Realisations of T defined by (4.8) with $J=10$ and k_j as in (b).

following the same distribution as k and define

$$T(x) = \sum_{j=1}^{J} V_j \zeta_{k_j}(x). \tag{4.8}$$

Since V_j is positive, T is increasing and as (V_j) sums up to unity T has mean identity. Realisations of these warp functions are given in Figures 4.4(b) and 4.4(c) for $J=2$ and $J=10$ respectively. The parameters (k_j) were chosen as symmetrised Poisson random variables: each k_j has the law of XY with X Poisson with mean 3 and $\mathbb{P}(Y=1)=\mathbb{P}(Y=-1)=1/2$ for Y and X independent. We see that when $J=10$ is large, the function T deviates only mildly from the identity, since a law of large numbers begins to take effect. In contrast, $J=2$ yields functions that are quite different from the identity. Thus, it can be said that the parameter J controls the variance of the random warp function T.

4.5.2 Bimodal Cox Processes

Let the structural mean measure λ be a mixture of a bimodal Gaussian distribution (restricted to $K=[-16,16]$) and a beta background on the interval $[-12,12]$, so that mass is added at the centre of K but not near the boundary. In symbols this is given as follows. Let φ be the standard Gaussian density and let $\beta_{\alpha,\beta}$ denote the density of a the beta distribution with parameters α and β. Then λ is chosen as the measure with density

$$f(x) = \frac{1-\epsilon}{2}[\varphi(x-8)+\varphi(x+8)] + \frac{\epsilon}{24}\beta_{1.5,1.5}\left(\frac{x+12}{24}\right), \qquad x\in[-16,16], \tag{4.9}$$

where $\epsilon\in[0,1]$ is the weight of the beta background. (We ignore the loss of a negligible amount of mass due to the restriction of the Gaussians to $[-16,16]$.) Plots of the density and distribution functions are given in Figure 4.5.

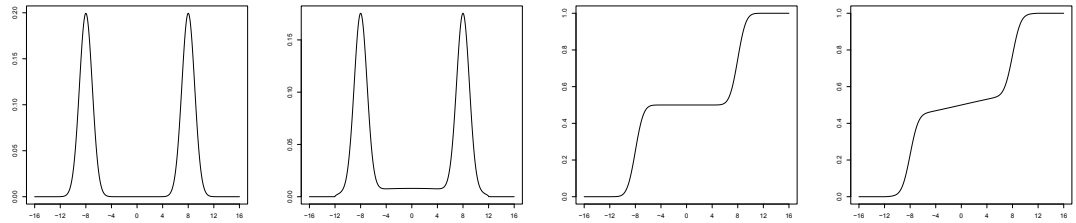


Figure 4.5: Density and distribution functions corresponding to (4.9) with $\epsilon = 0$ and $\epsilon = 0.15$.

The main criterion for the quality of our regularised Fréchet–Wasserstein estimator will be its success in discerning the two modes at ± 8; these will be smeared by the phase variation arising from the warp functions.

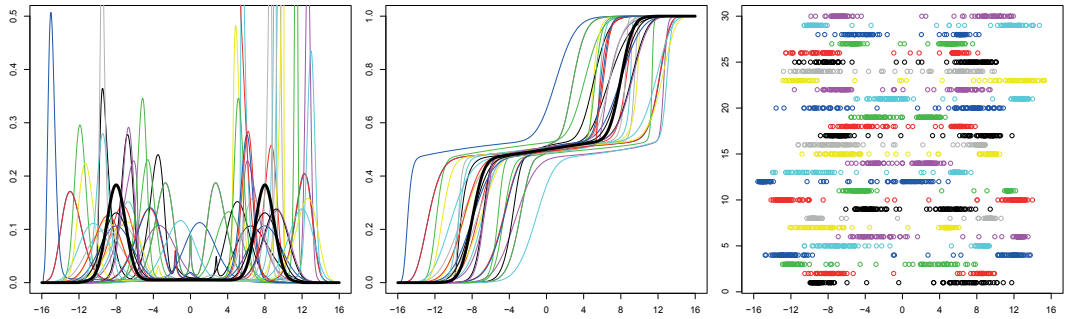
We next simulated 30 independent Poisson processes with mean measure λ, $\epsilon = 0.1$ and total intensity (expected number of points) $\tau = 93$. In addition, we generated warp functions as in (4.8) but rescaled to $[-16, 16]$; that is, having the same law as the functions

$$32\,T\left(\frac{x+16}{32}\right) - 16$$

from K to K. These cause rather violent phase variation, as can be seen by the plots of the densities and distribution functions of the conditional measures $\Lambda = T\#\lambda$ presented in Figures 4.6(a) and 4.6(b); the warped points themselves are displayed in Figure 4.6(c).

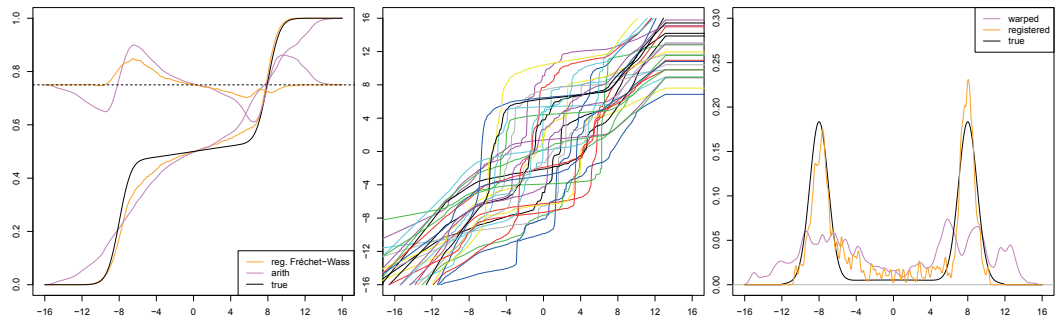
Using these warped point patterns, we construct the *regularised Fréchet–Wasserstein* estimator employing the procedure described in Section 4.3. Each $\widetilde{\Pi}_i$ was smoothed with a Gaussian kernel and bandwidth chosen by unbiased cross validation. We deviate slightly from the recipe presented in Section 4.3 by not restricting the resulting estimates to the interval $[-16, 16]$, but this has no essential effect on the finite sample performance. The regularised Fréchet–Wasserstein estimator $\widehat{\lambda}_n$ serves as the estimator of the structural mean λ and is shown in Figure 4.7(a). It is contrasted with λ at the level of distribution functions, as well as with the empirical arithmetic mean; the latter, the **naive estimator**, is calculated by ignoring the warping and simply averaging linearly the (smoothed) empirical distribution functions across the observations. We notice that $\widehat{\lambda}_n$ is rather successful at locating the two modes of λ, in contrast with the naive estimator that is more diffuse (the distribution function increases approximately linearly, suggesting a nearly constant density instead of the correct bimodal one).

Estimators of the warp maps $\widehat{T_i}$, depicted in Figure 4.7(b), and their inverses are defined as the optimal maps between $\widehat{\lambda}_n$ and the estimated conditional mean measures, as explained in Subsection 4.3.4. Then we register the point patterns by applying to them the inverse estimators $\widehat{T_i^{-1}}$ (Figure 4.8). Figure 4.7(c) gives two kernel estimators of the density of λ constructed from a superposition of all the warped points and all the registered ones. Notice



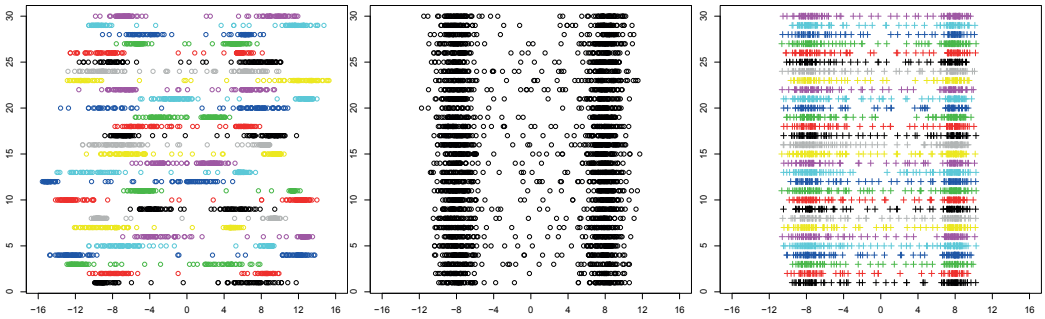
(a) (b) (c)

Figure 4.6: (a) 30 warped bimodal densities, with density of λ given by (4.9) in solid black; (b) Their corresponding distribution functions, with that of λ in solid black; (c) 30 Cox processes, constructed as warped versions of Poisson processes with mean intensity $93f$ using as warp functions the rescaling to $[-16.16]$ of (4.8).



(a) (b) (c)

Figure 4.7: (a) Comparison between the the regularised Fréchet–Wasserstein estimator, the empirical arithmetic mean, and the true distribution function, including residual curves centred at $y = 3/4$; (b) The estimated warp functions; (c) Kernel estimates of the density function f of the structural mean, based on the warped and registered point patterns.



(a) (b) (c)

Figure 4.8: Bimodal Cox processes: (a) The observed warped point processes; (b) The unobserved original point processes; (c) The registered point processes.

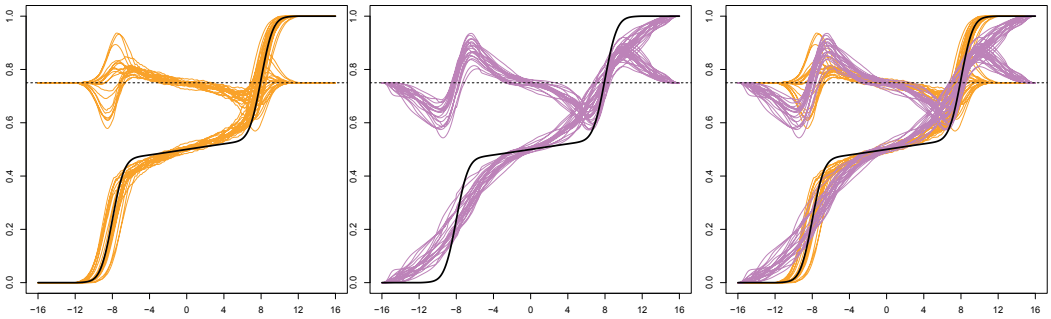
that the estimator that uses the registered points is much more successful than the one using the warped ones in discerning the two density peaks. This is not surprising after a brief look at Figure 4.8, where the unwarped, warped and registered points are displayed. Indeed, there is very high concentration of registered points around the true location of the peaks, ± 8. This is not the case for the warped points because of the phase variation that translates the centres of concentration for each individual observation. It is important to remark that the fluctuations in the density estimator in Figure 4.7(c) are not related to the registration procedure, and could be reduced by a better choice of bandwidth (note that our procedure does not attempt to estimate the density, but rather, the distribution function).

Figure 4.9 presents a superposition of the regularised Fréchet–Wasserstein estimators for 20 independent replications of the experiment, contrasted with a similar superposition for the naive estimator. The latter is clearly seen to be biased around the two peaks, while the regularised Fréchet–Wasserstein seems approximately unbiased, despite presenting fluctuations. It always captures the bimodal nature of the density, as is seen from the two clear elbows in each realisation.

To illustrate the consistency of the regularised Fréchet–Wasserstein estimator $\widehat{\lambda}_n$ for λ as shown in Theorem 4.4.1, we let the number of processes n as well as the expected number of observed point per process τ to vary. Figures 4.10 and 4.11 show the sampling variation of $\widehat{\lambda}_n$ for different values of n and τ. We observe that as either of these increases, the realisations $\widehat{\lambda}_n$ indeed approach λ. The figures suggest that, in this scenario, the amplitude variation plays a stronger role than the phase variation, as the effect of τ is more substantial.

4.5.3 Effect of the smoothing parameter

In order to work with measures of strictly positive density, the observed point patterns have been smoothed using a kernel function. This necessarily incurs an additional bias that depends



(a) (b) (c)

Figure 4.9: (a) Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment; (b) Sampling variation of the arithmetic mean, and the true mean measure λ for the same 20 replications; (c) Superposition of (a) and (b). For ease of comparison all three panels include residual curves centred at $y = 3/4$.

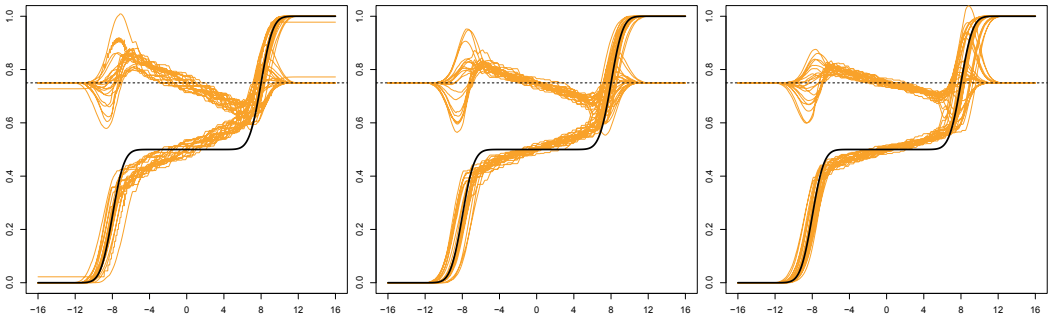


Figure 4.10: Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment, with $\epsilon = 0$ and $n = 30$. Left: $\tau = 43$; middle: $\tau = 93$; right: $\tau = 143$. For ease of comparison all three panels include residual curves centred at $y = 3/4$.

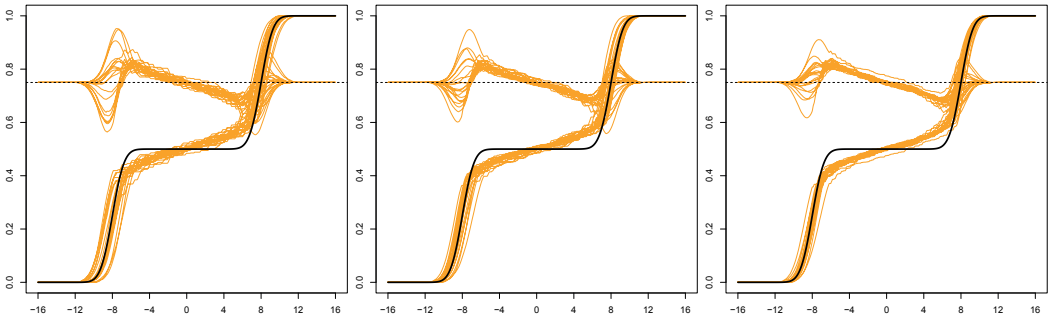


Figure 4.11: Sampling variation of the regularised Fréchet–Wasserstein mean $\widehat{\lambda}_n$ and the true mean measure λ for 20 independent replications of the experiment, with $\epsilon = 0$ and $\tau = 93$. Left: $n = 30$; middle: $n = 50$; right: $n = 70$. For ease of comparison all three panels include residual curves centred at $y = 3/4$.

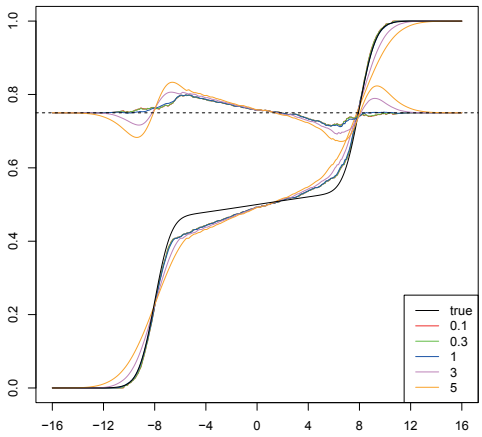


Figure 4.12: Regularised Fréchet–Wasserstein mean as a function of the smoothing parameter multiplier s, including residual curves. Here $n = 30$ and $\tau = 143$.

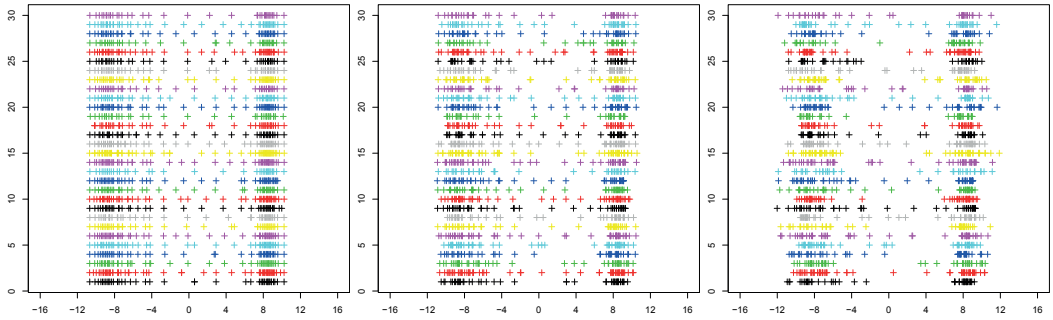


Figure 4.13: Registered point processes as a function of the smoothing parameter multiplier s. Left: $s = 0.1$; middle: $s = 1$; right: $s = 3$. Here $n = 30$ and $\tau = 43$.

on the bandwidth σ_i. Our asymptotic results (Theorem 4.4.1) guarantee the consistency of our estimators, in particular the regularised Fréchet–Wasserstein estimator $\widehat{\lambda}_n$, provided that $\max_{i=1}^{n} \sigma_i \to 0$. In our simulations, we choose σ_i in a data-driven way by employing unbiased cross validation. To gauge for the effect of the smoothing, we carry out the same estimation procedure but with σ_i multiplied by a parameter s. Figure 4.12 presents the distribution function of $\widehat{\lambda}_n$ as a function of s. Interestingly, the curves are nearly identical as long as $s \leq 1$, whereas when $s > 1$, the bias becomes more substantial.

These findings are reaffirmed in Figure 4.13, that show the registered point processes again as a function of s. We see that only minor differences are present as s varies from 0.1 to 1, for example in the grey (8), black (17) and green (19) processes. When $s = 3$, the distortion becomes quite more substantial. This phenomenon repeats itself across all combinations of n, τ and s tested.

4.6 Further results on the real line

4.6.1 Convergence rates and a central limit theorem

Since the conditional mean measures Λ_i are discretely observed, the rate of convergence of our estimators will be affected by the rate at which the number of observed points per process $N_i^{(n)}$ increases to infinity. The latter is controlled by the next lemma, which is in fact valid for any complete separable metric space $\mathscr{X}$.

Lemma 4.6.1 (number of points grows linearly)**.** *Let $N_i^{(n)} = \Pi_i^{(n)}(\mathscr{X})$ denote the total number of observed points. If $\tau_n / \log n \to \infty$, then there exists a constant $C_\Pi > 0$, depending only on the distribution of Π, such that almost surely*

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\tau_n} \ge C_\Pi.$$

In particular, there are no empty point processes, so the normalisation is well-defined. If Π is a Poisson process, then we have the more precise result

$$\lim_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\tau_n} = 1 \qquad \textit{almost surely.}$$

Remark 10. *One can also show that the limit superior of the same quantity is bounded by a constant C'_Π. If $\tau_n / \log n$ is bounded below, then the same result holds but with worse constants. If only $\tau_n \to \infty$, then the result holds for each i separately but in probability.*

Proof. If Π is a binomial process (i.e. $\widetilde{\Lambda}_i$ is the empirical measure), then $N_i^{(n)} = \tau_n$ for all i and all n and there is nothing to prove.

Let us begin with the Poisson case, in which case the argument is more transparent. In this case $N_1^{(n)}, \dots, N_n^{(n)}$ are independent Poisson random variables with parameter τ_n. We can then use a Chernoff bound as follows: if N has a Poisson(τ) distribution, then for any $c > 1$ and any $t \ge 0$,

$$\mathbb{P}(N \le \tau/c) = \mathbb{P}(e^{-Nt} \ge e^{-\tau t/c}) \le \frac{\mathbb{E}e^{-Nt}}{e^{-\tau t/c}} = \exp\left[\tau\left(e^{-t} + \frac{t}{c} - 1\right)\right].$$

The bound is optimised when $t = \log c$, yielding

$$\mathbb{P}(N \le \tau/c) = \exp -\tau\alpha, \qquad \alpha = \alpha(c) = c^{-1}[c - 1 - \log c] > 0.$$

Since $\tau_n \to \infty$, this in particular shows that the probability that $N_i^{(n)}/\tau_n < 1/c$ vanishes as $n \to \infty$.

By Bonferroni's inequality, and since $N_i^{(n)}$ have the same distribution,

$$\mathbb{P}\left(\min_{1\le i\le n} N_i^{(n)} \le \frac{\tau_n}{c}\right) \le n\mathbb{P}\left(N_1^{(n)} \le \frac{\tau_n}{c}\right) \le n\exp[-\alpha(c)\tau_n].$$

If $\tau_n/\log_n \to \infty$ then for n large, the expression in the exponent is smaller than $-3\log n$. Summation over n of the probability on the left-hand side is therefore convergent, and the Borel–Cantelli lemma gives

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\tau_n} \ge 1 \qquad \text{almost surely.}$$

One then shows the reverse inequalities by analogous calculations.

When Π is no longer Poisson, we replace the above argument with a Chernoff bound on binomial distributions, using a very crude bound.

Denote by p the probability that Π has no points. If τ is an integer, then $N_i^{(n)}$ is a sum of independent integer-valued random variables X_i. Since X_i is always an integer, we have the lower bound $X_i \ge \mathbf{1}\{X_i \ge 1\}$. Thus $N_i^{(n)}$ is stochastically larger than a random variable $N \sim B(\tau_n, 1-p)$. Set $q = 1-p$ and use the Chernoff bound as follows: for any $t \ge 0$

$$\mathbb{P}\left(N \le \frac{\tau q}{c}\right) = \mathbb{P}\left(\exp(-Nt) \ge \exp\left(-t\frac{\tau q}{c}\right)\right) \le \mathbb{E}\exp(-Nt)\exp\left(t\frac{\tau q}{c}\right) = \left[s^{q/c}\left(1-q+\frac{q}{s}\right)\right]^{\tau},$$

where $s = e^t \ge 1$. The bound is optimised when $s = (c-q)/(1-q) > 1$, and we obtain

$$\mathbb{P}\left(N_i^{(n)} \le \tau_n q/c\right) \le \beta^{\tau_n}, \qquad \beta = \beta(q,c) = c\left((1-q)/(c-q)\right)^{1-q/c} < 1.$$

One then concludes as before that if $\tau_n/\log n \to \infty$, then almost surely

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\tau_n} \ge 1-p.$$

Finally, we treat the case where τ_n are not integers. We claim that in any case, the probability that $N_1^{(n)} = 0$ is p^{τ_n}. Indeed, recall that the Laplace functional of $\Pi_1^{(n)}$ is

$$f \mapsto \mathbb{E}e^{-\Pi_1^{(n)} f} = [L_\Pi(f)]^{\tau_n} = \left[\mathbb{E}e^{-\Pi f}\right]^{\tau_n}, \qquad f : \mathscr{X} \to \mathbb{R}_+.$$

By the bounded convergence, we may recover the zero probabilities by taking $f \equiv m$ to be a constant function:

$$\mathbb{P}(N_i^{(n)} = 0) = \lim_{m\to\infty} \mathbb{E}e^{-mN_i^{(n)}} = \lim_{m\to\infty} [L_\Pi(m)]^{\tau_n} = \lim_{m\to\infty} [\mathbb{E}e^{-m\Pi(\mathscr{X})}]^{\tau_n} = p^{\tau_n}.$$

By infinite divisibility, $N_i^{(n)}$ has the same law as the sum of $\lfloor \tau_n \rfloor$ (the largest integer not larger than τ_n) independent integer valued random variables with zero probability $p' = p^{\tau_n/\lfloor \tau_n \rfloor} \le p$.

The same argument then gives

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\lfloor \tau_n \rfloor} \ge 1-p,$$

and as $\tau_n \to \infty$, we may replace $\lfloor \tau_n \rfloor$ by τ_n, which completes the proof. □

With Lemma 4.6.1 under our belt we can replace terms of the order $\min_i N_i^{(n)}$ by the more transparent order τ_n. As in the consistency proof, the idea is to write

$$F - \widehat{F}_n = (F - F_n) + (F_n - \tilde{F}_n) + (\tilde{F}_n - \widehat{F}_n)$$

and control each term separately. The first term is corresponds to the phase variation, and comes from the approximation of the theoretical expectation F by a sample mean F_n. The second term is associated with the amplitude variation resulting from observing Λ_i discretely. The third term can be viewed as the bias incurred by the smoothing procedure. Accordingly, the rate at which $\widehat{\lambda}_n$ converges to λ is a sum of three separate terms. We recall the standard $O_{\mathbb{P}}$ terminology: if X_n and Y_n are random variables, then $X_n = O_{\mathbb{P}}(Y_n)$ means that the sequence (X_n/Y_n) is **bounded in probability**, which by definition is the condition

$$\forall \epsilon > 0\ \exists M: \quad \sup_n \mathbb{P}\left(\left| \frac{X_n}{Y_n} \right| > M \right) < \epsilon.$$

Instead of $X_n = O_{\mathbb{P}}(Y_n)$, we will sometimes write $Y_n \ge O_{\mathbb{P}}(X_n)$. The former notation emphasises the condition that X_n grows no faster than Y_n, while the latter stresses out that Y_n grows at least as fast as X_n (which is of course the same assertion). Finally, $X_n = o_{\mathbb{P}}(Y_n)$ means that $X_n/Y_n \to 0$ in probability.

Theorem 4.6.2 (convergence rates in $\mathbb{R}$)**.** *Suppose in addition to Assumptions 4 that $d = 1$, $\tau_n/\log n \to \infty$ and that Π is either a Poisson process or a binomial process. Then*

$$W_2(\widehat{\lambda}_n, \lambda) \le O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt[4]{\tau_n}}\right) + O_{\mathbb{P}}(\sigma_n), \qquad \sigma_n = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^{(n)},$$

where all the constants in the $O_{\mathbb{P}}$ terms are explicit.

Remark 11. *Unlike classical density estimation, no assumptions on the rate of decay of σ_n are required, because we only need to estimate the distribution function and not the derivative. If the smoothing parameter is chosen to be $\sigma_i^{(n)} = [N_i^{(n)}]^{-\alpha}$ for some $\alpha > 0$ and $\tau_n/\log n \to \infty$, then by Lemma 4.6.1 $\sigma_n \le \max_{1\le i\le n} \sigma_i^{(n)} = O_{\mathbb{P}}(\tau_n^{-\alpha})$. For example, if* Rosenblatt's rule *$\alpha = 1/5$ is employed, then the $O_{\mathbb{P}}(\sigma_n)$ term can be replaced by $O_{\mathbb{P}}(1/\sqrt[5]{\tau_n})$.*

One can think about the parameter τ as separating the *sparse* and *dense* regimes as in classical functional data analysis (see also Wu, Müller, & Zhang [93]). If τ is bounded, then the setting is *ultra sparse* and consistency cannot be achieved. A sparse regime can be defined as the case where $\tau_n \to \infty$ but slower than $\log n$. In that case consistency is guaranteed, but some

point patterns will be empty. The *dense* regime can be defined as $\tau_n \gg n^2$, in which case the amplitude variation is negligible asymptotically when compared with the phase variation.

The exponent $-1/4$ of τ_n can be shown to be optimal without further assumptions, but it can be improved to $-1/2$ if $\mathbb{P}(f_\Lambda \geq \epsilon \text{ on } K) = 1$ for some $\epsilon > 0$, where f_Λ is the density of Λ (see Subsection 4.6.2). In terms of T, the condition is that $\mathbb{P}(T' \geq \epsilon) = 1$ for some ϵ and λ has a density bounded below. When this is the case, τ_n needs to compared with n rather than n^2 in the next paragraph and the next theorem.

Theorem 4.6.2 provides conditions for the optimal parametric rate $\sqrt{n}$ to be achieved: this happens if we set σ_n to be of the order $O_{\mathbb{P}}(n^{-1/2})$ or less and if τ_n is of the order n^2 or more. But if the last two terms in Theorem 4.6.2 are *negligible* with respect to $n^{-1/2}$, then a sort of *central limit theorem* holds for $\widehat{\lambda}_n$:

Theorem 4.6.3 (asymptotic normality)**.** *In addition to the conditions of Theorem 4.6.2, assume that $\tau_n / n^2 \to \infty$, $\sigma_n = o_{\mathbb{P}}(n^{-1/2})$ and λ possesses a (piecewise) continuous density that is bounded below on K. Then*

$$\sqrt{n}\left(\mathbf{t}_\lambda^{\widehat{\lambda}_n} - \mathbf{i}\right) \longrightarrow Z \quad \textit{narrowly in } L_2(K),$$

for a zero-mean Gaussian process Z with the same covariance operator of T (the latter viewed as a random element in $L_2(K)$), namely with covariance kernel

$$\kappa(x,y) = \operatorname{cov}\Big\{T(x), T(y)\Big\}.$$

In view of Section 3.3, Theorem 4.6.3 can be interpreted as asymptotic normality of $\widehat{\lambda}_n$ in the *tangential* sense: $\sqrt{n}\log_\lambda(\widehat{\lambda}_n)$ converges to a Gaussian random element in $L_2(K)$.

Proof of Theorem 4.6.2. Denote the quantile function of $\theta \in \mathcal{W}_2(K)$ by $g(\theta) = F_\theta^{-1} \in L_2(0,1)$ and recall that $\mathcal{W}_2(\gamma,\theta) = \|g(\theta) - g(\gamma)\|$ (Section 2.6). The empirical Fréchet mean λ_n that minimises F_n is found by averaging the quantile functions of Λ_i (see Subsection 3.5.2), so that

$$\sqrt{n}(g(\lambda_n) - g(\lambda)) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n F_{\Lambda_i}^{-1} - F_\lambda^{-1}\right).$$

By the central limit theorem in Hilbert spaces, the above expression converges narrowly to a Gaussian limit GP with $\mathbb{E}\|GP\|^2 < \infty$ as $n \to \infty$. In particular,

$$W_2(\lambda_n, \lambda) = \|g(\lambda_n) - g(\lambda)\| = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Replacing λ_n with $\widehat{\lambda}_n$, the minimiser of $\widehat{M}_n$, results in the error

$$\left\|g(\lambda_n) - g(\widehat{\lambda}_n)\right\| = \left\|\frac{1}{n}\sum_{i=1}^n F_{\Lambda_i}^{-1} - \frac{1}{n}\sum_{i=1}^n F_{\widehat{\Lambda}_i}^{-1}\right\| \leq \frac{1}{n}\sum_{i=1}^n \left\|F_{\Lambda_i}^{-1} - F_{\widehat{\Lambda}_i}^{-1}\right\| = \frac{1}{n}\sum_{i=1}^n W_2(\Lambda_i, \widehat{\Lambda}_i).$$

Invoking the triangle inequality, we split this again to the amplitude term and the smoothing term:

$$\frac{1}{n}\sum_{i=1}^{n} W_2(\Lambda_i, \widehat{\Lambda_i}) \le \frac{1}{n}\sum_{i=1}^{n} W_2\left(\Lambda_i, \widetilde{\Lambda}_i\right) + \frac{1}{n}\sum_{i=1}^{n} W_2\left(\widetilde{\Lambda}_i, \widehat{\Lambda_i}\right) \le \frac{1}{n}\sum_{i=1}^{n} W_2\left(\Lambda_i, \widetilde{\Lambda}_i\right) + \sqrt{C_{\psi,K}}\sigma_n$$

by Lemma 4.4.2.

Define (as in the proof of Theorem 4.4.1) $X_{ni} = W_2(\Lambda_i, \widetilde{\Lambda}_i)$ and recall that

$$\widetilde{\Lambda}_i = \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}} \quad \text{if } N_i^{(n)} > 0, \qquad \text{and } \lambda^{(0)} \text{ otherwise.}$$

Set $S_{ni} = \mathbf{1}\{N_i^{(n)} > 0\}$ and write

$$X_{ni} = W_2(\Lambda_i, \widetilde{\Lambda}_i)S_{ni} + W_2(\Lambda_i, \lambda^{(0)})(1 - S_{ni}) \le W_2(\Lambda_i, \widetilde{\Lambda}_i)S_{ni} + d_K(1 - S_{ni}).$$

The last term is zero for n large by Lemma 4.6.1 so converges at any rate: if $a_n \to \infty$ is any sequence, then

$$\mathbb{P}\left(a_n \sum_{i=1}^{n} 1 - S_{ni} > \epsilon\right) = \mathbb{P}\left(a_n \sum_{i=1}^{n} \mathbf{1}\{N_i^{(n)} = 0\} > \epsilon\right) \le \mathbb{P}\left(a_n \sum_{i=1}^{n} \mathbf{1}\{N_i^{(n)} = 0\} > 0\right) \to 0.$$

It remains to find the rate of the average of $X_{ni}S_{ni}$. As a first step, we replace probability calculations by expectations, using Markov's inequality:

$$\mathbb{P}\left(a_n \frac{1}{n}\sum_{i=1}^{n} X_{ni}S_{ni} > \epsilon\right) \le \frac{a_n \mathbb{E}\sum_{i=1}^{n} X_{ni}S_{ni}}{n\epsilon} = \frac{a_n \mathbb{E} X_{n1}S_{n1}}{\epsilon}.$$

The idea is now to replace W_2 by W_1 (using (3.2)), which one can evaluate in terms of distribution functions by Corollary 2.6.3. Let us introduce $a = \inf K$ and $b = \sup K$, so that $K = [a, b]$ (since K is compact and convex). For any measure θ on $\mathbb{R}$ denote $\theta((-\infty, t])$ by $\theta(t)$. Then

$$\mathbb{E} X_{n1}^2 S_{n1} \le d_K \mathbb{E} S_{n1} W_1\left(\Lambda_1, \frac{\widetilde{\Pi}_1^{(n)}}{N_1^{(n)}}\right) = d_K \int_a^b \mathbb{E}\left|\Lambda_1(t) - \frac{\widetilde{\Pi}_1^{(n)}(t)}{N_1^{(n)}}\right| S_{n1}\,\mathrm{d}t = d_K \int_a^b \mathbb{E}|B_t|\,\mathrm{d}t,$$

where B_t is defined by the above equation. Let us assume that Π is a Poisson process. Fix $t \in [a, b]$ and notice that conditional on Λ_1 and on the event $N_1^{(n)} = k$, $B_t = 0$ if $k = 0$ and otherwise follows a centred renormalised binomial distribution, of the form $B_t = B(k, \Lambda_1(t))/k - \Lambda_1(t)$. The variance of B_t is smaller than $1/(4k)$, and this does not depend on Λ_1. Thus $\mathbb{E}B_t^2 | N_1^{(n)} \le S_{n1}/(4N_1^{(n)})$.

The random variable $N_1^{(n)}$ follows a Poisson distribution with parameter $\tau = \tau_n$. Taking

expectations and noticing that $1/k \le 2/(k+1)$, we find

$$\mathbb{E}\frac{S_{n1}}{N_1^{(n)}} = \sum_{k=1}^{\infty} \frac{1}{k} e^{-\tau} \frac{\tau^k}{k!} \le \sum_{k=1}^{\infty} 2e^{-\tau} \frac{\tau^k}{(k+1)!} = 2\tau^{-1} \sum_{k=1}^{\infty} e^{-\tau} \frac{\tau^{k+1}}{(k+1)!} = \frac{2}{\tau}(1 - e^{-\tau} - \tau e^{-\tau}) \le \frac{2}{\tau},$$

so that $\mathbb{E}B_t^2 \le (2\tau_n)^{-1}$ and $\mathbb{E}|B_t| \le (2\tau_n)^{-1/2}$. Thus ($d_K = b - a$)

$$\mathbb{E}X_{n1}S_{n1} \le \sqrt{d_K \int_a^b (2\tau_n)^{-1/2} \, \mathrm{d}t} = d_K (2\tau_n)^{-1/4}.$$

If instead of a Poisson process Π is a binomial process, then $N_1^{(n)} = \tau_n$ and $\mathbb{E}B_t^2 \le (4\tau_n)^{-1}$ so the same result holds with an improved constant (and a shorter proof). We conclude that $W_2(\widehat{\lambda}_n, \lambda)$ is smaller than the sum of terms of orders $n^{-1/2}$, $d_K(2\tau_n)^{-1/4}$, $\sqrt{C_{\psi,K}}\sigma_n$ and a last one that is identically zero for n large. □

Proof of Theorem 4.6.3. The hypotheses guarantee that $\sqrt{n}(g(\widehat{\lambda}_n) - g(\lambda_n))$ is $o_{\mathbb{P}}(1)$ and so by Slutsky's theorem

$$\sqrt{n}\left(F_{\widehat{\lambda}_n}^{-1} - F_{\lambda}^{-1}\right) = \sqrt{n}(g(\widehat{\lambda}_n) - g(\lambda)) \to GP \quad \text{narrowly in } L_2(0,1),$$

where GP is the Gaussian process defined in the proof of Theorem 4.6.2. By the continuous mapping theorem, in order to conclude the narrow convergence

$$\sqrt{n}(\mathbf{t}_{\lambda}^{\widehat{\lambda}_n} - \mathbf{i}) = \sqrt{n}\left(F_{\widehat{\lambda}_n}^{-1} \circ F_{\lambda} - F_{\lambda}^{-1} \circ F_{\lambda}\right) = \left[\sqrt{n}\left(F_{\widehat{\lambda}_n}^{-1} - F_{\lambda}^{-1}\right)\right] \circ F_{\lambda} \to GP \circ F_{\lambda},$$

in $L_2(K)$, it suffices to show that $h \mapsto h \circ F_{\lambda}$ is continuous from $L_2(0,1)$ to $L_2(K)$. Once this is shown, we can also write $Z = GP \circ F_{\lambda}$ as the (narrow) limit of the process

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} F_{\Lambda_i}^{-1} \circ F_{\lambda} - F_{\lambda}^{-1} \circ F_{\lambda}\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{t}_{\lambda}^{\Lambda_i} - \mathbf{i}\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} T_i - \mathbf{i}\right),$$

which again by the central limit theorem in $L_2(K)$ is a mean zero Gaussian process and has covariance kernel

$$\kappa(s,t) = \mathbb{E}[(T(s) - s)(T(t) - t)] = \mathrm{cov}(T(s), T(t)), \qquad s, t \in \mathrm{int}K.$$

To prove the purported continuity of $h \mapsto h \circ F_{\lambda}$, we first notice that this map is linear, so one needs only show continuity at 0. This is a straightforward consequence of the change of variables formula: let $[a,b] = K$ and notice that F_{λ} is strictly increasing and piecewise continuously differentiable on $[a,b]$ with derivative bounded below by $\delta > 0$. Hence for all $p \ge 1$

$$\|h \circ F_{\lambda}\|_{L_p(K)}^p = \int_a^b |h^p(F_{\lambda}(s))| \, \mathrm{d}s = \int_{F_{\lambda}(a)}^{F_{\lambda}(b)} |h^p(t)| \frac{1}{F_{\lambda}'(F_{\lambda}^{-1}(t))} \, \mathrm{d}t \le \frac{1}{\delta}\|h\|_{L_p(0,1)}^p,$$

so when $p = 2$ this map is $\delta^{-1/2}$-Lipschitz. □

4.6.2 Optimality of the rates of convergence

One may find the term $O_{\mathbb{P}}(1/\sqrt[4]{\tau_n})$ in Theorem 4.6.2 to be somewhat surprising, and expect that it ought to be $O_{\mathbb{P}}(1/\sqrt{\tau_n})$. The goal of this section is to show why the rate $1/\sqrt[4]{\tau_n}$ is optimal without further assumptions and discuss conditions under which it can be improved to the optimal rate $1/\sqrt{\tau_n}$. For simplicity we concentrate on the case $\tau_n = n$ and assume that the point process Π is a binomial process; the Poisson case being easily obtained from the simplified one (using Lemma 4.6.1). We are thus led to study rates of convergence of empirical measures in the Wasserstein space. That is to say, for a fixed exponent $p \geq 1$ and a fixed measure $\mu \in \mathscr{W}_p(\mathbb{R})$, we consider independent random variables $X_1, \dots$ with law μ and the **empirical measure** $\mu_n = n^{-1} \sum_{i=1}^n \delta\{X_i\}$.

Rates of convergence of $\mathbb{E}W_p(\mu_n, \mu)$ to zero is a vast topic that we will only touch from a superficial point of view. When μ is a measure on $\mathbb{R}^d$, Barthe & Bordenave [10] give sufficient conditions for $W_p(\mu_n, \mu)$ to be (almost surely) of the order $n^{-1/d}$ when $d > 2p$. Boissard & Le Gouic [19] deal with measures on more general spaces in terms of covering numbers.

In the case of the real line, these rates have been extensively studied by Bobkov & Ledoux [18]. We first give a lower bound on the rate and sketch some of their results that are relevant for our particular application. We then show that the rate $n^{-1/4}$ is optimal for W_2 over the class of compactly supported measures in $\mathbb{R}$. For simplicity, we assume throughout that μ is nondegenerate and compactly supported:

$$\text{conv}(\text{supp}\mu) = [a, b] \text{ is compact.}$$

First and foremost, by Proposition 3.2.5, we know that $W_p(\mu_n, \mu) \to 0$ almost surely. Let F_n be the distribution function of μ_n and F that of μ.

Lemma 4.6.4. *There exists a constant c such that for all $p \geq 1$ and all n*

$$\mathbb{E}W_p(\mu_n, \mu) \geq \frac{c}{\sqrt{n}}.$$

Sketch of proof. It suffices to consider $p = 1$, because $W_p \geq W_1$. Let (t_0, t_1) be a nonempty interval such that $0 < F(t_0) \leq F(t_1) < 1$. By Proposition 2.6.2 and Fubini's theorem

$$W_1(\mu_n, \mu) = \int_{\mathbb{R}} \mathbb{E}|F_n(t) - F(t)| \, \mathrm{d}t \geq \int_{t_0}^{t_1} \mathbb{E}|F_n(t) - F(t)| \, \mathrm{d}t.$$

On that interval the random variable $F_n(t)$ is a binomial with success parameter bounded away from zero and one and so the its absolute deviation from its mean is bounded below by a constant times $\sqrt{n}$, uniformly over t. □

As has been shown in the proof of Theorem 4.6.2, for any n and any t

$$\mathbb{E}|F_n(t) - F(t)| \le \sqrt{\mathrm{var}F_n(t)} \le \frac{1}{2}\frac{1}{\sqrt{n}},$$

and so when μ is supported on $[a,b]$, $\mathbb{E}W_1(\mu_n,\mu) \le (b-a)/(2\sqrt{n})$ and the optimal rate of $\mathbb{E}W_1$ is attained. Since $\mathrm{var}F_n(t) = F(t)(1-F(t))/n$, a more careful reflection shows that this rate is attained if

$$J_1(\mu) = \int_{\mathbb{R}} \sqrt{F(x)(1-F(x))}\,\mathrm{d}x < \infty,$$

which is far weaker than μ having a compact support. Bobkov & Ledoux [18, Corollary 3] show that this condition is also necessary, and extends to general p as follows. Let f denote the density of μ if the latter is absolutely continuous and let $f = 0$ if μ is discrete[1]. Then they show in [18, Theorem 5.10]:

Theorem 4.6.5 (rate of convergence of empirical measures)**.** *Let $p \ge 1$. The condition*

$$J_p(\mu) = \int_{\mathbb{R}} \frac{[F(t)(1-F(t))]^{p/2}}{[f(t)]^{p-1}}\,\mathrm{d}t < \infty$$

is necessary and sufficient for the existence of a constant c_p such that $\mathbb{E}W_p(\mu_n,\mu) \le c_p/\sqrt{n}$ for all n. When it is satisfied, we actually have $[\mathbb{E}W_p^p(\mu_n,\mu)]^{1/p} \le \tilde{c}_p/\sqrt{n}$ for all n.

This condition is satisfied when f is bounded below (in which case the support of μ must be compact). We also have the formulae

$$\tilde{c}_p = 5pJ_p^{1/p}(\mu), \qquad \tilde{c}_2 = \sqrt{2J_2(\mu)}.$$

(see [18, Theorem 5.3] for a stronger result and [18, Theorem 5.1] for the better constant when $p = 2$).

Let us now put this in the context of Theorem 4.6.2. In the binomial case, since each $\Pi_i^{(n)}$ and each Λ_i are independent, we have

$$\mathbb{E}W_2(\Lambda_i, \widetilde{\Lambda}_i)|\Lambda_i \le \sqrt{2J_2(\Lambda_i)}\frac{1}{\sqrt{\tau_n}}.$$

(In the Poisson case we need to condition on $N_i^{(n)}$ and then estimate its inverse square root as is done in the proof of Theorem 4.6.2.) Therefore, a sufficient condition for the rate $1/\sqrt{\tau_n}$ to hold is that $\mathbb{E}\sqrt{J_2(\Lambda)} < \infty$ and a necessary condition is that $\mathbb{P}(\sqrt{J_2(\Lambda)} = \infty) = 1$. These of course hold if there exists $\delta > 0$ such that with probability one Λ has a density bounded below by δ. Since $\Lambda = T\#\lambda$, this will happen provided that λ itself has a bounded below density and T has a bounded below derivative. Bigot, Gouet, Klein & Lópes [15] show that the rate $\sqrt{\tau_n}$ cannot be improved.

[1] More generally, f is the density of the absolutely continuous part of μ.

In the remainder of this subsection we illustrate how the rate of $\mathbb{E}W_p$ can be slow, even when μ has a smooth and strictly positive density, and why the rate $1/\sqrt[4]{\tau_n}$ is optimal in Theorem 4.6.2. By Jensen's inequality and (3.2), we have the upper bound for any $\mu \in P([0,1])$,

$$\mathbb{E}W_p(\mu_n,\mu) \leq \left[\mathbb{E}W_p^p(\mu_n,\mu)\right]^{1/p} \leq \left[\mathbb{E}W_1(\mu_n,\mu)\right]^{1/p} \leq 2^{1/p} n^{-1/(2p)}.$$

When μ is compactly supported the right-hand side is scaled by the length of the convex hull of the support of μ. We show that this cannot be improved:

Proposition 4.6.6. *For any rate $\epsilon_n \to 0$ there exists a measure μ on $[-1,1]$ with positive density there, and such that for all n*

$$\mathbb{E}W_p(\mu_n,\mu) \geq C(p,\mu) n^{-1/(2p)} \epsilon_n.$$

Proof. We first show that there exists such a discrete measure, since this example (taken from [18, Example 2.3]) provides the fundamental idea of what can go wrong. Let μ be uniform on the two points $\{0,1\}$. The empirical measure μ_n is concentrated on 0 and 1 with weights k/n and $1-k/n$. If $k \leq n/2$, then the optimal coupling from μ to μ_n is to send k/n mass to 0 and leave the rest in place, yielding a total cost of $|k/n - 1/2|$. The case $k \geq n/2$ is analogous, and we obtain

$$\mathbb{E}W_p(\mu_n,\mu) = \mathbb{E}\sqrt[p]{\frac{1}{n}\left|\frac{n}{2}-k\right|} = n^{-1/p}\mathbb{E}\sqrt[p]{\left|\frac{n}{2}-k\right|}.$$

Since k follows a binomial distribution $Z_n = 2(n/2-k)/\sqrt{n}$ converges narrowly to a standard Gaussian random variable Z by the central limit theorem. The convergence holds also in $\mathcal{W}_2$ by Theorem 3.2.1 and it also follows that $\mathbb{E}|Z_n|^{1/p} \to \mathbb{E}|Z|^{1/p}$. Thus

$$\mathbb{E}W_p(\mu_n,\mu) = 2^{-1/p} n^{-1/(2p)} \mathbb{E}|Z_n|^{1/p} \approx c_p n^{-1/(2p)}, \quad \left(c_p = 2^{-1/p}\mathbb{E}|Z|^{1/p} = \frac{\Gamma((1+p^{-1})/2)}{\sqrt{\pi}2^{-1/(2p)}}\right),$$

in the sense that the ratio between both sides converges to 1 as $n \to \infty$. This proves the existence of some compactly supported μ satisfying the conclusion of the proposition, even with $\epsilon_n = 1$.

To make μ absolutely continuous we replace the Dirac masses by uniform random variables on $[0,\theta]$ for $\theta < 1/2$, so that μ is uniform on $[0,\theta] \cup [1-\theta,1]$. Assume again that k of the points in μ_n are in the first interval, and that $k \leq n/2$. Then the optimal coupling between μ and μ_n will spend k/n mass from $[0,\theta]$ to send to these k points, and the remaining $1/2 - k/n$ mass must travel to somewhere in $[1-\theta,1]$ which is a distance of at least $1-2\theta$. Thus

$$\mathbb{E}W_p(\mu_n,\mu) \geq \mathbb{E}\sqrt[p]{\frac{1}{n}\left|\frac{n}{2}-k\right|(1-2\theta)^p} = (1-2\theta)\mathbb{E}\sqrt[p]{\frac{1}{n}\left|\frac{n}{2}-k\right|} \approx (1-2\theta)c_p n^{-1/(2p)}.$$

In order to construct a measure with strictly positive density we will put a density that decays very rapidly to zero in $[\theta,1-\theta]$ and consequently the factor $1-2\theta$ will depend on n but vanish

very slowly. For convenience we change the scale of μ so the centre of the mass is at the origin. Let μ be a measure on $[-C, C]$ with distribution F and symmetric density around zero such that $F(\epsilon_n) = 1/2 + 1/n^2$ and $F(-\epsilon_n) = 1/2 - 1/n^2$. Suppose that in μ_n, k points are in $[-1, -\epsilon_n]$, $n-k$ points in $[\epsilon_n, 1]$ and none in $(-\epsilon_n, \epsilon_n)$. Then at least $|1/2 - k/n|$ mass must travel at least a distance of ϵ_n. We see that

$$\begin{aligned}\mathbb{E}W_p(\mu_n, \mu) &\geq \sum_{k=0}^{n} \left(\frac{1}{2} - \frac{1}{n^2}\right)^k \left(\frac{1}{2} - \frac{1}{n^2}\right)^{n-k} \binom{n}{k} \sqrt[p]{\left|\frac{1}{2} - \frac{k}{n}\right| \epsilon_n} \\ &= \left(1 - \frac{2}{n^2}\right)^n \epsilon_n^{1/p} \quad 2^{-n} \sum_{k=0}^{n} \binom{n}{k} \sqrt[p]{\frac{1}{n}\left|\frac{n}{2} - k\right|} \approx c_p n^{-1/(2p)} \epsilon_n^{1/p}.\end{aligned}$$

Since $\epsilon_n^{1/p} \geq \epsilon_n$ this completes the proof. For example, $\epsilon_n = \sqrt{\log n}$ gives $F(x) = 1/2 + \exp(-1/x^2)$ for $x > 0$ close to zero. □

5 Computation of multivariate Fréchet means

When given measures $\mu^1, \dots, \mu^N$ are supported on the real line, computing their Fréchet mean $\bar{\mu}$ is straightforward (Subsection 3.5.2). This is in contrast to the multivariate case, where, besides the important yet special case of *compatible measures*, closed-form formulae are not available. Important advances in this direction have been made by restricting or approximating the problem (Bonneel, Rabin, Peyré & Pfister [21]; Cuturi & Doucet [27]); the *iterative barycentre* of Boissard, Le Gouic & Loubes [20] solves the compatible case. In this chapter, we propose an iterative algorithm that provably converges at least to a Karcher mean without restrictions on the measures $\mu^1, \dots, \mu^N$. Our algorithm is based on the differentiability properties of the Fréchet functional developed in Subsection 3.5.5 and can be interpreted as classical steepest descent in the Wasserstein space $\mathcal{W}_2(\mathbb{R}^d)$. It reduces the problem of finding the Fréchet mean to *pairwise problems*, involving only the Monge–Kantorovich problem between two measures. In the Gaussian case, the latter can be done explicitly, rendering the algorithm particularly appealing (see Subsection 5.4.1). For more general measures the optimal maps cannot be found analytically, but can at least be approximated numerically (Benamou & Brenier [11]; Chartrand, Wohlberg, Vixie & Bollt [24]; Haber, Rehman & Tannenbaum [46]).

This chapter can been seen as complementary to Chapter 4. On one hand, one can use the proposed algorithm to construct the regularised Fréchet–Wasserstein estimator $\widehat{\lambda}_n$ that approximates a population version (see Section 4.3). On the other hand, it could be that the object of interest is the sample $\mu^1, \dots, \mu^N$ itself, but that the latter is observed with some amount of noise. If one only has access to proxies $\widehat{\mu^1}, \dots, \widehat{\mu^N}$, then it is natural to use their Fréchet mean $\widehat{\bar{\mu}}$ as an estimator of $\bar{\mu}$. The proposed algorithm can then be used in order to compute $\bar{\mu}$, and the consistency framework of Section 4.4 then allows to conclude that if each $\widehat{\mu^i}$ is consistent, then so is $\widehat{\bar{\mu}}$.

After presenting the algorithm in Section 5.1, we make some connections to Procrustes analysis in Section 5.2. A convergence analysis of the algorithm is carried out in Section 5.3, after which examples are given in Section 5.4. Some improvements of intermediary results in the convergence analysis are postponed to Section 5.5, and an extension to infinitely many measures is sketched in Section 5.6.

The algorithm we discuss here was outlined in Zemel & Panaretos [94]. It was independently and concurrently discovered by Álvarez-Esteban, del Barrio, Cuesta-Albertos & Matrán [5], and we comment on similarities and differences in the end of Section 5.1.

5.1 A steepest descent algorithm for the computation of Fréchet means

Throughout this section, we assume that N is a fixed integer and consider a fixed collection

$$\mu^1,\ldots,\mu^N \in \mathscr{W}_2(\mathbb{R}^d) \qquad \text{with } \mu^1 \text{ absolutely continuous with bounded density,}$$

to which it is desired to find the unique (Proposition 3.5.17) Fréchet mean $\bar{\mu}$. It has been established that if γ is absolutely continuous then the associated Fréchet functional

$$F(\gamma) = \frac{1}{2N}\sum_{i=1}^{n} W_2^2(\mu^i,\gamma), \qquad \gamma \in \mathscr{W}_2(\mathbb{R}^d),$$

has Fréchet derivative (Theorem 3.5.13)

$$F'(\gamma) = -\frac{1}{N}\sum_{i=1}^{N}\log_\gamma(\mu^i) = -\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{t}_\gamma^{\mu_i}-\mathbf{i}\right) \in \mathrm{Tan}_\gamma \tag{5.1}$$

at γ. Let $\gamma_j \in \mathscr{W}_2(\mathbb{R}^d)$ be an absolutely continuous measure, representing our current estimate of the Fréchet mean at step j. Then it makes sense to introduce a step size $\tau_j > 0$, and to follow the steepest descent of F given by the negative of the gradient:

$$\gamma_{j+1} = \exp_{\gamma_j}\left(-\tau_j F'(\gamma_j)\right) = \left[\mathbf{i} + \tau_j \frac{1}{N}\sum_{i=1}^{N}\log_\gamma(\mu^i)\right]\#\gamma_j = \left[\mathbf{i} + \tau_j \frac{1}{N}\sum_{i=1}^{N}(\mathbf{t}_{\gamma_j}^{\mu^i}-\mathbf{i})\right]\#\gamma_j.$$

In order to employ further descent at γ_{j+1}, it needs to be verified that F is differentiable at γ_{j+1}, which amounts to showing that the latter stays absolutely continuous. This will happen for all but countably many values of the step size τ_j, but necessarily if the latter is contained in $[0,1]$:

Lemma 5.1.1 (regularity of the iterates)**.** *If γ_0 is absolutely continuous and $\tau = \tau_0 \in [0,1]$ then so is γ_1.*

The idea is that injective push-forwards of absolutely continuous measures stay absolutely continuous and optimal maps between absolutely continuous measures are injective. An average of injective maps does not have to be injective in general, but it is if the injectivity holds in a “compatible" way. Here are the details:

Proof of Lemma 5.1.1. By [6, Proposition 6.2.12] there exist γ_0-null sets $\mathscr{N}_1$ such that on $\mathbb{R}^d \setminus$

$\mathcal{N}_1$, $\mathbf{t}_{\gamma_0}^{\mu^1}$ is differentiable, $\nabla \mathbf{t}_{\gamma_0}^{\mu^1} > 0$ (positive definite), and $\mathbf{t}_{\gamma_0}^{\mu^1}$ is strictly monotone:

$$\left\langle \mathbf{t}_{\gamma_0}^{\mu^1}(x) - \mathbf{t}_{\gamma_0}^{\mu^1}(x'), x - x' \right\rangle > 0, \qquad x, x' \notin \mathcal{N}_1, \quad x \neq x',$$

and with weak inequalities on $\mathcal{N}_2, \ldots, \mathcal{N}_N$. Since $\mathbf{t}_{\gamma_0}^{\gamma_1} = (1-\tau)\mathbf{i} + \tau N^{-1} \sum_{i=1}^N \mathbf{t}_{\gamma_0}^{\mu^i}$, it stays strictly monotone (hence injective) and $\nabla \mathbf{t}_{\gamma_0}^{\gamma_1} > 0$ outside $\mathcal{N} = \cup \mathcal{N}_i$, which is a γ_0-null set.

Let h_0 denote the density of γ_0 and set $\Sigma = \mathbb{R}^d \setminus \mathcal{N}$. Then $\mathbf{t}_{\gamma_0}^{\gamma_1}|_\Sigma$ is injective and $\{h_0 > 0\} \setminus \Sigma$ is Lebesgue negligible because

$$0 = \gamma_0(\mathcal{N}) = \gamma_0(\mathbb{R}^d \setminus \Sigma) = \int_{\mathbb{R}^d \setminus \Sigma} h_0(x)\,\mathrm{d}x = \int_{\{h_0 > 0\} \setminus \Sigma} h_0(x)\,\mathrm{d}x,$$

and the integrand is strictly positive. Since $|\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}| > 0$ on Σ we obtain that $\gamma_1 = \mathbf{t}_{\gamma_0}^{\mu^i} \# \gamma_0$ is absolutely continuous by [6, Lemma 5.5.3]. □

Lemma 5.1.1 suggests that the step size should be restricted to $[0,1]$. The next result suggests that the objective function essentially tells us that the *optimal* step size, achieving the maximal reduction of the objective function (thus corresponding to an approximate line search), is exactly equal to 1. The lemma does not use the Euclidean structure and holds when $\mathbb{R}^d$ is replaced by any separable Hilbert space.

Lemma 5.1.2 (optimal stepsize)**.** *If $\gamma_0 \in \mathcal{W}_2(\mathbb{R}^d)$ is absolutely continuous then*

$$F(\gamma_1) - F(\gamma_0) \leq -\|F'(\gamma_0)\|^2 \left[\tau - \frac{\tau^2}{2}\right]$$

and the bound on the right-hand side of the last display is minimised when $\tau = 1$.

Proof of Lemma 5.1.2. Let $S_i = \mathbf{t}_{\gamma_0}^{\mu^i}$ be the optimal map from γ_0 to μ^i, and set $W_i = S_i - \mathbf{i}$. Then

$$2NF(\gamma_0) = \sum_{i=1}^N W_2^2(\gamma_0, \mu^i) = \sum_{i=1}^N \int_{\mathbb{R}^d} \|S_i - \mathbf{i}\|^2\,\mathrm{d}\gamma_0 = \sum_{i=1}^N \langle W_i, W_i \rangle = \sum_{i=1}^N \|W_i\|^2, \tag{5.2}$$

with the inner product being in $L^2(\gamma_0)$. Both γ_1 and μ^i can be written as push-forwards of γ_0 and (3.3) gives the bound

$$W_2^2(\gamma_1, \mu^i) \leq \int_{\mathbb{R}^d} \left\| \left[(1-\tau)\mathbf{i} + \frac{\tau}{N} \sum_{j=1}^N S_j \right] - S_i \right\|_{\mathbb{R}^d}^2 \mathrm{d}\gamma_0 = \left\| -W_i + \frac{\tau}{N} \sum_{j=1}^N W_j \right\|_{L^2(\gamma_0)}^2.$$

The norm is always in $L^2(\gamma_0)$, regardless of i. Developing the squares, summing over $i = 1, \ldots, N$

and using (5.2) gives

$$\begin{aligned} 2NF(\gamma_1) &\leq \sum_{i=1}^{N} \|W_i\|^2 - 2\frac{\tau}{N} \sum_{i,j=1}^{N} \langle W_i, W_j \rangle + \frac{\tau^2}{N^2} \sum_{i,j,k=1}^{N} \langle W_j, W_k \rangle \\ &= 2NF(\gamma_0) - 2N\tau \left\| \sum_{i=1}^{N} \frac{1}{N} W_i \right\|^2 + N\tau^2 \left\| \sum_{i=1}^{N} \frac{1}{N} W_i \right\|^2, \end{aligned}$$

and recalling that $W_i = S_i - \mathbf{i}$ yields

$$F(\gamma_1) - F(\gamma_0) \leq \frac{\tau^2 - 2\tau}{2} \left\| \frac{1}{N} \sum_{i=1}^{N} W_i \right\|^2 = -\|F'(\gamma_0)\|^2 \left[\tau - \frac{\tau^2}{2} \right].$$

Since $\tau - \tau^2/2$ is clearly maximised at $\tau = 1$, the proof is complete. □

In light of Lemmas 5.1.1 and 5.1.2, we will always take $\tau_j = 1$. We then know that the sequence $(F(\gamma_j))$ is nonincreasing and that for any integer k,

$$\frac{1}{2} \sum_{j=0}^{k} \|F'(\gamma_j)\|^2 \leq \sum_{j=0}^{k} F(\gamma_j) - F(\gamma_{j+1}) = F(\gamma_0) - F(\gamma_{k+1}) \leq F(\gamma_0).$$

As $k \to \infty$, the infinite sum on the left-hand side converges, so $\|F'(\gamma_j)\|^2$ must vanish as $j \to \infty$. Without this fact, convergence of (γ_j) to a Karcher mean would have been hopeless.

The steepest descent iteration is presented succinctly as Algorithm 1 (the notion of Procrustes analysis will be explained in the next section). It was also discovered concurrently by Álvarez-Esteban, del Barrio, Cuesta-Albertos & Matrán [5], who carry out a similar convergence analysis (their results are equivalent to Theorem 5.3.1). Both the motivation and the techniques of proofs differ substantially between the two approaches. Firstly, Algorithm 1 is motivated by the geometry of the Wasserstein space, and is obtained as steepest descent; while the one in [5] is motivated as a fixed point iteration through the special case of Gaussian measures, where it is known that the Fréchet mean is the unique solution to a certain matrix equation (see Section 5.4). Also, rather than directly use the geometry of monotone operators in $\mathbb{R}^d$ as we do in Proposition 5.3.6, the authors of [5] take advantage of an almost-sure representation result on the optimal transportation maps in order to prove convergence of their algorithm.

One advantage of our approach is that it almost automatically gives uniform convergence of the optimal maps (Theorem 5.3.3), required for determining the optimal multicoupling of $\mu^1, \ldots, \mu^N$ by means of Theorem 3.5.20.

5.2 Relationship to shape theory and Procrustes analysis

Algorithm 1 is similar in spirit to another procedure, **generalised Procrustes analysis**, that is used in the field of **shape theory**. Given a subset $B \subseteq \mathbb{R}^d$, most commonly a finite collection of

Algorithm 1 Steepest descent via Procrustes analysis.

(A) Set a tolerance threshold $\epsilon > 0$.

(B) For $j = 0$, let γ_j be an arbitrary absolutely continuous measure.

(C) For $i = 1, \dots, N$ solve the (pairwise) Monge problem and find the optimal transport map $\mathbf{t}_{\gamma_j}^{\mu^i}$ from γ_j to μ^i.

(D) Define the map $T_j = N^{-1} \sum_{i=1}^{N} \mathbf{t}_{\gamma_j}^{\mu^i}$.

(E) Set $\gamma_{j+1} = T_j \# \gamma_j$, i.e. push-forward γ_j via T_j to obtain γ_{j+1}.

(F) If $\|F'(\gamma_{j+1})\| < \epsilon$, stop, and output γ_{j+1} as the approximation of $\bar{\mu}$ and $\mathbf{t}_{\gamma_{j+1}}^{\mu^i}$ as the approximation of $\mathbf{t}_{\bar{\mu}}^{\mu^i}$, $i = 1, \dots, N$. Otherwise, return to step (C).

labelled points called **landmarks**, an interesting question is how to mathematically define the *shape* of B. One way to reach such a definition is is to "subtract" from B properties deemed irrelevant for what one considers this shape should be; typically, these would include its location, its orientation and/or its scale. Accordingly, the shape of B can be defined as the equivalence class containing all sets obtained as gB, where g belongs to a collection $\mathscr{G}$ of transformations of $\mathbb{R}^d$ containing all combinations of rotations, translations, dilations and/or reflections (Dryden & Mardia [30, Chapter 4]).

If B_1 and B_2 are two collections of k landmarks, one may define the distance between their shapes as the infimum of $\|B_1 - gB_2\|^2$ over the group $\mathscr{G}$. In other words, one seeks to *register* B_2 as close as possible to B_1 by using elements of the group $\mathscr{G}$, with distance being measured as the sum of squared Euclidean distances between the transformed points of B_2 and those of B_1. In a sense, one can think about the shape problem and the Monge problem as dual to each other. In the former, one is given constraints on how to optimally carry out the registration of the points with the cost being judged by how successful the registration procedure is. In the latter, one imposes that the registration be done *exactly*, and evaluates the cost by how much the space must be deformed in order to achieve this.

The optimal g and the resulting distance can be found in closed-form by means of **ordinary Procrustes analysis** [30, Section 5.2]. Suppose now that we are given $N > 2$ collections of points, $B_1, \dots, B_N$, with the goal of minimising the sum of squares $\|g_i B_i - g_j B_j\|^2$ over $g_i \in \mathscr{G}$[1]. As in the case of Fréchet means in $\mathscr{W}_2(\mathbb{R}^d)$ (Subsection 3.5.6), there is a formulation in terms of sum of squares from the average $N^{-1} \sum g_j B_j$. Unfortunately, there is no explicit solution for this problem when $d \geq 3$. Like Algorithm 1, **generalised Procrustes analysis** (Gower [43]; Dryden & Mardia [30, p. 90]) tackles this multimarginal setting by iteratively solving the pairwise problem, as follows. Choose one of the configurations as an initial estimate/template,

[1] One needs to add an additional constraint to prevent registering all the collection to the origin.

then register every other configuration to the template, employing ordinary Procrustes analysis. The new template is then given by the linear average of the registered configurations, and the process is iterated subsequently.

In parallel, given the current template γ_j, Algorithm 1 iterates the two steps of registration and linear averaging, but in a different manner:

(1) **Registration**: by finding the optimal transportation maps $\mathbf{t}_{\gamma_j}^{\mu^i}$, we identify each μ^i with the element $\mathbf{t}_{\gamma_j}^{\mu^i} - \mathbf{i} = \log_{\gamma_j}(\mu^i)$. In this sense, the collection $(\mu^1, \ldots, \mu^N)$ is viewed in the common coordinate system given by the tangent space at the template γ_j and is in this sense registered to it.

(2) **Averaging**: the registered measures are averaged linearly, using the common coordinate system of the registration step (1), as elements in the linear space Tan_{γ_j}. The linear average is then retracted back onto the Wasserstein space via the exponential map to yield the estimate at the $(j+1)$-th step, γ_{j+1}.

Notice than in the Procrustes sense, the maps that register each μ^i to the template γ_j are $\mathbf{t}_{\mu^i}^{\gamma_j}$, the *inverses* of $\mathbf{t}_{\gamma_j}^{\mu^i}$. We will not use the term “registration maps" in the sequel, to avoid possible confusion.

5.3 Convergence of the algorithm

In order to tackle the issue of convergence, we will use an approach that is specific to the nature of optimal transportation. This is because the Hessian-type arguments that are used to prove similar convergence results for steepest descent on Riemannian manifolds (Afsari, Tron & Vidal [1]) or Procrustes algorithms (Le [61]; Groisser [44]) do not apply here, since the Fréchet functional may very well fail to be twice differentiable.

In fact, even in Euclidean spaces, convergence of steepest descent usually requires a Lipschitz bound on the derivative of F (Bertsekas [12, Subsection 1.2.2]). Unfortunately, F is not known to be differentiable at discrete measures, and these constitute a dense set in $\mathcal{W}_2$; consequently this Lipschitz condition is very unlikely to hold. Still, this specific geometry of the Wasserstein space affords some advantages; for instance, we will place no restriction on the starting point for the iteration, except that it be absolutely continuous; and no assumption on the spread of $\mu^1, \ldots, \mu^N$ is necessary as in, for example, [1, 44, 61].

Theorem 5.3.1 (limit points are Karcher means)**.** *Let $\mu^1, \ldots, \mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ be probability measures and suppose that one of them is absolutely continuous with a bounded density. Then, the sequence generated by Algorithm 1 stays in a compact set of the Wasserstein space $\mathcal{W}_2(\mathbb{R}^d)$, and any limit point of the sequence is a Karcher mean of $(\mu^1, \ldots, \mu^N)$.*

Since the Fréchet mean $\bar{\mu}$ is a Karcher mean (Proposition 3.5.17), we obtain immediately:

Corollary 5.3.2 (Wasserstein convergence of steepest descent)**.** *Under the conditions of Theorem 5.3.1, if F has a unique stationary point, then the sequence $\{\gamma_j\}$ generated by Algorithm 1 converges to the Fréchet mean of $\{\mu^1,\dots,\mu^N\}$ in the Wasserstein metric,*

$$W_2(\gamma_j,\bar{\mu})\longrightarrow 0, \qquad j\to\infty.$$

Alternatively, combining Theorem 5.3.1 with the optimality criterion Theorem 3.5.18 shows that the algorithm converges to $\bar{\mu}$ when the appropriate assumptions on $\{\mu^i\}$ and the Karcher mean $\mu=\lim\gamma_j$ are satisfied. This allows to conclude that Algorithm 1 converges to the unique Fréchet mean when μ^i are Gaussian measures (see Theorem 5.4.1).

The proof of Theorem 5.3.1 is rather elaborate, since we need to use specific methods that are tailored to the Wasserstein space. Before giving the proof, we state one more important consequence, uniform convergence of the optimal maps $\mathbf{t}_{\gamma_j}^{\mu^i}$ to $\mathbf{t}_{\bar{\mu}}^{\mu^i}$ on compacta. These maps are important for the solution of the multicoupling problem (as established in Theorem 3.5.20), and their convergence does not immediately follow from the Wasserstein convergence of γ_j to $\bar{\mu}$. We in addition automatically obtain convergence of the inverses. Both the formulation and the proof of this result are similar to those of Theorem 4.4.3.

Theorem 5.3.3 (uniform convergence of optimal maps)**.** *Under the conditions of Corollary 5.3.2, there exist sets $A,B^1,\dots,B^N\subseteq\mathbb{R}^d$ such that $\bar{\mu}(A)=1=\mu^1(B^1)=\cdots=\mu^N(B^N)$ and*

$$\sup_{\Omega_1}\left\|\mathbf{t}_{\gamma_j}^{\mu^i}-\mathbf{t}_{\bar{\mu}}^{\mu^i}\right\|\overset{j\to\infty}{\longrightarrow}0, \qquad \sup_{\Omega_2}\left\|\mathbf{t}_{\mu^i}^{\gamma_j}-\mathbf{t}_{\mu^i}^{\bar{\mu}}\right\|\overset{j\to\infty}{\longrightarrow}0, \qquad i=1,\dots,N,$$

for any pair of compacta $\Omega_1\subseteq A$, $\Omega_2\subseteq B^i$. If in addition all the measures $\mu^1,\dots,\mu^N$ have the same support, then one can choose all the sets B^i to be the same.

Proof of Theorem 5.3.1 We will prove the theorem by establishing the following facts:

1. The sequence (γ_j) stays in a compact subset of $\mathcal{W}_2(\mathbb{R}^d)$.
2. Any limit of (γ_j) is absolutely continuous.
3. The mapping $\gamma\mapsto\|F'(\gamma)\|^2$ is continuous.

Since it has already been established that $\|F'(\gamma_j)\|\to 0$, these three facts indeed suffice.

Lemma 5.3.4. *The sequence generated by Algorithm 1 stays in a compact subset of the Wasserstein space $\mathcal{W}_2(\mathbb{R}^d)$.*

Proof. Since $F(\gamma_j)$ is bounded, γ_j stay bounded in $\mathcal{W}_2(\mathbb{R}^d)$. It was shown in the proof of Proposition 3.5.4 that as a result of this (γ_j) is narrowly tight. We give a more direct proof of this claim, that is valid even if $\mathbb{R}^d$ is replaced by a separable Hilbert space.

For any $\epsilon > 0$ there exists a compact convex set K_ϵ such that $\mu^i(K_\epsilon) > 1 - \epsilon/N$ for $i = 1, \dots, N$. Let $A_j^i = (\mathbf{t}_{\gamma_j}^{\mu^i})^{-1}(K_\epsilon)$, $A_j = \cap_{i=1}^N A_j^i$. Then $\gamma_j(A_j^i) > 1 - \epsilon/N$, so that $\gamma_j(A_j) > 1 - \epsilon$. Since K_ϵ is convex, $T_j(x) \in K_\epsilon$ for any $x \in A_j$, so that

$$\gamma_{j+1}(K_\epsilon) = \gamma_j(T_j^{-1}(K_\epsilon)) \geq \gamma_j(A_j) > 1 - \epsilon, \qquad j = 0, 1, \dots.$$

We shall now show that any narrowly convergent subsequence of $\{\gamma_j\}$ is in fact convergent in the Wasserstein space. By Theorem 3.2.1, it suffices to show that

$$\lim_{R \to \infty} \sup_{j \in \mathbb{N}} \int_{\{x : \|x\| > R\}} \|x\|^2 \, \mathrm{d}\gamma_j(x) = 0. \tag{5.3}$$

Assume momentarily that $\mu^1, \dots, \mu^N$ have finite third moments:

$$\int_{\mathbb{R}^d} \|x\|^3 \, \mathrm{d}\mu^i(x) \leq M, \qquad i = 1, \dots, N.$$

Then for any $j \geq 1$ it holds that

$$\begin{aligned} \int_{\mathbb{R}^d} \|x\|^3 \, \mathrm{d}\gamma_j(x) = \int_{\mathbb{R}^d} \left\| \frac{1}{N} \sum_{i=1}^N \mathbf{t}_{\gamma_{j-1}}^{\mu^i}(x) \right\|^3 \mathrm{d}\gamma_{j-1}(x) &\leq \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \|\mathbf{t}_{\gamma_{j-1}}^{\mu^i}(x)\|^3 \, \mathrm{d}\gamma_{j-1}(x) \\ &= \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^d} \|x\|^3 \, \mathrm{d}\mu^i(x) \leq M. \end{aligned}$$

This implies that for any $R > 0$ and any $j > 0$,

$$\int_{\{x : \|x\| > R\}} \|x\|^2 \, \mathrm{d}\gamma_j(x) \leq \frac{1}{R} \int_{\{x : \|x\| > R\}} \|x\|^3 \, \mathrm{d}\gamma_j(x) \leq \frac{1}{R} M,$$

and (5.3) follows. If instead of third moments μ^i have a moment of order $2 + \epsilon$, then the same reasoning works with R replaced by R^ϵ and a different constant M. More generally, if for some (nondecreasing) function H that diverges to infinity (like $H(x) = \log\log(x + 10)$)

$$\int_{\mathbb{R}^d} \|x\|^2 H(x) \, \mathrm{d}\mu^i(x) \leq M, \qquad i = 1, \dots, N,$$

then still (3.4) holds by the same argument. That such H must exist is a consequence of the finiteness of the collection $(\mu^1, \dots, \mu^N)$, a result whose proof is postponed to Subsection 5.3.1. □

A closer look at the proof shows that the structure of (γ_j) as a sequence of iterates does not really play a role and a more general result can be established. Let us denote by $\mathscr{A}$ the steepest descent iteration that maps γ_j to γ_{j+1}. Then $\mathscr{A}$ is a function from the set of absolutely continuous measures of $\mathscr{W}_2(\mathbb{R}^d)$ to itself. What we have just shown is that the image of $\mathscr{A}$ is Wasserstein-tight, if we replace (5.3) by its more general counterpart (3.6).

In order to show that a narrowly convergent sequence (γ_j) of absolutely continuous measures has an absolutely continuous limit γ, it suffices to show that the densities of γ_j are uniformly bounded. Indeed, if C is such a bound, then for any open $O \subseteq \mathbb{R}^d$, $\liminf \gamma_k(O) \leq C\mathrm{Leb}(O)$, so $\gamma(O) \leq C\mathrm{Leb}(O)$ by the portmanteau lemma 2.9.1. It follows that γ is absolutely continuous with density bounded by C. We now show that such C can be found that applies to all measures in the image of $\mathscr{A}$, hence to all sequences resulting from iterations of Algorithm 1.

Proposition 5.3.5 (uniform density bound)**.** *Let the first k measures ($1 \leq k \leq N$) of $(\mu^1, \dots, \mu^N)$ be absolutely continuous with densities g^i and let $\gamma_1 = \mathscr{A}(\gamma_0)$ be any (absolutely continuous) probability measure. Then the density of γ_1 is bounded by $C_\mu = N^d \min_i \|g^i\|_\infty$.*

The constant C_μ of course depends only on the measures $(\mu^1, \dots, \mu^N)$, and is finite as long as one μ^i has a bounded density. We later discuss how it can be improved (Corollary 5.5.2), as will be necessary in order to obtain a population version of this result.

Proof. Let h_i be the density of γ_i. By the change of variables formula, for γ_0-almost any x

$$h_1(\mathbf{t}_{\gamma_0}^{\gamma_1}(x)) = \frac{h_0(x)}{\det \nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)}; \qquad g^i(\mathbf{t}_{\gamma_0}^{\mu^i}(x)) = \frac{h_0(x)}{\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)}, \qquad i = 1, \dots, k.$$

We seek a lower bound on the determinant of $\nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)$, which by definition equals

$$N^{-d} \det \sum_{i=1}^{N} \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x).$$

Such a bound is provided by the Brunn–Minkowski inequality (Stein & Shakarchi [85, Section 1.5]) for symmetric positive semidefinite matrices

$$[\det(A+B)]^{1/d} \geq [\det A]^{1/d} + [\det B]^{1/d}$$

that when applied inductively yields

$$\left[\det \nabla \mathbf{t}_{\gamma_0}^{\gamma_1}(x)\right]^{1/d} \geq \frac{1}{N} \sum_{i=1}^{N} \left[\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)\right]^{1/d} \geq \frac{1}{N} \sum_{i=1}^{k} \left[\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)\right]^{1/d}.$$

From this we easily obtain an upper bound for h_1:

$$\frac{1}{h_1^{1/d}(\mathbf{t}_{\gamma_0}^{\gamma_1}(x))} = \frac{\det^{1/d} \sum_{i=1}^{k} \nabla \mathbf{t}_{\gamma_0}^{\mu^i}(x)}{N h_0^{1/d}(x)} \geq \frac{1}{N} \sum_{i=1}^{k} \frac{1}{[g^i(\mathbf{t}_{\gamma_0}^{\mu^i}(x))]^{1/d}} \geq \frac{1}{N} \sum_{i=1}^{k} \frac{1}{\|g^i\|_\infty^{1/d}} \geq \frac{1}{N} \frac{1}{\|g^i\|_\infty^{1/d}},$$

for any i. Let Σ be the set of points where this inequality holds; then $\gamma_0(\Sigma) = 1$. Hence

$$\gamma_1(\mathbf{t}_{\gamma_0}^{\gamma_1}(\Sigma)) = \gamma_0[(\mathbf{t}_{\gamma_0}^{\gamma_1})^{-1}(\mathbf{t}_{\gamma_0}^{\gamma_1}(\Sigma))] \geq \gamma_0(\Sigma) = 1.$$

Thus γ_1-almost surely and for all i,

$$h_1(y) \le N^d \|g^i\|_\infty,$$

in particular for the i minimising $\|g^i\|_\infty$. □

The third statement (continuity of the gradient) is much more subtle to establish, and its rather lengthy proof is given next. In view of Proposition 5.3.5, the uniform bound on the densities is not a hindrance for the proof of convergence of Algorithm 1.

Proposition 5.3.6 (continuity of F')**.** *Let (γ_n) be a sequence of absolutely continuous measures with uniformly bounded densities and suppose that $W_2(\gamma_n, \gamma) \to 0$. Then $\|F'(\gamma_n)\|^2 \to \|F'(\gamma)\|^2$.*

Proof. As has been established in the discussion before Proposition 5.3.5, the limit γ must be absolutely continuous. Consequently, $F'(\gamma)$ is well-defined what needs to be shown is that

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{t}_{\gamma_n}^{\mu^i} - \mathbf{i} \right\|_{L^2(\gamma_n)}^2 \longrightarrow \left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{t}_{\gamma}^{\mu^i} - \mathbf{i} \right\|_{L^2(\gamma)}^2, \qquad n \to \infty.$$

We denote the integrands by g_n and g respectively and divide the proof into several steps. It is perhaps instructive to assume in first reading that g_n and g are bounded and continuous, in which case one can jump directly to Step 2. The first of these assumptions is satisfied when the μ^i have bounded supports, and the second can be obtained under the regularity conditions in Theorem 2.8.2.

Step 0: redefinition on null sets. At a given $x \in \mathbb{R}^d$, $g_n(x)$ can be undefined, either because some $\mathbf{t}_{\gamma_n}^{\mu^i}(x)$ is empty, or because it can be multivalued (see Subsection 2.9.2, p. 35). Redefine $g_n(x)$ at such points by setting it to 0 in the former case and choosing an arbitrary representative otherwise. Apply the same procedure for g. Then g_n and g are finite, nonnegative functions (in the proper sense) throughout $\mathbb{R}^d$. We claim that this modification is inconsequential and does not affect $\int g \, \mathrm{d}\gamma$. Indeed, the set of ambiguity points is a γ-null set: this is a consequence of the absolute continuity of γ, together with Remark 2.3 and Corollary 1.3 in Alberti & Ambrosio [3] (see the paragraphs preceding Assumptions 1 for a more detailed discussion). Similarly, the value of the integral $\int g_n \, \mathrm{d}\gamma_n$ remains unaltered after this modification. Finally, by Proposition 2.9.8, the set of points where g is not continuous is a γ-null set, before and after the modification.

Step 1: approximation by bounded functions. Since γ_n converge in the Wasserstein space, they satisfy the uniform integrability condition (5.3) by Theorem 3.2.1, and hence the uniform absolute continuity (3.7) that we repeat here for convenience:

$$\forall \epsilon > 0 \exists \delta > 0 \forall n \ge 1 \forall A \subseteq \mathbb{R}^d \text{ Borel}: \quad \gamma_n(A) \le \delta \implies \int_A \|x\|^2 \, \mathrm{d}\gamma_n(x) < \epsilon. \tag{5.4}$$

The δ's can be chosen in such a way that (5.4) holds true for the finite collection $\{\mu^1,\dots,\mu^N\}$ as well. Fix $\epsilon>0$, set $\delta=\delta_\epsilon$ as in (5.4), and invoke (5.3) to find an $R=R_\epsilon\geq 1$ such that

$$\forall i\ \forall n:\quad \int_{\{\|x\|^2>R\}}\|x\|^2\,\mathrm{d}\gamma_n(x)+\int_{\{\|x\|^2>R\}}\|x\|^2\,\mathrm{d}\mu^i(x)<\frac{\delta}{2N}.$$

The bound (holding γ_n-almost surely)

$$g_n(x)\leq 2\|x\|^2+\frac{2}{N}\sum_{i=1}^N\|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2$$

implies that the sets $A_n=\{x:g_n(x)\geq 4R\}$ satisfy

$$A_n\subseteq\{x:\|x\|^2>R\}\cup\bigcup_{i=1}^N\{x:\|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2>R\}.$$

To deal with the sets in the union observe that

$$\gamma_n(\{x:\|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2>R\})=\mu^i(\{x:\|x\|^2>R\})<\frac{\delta}{2N},$$

so that $\gamma_n(A_n)<\delta$. We use this in conjunction with (5.4) to bound

$$\begin{aligned}\int_{A_n}g_n(x)\,\mathrm{d}\gamma_n(x)&\leq 2\int_{A_n}\|x\|^2\,\mathrm{d}\gamma_n(x)+\frac{2}{N}\sum_{i=1}^N\int_{A_n}\|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2\,\mathrm{d}\gamma_n(x)\\&\leq 2\epsilon+\frac{2}{N}\sum_{i=1}^N\int_{\mathbf{t}_{\gamma_n}^{\mu^i}(A_n)}\|x\|^2\,\mathrm{d}\mu^i(x)\leq 4\epsilon,\end{aligned}$$

where we have used the measure-preservation property $\mu^i(\mathbf{t}_{\gamma_n}^{\mu^i}(A_n))=\gamma_n(A_n)<\delta$.

Define the truncation $g_{n,R}(x)=\min(g_n(x),4R)$. Then $0\leq g_n-g_{n,R}\leq g_n\mathbf{1}\{g_n>4R\}$, so

$$\int[g_n(x)-g_{n,R}(x)]\,\mathrm{d}\gamma_n(x)\leq\int_{A_n}g_n(x)\,\mathrm{d}\gamma_n(x)\leq 4\epsilon,\qquad n=1,2,\dots.$$

The analogous truncated function g_R satisfies

$$0\leq g_R(x)\leq 4R\quad\forall x\in\mathbb{R}^d\quad\text{and}\quad\{x:g_R\text{ is continuous }\}\text{ is of }\gamma\text{-full measure.}\tag{5.5}$$

Step 2: convergence of g_n to g. Let $E=\mathrm{supp}\gamma$. The sets

$$\mathscr{N}^i=(E\setminus E^{\mathrm{den}})\cup\{x:\mathbf{t}_\gamma^{\mu^i}(x)\text{ contains more than one element}\},\qquad i=1,\dots,N,$$

are γ-negligible and so is their union $\mathscr{N}$. As $n\to\infty$, Proposition 2.9.11 implies pointwise convergence (in a set-valued sense) of $\mathbf{t}_{\gamma_n}^{\mu^i}(x)$ to $\mathbf{t}_\gamma^{\mu^i}(x)$ for any $i=1,\dots,N$ and any $x\in E\setminus\mathscr{N}$. Thus $g_n\to g$ pointwise on $x\in E\setminus\mathscr{N}$ (for whatever choice of representatives selected to define g_n); consequently, $g_{n,R}\to g_R$ on $E\setminus\mathscr{N}$.

If E were compact, we could strengthen this to uniform convergence by Egorov's theorem. In order to restrict the integrands to a bounded set we invoke the tightness of the sequence (γ_n) and introduce a compact set K_ϵ such that $\gamma_n(\mathbb{R}^d \setminus K_\epsilon) < \epsilon/R$ for all n. Clearly, $g_{n,R} \to g_R$ on $E' = K_\epsilon \cap E \setminus \mathcal{N}$, and by Egorov's theorem (valid as $\mathrm{Leb}(E') \leq \mathrm{Leb}(K_\epsilon) < \infty$), there exists a Borel set $\Omega = \Omega_\epsilon \subseteq E'$ on which the convergence is uniform, and $\mathrm{Leb}(E' \setminus \Omega) < \epsilon/R$. Let us write

$$\int g_{n,R}\,\mathrm{d}\gamma_n - \int g_R\,\mathrm{d}\gamma = \int g_R\,\mathrm{d}(\gamma_n - \gamma) + \int_\Omega (g_{n,R} - g_R)\,\mathrm{d}\gamma_n + \int_{\mathbb{R}^d\setminus\Omega} (g_{n,R} - g_R)\,\mathrm{d}\gamma_n,$$

and bound each of the three integrals on the right-hand side as $n \to \infty$.

Step 3: bounding the first two integrals. The first integral vanishes as $n \to \infty$, by (5.5) and the portmanteau lemma 2.9.1, as the bounded function g_R is continuous besides a γ-null set. The second integral obviously tends to 0 as $n \to \infty$, since $g_{n,R}$ converge to g_R uniformly on Ω.

Step 4: bounding the third integral. The integrand is smaller than $8R$, so the integral is bounded by $8R\gamma_n(\mathbb{R}^d \setminus \Omega)$. The complement of $\Omega \subseteq E' = E \cap K_\epsilon \setminus \mathcal{N}$ is included in the union $\mathcal{N} \cup (E' \setminus \Omega) \cup (\mathbb{R}^d \setminus E) \cup (\mathbb{R}^d \setminus K_\epsilon)$, where the first set is Lebesgue-negligible and the second has Lebesgue measure smaller than ϵ/R. The hypothesis of the densities of γ_n implies that $\gamma_n(A) \leq C\mathrm{Leb}(A)$ for any Borel set $A \subseteq \mathbb{R}^d$ and any $n \in \mathbb{N}$; it follows from this and $\gamma_n(\mathbb{R}^d \setminus K_\epsilon) < \epsilon/R$ that

$$\left| \int_{\mathbb{R}^d\setminus\Omega} (g_{n,R} - g_R)\,\mathrm{d}\gamma_n \right| \leq 8R(C\epsilon/R + \gamma_n(\mathbb{R}^d \setminus E) + \epsilon/R) = 8\Big(R\gamma_n(\mathbb{R}^d \setminus E) + C\epsilon + \epsilon\Big).$$

The narrow (or even Wasserstein) convergence of γ_n to γ alone does not suffice for bounding the limit $\gamma_n(\mathbb{R}^d \setminus E)$, because E is closed and the portmanteau lemma gives the inequality in the wrong direction. Once again the uniform bound on the densities comes to our rescue. Write the open set $E_1 = \mathbb{R}^d \setminus E$ as a countable union of closed sets A_k with[2] $\mathrm{Leb}(E_1 \setminus A_k) < 1/k$, and conclude that

$$\limsup_{n\to\infty} \gamma_n(E_1) \leq \limsup_{n\to\infty} \gamma_n(A_k) + \limsup_{n\to\infty} \gamma_n(E_1 \setminus A_k) \leq \gamma(A_k) + \frac{C}{k} = \frac{C}{k},$$

where we have used the portmanteau lemma again, $A_k \cap \mathrm{supp}(\gamma) = \emptyset$ and $\gamma_n(A) \leq C\mathrm{Leb}(A)$.

Step 5: concluding. By Steps 3 and 4, we have for all k

$$\limsup_{n\to\infty} \left| \int g_{n,R}\,\mathrm{d}\gamma_n - \int g_R\,\mathrm{d}\gamma \right| \leq \limsup_{n\to\infty} \left| \int_{\mathbb{R}^d\setminus\Omega} (g_{n,R} - g_R)\,\mathrm{d}\gamma_n \right| \leq \frac{8R_\epsilon C}{k} + 8(C+1)\epsilon.$$

Letting $k \to \infty$, then incorporating the truncation error yields

$$\limsup_{n\to\infty} \left| \int g_n\,\mathrm{d}\gamma_n - \int g\,\mathrm{d}\gamma \right| \leq 8(C+1)\epsilon + 8\epsilon.$$

[2] This is possible even if E_1 is unbounded: let $E_1^m = E_1 \cap [-m,m]^d$, find a closed set $A_k^m \subseteq E_1^m$ with $\mathrm{Leb}(E_1^m \setminus A_k^m) < 2^{-m}/k$ and choose $A_k = \cup_m A_k^m$, which stays closed even though the union is countable.

The proof is complete upon noticing that ϵ is arbitrary. □

Remark 12 (continuity of $\mathscr{A}$)**.** *One can similarly show that if $W_2(\gamma_n,\gamma)\to 0$ and γ_n have uniformly bounded densities, then $\mathscr{A}(\gamma_n)\to\mathscr{A}(\gamma)$. Indeed, it is sufficient to show that for all bounded uniformly continuous f,*

$$\int_{\mathbb{R}^d} f\left(\frac{1}{N}\sum_{i=1}^N \mathbf{t}_{\gamma_n}^{\mu^i}(x)\right)\mathrm{d}\gamma_n(x)\to\int_{\mathbb{R}^d} f\left(\frac{1}{N}\sum_{i=1}^N \mathbf{t}_{\gamma}^{\mu^i}(x)\right)\mathrm{d}\gamma(x),\qquad n\to\infty$$

and this is done as in Steps 3,4 and 5 in Proposition 5.3.6.

Proof of Theorem 5.3.3. Let $E=\mathrm{supp}\bar{\mu}$ and set $A^i=E^{\mathrm{den}}\cap\{x:\mathbf{t}_{\bar{\mu}}^{\mu^i}(x)\text{ is univalued}\}$. As $\bar{\mu}$ is absolutely continuous, $\bar{\mu}(A^i)=1$, and the same is true for $A=\cap_{i=1}^N A^i$. The first assertion then follows from Proposition 2.9.11.

The second statement is proven similarly. Let $E^i=\mathrm{supp}\mu^i$ and notice that by absolute continuity the $B^i=(E^i)^{\mathrm{den}}\cap\{x:\mathbf{t}_{\mu^i}^{\bar{\mu}}(x)\text{ is univalued}\}$ has measure 1 with respect to μ^i. Apply Proposition 2.9.11. If in addition $E^1=\cdots=E^N$ then $\mu^i(B)=1$ for $B=\cap B^i$. □

5.3.1 A complete proof of Lemma 5.3.4

In this subsection we fill in the gap left at the end of the proof of Lemma 5.3.4 by showing in Lemma 5.3.8 that if a measure has a finite second moment, then it always has a tiny bit more then that. The idea is that if for a random variable X and a nonnegative function f, $\mathbb{E}f(X)<\infty$, then there always exists a function g that diverges to infinity but still $\mathbb{E}f(X)g(X)<\infty$. In other words, there is no "largest moment" in a generalised sense. Of course it may happen that $\mathbb{E}X^2<\infty$ but $\mathbb{E}X^{2+\epsilon}=\infty$ for all $\epsilon>0$, but this is not a contradiction because g can be $\log(x+1)$. For concreteness we take $f(x)$ to be x^2, since this is the application we have in Lemma 5.3.4, but the the idea is valid more generally. To alleviate the notation, we assume in this subsection that all functions and random variables are nonnegative (possibly infinite-valued). We write $f(x)\in\omega(g(x))$ or $f\in\omega(g)$ if $f(x)/g(x)\to\infty$ as $x\to\infty$. In particular $f\in\omega(1)$ means $f(x)\to\infty$.

Lemma 5.3.7. *Let f be integrable and nonincreasing. Then there exists a continuous nondecreasing function $g\in\omega(1)$ such that fg is integrable.*

Proof. Set $F(x)=\int_x^\infty f(t)\,\mathrm{d}t$ and $g(x)=[F(x)]^{-1/2}$. Then a change of variables gives

$$\int_0^\infty f(x)g(x)\,\mathrm{d}x=\int_0^\infty f(x)[F(x)]^{-1/2}\,\mathrm{d}x=\int_0^{F(0)}u^{-1/2}\,\mathrm{d}u=2\sqrt{\|f\|_1}<\infty,$$

and $g(x)\to\infty$ because $F(x)\to 0$ as $x\to\infty$ by dominated convergence. □

Lemma 5.3.8. *Let X be a random variable with $\mathbb{E}X^2<\infty$. Then there exists a convex nondecreasing function $H(x)\in\omega(x^2)$ such that $\mathbb{E}H(X)<\infty$.*

Proof. Since

$$\infty > \mathbb{E}X^2 = \int_0^\infty \mathbb{P}(X^2 > t)\,\mathrm{d}t,$$

there exists a function g as in Lemma 5.3.7 such that

$$\infty > \int_0^\infty \mathbb{P}(X^2 > t)g(t)\,\mathrm{d}t = \int_0^\infty \mathbb{P}(X^2 > G^{-1}(u))\,\mathrm{d}u = \int_0^\infty \mathbb{P}(G(X^2) > u)\,\mathrm{d}u = \mathbb{E}G(X^2),$$

where G is the primitive of g and $G(0) = 0$. The properties of g imply that G is convex and invertible, and that for $y < x$,

$$G(x) \geq \int_y^x g(t)\,\mathrm{d}t \geq \int_y^x g(y)\,\mathrm{d}t = (x-y)g(y),$$

which, combined with $g \in \omega(1)$, yields

$$\liminf_{x\to\infty} \frac{G(x)}{x} \geq g(y) \to \infty, \qquad y \to \infty,$$

so that $G(x) \in \omega(x)$. The function $H(x) = G(x^2)$ then has all the desired properties. □

We can now complete the proof of Lemma 5.3.4. Let $Z^i \sim \mu^i$ and define $X^i = \|Z^i\|$, so that

$$\int_{\mathbb{R}^d} \|x\|^2\,\mathrm{d}\mu^i(x) < \infty \qquad \Longrightarrow \qquad \int_0^\infty \mathbb{P}(X_i^2 > t)\,\mathrm{d}t < \infty, \qquad i = 1,\dots,N.$$

There exist functions g^i as in Lemma 5.3.7 with

$$\int_0^\infty \mathbb{P}(X_i^2 > t)g^i(t)\,\mathrm{d}t < \infty, \qquad i = 1,\dots,N.$$

The same holds with g^i replaced by $g = \min_i g^i$, which is still continuous, nondecreasing and divergent. Setting H as in Lemma 5.3.8, we see that $H(x) \in \omega(x^2)$ and

$$M^i = \mathbb{E}H(X^i) = \int_{\mathbb{R}^d} H(\|x\|)\,\mathrm{d}\mu^i(x) < \infty, \qquad i = 1,\dots,N.$$

Convexity of H and $\|\cdot\|$ combined with monotonicity of H yield

$$\begin{aligned}\int_{\mathbb{R}^d} H(\|x\|)\,\mathrm{d}\gamma_j(x) &= \int_{\mathbb{R}^d} H\left(\left\|\frac{1}{N}\sum_{i=1}^N \mathbf{t}^{\mu^i}_{\gamma_{j-1}}(x)\right\|\right)\mathrm{d}\gamma_{j-1}(x)\\ &\leq \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} H\left(\left\|\mathbf{t}^{\mu^i}_{\gamma_{j-1}}(x)\right\|\right)\mathrm{d}\gamma_{j-1}(x) = \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} H(\|x\|)\,\mathrm{d}\mu^i(x) \leq M,\end{aligned}$$

where $M = \sum_{i=1}^{N} M^i/N$. This implies that for any $R > 0$ and any $j > 0$,

$$\int_{\{x:\|x\|>R\}} \|x\|^2 \, \mathrm{d}\gamma_j(x) \leq \sup_{y>R} \frac{y^2}{H(y)} \int_{\{x:\|x\|>R\}} H(\|x\|) \, \mathrm{d}\gamma_j(x) \leq M \sup_{y>R} \frac{y^2}{H(y)},$$

and this vanishes with $R \to \infty$ because $H(y) \in \omega(y^2)$.

5.4 Illustrative examples

As an illustration, we implement Algorithm 1 in several scenarios for which pairwise optimal maps can be calculated explicitly at every iteration, allowing for fast computation without error propagation. In each case we give some theory first, describing how the optimal maps are calculated, and then carry out Algorithm 1 on simulated examples. One prospect of future work is to incorporate numerical schemes such as those given by [11, 24, 46] and apply the algorithm in more general settings.

5.4.1 Gaussian measures

No example illustrates the use of Algorithm 1 better than the Gaussian case. This is so because optimal maps between centred Gaussian measures have the explicit form (see Section 2.7)

$$\mathbf{t}_A^B(x) = A^{-1/2}[A^{1/2}BA^{1/2}]^{1/2}A^{-1/2}x, \qquad x \in \mathbb{R}^d,$$

with a slight abuse of notation. In contrast, the Fréchet mean of a collection of Gaussian measures does not admit a closed-form formula and is only known to be a Gaussian measure whose covariance matrix Γ is the unique root of the matrix equation

$$\Gamma = \frac{1}{N} \sum_{i=1}^{N} \left[\Gamma^{1/2} S_i \Gamma^{1/2}\right]^{1/2}, \tag{5.6}$$

where S_i is the covariance matrix of μ^i.

Given the formula for $\mathbf{t}_A^B$, application of Algorithm 1 to Gaussian measures is straightforward. The next result shows that the iterates must converge to the unique Fréchet mean, and that (5.6) can be derived from the characterisation of Karcher means. This example was studied independently by Álvarez-Esteban et al. [5, Section 4], who give an alternative proof. Our proof is shorter and arguably simpler, but the proof in [5] shows the additional property that the traces of the matrix iterates are monotonically increasing.

Theorem 5.4.1 (convergence in Gaussian case)**.** *Let $\mu^1, \ldots, \mu^N$ be Gaussian measures with zero means and covariance matrices S_i with S_1 nonsingular, and let the initial point γ_0 be $\mathcal{N}(0, \Gamma_0)$ with Γ_0 nonsingular. Then the sequence of iterates generated by Algorithm 1 converges to the unique Fréchet mean of $(\mu^1, \ldots, \mu^N)$.*

Proof. Since the optimal maps are linear, so is their mean and therefore γ_k is a Gaussian measure for all k, say $\mathcal{N}(0,\Gamma_k)$ with Γ_k nonsingular. Any limit point of γ is a Karcher mean by Theorem 5.3.1. If we knew that γ itself were Gaussian, then it actually must be the Fréchet mean because $N^{-1}\sum \mathbf{t}_\gamma^{\mu^i}$ equals the identity everywhere on $\mathbb{R}^d$ (see the discussion before Theorem 3.5.18).

Let us show that every limit point γ is indeed Gaussian. It suffices to prove that (Γ_k) is a bounded sequence, because if $\Gamma_k \to \Gamma$ then $\mathcal{N}(0,\Gamma_k) \to \mathcal{N}(0,\Gamma)$ narrowly, as can be seen from either Lehmann–Scheffé's theorem (the densities converge) or Lévy's continuity theorem (the characteristic functions converge).

To see that (Γ_k) is bounded, observe first that for any centred (Gaussian or not) measure μ with covariance matrix S,

$$W_2^2(\mu,\delta_0) = \mathrm{tr}S,$$

where δ_0 is a Dirac mass at the origin. (This follows from the singular value decomposition of S.) Therefore

$$0 \le \mathrm{tr}\Gamma_k = W_2^2(\gamma_k,\delta_0)$$

is bounded uniformly, because $\{\gamma_k\}$ stays in a Wasserstein-compact set by Lemma 5.3.4. If we define $C = \sup_k \mathrm{tr}\Gamma_k < \infty$, then all the diagonal elements of Γ_k are bounded uniformly. When A is symmetric and positive semidefinite, $2|A_{ij}| \le A_{ii} + A_{jj}$. Consequently all the entries of Γ_k are bounded uniformly by C, which means that (Γ_k) is a bounded sequence. □

From the formula for the optimal maps, we see that if Γ is the covariance of the Fréchet mean, then

$$I = \sum_{i=1}^N \Gamma^{-1/2}\left[\Gamma^{1/2} S_i \Gamma^{1/2}\right]^{1/2} \Gamma^{-1/2}$$

and we recover the fixed point equation (5.6).

If the means are nonzero, then the optimal maps are affine and the same result applies; the Fréchet mean is still a Gaussian measure with covariance matrix Γ and mean that equals the average of the means of μ^i, $i = 1,\dots,N$.

Figure 5.1 shows density plots of $N = 4$ centred Gaussian measures on $\mathbb{R}^2$ with covariances $S_i \sim \mathrm{Wishart}(I_2,2)$, and Figure 5.2 shows the density of the resulting Fréchet mean. In this particular example, the algorithm needed 11 iterations starting from the identity matrix. The corresponding optimal maps are displayed in Figure 5.3. It is apparent from the figure that these maps are linear, and after a more careful reflection one can be convinced that their average is the identity. The four plots in the figure are remarkably different, in accordance with the measures themselves having widely varying condition numbers and orientations; μ^3

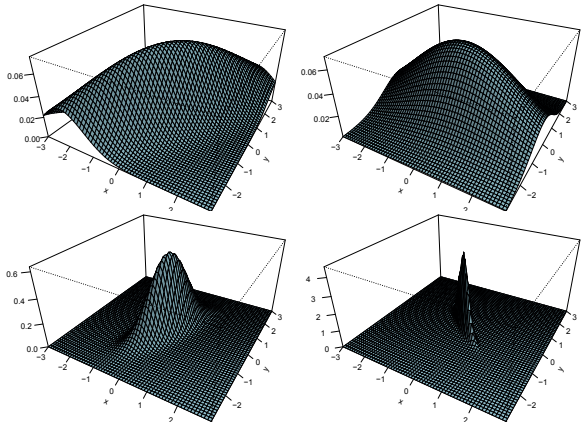


Figure 5.1: Density plot of four Gaussian measures in $\mathbb{R}^2$.

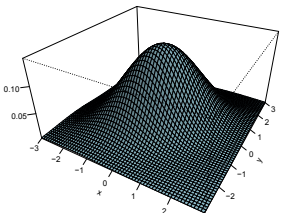


Figure 5.2: Density plot of the Fréchet mean of the measures in Figure 5.1.

and more so μ^4 are very concentrated, so the optimal maps "sweep" the mass towards zero. In contrast, the optimal maps to μ^1 and μ^2 spread the mass out away from the origin.

5.4.2 Compatible measures

We next discuss the behaviour of the algorithm when the measures are compatible. Recall that a collection $\mathscr{C} \subseteq \mathscr{W}_2(\mathscr{X})$ is *compatible* if for all $\gamma, \rho, \mu \in \mathscr{C}$, $\mathbf{t}_\mu^\nu \circ \mathbf{t}_\gamma^\mu = \mathbf{t}_\gamma^\nu$ in $L_2(\gamma)$ (Definition 3.3.1). Boissard, Le Gouic & Loubes [20] showed that when this condition holds, the Fréchet mean of $(\mu^1, \ldots, \mu^N)$ can be found by simple computations involving the *iterated barycentre*. We again denote by γ_0 the initial point of Algorithm 1, which can be any absolutely continuous measure.

Lemma 5.4.2 (compatibility and convergence)**.** *If $\gamma_0 \cup \{\mu^i\}$ is compatible then Algorithm 1 converges to the Fréchet mean of (μ^i) after a single step.*

Proof. By definition, the next iterate

$$\gamma_1 = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{t}_{\gamma_0}^{\mu^i} \right] \# \gamma_0$$

is the Fréchet mean by Theorem 3.5.21. □

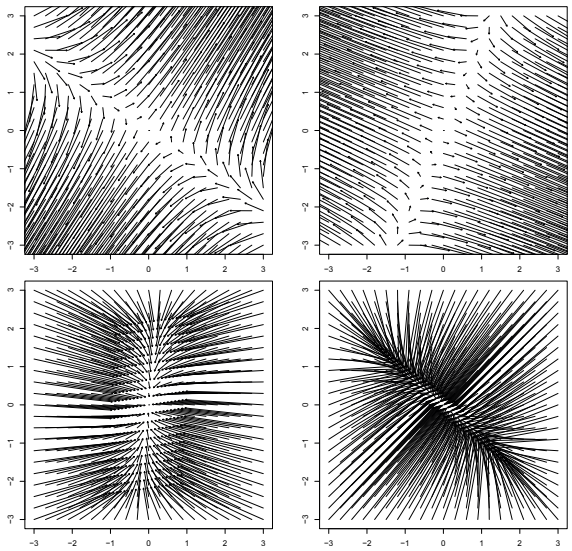


Figure 5.3: Gaussian example: vector fields depicting the optimal maps $x \mapsto \mathbf{t}_{\bar{\mu}}^{\mu^i}(x)$ from the Fréchet mean $\bar{\mu}$ of Figure 5.2 to the four measures $\{\mu^i\}$ of Figure 5.1. The order corresponds to that of Figure 5.1.

In this case, Algorithm 1 requires the calculation of N pairwise optimal maps, and this can be reduced to $N-1$ if the initial point is chosen to be μ^1. This is the same computational complexity as the calculation of the iterated barycentre proposed in [20].

When the measures have a common copula, finding the optimal maps reduces to finding the optimal maps between the one-dimensional marginals (see Lemma 3.3.2) and this can be done using quantile functions as described in Section 2.6. We next illustrate Algorithm 1 in three such scenarios.

The one-dimensional case

When the measures are supported on the real line, there is no need to use the algorithm since the Fréchet mean admits a closed-form expression in terms of quantile functions (see Subsection 3.5.2). We nevertheless discuss this case briefly because we build upon this construction in subsequent examples. Given that $\mathbf{t}_\mu^\nu = F_\nu^{-1} \circ F_\mu$, we may apply Algorithm 1 starting from one of these measures (or any other measure). Figure 5.4 plots $N = 4$ univariate densities and the Fréchet mean yielded by the algorithm in two different scenarios. At the left, the densities were generated as

$$f^i(x) = \frac{1}{2}\phi\left(\frac{x-m_1^i}{\sigma_1^i}\right) + \frac{1}{2}\phi\left(\frac{x-m_2^i}{\sigma_2^i}\right), \tag{5.7}$$

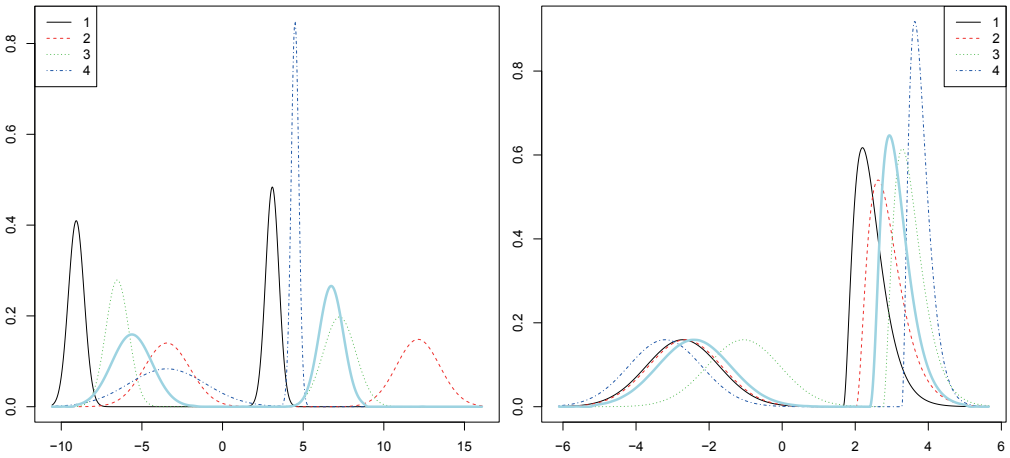


Figure 5.4: Densities of a bimodal Gaussian mixture (left) and a mixture of a Gaussian with a gamma (right), with the Fréchet mean density in light blue.

with ϕ the standard normal density, and the parameters generated independently as

$$m_1^i \sim U[-13,-3], \quad m_2^i \sim U[3,13], \quad \sigma_1^i, \sigma_2^i \sim Gamma(4,4).$$

At the right of Figure 5.4, we used a mixture of a shifted gamma and a Gaussian:

$$f^i(x) = \frac{3}{5}\frac{\beta_i^3}{\Gamma(3)}(x-m_3^i)^2 e^{-\beta_i(x-m_3^i)} + \frac{2}{5}\phi(x-m_4^i), \tag{5.8}$$

with

$$\beta^i \sim Gamma(4,1), \quad m_3^i \sim U[1,4], \quad m_4^i \sim U[-4,-1].$$

The resulting Fréchet mean density for both settings is shown in thick light blue, and can be seen to capture the bimodal nature of the data. Even though the Fréchet mean of Gaussian mixtures is not a Gaussian mixture itself, it is approximately so, provided that the peaks are separated enough. Figure 5.5 shows the optimal maps pushing the Fréchet mean $\bar{\mu}$ to the measures $\mu^1, \ldots, \mu^N$ in each case. If one ignores the “middle part" of the x axis, the maps appear (approximately) affine for small values of x and for large values of x, indicating how the peaks are shifted. In the middle region, the maps need to “bridge the gap" between the different slopes and intercepts of these affine maps.

Independence

We next take measures μ^i on $\mathbb{R}^2$, having independent marginal densities f_X^i as in (5.7), and f_Y^i as in (5.8). Figure 5.6 shows the density plot of $N = 4$ such measures, constructed as the product of the measures from Figure 5.4. One can distinguish the independence by the “parallel" structure of the figures: for every pair (y_1, y_2), the ratio $g(x, y_1)/g(x, y_2)$ does not

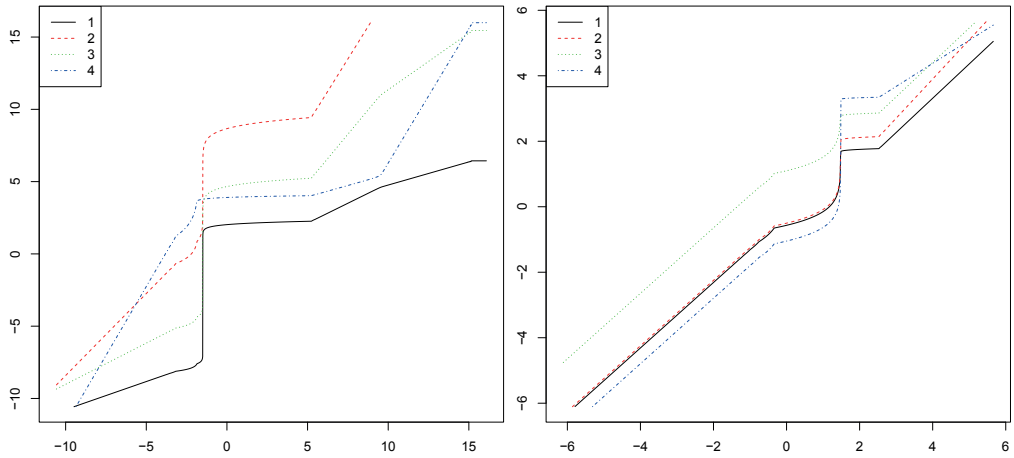


Figure 5.5: Optimal maps $\mathbf{t}_{\bar{\mu}}^{\mu^i}$ from the Fréchet mean $\bar{\mu}$ to the four measures $\{\mu^i\}$ in Figure 5.4. The left plot corresponds to the bimodal Gaussian mixture, and the right plot to the Gaussian/gamma mixture.

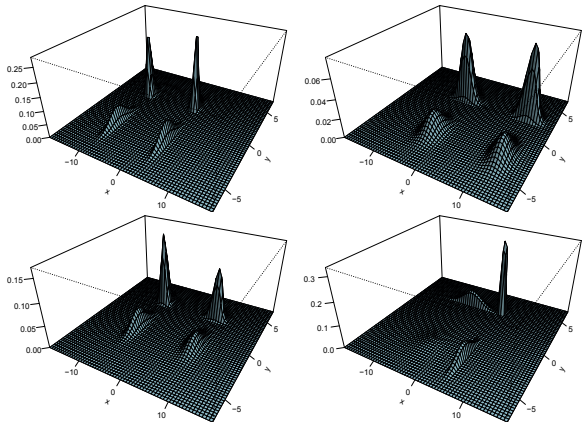


Figure 5.6: Density plots of the four product measures of the measures in Figure 5.4.

depend on x (and vice versa, interchanging x and y). Figure 5.7 plots the density of the resulting Fréchet mean. We observe that the Fréchet mean captures the four peaks, and their location. Furthermore, the parallel nature of the figure is preserved in the Fréchet mean. Indeed, by Lemma 3.5.10 the Fréchet mean is a product measure. The optimal maps, in Figure 5.10, are the same as in the next example, and will be discussed there.

Common copula

Let μ^i be a measure on $\mathbb{R}^2$ with density

$$g^i(x,y) = c(F_X^i(x), F_Y^i(y)) f_X^i(x) f_Y^i(y),$$

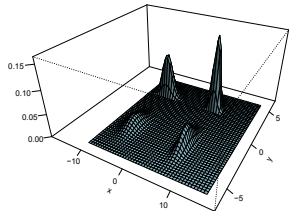


Figure 5.7: Density plot of the Fréchet mean of the measures in Figure 5.6.

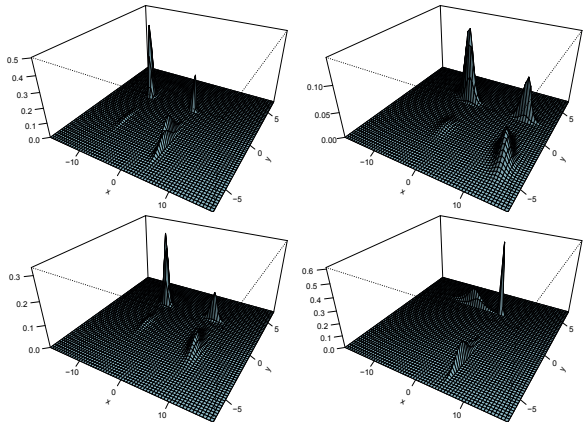


Figure 5.8: Density plots of four measures in $\mathbb{R}^2$ with Frank copula of parameter -8.

where f_X^i and f_Y^i are random densities on the real line with distribution functions F_X^i and F_Y^i, and c is a copula density. Figure 5.8 shows the density plot of $N = 4$ such measures, with f_X^i generated as in (5.7), f_Y^i as in (5.8), and c is the Frank(-8) copula density, while Figure 5.9 plots the density of the Fréchet mean obtained. (For ease of comparison we use the same realisations of the densities that appear in Figure 5.4.) The Fréchet mean can be seen to preserve the shape of the density, having four clearly distinguished peaks. Figure 5.10, depicting the resulting optimal maps, allows for a clearer interpretation: for instance the leftmost plot (in black) shows more clearly that the map splits the mass around $x = -2$ to a much wider interval; and conversely a very large amount mass is sent to $x \approx 2$. This rather extreme behaviour matches the peak of the density of μ^1 located at $x = 2$.

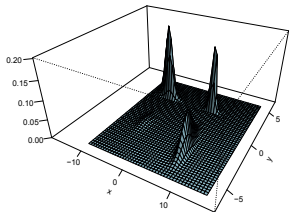


Figure 5.9: Density plot of the Fréchet mean of the measures in Figure 5.8.

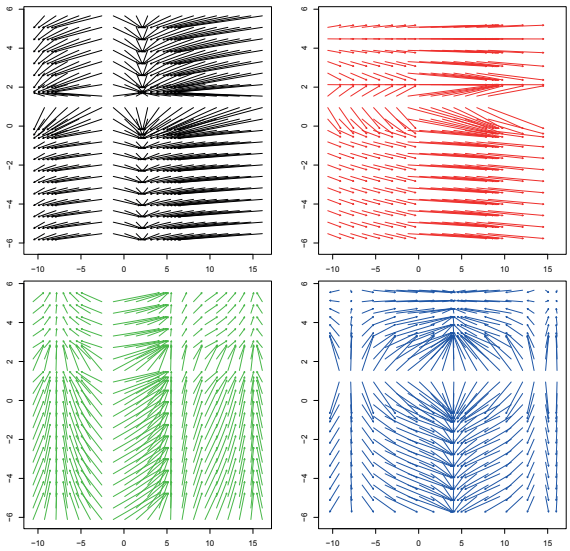


Figure 5.10: Frank copula example: vector fields of the optimal maps $\mathbf{t}_{\bar{\mu}}^{\mu^i}$ from the Fréchet mean $\bar{\mu}$ of Figure 5.9 to the four measures $\{\mu^i\}$ of Figure 5.8. The colours match those of Figure 5.4.

5.4.3 Partially Gaussian trivariate measures

We now apply Algorithm 1 in a situation that entangles two of the previous settings. Let U be a 3×3 real orthogonal matrix with columns U_1, U_2, U_3 and let μ^i have density

$$g^i(y_1, y_2, y_3) = g^i(y) = f^i(U_3^t y) \frac{1}{2\pi\sqrt{\det S^i}} \exp\left[-\frac{(U_1^t y, U_2^t y)(S^i)^{-1}\binom{U_1^t y}{U_2^t y}}{2} \right],$$

with f^i bounded density on the real line and $S^i \in \mathbb{R}^{2\times 2}$ positive definite. We simulated $N = 4$ such densities with f^i as in (5.7) and $S^i \sim \text{Wishart}(I_2, 2)$. We apply Algorithm 1 to this collection of measures and find their Fréchet mean (see the end of this subsection for precise details on how the optimal maps were calculated). Figure 5.11 shows level set of the resulting densities for some specific values. The bimodal nature of f^i implies that for most values of a, $\{x : f^i(x) = a\}$ has four elements. Hence the level sets in the figures are unions of four separate parts, with each peak of f^i contributing two parts that form together the boundary of an ellipsoid in $\mathbb{R}^3$ (see Figure 5.12). The principal axes of these ellipsoids and their position in $\mathbb{R}^3$ differ between the measures, but the Fréchet mean can be viewed as an average of those in some sense.

In terms of orientation (principal axes) of the ellipsoids, the Fréchet mean is most similar to μ^1 and μ^2, whose orientations are similar to one another.

Let us now see how the optimal maps are calculated. If $Y = (y_1, y_2, y_3) \sim \mu^i$, then the random

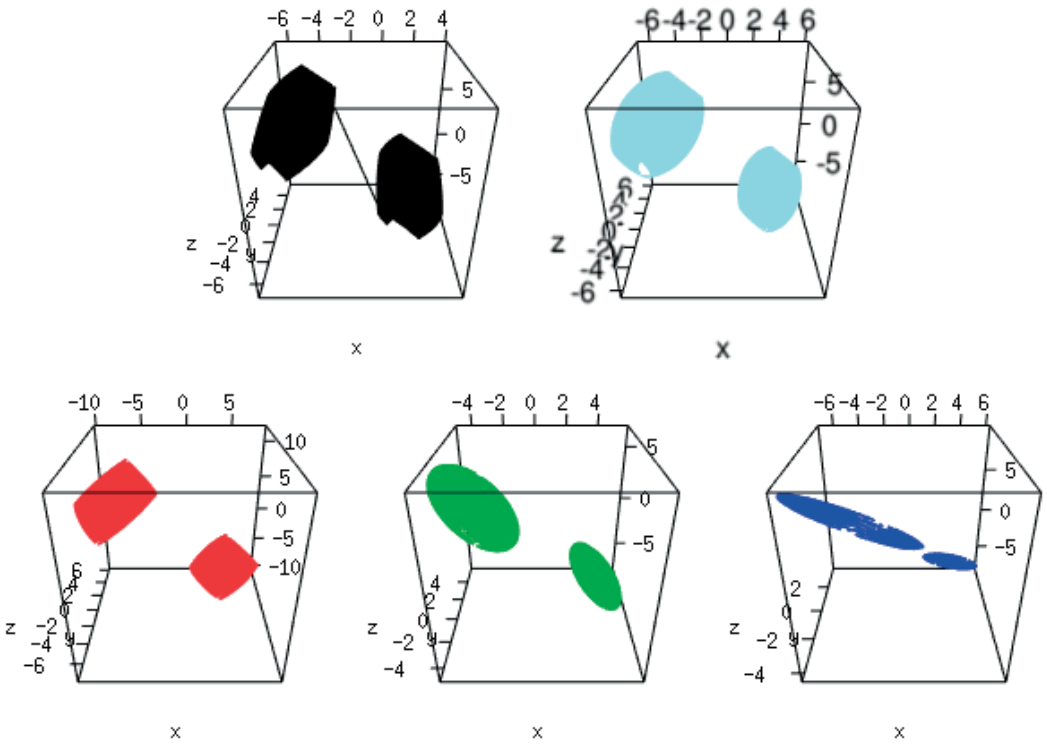


Figure 5.11: The set $\{v \in \mathbb{R}^3 : g^i(v) = 0.0003\}$ for $i = 1$ (black), the Fréchet mean (light blue), $i = 2,3,4$ in red, green and dark blue respectively.

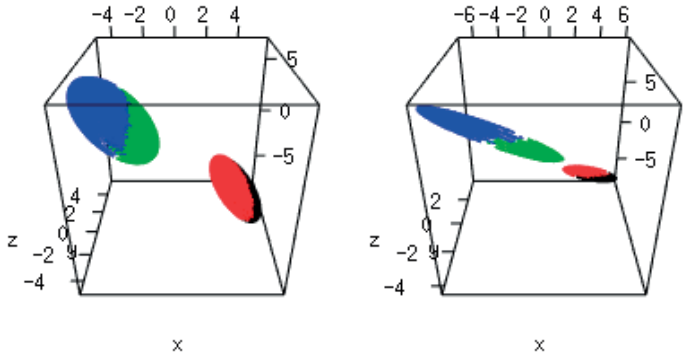


Figure 5.12: The set $\{v \in \mathbb{R}^3 : g^i(v) = 0.0003\}$ for $i = 3$ (left) and $i = 4$ (right), with each of the four different inverses of the bimodal density f^i corresponding to a colour.

vector $(x_1, x_2, x_3) = X = U^{-1}Y$ has joint density

$$f^i(x_3)\exp\left[-\frac{(x_1,x_2)(\Sigma^i)^{-1}\binom{x_1}{x_2}}{2}\right]\frac{1}{2\pi\sqrt{\det\Sigma^i}},$$

so the probability law of X is $\rho^i \otimes \nu^i$ with ρ^i centred Gaussian with covariance matrix Σ^i and ν^i having density f^i on $\mathbb{R}$. By Lemma 3.5.10, the Fréchet mean of $(U^{-1}\#\mu^i)$ is the product measure of that of (ρ^i) and that of (ν^i); by Lemma 3.5.11, the Fréchet mean of (μ^i) is therefore

$$U\#(\mathcal{N}(0,\Sigma)\otimes f), \qquad f = F', \quad F^{-1}(q) = \frac{1}{N}\sum_{i=1}^{N} F_i^{-1}(q), \quad F_i(x) = \int_{-\infty}^{x} f^i(s)\,\mathrm{d}s,$$

where Σ is the Fréchet–Wasserstein mean of $\Sigma_1, \dots, \Sigma_N$.

Starting at an initial point $\gamma_0 = U\#(\mathcal{N}(0,\Sigma_0)\otimes\nu_0)$, with ν_0 having continuous distribution F_{ν_0}, the optimal maps are $U\circ\mathbf{t}_0^i\circ U^{-1} = \nabla(\varphi_0^i\circ U^{-1})$ with

$$\mathbf{t}_0^i(x_1,x_2,x_3) = \begin{pmatrix} \mathbf{t}_{\Sigma_0}^{\Sigma^j}(x_1,x_2) \\ F_j^{-1}\circ F_{\nu_0}(x_3)\end{pmatrix}$$

the gradients of the convex function

$$\varphi_0^i(x_1,x_2,x_3) = (x_1,x_2)\mathbf{t}_{\gamma_0}^{\Sigma^i}\begin{pmatrix}x_1\\x_2\end{pmatrix} + \int_0^{x_3} F_j^{-1}(F_{\nu_0}(s))\,\mathrm{d}s,$$

where we identify $\mathbf{t}_{\gamma_0}^{\Sigma^i}$ with the positive definite matrix $(\Sigma^i)^{1/2}[(\Sigma^i)^{1/2}\Sigma_0(\Sigma^i)^{1/2}]^{-1/2}(\Sigma^i)^{1/2}$ that pushes forward $\mathcal{N}(0,\Sigma_0)$ to $\mathcal{N}(0,\Sigma^i)$. Due to the one-dimensionality, the algorithm finds the third component of the rotated measures after one step, but the convergence of the Gaussian component requires further iterations.

5.5 Further properties of Karcher means

The convergence proof of Algorithm 1 reveals further properties that are worth mentioning. Recall that $\mathscr{A}$ is the function that takes an absolutely continuous $\gamma \in \mathscr{W}_2(\mathbb{R}^d)$ and applies to it one iteration. We will prove in this section results pertaining to any measure γ in the image of $\mathscr{A}$. Such a γ will be called a **descent iterate**. If γ is a Karcher mean, then $\mathscr{A}(\gamma) = \gamma$, so in particular these results apply to Karcher means; in view of Propositions 3.5.16 and 3.5.17, they also hold for Fréchet means provided that

$$\mu^1 \text{ is absolutely continuous and has a bounded density.} \tag{5.9}$$

Let us begin with the support.

Corollary 5.5.1 (support of algorithm iterates)**.** *The support of any descent iterate γ is included*

in the set

$$E = \frac{1}{N}\left(\mathrm{supp}\mu^1 + \cdots + \mathrm{supp}\mu^N\right).$$

This generalises Proposition 3.5.5 (when $\mathscr{X} = \mathbb{R}^d$) in which (μ^i) have a common convex support.

Proof. Write $\gamma = \mathscr{A}(\rho)$, so that $\mathbf{t}_\rho^\gamma = N^{-1}\sum \mathbf{t}_\rho^{\mu^i}$. But for ρ-almost every x, $\mathbf{t}_\rho^{\mu^i}(x) \in \mathrm{supp}\mu^i$, and so $\mathbf{t}_\rho^\gamma(x) \in E$. □

We next discuss an improvement of the constant in Proposition 5.3.5 that will be fundamental in order to obtain a population version of Algorithm 1. This amelioration comes from replacing the minimum of the density bounds by their harmonic mean. Let γ be a descent iterate and denote its density by h. It has been established in the proof of Proposition 5.3.5 that

$$\frac{1}{\|h\|_\infty^{1/d}} \geq \frac{1}{N}\sum_{i=1}^N \frac{1}{\|g^i\|_\infty^{1/d}},$$

if the measures $\mu^1,\dots,\mu^N$ have densities $g^1,\dots,g^N$. To avoid the need to introduce an extra symbol we write $\|g^i\|_\infty = \infty$ if μ^i is not absolutely continuous, even though g^i does not exist; the above inequality then holds in full generality.

Corollary 5.5.2 (improved density bound)**.** *Suppose that a fraction $q = n/N$ $(1 \leq n \leq N)$ of the measures possess densities that are bounded by M. Then $\|h\|_\infty \leq M/q^d$.*

Proof. Without loss generality the bound is satisfied by $g^1,\dots,g^n$. Then

$$\frac{1}{\|h\|_\infty^{1/d}} \geq \frac{1}{N}\sum_{i=1}^N \frac{1}{\|g^i\|_\infty^{1/d}} \geq \frac{1}{N}\sum_{i=1}^n \frac{1}{\|g^i\|_\infty^{1/d}} \geq \frac{1}{N}\sum_{i=1}^n \frac{1}{M^{1/d}} = \frac{n}{N}\frac{1}{M^{1/d}} = \frac{q}{M^{1/d}},$$

and the result follows from taking both sides to the power $-d$. □

As observed by Pass [71, Subsection 3.3], the fact that the number of measures N does not appear in the bound opens the door for a population version of this result, which is the topic of the next section.

5.6 Population version of Algorithm 1

Let $\Lambda \in \mathscr{W}_2(\mathbb{R}^d)$ be a random measure with finite Fréchet functional. The population version of (5.9) is

$$q = \mathbb{P}(\Lambda \text{ absolutely continuous with density bounded by } M) > 0 \quad \text{for some } M < \infty, \tag{5.10}$$

which we assume henceforth. This condition is satisfied if and only if

$$\mathbb{P}(\Lambda \text{ absolutely continuous with bounded density}) > 0.$$

These probabilities are well-defined because the set

$$\mathscr{W}_2(\mathbb{R}^d; M) = \{\mu \in \mathscr{W}_2(\mathbb{R}^d) : \mu \text{ absolutely continuous with density bounded by } M\}$$

is narrowly closed (see the paragraph before Proposition 5.3.5), hence a Borel set of $\mathscr{W}_2(\mathbb{R}^d)$.

In light of Theorem 3.5.15, we can define a population version of Algorithm 1 with the iteration function

$$\mathscr{A}(\gamma) = \mathbb{E}\mathbf{t}_\gamma^\Lambda, \qquad \gamma \in \mathscr{W}_2(\mathbb{R}^d) \text{ absolutely continuous.}$$

The (Bochner) expectation is well-defined in $\mathscr{L}_2(\gamma)$ because the random map $\mathbf{t}_\gamma^\Lambda$ is measurable (Lemma 3.4.5). Since $\mathscr{L}_2(\gamma)$ is a Hilbert space, the law of large numbers applies there, and results for the empirical version carry over to the population version by means of approximations. In particular:

Lemma 5.6.1. *Any descent iterate γ has density bounded by $q^{-d}M$.*

Proof. Let $\Lambda_1, \dots$ be a sample from Λ and let q_n be the proportion of measures in $(\Lambda_1, \dots, \Lambda_n)$ that have density bounded by M. Then both $n^{-1}\sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i} \to \mathbb{E}\mathbf{t}_\gamma^\Lambda$ and $q_n \to q$ almost surely by the law of large numbers. Pick any ω in the probability space for which this happens and notice that by Lemma 3.4.4 and Corollary 5.5.2

$$\mathscr{A}(\gamma) = \left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i}\right] \#\gamma = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^n \mathbf{t}_\gamma^{\Lambda_i}\right]\#\gamma$$

has a density smaller than $q^{-d}M$ (see the next proof for a more detailed reasoning). □

Though it follows that every Karcher mean of Λ has a bounded density, we cannot yet conclude that the same bound holds for the Fréchet mean, because we need an a-priori knowledge that the latter is absolutely continuous. With tools previously developed, this assertion can be established by approximation of the empirical analogue under compactness assumptions. For a cleaner statement we restrict Λ to a compact set but this can be considerably relaxed, see Remark 13. Incorporating Theorem 2 in Le Gouic & Loubes [42], the results are in fact valid for any Λ with finite Fréchet functional.

Theorem 5.6.2 (bounded density for population Fréchet mean)**.** *Let $K \subset \mathbb{R}^d$ be a bounded Borel set and $\Lambda \in \mathscr{W}_2(K)$ be a random measure. If Λ has a bounded density with positive probability then the Fréchet mean of Λ is absolutely continuous with a bounded density.*

Proof. Clearly Λ has a finite Fréchet functional. Let q and M be as in (5.10) and consider the same construction of the preceding lemma. For almost every ω the empirical Fréchet mean λ_n of the sample $(\Lambda_1, \dots, \Lambda_n)$ has a density bounded by $q_n^{-d}(\omega)M$. The Fréchet mean λ of Λ is unique by Proposition 3.5.8, and consequently $\lambda_n \to \lambda$ in $\mathcal{W}_2(K)$, as has been established in the proof of Theorem 4.4.1. For any $C > \limsup q_n^{-d} M$, the density of λ is bounded by C by the portmanteau lemma 2.9.1 (p. 31). Thus the density is bounded by $q^{-d}M$. □

In the same way, one shows the population version of Theorem 3.5.21:

Theorem 5.6.3 (Fréchet mean of compatible measures)**.** *Let $K \subset \mathbb{R}^d$ be a bounded Borel set and $\Lambda \in \mathcal{W}_2(K)$ be a random measure, defined on a probability space Ω and absolutely continuous with positive (inner) probability. If the collection $\{\gamma\} \cup \Lambda(\Omega)$ is compatible and γ is absolutely continuous, then $[\mathbb{E}\mathbf{t}_\gamma^\Lambda]\#\gamma$ is the Fréchet mean of Λ.*

It is of course sufficient that $\{\gamma\} \cup \Lambda(\Omega \setminus \mathcal{N})$ be compatible for some null set $\mathcal{N} \subset \Omega$.

Remark 13 (balls in $\mathcal{W}_{p+\epsilon}$ are compact in $\mathcal{W}_p$)**.** *As the proof shows, we may replace the compact set $\mathcal{W}_2(K)$ by a compact set $\mathcal{K} \subset \mathcal{W}_2(\mathbb{R}^d)$, provided that we know that $\widehat{\lambda}_n$ stay in $\mathcal{K}$. An example of such $\mathcal{K}$ is the set of measures μ such that for some $M > 0$,*

$$\int_{\mathbb{R}^d} \|x\|^3 \, \mathrm{d}\mu(x) \le M.$$

The power 3 can be replaced by $2+\epsilon$ and in fact $\|x\|^3$ can be replaced by $H(\|x\|)$ with H any function that goes to infinity faster than x^2, as is evident from (3.6).

6 Outlook

We conclude the thesis with what we believe are interesting prospects for future work.

6.1 Extensions of Algorithm 1

Implementation. As mentioned in Section 5.4, writing a full implementation of Algorithm 1 that incorporates a numerical scheme for the solution of the pairwise problem is an important project. Since these numerical schemes are themselves iterative, such implementation would need to take care in managing propagation of errors.

Conditions for uniqueness of Karcher means. In general, the Fréchet functional F associated to measures $\mu^1, \dots, \mu^N$ may have more than one local minimum (Karcher mean). We have given a criterion for a local minimum to be the global minimum (Theorem 3.5.18). We conjecture that in the setting of this result, F will in fact have only one local minimum, which is then the Fréchet mean. More precisely, we believe that a result in the flavour of the following should be true:

Conjecture. Suppose that $\mu^1, \dots, \mu^N$ have densities bounded above and below (perhaps smooth) on a (possibly smooth) convex compact $K \subset \mathbb{R}^d$. Then F has a unique Karcher mean, which is the Fréchet mean of $\mu^1, \dots, \mu^N$.

Discrete measures. Suppose that each $\mu^1, \dots, \mu^N$ has a finite support of the same size M, with equal weights. One can still apply Algorithm 1 in this setting. Empirical evidence shows that there are many Karcher means that are not Fréchet means in this setup. But, can we at least show that the Fréchet mean (in case it is unique) is also concentrated on M points? Unlike the discrete case in Section 2.3, the argument must be related to the cost function, because the polytope resulting from the constraints has many extreme points, and most of them are not associated with measures supported on M points. Empirical evidence shows, however, that the support of the Fréchet mean is indeed only M points. A more general setup is explored by Anderes, Borgwardt & Miller [9], and perhaps their work can shed some light on this question.

Population version. We have sketched in Section 5.6 a version of the algorithm for infinitely many measures. It would be of interest to see under which condition the analogue of Theorem 5.3.1 holds true. We believe that merely having a finite Fréchet functional would be sufficient.

Convergence rates for Algorithm 1. We observed in the Gaussian case very rapid convergence of Algorithm 1 to the Fréchet mean. Nevertheless, no analytic results in this direction are known.

6.2 Generalising the consistency framework of Chapter 4

Beyond compactness. In [42] Le Gouic & Loubes prove a general consistency result for Fréchet means in $\mathcal{W}_2$. It should therefore be possible to remove the compactness assumption in Theorem 4.4.1 and, more generally, Section 4.4. Perhaps an assumption on the measures lying in some finite Wasserstein ball will be required, as the uniform Lipschitz bounds (4.6) on F and F_n still hold in such setup.

Convergence rates. Another interesting development of this work would be to extend the convergence rates to a multivariate setting. With the results of Barthe & Bordenave [10], controlling the rate of convergence of $\widehat{F}_n$ to F should not pose a major difficulty. Rather, relating that rate to convergence of minimisers is not straightforward, due to the curvature of the Wasserstein space. Finding an upper bound for the sectional curvature seems to be crucial for the establishment of such results.

Bibliography

[1] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.

[2] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *Society for Industrial and Applied Mathematics*, 43(2):904–924, 2011.

[3] G. Alberti and L. Ambrosio. A geometrical approach to monotone functions in $\mathbb{R}^n$. *Math. Z.*, 230(2):259–316, 1999.

[4] P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos, and C. Matrán. Uniqueness and approximate computation of optimal incomplete transportation plans. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(2):358–375, 2011.

[5] P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos, and C. Matrán. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2):744–762, 2016.

[6] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics. ETH Zürich. Springer Science & Business Media, 2nd edition, 2008.

[7] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1-theory of optimal transportation. In *Optimal Transportation and Applications*, pages 123–160. Springer, 2003.

[8] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86(414):376–387, 1991.

[9] E. Anderes, S. Borgwardt, and J. Miller. Discrete Wasserstein barycenters: Optimal transport for discrete data. *Mathematical Methods of Operations Research*, 84(2):1–21, 2016.

[10] F. Barthe and C. Bordenave. Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*, pages 483–535. Springer International Publishing, Heidelberg, 2013.

[11] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[12] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.

[13] R. Bhatia. *Positive definite matrices*. Princeton university press, 2009.

[14] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.

[15] J. Bigot, R. Gouet, T. Klein, and A. López. Minimax convergence rate for estimating the Wasserstein barycenter of random measures on the real line. *arXiv preprint arXiv:1606.03933*, 2016.

[16] J. Bigot and T. Klein. Characterization of barycenters in the Wasserstein space by averaging optimal transport maps. *arXiv preprint arXiv:1212.2562*, 2012.

[17] P. Billingsley. *Convergence of probability measures*. John Wiley&Sons Inc., New York, 2nd edition, 1999.

[18] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics and Kantorovich transport distances. *preprint*, 2014.

[19] E. Boissard and T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(2):539–563, 2014.

[20] E. Boissard, T. Le Gouic, and J.-M. Loubes. Distribution's template estimate with Wasserstein metrics. *Bernoulli*, 21(2):740–759, 2015.

[21] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. Sliced and radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[22] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[23] L. A. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.

[24] R. Chartrand, B. Wohlberg, K. R. Vixie, and E. M. Bollt. A gradient descent solution to the Monge–Kantorovich problem. *Applied Mathematical Sciences*, 3(22):1071–1080, 2009.

[25] P. Clément and W. Desch. An elementary proof of the triangle inequality for the Wasserstein metric. *Proceedings of the American Mathematical Society*, 136(1):333–339, 2008.

[26] J. A. Cuesta-Albertos and C. Matrán. Notes on the Wasserstein metric in Hilbert spaces. *The Annals of Probability*, 17(3):1264–1276, 1989.

[27] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. *Proceedings of the International Conference on Machine Learning 2014, JMLR W&CP*, 32(1):685–693, 2014.

[28] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[29] D. Dowson and B. Landau. The Fréchet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12(3):450–455, 1982.

[30] I. L. Dryden and K. V. Mardia. *Statistical shape analysis*, volume 4. J. Wiley Chichester, 1998.

[31] R. M. Dudley. *Real analysis and probability*, volume 74. Cambridge University Press, 2002.

[32] N. Dunford, J. T. Schwartz, W. G. Bade, and R. G. Bartle. *Linear operators*. Wiley-interscience New York, 1971.

[33] R. Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[34] J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19(2):248–264, 1972.

[35] J. Fontbona, H. Guérin, and S. Méléard. Measurability of optimal transportation and strong coupling of martingale measures. *Electron. Commun. Probab*, 15:124–133, 2010.

[36] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310, 1948.

[37] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[38] W. Gangbo and A. Święch. Optimal maps for the multidimensional Monge–Kantorovich problem. *Communications on pure and applied mathematics*, 51(1):23–45, 1998.

[39] T. Gasser and A. Kneip. Searching for structure in curve samples. *Journal of the american statistical association*, 90(432):1179–1188, 1995.

[40] D. Gervini and T. Gasser. Self-modelling warping functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):959–971, 2004.

[41] D. Gervini and T. Gasser. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika*, 92(4):801–820, 2005.

[42] T. L. Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *arXiv preprint arXiv:1506.04153*, 2015.

[43] J. C. Gower. Generalized Procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[44] D. Groisser. On the convergence of some Procrustean averaging algorithms. *Stochastics An International Journal of Probability and Stochastic Processes*, 77(1):31–60, 2005.

[45] A. Gut. *Probability: a graduate course*, volume 75. Springer Science & Business Media, 2012.

[46] E. Haber, T. Rehman, and A. Tannenbaum. An efficient numerical method for the solution of the L_2 optimal mass transfer problem. *SIAM Journal on Scientific Computing*, 32(1):197–211, 2010.

[47] T. Hildebrandt. Integration in abstract spaces. *Bulletin of the American Mathematical Society*, 59(2):111–139, 1953.

[48] L. Horváth and P. Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

[49] T. Hsing and R. Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.

[50] G. M. James. Curve alignment by moments. *The Annals of Applied Statistics*, 1(2):480–501, 2007.

[51] H. E. Jones and N. Bayley. The Berkeley growth study. *Child development*, 12(2):167–173, 1941.

[52] O. Kallenberg. *Random measures*. Academic Pr, 3rd edition, 1983.

[53] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, 2nd edition, 1997.

[54] L. V. Kantorovich. On the translocation of masses. *(Dokl.) Acad. Sci. URSS 37*, 3:199–201, 1942.

[55] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

[56] A. Karr. *Point processes and their statistical inference*, volume 7. CRC press, 1991.

[57] A. Kneip and J. O. Ramsay. Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165, 2008.

[58] M. Knott and C. S. Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1):39–49, 1984.

[59] S. G. Krantz. *Convex Analysis*. Textbooks in Mathematics. CRC Press, 2014.

[60] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2:83–97, 1955.

[61] H. Le. Locating Fréchet means with application to shape spaces. *Advances in Applied Probability*, 33(2):324–338, 2001.

[62] E. L. Lehmann. A general concept of unbiasedness. *The Annals of Mathematical Statistics*, 22(4):587–592, 1951.

[63] D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, 2008.

[64] J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484, 2015.

[65] R. J. McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.

[66] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 177:666–704, 1781.

[67] J. Munkers. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5:32–38, 1957.

[68] R. B. Nelsen. *An introduction to copulas*, volume 139. Springer Science & Business Media, 2013.

[69] I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.

[70] V. M. Panaretos and Y. Zemel. Amplitude and phase variation of point processes. *The Annals of Statistics*, 44(2):771–812, 2016.

[71] B. Pass. Optimal transportation with infinitely many marginals. *Journal of Functional Analysis*, 264(4):947–963, 2013.

[72] D. Pollard. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

[73] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory, Volume II: Applications*. Springer Science & Business Media, 1998.

[74] J. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):351–363, 1998.

[75] J. O. Ramsay and B. W. Silverman. *Applied functional data analysis: methods and case studies*, volume 77. Citeseer, 2002.

[76] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2nd edition, 2005.

[77] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497–510, 1966.

[78] R. T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, NJ, 1970.

[79] B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):243–259, 2001.

[80] L. Rüschendorf. On c-optimal random variables. *Statistics & probability letters*, 27(3):267–270, 1996.

[81] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *Journal of Multivariate Analysis*, 32(1):48–54, 1990.

[82] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978.

[83] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87. Springer, 2015.

[84] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge–Kantorovich problem. *Proceedings of the American Mathematical Society*, 137(2):519–529, 2009.

[85] E. M. Stein and R. Shakarchi. *Real Analysis: Measure Theory, Integration & Hilbert Spaces*. Princeton University Press, 2005.

[86] R. Tang and H.-G. Müller. Pairwise curve synchronization for functional data. *Biometrika*, 95(4):875–889, 2008.

[87] J. D. Tucker, W. Wu, and A. Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

[88] C. Villani. *Topics in Optimal Transportation*, volume 58. American Mathematical Society, 2003.

[89] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[90] J.-L. Wang, J.-M. Chiou, and H.-G. Müller. Review of functional data analysis. *arXiv preprint arXiv:1507.05135*, 2015.

[91] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3):1251–1276, 1997.

[92] K. Wang and T. Gasser. Synchronizing sample curves nonparametrically. *Annals of Statistics*, 27:439–460, 1999.

[93] S. Wu, H.-G. Müller, and Z. Zhang. Functional data analysis for point processes with rare events. *Statistica Sinica*, 23(1):1–23, 2013.

[94] Y. Zemel and V. M. Panaretos. Fréchet means in Wasserstein space: gradient descent and Procrustes analysis. *Tech. Report #01/16*. http://smat.epfl.ch/reports/1-16.pdf. Chair of Mathematical Statistics, EPFL (February 2016).

Yoav ZEMEL

32 years (23.3.1985)
12 ch. François-de-Lucinge
1006 Lausanne
Switzerland

Citizenship: Israeli
+41 76 403 5543
yoav.zemel@epfl.ch

EDUCATION

PhD, Ecole polytechnique fédérale de Lausanne (EPFL) 2012-2017

PhD candidate at the Chair of Mathematical Statistics
Provisional title: *Fréchet means in Wasserstein space: Theory and Algorithms*
Graduation expected in April 2017

MSc, Ecole polytechnique fédérale de Lausanne (EPFL) 2010-2012

Master in applied mathematics, orientated in Statistics and Financial Mathematics
Thesis title: *Optimal Transportation: Continuous and Discrete*
GPA: *5.8 out of 6*

BSc, Hebrew University of Jerusalem, Israel 2007-2010

Bachelor in Mathematics and Economics, *Summa cum laude*
GPA: 97.84 out of 100

AWARDS & HONOURS

EPFL teaching award	2015
Swiss Government Scholarship	2010/2011
EPFL Excellence Scholarship (declined)	2010/2011
Participant in workshops for honor students in mathematics and economics	2009/2010
Hebrew University "Amirim" Scholarship	2008-2010
Hebrew University Dean's prize	2008/2009
Hebrew University Rector prize	2007/2008

PAPERS, PREPRINTS AND MONOGRAPHS

Panaretos, V. M. and Zemel, Y. (2016) **Amplitude and Phase Variation of Point Processes**. *Annals of Statistics, **44 (2)**: 771-812.*
http://projecteuclid.org/euclid.aos/1458245735

Henshaw, J. M. and Zemel, Y. (2016) **A unified measure of linear and nonlinear selection on quantitative traits**. *Methods in Ecology and Evolution (doi:10.1111/2041-210X.12685).*
http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12685/full

Zemel, Y. and Panaretos, V. M. (2017) **Fréchet Means and Procrustes Analysis in Wasserstein Space**. *In review.*
https://arxiv.org/pdf/1701.06876.pdf

Panaretos, V. M. and Zemel, Y. (2017) **Foundations of Statistics in the Wasserstein Space**. *In preparation.*

PRESENTATIONS

Invited talk: Workshop on Statistics of Time Warping and Phase Variations, Mathematical Biosciences Institute, Ohio State University 11/2012

EPFL internal PhD mathematics colloquium 03/2014

Invited talk: ISNPS meeting, Biosciences, Medicine, and novel Non-Parametric Methods, Medical University of Graz, Austria 07/2015

Invited talk: 8th International conference of the ERCIM WG on Computational and Methodological Statistics, University of London, UK 12/2015

Invited talk: John Aston group meeting, University of Cambridge, UK 12/2015

Invited PhD presentation: Workshop on Statistical Recovery of Discrete, Geometric and Invariant Structures, Oberwolfach, Germany 03/2017

Invited contributed talk: European Meeting of Statisticians, Helsinki, Finland 07/2017

TEACHING

Teaching assistant in the following courses:
Measure theory (3rd year Bachelor) *(2012-2014)*
Time series (3rd year Bachelor) (2013)
Statistics (2nd year Bachelor) (2014-2016)
Linear models (3rd year Bachelor) (2015)

Student supervision:
Semester project on principal components analysis (2014)

ACADEMIC SERVICE

Served as a referee for: *Biometrika*, *Electronic Journal of Statistics*, *Annals of Statistics*

WORK EXPERIENCE

Hirslanden, Clinique Cecil, Centre de la douleur 10/11-2/12
Statistical analysis of medical data (full time internship)

Berel Ginges Computer Centre, Hebrew University of Jerusalem 2008-2010
Technical support and assistance at the Computer Centre,
serving students, academic staff and visitors (15-20 hours/week)

LANGUAGES

English	Fluent (C2)
French	Fluent (C2)
Hebrew	Native language
Spanish	Intermediate (B1)
German	Intermediate (B1)
Russian	Basic (A2)
Hungarian	Basic (A2)

IT AND PROGRAMMING

Programming	C, C++, basic knowledge of VBScript, ASP, JavaScript (for web)
Technical	R (statistical programming), basic knowledge of MATLAB
Other	Latex, basic knowledge of Linux

HOBBIES AND INTERESTS

Hiking, football, skiing, running, travelling, lindy hop dancing, history