Metric two-level measure spaces: A state space for modeling evolving genealogies in host-parasite systems

Von der Fakultät für Mathematik der Universität Duisburg-Essen
zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)
genehmigte Dissertation

von

Roland Meizis

Datum der mündlichen Prüfung:
05. März 2019

Gutachterin: Prof. Dr. Anita Winter
Gutachter: Prof. Dr. Peter Pfaffelhuber

Zusammenfassung

Wir verallgemeinern den Begriff der metrischen Maßräume zu sogenannten metrischen zwei-Level Maßräumen (abgekürzt als m2m-Räume). Dabei definieren wir einen m2m-Raum als die Isomorphieklasse eines Tripels (X,r,ν), wobei (X,r) ein polnischer metrischer Raum und $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ ein endliches Maß auf der Menge der endlichen Maße auf X ist. Wir betrachten die Menge aller m2m-Räume und führen eine separierende Klasse von Testfunktionen auf dieser Menge ein. Die Definition der Testfunktionen basiert dabei auf der Idee endliche Teilräume von (X,r) mit Hilfe von ν zu samplen, d. h. wir samplen Maße aus $\mathcal{M}_f(X)$ mittels ν und benutzen diese Maße, um endliche viele Punkte aus X zu samplen. Wir untersuchen die von diesen Testfunktionen erzeugte Topologie (auf der Menge der m2m-Räume) und zeigen unter anderem, dass diese Topologie polnisch ist, indem wir eine vollständige Metrik für diese Topologie angeben.

Die Definition der m2m-Räume und der obigen Topologie ist motiviert durch Anwendungen in der mathematische Biologie. Die Menge der m2m-Räume ist geeignet als Zustandsraum für stochastische Prozesse, welche die Genealogie einer Population in einem hierarchischen System mit zwei Leveln modelliert, wie z. B. in einem Wirt-Parasit System. Beispielhaft wenden wir unsere Resultate an, um einen zufälligen m2m-Raum zu konstruieren, welcher die Genealogie in einem zwei-Level Kingman Koaleszenten modelliert.

Abstract

We extend the notion of metric measure spaces to so-called metric two-level measure spaces (m2m spaces for short). An m2m space is an isomorphism class of a triple (X,r,ν), where (X,r) is a Polish metric space and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$, i. e. ν is a finite measure on the set of finite measures on X. We define a point separating class of test functions on the set of m2m spaces based on the idea of sampling finite subspaces of (X,r) by means of ν (we use ν to sample measures from $\mathcal{M}_f(X)$ and then sample a finite subset of X with the sampled measures). We then study the topology which is induced by these test functions and show that this topology is Polish by providing a complete metric.

The framework introduced in this thesis is motivated by possible applications in mathematical biology. It is designed for modeling the random evolution of the genealogy of a population in a hierarchical system with two levels, for example, a host-parasite system or a population which is divided into colonies. As an example we apply our theory to construct a random m2m space which is modeling the genealogy of a nested Kingman coalescent.

Acknowledgments

First of all I would like to gratefully acknowledge the guidance and encouragement of my supervisor Anita Winter. Without her constant support over the last few years, this thesis would not have been possible. Besides my supervisor, I am grateful to Wolfgang Löhr for several helpful comments and discussions and for being my temporary supervisor in the sabbatical semester of Anita Winter. Moreover, I would like to thank Volker Krätschmer for several helpful comments about nets. Last but not least, I would like to thank my fellow PhD student Fabian Gerle for many stimulation conversations, his helpful comments and for proofreading this thesis.

Contents

1 Introduction **1**

2 Preliminaries **9**
2.1 Nets in topological spaces 9
2.2 Initial topology 13
2.3 A note on Polish spaces 14
2.4 The weak topology and the Prokhorov metric 15
2.5 Separating and convergence determining sets 17
2.6 Push-forward operators 19

3 Metric measure spaces **23**
3.1 The reconstruction theorem for mm spaces 24
3.2 Mm-monomials and the Gromov-weak topology 29
3.3 The Gromov-Prokhorov metric and its topology 34
3.4 Distance distribution and modulus of mass distribution 36
3.5 Compact sets 43
3.6 Equivalence of Gromov-weak and Gromov-Prokhorov topology 44
3.7 Distributions on $\mathbb{M}$ 45

4 Metric two-level measure spaces **47**
4.1 Motivation and basic definitions 48
4.2 Two-level measures and moment measures 51
4.3 Metric two-level probability measure spaces 60
4.3.1 The reconstruction theorem for m2pm spaces 60
4.3.2 M2pm-polynomials and the two-level Gromov-weak topology . 65
4.3.3 The two-level Gromov-Prokhorov metric and its topology . . . 68
4.3.4 Distance distribution and modulus of mass distribution 72
4.3.5 Compact sets 75
4.3.6 Equivalence of both topologies on $\mathbb{M}_1^{(2)}$ 76
4.3.7 Distributions on $\mathbb{M}_1^{(2)}$ 77
4.4 Metric two-level measure spaces 78
4.4.1 Reconstruction of two-level measures 78
4.4.2 M2m-monomials and the two-level Gromov-weak topology . . . 81
4.4.3 The two-level Gromov-Prokhorov metric and its topology . . . 86
4.4.4 Distance distribution and modulus of mass distribution 88

4.4.5 Compact sets 91
4.4.6 Equivalence of both topologies on $\mathbb{M}^{(2)}$ 97
4.4.7 Distributions on $\mathbb{M}^{(2)}$ 98

5 The nested Kingman coalescent measure tree **101**
5.1 The nested Kingman coalescent 101
5.2 The nested Kingman coalescent measure tree 103
5.3 Proof of Theorem 5.5 105
5.3.1 Weak convergence of $(\mathbb{Q}_{M,N})_N$ for fixed M 107
5.3.2 Weak convergence of $(\mathbb{Q}_M)_M$ 109

A Supplementary facts **113**
A.1 Gromov-Hausdorff metric and topology 113
A.2 Semi-continuous functions 115

Bibliography **117**

Index **121**

List of symbols **123**

Chapter 1

Introduction

The aim of this thesis is to provide a state space for stochastic processes which model the evolving genealogy of a population in a hierarchical system with two-levels, for example in a host-parasite system or a population which is divided into colonies. To this end, we introduce metric two-level measure spaces and study the convergence of these objects. The notion of metric two-level measure spaces is a generalization of the notion of metric measure spaces. A metric measure space $[X,r,\mu]$ is defined as the isomorphism class of a triple (X,r,μ) where (X,r) is a Polish metric space (i. e. a complete and separable metric space) and μ is a finite Borel measure on X. Two such triples (X_1,r_1,μ_1) and (X_2,r_2,μ_2) are said to be isomorphic if there is a function $f\colon X_1 \to X_2$ such that $\mu_2 = f_*\mu_1 := \mu_1 \circ f^{-1}$ and f is isometric on the support of μ_1. Roughly speaking, (X_1,r_1,μ_1) and (X_2,r_2,μ_2) are isomorphic if the measures μ_1 and μ_2 have the same structure, even though they are defined on different sets. For this reason, the isomorphism f is sometimes called a measure-preserving isometry.

Metric measure spaces, hereinafter abbreviated as mm spaces, are central objects in probability theory. Every random variable on a Polish metric space (X,r) can be identified with its probability distribution μ on X. Hence, the random variable can be represented by the mm space $[X,r,\mu]$. Therefore, metric measure spaces occur almost everywhere in probability theory, although most of the time only implicitly. On the other hand, notions of convergence of mm spaces and metrics on the set of mm spaces have been of interest in geometric analysis (see for example [Gro99, Stu06, LV09]), optimal transport (see [Vil09] and the references given therein), mathematical biology (see for instance [GPW09, GPW13, DGP12]) and general probability theory (see for example [ALW17, Cro16]). The introduction of these notions of convergence has allowed for the study of mm space-valued stochastic processes. This is of particular interest in mathematical biology, where such processes are used to study the evolution of genealogical (or phylogenetic) trees (cf. [GPW13, DGP12, Glö12, KW17, Guf17]). Typically, the metric space (X,r) represents a population together with its genealogical tree and μ is a sampling measure on the set of individuals.

The basis of this thesis is an article of Greven, Pfaffelhuber and Winter ([GPW09]). The authors equip the set of metric *probability* measure spaces with a Polish topology, making it into a useful state space for tree-valued stochastic processes. A metric probability measure space (mpm space for short) is a metric measure space $[X,r,\mu]$ where μ is a probability measure. It follows from Gromov's reconstruction theorem ([Gro99, Theorem $3\frac{1}{2}.5$]) that an mpm space $[X,r,\mu]$ is uniquely determined by its distance matrix distribution

$$R_*^X \mu^{\otimes\infty} \in \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}\times\mathbb{N}}),$$

where $R_*^X \mu^{\otimes\infty}$ denotes the push-forward of the product measure $\mu^{\otimes\infty}$ under the distance map R^X which maps a sequence $(x_i)_{i\in\mathbb{N}} \in X^{\mathbb{N}}$ to its distance matrix

$$R^X((x_i)_i) := (r(x_i,x_j))_{i,j\in\mathbb{N}}.$$

An mpm space can therefore be identified with its distance matrix distribution. This idea is used in [GPW09] to define a topology on the set $\mathbb{M}_1$ of mpm spaces. The Gromov-weak topology on $\mathbb{M}_1$ is defined as the topology which is induced by all test functions $\Phi\colon \mathbb{M}_1 \to \mathbb{R}$ of the form

$$\Phi([X,r,\mu]) = \int \varphi \circ R^X \,\mathrm{d}\mu^{\otimes m},$$

where $m \in \mathbb{N}_{\geq 2}$ and $\varphi \in \mathcal{C}_b(\mathbb{R}_+^{\mathbb{N}\times\mathbb{N}})$. Convergence with respect to the Gromov-weak topology is the same as weak convergence of the associated distance matrix distributions. That is, a sequence $([X_n,r_n,\mu_n])_n$ in $\mathbb{M}_1$ converges to $[X,r,\mu] \in \mathbb{M}_1$ if and only if $R_*^{X_n}\mu_n^{\otimes\infty} \xrightarrow{w} R_*^X \mu^{\otimes\infty}$.

Greven, Pfaffelhuber and Winter also provide a complete metric on $\mathbb{M}_1$, the so-called Gromov-Prokhorov metric d_{GP}. To obtain the Gromov-Prokhorov distance between two mpm spaces $[X,r,\mu]$ and $[Y,d,\eta]$, one embeds both spaces X and Y into a common Polish space (Z,r_Z) using isometries $i_X\colon X \to Z$ and $i_Y\colon Y \to Z$. This allows to consider the push-forward measures $i_{X*}\mu := \mu \circ i_X^{-1}$ and $i_{Y*}\eta := \mu \circ i_Y^{-1}$. Since these are measures on the same space Z, one can use the Prokhorov metric d_{P} on $\mathcal{M}_1(Z)$ to compute the distance between the push-forward measures. The value of

$$d_{\mathrm{P}}(i_{X*}\mu, i_{Y*}\eta) \tag{E1.1}$$

provides an estimate of the distance between the two mpm spaces $[X,r,\mu]$ and $[Y,d,\eta]$. Of course this value is very dependent on the choice of the space Z and of the embedding functions i_X and i_Y. To obtain a well-defined value, one defines the Gromov-Prokhorov distance $d_{\mathrm{GP}}([X,r,\mu],[Y,d,\eta])$ as the infimum of all values (E1.1), where the infimum is taken over all Polish spaces Z and embeddings of X and Y into Z.

It was proved in [GPW09] that the Gromov-weak topology is in fact metrized by the Gromov-Prokhorov metric. It follows that $\mathbb{M}_1$ equipped with the Gromov-weak topology is Polish and thus a suitable state space for stochastic processes.

The results of [GPW09] have been generalized to metric measure spaces with finite measures ([Glö12]) and extended to marked metric measure spaces (see [DGP11] for probability measures and [KW17] for finite measures). We also seek to extend the results in [GPW09] by replacing the measure μ on the metric space (X,r) with a two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$, i. e. with a finite measure on the set of finite measures on X. This extension is motivated by the study of two-level branching systems in biology, e. g. host-parasite systems, where individuals of the first level are grouped together to form the second level and both levels are subject to branching or resampling mechanisms.

Let us give a few examples of models of such two-level systems that can be found in the mathematical literature:

(a) Dawson, Hochberg and Wu develop a two-level branching process in [DHW90, Wu91]. They consider particles which move in $\mathbb{R}^d$ and are subject to a birth-and-death process. Moreover, the particles are grouped into so-called superparticles, which are subject to another birth-and-death process. The state of this process is given by a two-level measure $\nu \in \mathcal{M}(\mathcal{M}(\mathbb{R}^d))$, i. e. a Borel measure on the set of Borel measures on $\mathbb{R}^d$. The authors also consider the small mass, high density limit of the discrete process. This leads to a two-level diffusion process.

(b) The authors in [BT11] provide a continuous-time two-level branching model for parasites in cells. The parasites live and reproduce (i. e. branch) inside of cells which are subject to cell division. At division of a cell the parasites inside are distributed randomly between the two daughter cells. The model originates from the discrete processes in [Kim97, Ban08] and can be seen as a diffusion limit of these processes.

(c) Another example of a two-level process is given in [MR13]. The authors model the evolution of a population together with different kinds of cells that proliferate inside of the individuals of the population. The individuals follow a birth-and-death mechanism (including mutation and selection) and the cells inside the individuals follow another birth-and-death mechanism.

(d) Dawson studies two-level resampling models in [Daw18]. He considers a random process that models a population which is divided into colonies. The type space X of the individuals is finite and the state of the process is given by a two-level probability measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$, i. e. by a Borel probability measure on the set of Borel probability measures on X. The individuals are subject to mutation, selection, resampling and migration mechanisms while at the same time the colonies are also subject to selection and resampling mechanisms. The author considers the finite case in which the number of colonies and the number of individuals per colony are fixed as well as the limit case when both of these numbers go to infinity. The limit is a two-level diffusion process called the two-level Fleming-Viot process.

We want to extend the theory of metric measure spaces in such a way that the new framework is suitable for modeling the aforementioned examples. The hierarchical two-level structure is of particular interest for us. That is why we study triples $(X, r, \nu)^{(2)}$ where (X, r) is a Polish metric space and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. We call such a triple a metric two-level measure triple (abbreviated as m2m triple).

Let us explain how we intend to use an m2m triple $(X,r,\nu)^{(2)}$ to model the genealogy of a parasite population in a host-parasite system: The metric space (X,r) represents the set X of parasites together with its genealogical tree, which is encoded in the metric r. A single host is represented by a measure $\mu \in \mathcal{M}_f(X)$ (i. e. by a sampling measure on the parasites inside this host) and the two-level measure ν is a sampling measure on the hosts. As such it encodes the distribution of individuals among the host. We can use ν to sample a host which is represented by a measure μ and then use μ to sample individuals within this host. This approach makes sense from the point of view of applications: If one wants to gain information about a parasite population, one would first sample hosts and then sample parasites within these hosts.

In the context of our theory we are only interested in the structure of the (genealogical) trees and not in their labels. To get rid of labels, we consider isomorphism classes of m2m triples. Our definition of isomorphisms focuses on the structure of the measure ν and its "effective support in X". By effective support in X we mean the smallest closed subset $C \subset X$ with $\operatorname{supp}\mu \subset C$ for ν-almost every $\mu \in \mathcal{M}_f(X)$. This set is equal to the support of the first moment measure $\mathfrak{M}_\nu(\,\cdot\,) := \int \mu(\,\cdot\,)\,\mathrm{d}\nu(\mu)$. Roughly speaking, we identify two m2m triples $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ if λ can be mapped into ν with a function that is isometric on the effective supports of the measures. To be precise, $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ are said to be isomorphic if there is a function $f\colon X \to Y$ which is isometric on $\operatorname{supp}\mathfrak{M}_\nu$ (the effective support in X of ν) and measure preserving in the sense that $\nu = f_{**}\lambda$. Here, $f_{**}\lambda$ denotes the two-level push-forward of λ. It is the push-forward of the push-forward operator of f. That is, $f_{**}\lambda := \lambda \circ f_*^{-1}$, where f_* is the function from $\mathcal{M}_f(X)$ to $\mathcal{M}_f(Y)$ that maps a finite Borel measure μ to its push-forward $f_*\mu := \mu \circ f^{-1}$.

The central object of this thesis are the isomorphism classes of m2m triples. The isomorphism class of an m2m triple $(X,r,\nu)^{(2)}$ is called a metric two-level measure space (m2m space for short) and denoted by $[X,r,\nu]^{(2)}$. We define a Polish topology on the set $\mathbb{M}^{(2)}$ of all m2m spaces (i. e. the set of all isomorphism classes), thus providing a possible state space for tree-valued two-level processes. We proceed in similar manner as Greven, Pfaffelhuber and Winter did for mm spaces in [GPW09].

It turns out that generalizing the results of [GPW09] is considerably easier if we restrict ourselves to probability measures only. That is why we first consider the subset $\mathbb{M}_1^{(2)} := \{\,[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)} \mid \nu \in \mathcal{M}_1(\mathcal{M}_1(X))\,\} \subset \mathbb{M}^{(2)}$. Elements of $\mathbb{M}_1^{(2)}$ are called metric two-level *probability* measure spaces (m2pm spaces for short).

We provide a reconstruction theorem for m2pm spaces, which shows that an m2pm space $[X,r,\nu]^{(2)} \in \mathbb{M}_1^{(2)}$ is uniquely determined by its distance array distribution

$$R_*^X \mathfrak{M}_\nu^{\infty,\infty} \in \mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}^4}),$$

where $R_*^X \mathfrak{M}_\nu^{\infty,\infty}$ denotes the push-forward of the infinite mixed moment measure

$$\mathfrak{M}_\nu^{\infty,\infty} := \int \int \bigotimes_{i=1}^{\infty} \mu_i^{\otimes\infty}\,\mathrm{d}\nu^{\otimes\infty}(\mu_1,\mu_2,\dots) \in \mathcal{M}_1(X^{\mathbb{N}\times\mathbb{N}})$$

under the distance map R^X which maps an infinite matrix $(x_{ij})_{ij} \in X^{\mathbb{N}\times\mathbb{N}}$ to its distance array

$$R^X((x_{ij})_{ij}) := \big(r(x_{ij}, x_{kl})\big)_{i,j,k,l\in\mathbb{N}}. \tag{E1.2}$$

Heuristically, the infinite mixed moment measure samples infinitely many measures $(\mu_i)_i \subset \mathcal{M}_1(X)$ with distribution ν and then samples with each of these measures μ_i infinitely many points $(x_{ij})_j \subset X$. Then, the distance array distribution is the distribution of the random distance array in (E1.2).

We introduce the two-level Gromov-weak topology τ'_{2Gw} on $\mathbb{M}_1^{(2)}$ as the topology which is induced by all test functions $\Phi\colon \mathbb{M}_1^{(2)} \to \mathbb{R}$ of the form

$$\Phi([X,r,\nu]^{(2)}) = \int\int \varphi \circ R^X \,\mathrm{d}\boldsymbol{\mu}^{\otimes \boldsymbol{n}} \,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) \tag{E1.3}$$

with $m \in \mathbb{N}$, $\boldsymbol{n} = (n_1,\dots,n_m) \in \mathbb{N}^m$, $\varphi \in \mathcal{C}_b(\mathbb{R}^{|\boldsymbol{n}|\times|\boldsymbol{n}|})$ and $\boldsymbol{\mu}^{\otimes\boldsymbol{n}} := \bigotimes_{i=1}^m \mu_i^{\otimes n_i}$. Loosely speaking, a test function Φ samples finitely many measures with ν, then samples finitely many points with each of these measures and finally evaluates the metric space spanned by all sampled points. Convergence with respect to τ'_{2Gw} is the same as weak convergence of the corresponding distance array distributions. That is, a sequence $([X_n,r_n,\nu_n]^{(2)})_n$ in $\mathbb{M}_1^{(2)}$ converges two-level Gromov-weakly to $[X,r,\nu]^{(2)} \in \mathbb{M}_1^{(2)}$ if and only if

$$R_*^{X_n}\mathfrak{M}_{\nu_n}^{\infty,\infty} \xrightarrow{w} R_*^X\mathfrak{M}_{\nu}^{\infty,\infty}.$$

We show that $(\mathbb{M}_1^{(2)}, \tau'_{2Gw})$ is Polish by providing a complete metric, the so-called two-level Gromov-Prokhorov metric $d_{2\mathrm{GP}}$. This metric is a straight-forward generalization of the (one-level) Gromov-Prokhorov metric. To obtain the distance between two m2pm spaces $[X,r,\nu]^{(2)}$ and $[Y,d,\lambda]^{(2)}$ we embed both metric spaces into a common Polish metric space Z using isometries $i_X\colon X \to Z$ and $i_Y\colon Y \to Z$ and use the Prokhorov metric d_{P} on $\mathcal{M}_1(\mathcal{M}_1(Z))$ to compute the distance

$$d_{\mathrm{P}}(i_{X**}\nu, i_{Y**}\lambda) \tag{E1.4}$$

between the two-level push-forward measures. The two-level Gromov-Prokhorov distance $d_{2\mathrm{GP}}([X,r,\nu]^{(2)}, [Y,d,\lambda]^{(2)})$ is defined as the infimum of all the values (E1.4), where the infimum is taken over all Polish metric spaces Z and all isometric embeddings of X and Y into Z.

Because the two-level Gromov-weak topology is induced by the test functions (E1.3), we can use a theorem due to Le Cam to show that these test functions are also convergence determining for random m2pm spaces. This means that a sequence of distributions $\mathbb{P}_n$ on $\mathbb{M}_1^{(2)}$ converges weakly to a distribution $\mathbb{P}$ if and only if

$$\int \Phi \,\mathrm{d}\mathbb{P}_n \to \int \Phi \,\mathrm{d}\mathbb{P}$$

for every test function of the form (E1.3). This makes these test functions particularly useful for defining generators of Markov processes with values in $\mathbb{M}_1^{(2)}$.

After restricting ourselves to probability measures only, we go over to the more general case. It turns out that for general m2m spaces there is no reconstruction theorem similar to the one for m2pm spaces. That is, we cannot identify m2m spaces with measures on distance arrays. Roughly speaking, the problem is that the two-level measure ν may have $\nu(\{o\})>0$ and that we cannot sample points with the null measure o. For this reason we can only reconstruct $\nu-\nu(\{o\})\delta_o$ (i. e. ν without its possible atom at o) by sampling points from the underlying metric space X.

However, we can still define a point separating class of test functions on $\mathbb{M}^{(2)}$. Because we are now dealing with finite measures, we also have to gain control over the mass of the measures. For this reason we decompose finite measures $\mu\in\mathcal{M}_f(X)$ into their mass $\mathfrak{m}(\mu):=\mu(X)$ and their normalization $\overline{\mu}:=\frac{\mu}{\mathfrak{m}(\mu)}$. To determine an m2m space uniquely we need three different types of test functions. We define $\mathcal{T}^{(2)}$ as the set of all functions $\Phi\colon\mathbb{M}^{(2)}\to\mathbb{R}$ which can be written in one of the following three forms:

$$\Phi([X,r,\nu]^{(2)})=\chi(\mathfrak{m}(\nu)), \tag{TF1}$$

$$\Phi([X,r,\nu]^{(2)})=\chi(\mathfrak{m}(\nu))\int\psi(\mathfrak{m}(\boldsymbol{\mu}))\,\mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu}), \tag{TF2}$$

$$\Phi([X,r,\nu]^{(2)})=\chi(\mathfrak{m}(\nu))\int\psi(\mathfrak{m}(\boldsymbol{\mu}))\int\varphi\circ R(\boldsymbol{x})\,\mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}(\boldsymbol{x})\,\mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu}), \tag{TF3}$$

where $m\in\mathbb{N}$, $\boldsymbol{n}=(n_1,\dots,n_m)\in\mathbb{N}^m$, $\chi\in\mathcal{C}_b(\mathbb{R}_+)$ with $\chi(0)=0$, $\psi\in\mathcal{C}_b(\mathbb{R}_+^m)$ with $\psi(\boldsymbol{a})=0$ whenever any of the components of the vector $\boldsymbol{a}\in\mathbb{R}_+^m$ is 0 and where $\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}:=\bigotimes_{i=1}^m\overline{\mu_i}^{\otimes n_i}$. The test functions of $\mathcal{T}^{(2)}$ determine different parts of the evaluated m2m space $[X,r,\nu]^{(2)}$. Functions of type (TF1) determine the mass $\mathfrak{m}(\nu)$, whereas type (TF2) determines the mass distribution $\mathfrak{m}_*\nu$, i. e. the push-forward of ν under the functions $\mathfrak{m}\colon\mathcal{M}_f(X)\to\mathbb{R}_+$ with $\mathfrak{m}(\mu):=\mu(X)$. The structure of the two-level measure ν and the space (X,r) is determined by test functions of type (TF3).

We define the two-level Gromov-weak topology τ_{2Gw} on $\mathbb{M}^{(2)}$ as the topology which is induced by all test functions of $\mathcal{T}^{(2)}$. Note that we use the same name for both topologies τ_{2Gw} and τ'_{2Gw} because they coincide on $\mathbb{M}_1^{(2)}$. The two-level Gromov-Prokhorov metric $d_{2\mathrm{GP}}$ can easily be extended from $\mathbb{M}_1^{(2)}$ to $\mathbb{M}^{(2)}$. We show that $d_{2\mathrm{GP}}$ is complete and metrizes the two-level Gromov-weak topology. Thus, $(\mathbb{M}^{(2)},\tau_{2Gw})$ is a Polish space.

We also show that the test functions of $\mathcal{T}^{(2)}$ are convergence determining for distributions on $\mathbb{M}^{(2)}$ (in the same way as we did for test functions (E1.3) on $\mathbb{M}_1^{(2)}$). Hence, they are a useful class of functions for defining generators of Markov processes with values in $\mathbb{M}^{(2)}$. In particular, this will allow us to create m2m space-valued analogs of the two-level examples given above in future research articles.

Let us briefly summarize the two main obstacles which arise when we extend the theory from m2pm spaces to m2m spaces:

(a) The space $(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ can be embedded into $\mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}^4})$ using distance array distributions. It follows directly that the two-level Gromov-weak topology on

$\mathbb{M}_1^{(2)}$ is metrizable and thus we can use sequences to show compactness of subsets or continuity of functions on $\mathbb{M}_1^{(2)}$. As we mentioned before, such an embedding is not possible anymore for $\mathbb{M}^{(2)}$. Therefore, it is not a priori clear whether the two-level Gromov-weak topology on $\mathbb{M}^{(2)}$ is metrizable and our proofs (for continuity, compactness, etc.) must not rely on sequences. Instead, we will work with nets. Nets are a generalization of sequences and most of the theorems for metric spaces using sequences (e. g. continuity of functions, closedness of sets, compactness of sets) hold true for general topological spaces when sequences are replaced by nets. After proving that $\mathbb{M}^{(2)}$ equipped with the two-level Gromov-weak topology τ_{2Gw} is in fact Polish, we go back using ordinary sequences.

(b) For characterizing compact sets of m2m spaces it is necessary to work with finite first moment measures. Unfortunately, the first moment measure $\mathfrak{M}_\nu(\,\cdot\,) = \int \mu(\,\cdot\,)\,\mathrm{d}\nu(\mu)$ of a two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ may be infinite. We can overcome this problem by approximating ν sufficiently close by a two-level measure with finite moment measures. This is done by restricting ν to $\mathcal{M}_{\leq K}(X) = \{\, \mu \in \mathcal{M}_f(X) \mid \mu(X) \leq K \,\}$ using a smooth density function f_K. Then, $f_K \cdot \nu$ is an element of $\mathcal{M}_f(\mathcal{M}_{\leq K}(X))$ and has finite moment measures. Moreover, $f_K \cdot \nu$ converges weakly to ν for $K \nearrow \infty$.

Finally, we apply our framework to an example. We define a two-level coalescent process called the nested Kingman coalescent, which is a coalescent model for individuals of different species. The nested Kingman coalescent is a special case of the nested coalescent model, which was recently introduced in [Bla16, BDLSJ18]. We equip the genealogical tree stemming from the nested Kingman coalescent with a two-level probability measure which contains the two-level structure of the coalescent. The measure is defined based on the idea of sampling species and then sampling individuals of this species. The result is a random m2pm space called the nested Kingman coalescent measure tree.

Outline: The rest of this thesis is organized as follows: We start with some preliminaries in chapter 2. We discuss nets in topological spaces and how they can be used to determine a topology. Moreover, we show how a topology can be induced by functions. We also study weak convergence of finite measures together with the corresponding topology. Then we explain how to push-forward one- and two-level measures and study the continuity properties of these push-forward operators.

Chapter 3 is devoted to metric measure spaces with finite measures. We proceed in a similar way as the authors of [GPW09] did for metric probability measure spaces. We define the Gromov-weak topology on the set $\mathbb{M}$ of metric measure spaces by means of test functions. Then we define and study the Gromov-Prokhorov metric on $\mathbb{M}$ and characterize compact subsets of $\mathbb{M}$. Finally, we show that the Gromov-weak topology is metrized by the Gromov-Prokhorov metric and thus Polish.

In chapter 4 we present the main results of this thesis. First, we discuss two-level measures on a fixed Polish metric space and study the interrelationship with their moment measures. Then, we present our results for metric two-level probability measure spaces. We provide a reconstruction theorem for m2pm spaces and a point separating class of test functions on the set $\mathbb{M}_1^{(2)}$ of m2pm spaces. Furthermore, we introduce the two-level Gromov-weak topology on $\mathbb{M}_1^{(2)}$ as the topology which is induced by these test functions. We then define the two-level Gromov-Prokhorov metric on $\mathbb{M}_1^{(2)}$ and show that this metric is complete. Finally, we characterize compact subsets of $\mathbb{M}_1^{(2)}$ and prove that the two-level Gromov-weak topology is metrized by the two-level Gromov-Prokhorov metric and thus Polish. In the last part of this chapter we discuss general metric two-level measure spaces. We provide a point separating class of test functions on the set $\mathbb{M}^{(2)}$ of m2m spaces and study the induced topology, which is also called the two-level Gromov-weak topology. Moreover, we extend the two-level Gromov-Prokhorov metric from $\mathbb{M}_1^{(2)}$ to $\mathbb{M}^{(2)}$. We characterize compact nets and compact subsets in $\mathbb{M}^{(2)}$ and eventually prove that the Gromov-Prokhorov metric on $\mathbb{M}^{(2)}$ metrizes the two-level Gromov-weak topology.

In chapter 5 we construct a random m2pm space called the nested Kingman coalescent measure tree. Its distribution is obtained as the weak limit of distributions of finite m2pm spaces. To prove convergence, we apply the tightness results of the previous chapter.

Additionally, in chapter A of the appendix we present some supplementary facts about the Gromov-Hausdorff topology and semi-continuous functions.

Parts of this thesis (mainly section 4.4 and chapter 5, but also parts of this introduction and the preliminary chapter) have been published in [Mei18].

Chapter 2

Preliminaries

In this chapter we introduce the topological and probabilistic tools which are necessary for the main parts of our thesis.

Note that throughout the whole thesis we use bold letters to denote tuples, vectors or matrices. For instance, $\boldsymbol{x}$ denotes a tuple (or matrix) of points from a set X and $\boldsymbol{\mu}$ denotes a tuple (or matrix) of measures. The symbol $\complement$ is used to denote the complement of a set, i. e. $\complement A = X \setminus A$ for a subset $A \subset X$, and the symbol o is used to denote the null measure, which is 0 on all sets.

Sometimes we call finite measures distributions even though they are not probability measures (e. g. distance distribution or mass distribution). This is due to the fact that we generalize results for probability measures to finite measures and do not want to change the names. Also, a finite measure is only a probability measure multiplied by a constant, hence we can always normalize a finite measure to obtain a distribution.

Outline of this chapter: In section 2.1 we define the notion of nets and explain how nets can be used to determine the topology of general topological spaces. In the subsequent section 2.2 we introduce the initial topology, a way of defining a topology by means of continuous functions. Also, we show how the initial topology can be characterized in terms of converging nets. Section 2.3 is a short note about Polish spaces and their embedding properties. In section 2.4 we provide a short summary about the weak topology, which can be defined as an initial topology, and the Prokhorov metric. The subsequent section 2.5 deals with concepts related to weak convergence, namely separating and convergence determining sets of functions. Finally, in section 2.6 we introduce the one-level and two-level push-forward operators.

2.1 Nets in topological spaces

This section is a short introduction to nets. Nets are a generalization of sequences suitable for describing the topology of general topological spaces (in which the topology is not uniquely determined by its convergent sequences). A more comprehensive survey about nets can be found in the topology book of Kelley ([Kel55]), which is also the origin of most of the results of this section.

We start with a brief motivation for the notion of nets. Let X and Y be topological spaces. Recall that a function $f\colon X\to Y$ is called *continuous* if the preimage $f^{-1}(O)$ is open in X for every open set $O\subset Y$. Moreover, a subset $A\subset X$ is called *compact* if every open cover of A has a finite subcover. Both of these definitions are quite abstract and sometimes hard to verify directly. Therefore, it is desirable to have alternative characterizations of continuity and compactness which are easier to prove. If X is in fact a metric space, it is possible to characterize both properties in terms of convergence of sequences. In this case we have the equivalences

$$\begin{aligned} f \text{ is continuous } &\Leftrightarrow f \text{ is sequentially continuous,}\\ A \text{ is compact } &\Leftrightarrow A \text{ is sequentially compact,}\end{aligned} \tag{E2.1}$$

where f is said to be *sequentially continuous*, if for every convergent sequence $(x_n)_n$ in X with $x_n\to z\in X$ we have $f(x_n)\to f(z)$ and where A is said to be *sequentially compact*, if every sequence in A has a convergent subsequence with limit in A.

However, in general topological spaces the equivalences in (E2.1) fail to be true. This is due to the fact that a topology may be so complex and rich that it cannot be determined by sequences anymore. We illustrate this with a simple example, which shows that different topologies may have the same convergent sequences.

Example 2.1
Let X be an uncountable set and let τ_d denote the discrete topology on X. Furthermore, let τ_c denote the topology on X in which a set $A\subset X$ is open if and only if either $\complement A$ is countable or $A=\emptyset$. The topology τ_c is called the cocountable topology of X. It is easy to see that a sequence $(x_n)_n\subset X$ converges to a point $x\in X$ with respect to the cocountable topology if and only if $x_n=x$ for all but finitely many n. The same is true for convergence with respect to the discrete topology. Hence, τ_c and τ_d admit the same convergent sequences, but obviously these topologies do not coincide.

Note that a sequence $(x_n)_n$ in X is formally defined as a function from the set $\mathbb{N}$ to X and that its domain $\mathbb{N}$ is countable and linearly ordered. The notion of a sequence can be generalized by allowing more general domains than $\mathbb{N}$ (namely directed, partially ordered sets) and these generalized sequences are called nets. It turns out that the notion of nets is broad enough to determine a topology, i. e. two topologies coincide if and only if they have the same convergent nets. Moreover, it is possible to characterize continuity and compactness in general topological spaces in terms of convergence of nets (similar to the equivalences in (E2.1)).

Let us now start to properly define the notion of nets. A non-empty set $\mathcal{A}$ with a partial order $\preceq$ is called *directed* if every pair $\alpha_1,\alpha_2\in\mathcal{A}$ has a common successor $\alpha\in\mathcal{A}$ (i. e. $\alpha_1\preceq\alpha$ and $\alpha_2\preceq\alpha$). A map x from a directed set $(\mathcal{A},\preceq)$ to a topological space (X,τ) is called a *net in* X. Similar to sequences we will denote this map by $(x_\alpha)_{\alpha\in\mathcal{A}}$ or $(x_\alpha)_\alpha$. Observe that every sequence $(x_n)_{n\in\mathbb{N}}$ is a net on the directed set $(\mathbb{N},\leq)$. This shows that nets are a generalization of sequences.

We say that the net $(x_\alpha)_\alpha$ is *eventually* in a set $A\subset X$ if there is an $\alpha_0\in\mathcal{A}$ such that $x_\alpha\in A$ for all $\alpha\succeq\alpha_0$. The set $\{\,x_\alpha \mid \alpha\succeq\alpha_0\,\}$ is also called a *tail of* $(x_\alpha)_\alpha$. We say that

$(x_\alpha)_\alpha$ is *frequently* in A if every $\alpha_0 \in \mathcal{A}$ has a successor $\alpha \succeq \alpha_0$ with $x_\alpha \in A$. Likewise, we say that a net *eventually (resp. frequently)* has a certain property if it eventually (resp. frequently) takes values in the set of elements of X with this property. For example, we say that a real-valued net $(x_\alpha)_\alpha$ is eventually bounded by $M > 0$, if there is an α_0 such that $x_\alpha \leq M$ for all $\alpha \succeq \alpha_0$

A net $(x_\alpha)_\alpha$ is said to *converge* to a point $z \in X$ if for every neighborhood $N \subset X$ of z there is an $\alpha_0 \in \mathcal{A}$ such that $x_\alpha \in N$ for all $\alpha \succeq \alpha_0$ (i. e. the net is eventually in N). We denote this convergence by $x_\alpha \to z$ and say that z is a *limit* of the net $(x_\alpha)_\alpha$. If $(x_\alpha)_\alpha$ has only one limit point, we write $\lim_\alpha x_\alpha = z$ and say that z is *the* limit of $(x_\alpha)_\alpha$. A topological space X is a Hausdorff space if and only if every net in X has at most one limit. Note that in this thesis we will only be concerned with Hausdorff spaces.

The following propositions are taken from [Kel55]. The first statements shows that closed sets can be characterize by convergent nets.

Proposition 2.2
Let A be a subset of a topological space X. The set A is closed if and only if the limits of every convergent net in A are also in A.

Since a topology is determined by its closed sets, this proposition shows that a topology is uniquely determined by its convergent nets. That is, two topologies coincide if and only if they admit the same convergent nets.

As we claimed in the motivational part at the beginning, nets can be used to characterize the continuity of functions.

Proposition 2.3
Let X and Y be topological spaces. A function $f\colon X \to Y$ is continuous if and only if for every convergent net $(x_\alpha)_\alpha$ in X with $x_\alpha \to z \in X$ we have $f(x_\alpha) \to f(z)$.

It is of great interest to know when sequences are sufficient to determine the topology of a space. The most important class of topological spaces in which this is true are the first-countable spaces. Recall that a topological space X is said to be *first-countable* if each point in X has a countable neighborhood base. Every metric space is first-countable since the collection of balls $B(x, 1/n)$ with $n \in \mathbb{N}$ form a countable neighborhood base for each point $x \in X$.

Proposition 2.4
Let X be a first-countable topological space.

(a) Let A be a subset of X. The set A is closed if and only if the limits of every convergent sequence in A are also in A.

(b) A function $f\colon X \to Y$ from X to another topological space Y is continuous if and only if f is sequentially continuous.

The first statement implies that two first-countable (e. g. metrizable) topologies coincide if and only if they have the same convergent sequences. Note that this is true only if *both* topologies are first-countable.

The analog of subsequences for nets are subnets. Let $(\mathcal{B},\preceq_{\mathcal{B}})$ be another directed set and $(y_\beta)_{\beta\in\mathcal{B}}$ be another net in X. We say that $(y_\beta)_\beta$ is a *subnet* of $(x_\alpha)_\alpha$ if there is a function T from $\mathcal{B}$ to $\mathcal{A}$ with $y = x \circ T$ (i. e. $y_\beta = x_{T(\beta)}$ for every β) and if for every $\alpha_0 \in \mathcal{A}$ there is a $\beta_0 \in \mathcal{B}$ such that $T(\beta) \succeq \alpha_0$ for every $\beta \succeq_{\mathcal{B}} \beta_0$. Note that a subsequence is always a subnet of the original sequence, but that a subnet of a sequence is not necessarily a subsequence.

A point $z \in X$ is called a *cluster point* of $(x_\alpha)_\alpha$ if for every neighborhood $N \subset X$ of z and every $\alpha_0 \in \mathcal{A}$ there is an $\alpha_1 \succeq \alpha_0$ with $x_{\alpha_1} \in N$ (i. e. the net is frequently in N). It can be shown that z is a cluster point of $(x_\alpha)_\alpha$ if and only if there is a subnet which converges to z.

As mentioned in the beginning of this section, we can use nets and subnets to characterize compact subsets.

Proposition 2.5
A subset A of a topological space X is compact if and only if every net in A has a convergent subnet with limit in A.

Again, it is desirable to replace nets by sequences. However, for characterizing compactness by sequences we need a stronger condition than first-countability.

Proposition 2.6
A subset A of a metrizable space X is compact if and only if it is sequentially compact.

A frequently used property of sequences is that a sequence converges if and only if every subsequence has a convergent subsequence and every convergent subsequence has the same limit. This property is easily generalized to nets: A net is convergent if and only if every subnet has a convergent subnet and every convergent subnet has the same limit. We call a net *compact* if every subnet has a convergent subnet. Thus, a net is convergent if and only if it is compact and every convergent subnet has the same limit.

It follows easily from Proposition 2.3 that $f(x_\alpha)$ is a compact net if $(x_\alpha)_\alpha$ is a compact net and f is a continuous function. Note that the elements of a compact net do not necessarily form a relatively compact set. We illustrate this fact with an example.

Example 2.7
Let $\mathcal{A} = (-\infty, 0]$ be equipped with the usual ordering '$\leq$' and define $x_\alpha = \alpha$ for every $\alpha \in \mathcal{A}$. The net $(x_\alpha)_\alpha$ converges to 0 and is therefore a compact net. However, the set of its elements $\{\, x_\alpha \mid \alpha \in \mathcal{A} \,\} = (-\infty, 0]$ is not relatively compact.

We can use compact nets to characterize relatively compact subsets. The following lemma is based on [Top74, Lemma 2.3].

Lemma 2.8
Let C be a subset of a regular topological space X. The following are equivalent:

(a) C is relatively compact.

(b) Every net in C has a converging subnet.

(c) Every net in C has a cluster point.

(d) Every net in C is a compact net.

Recall that a topological space X is called regular if every closed set $A \subset X$ and disjoint point $x \notin A$ can be separated by open neighborhoods, i. e. there are disjoint open sets $O_A, O_x \subset X$ with $A \subset O_A$ and $x \in O_x$. It is easy to see that every metric space (X, r) is regular (take $O_x := B(x, \varepsilon)$ and $O_A := B(A, \varepsilon)$ with $\varepsilon := r(x, A)$).

We define the limit superior of a real-valued net $(x_\alpha)_\alpha$ by

$$\limsup_\alpha x_\alpha = \lim_\alpha \sup_{\alpha' \succeq \alpha} x_{\alpha'},$$

i. e. it is the limit of the supremum of the tails of the net. The limit superior is the largest cluster point of the net or ∞ if there is no largest cluster point.

Sometimes we will be concerned with measure-valued nets $(\mu_\alpha)_\alpha \subset \mathcal{M}_f(\mathbb{R})$, where $\mathcal{M}_f(\mathbb{R})$ denotes the set of finite Borel measures on $\mathbb{R}$. We call $(\mu_\alpha)_\alpha$ *tight* if for every $\varepsilon > 0$ there is a compact set $C \subset \mathbb{R}$ such that $\limsup_\alpha \mu_\alpha(\complement C) < \varepsilon$ (i. e. $\mu_\alpha(\complement C) < \varepsilon$ eventually). If $(\mu_\alpha)_\alpha$ is a compact net (e. g. a convergent net), then it is also tight (cf. [Top70, p.44]).

Finally, as a preparation for one of the main proofs of this thesis, we prove a version of Cantor's diagonal scheme for nets.

Lemma 2.9 (Diagonal argument for nets)
Let $(x_\alpha)_\alpha$ be a net in a non-empty set X and let $(y^{(1)}_\beta)_{\beta \in \mathcal{B}_1}, (y^{(2)}_\beta)_{\beta \in \mathcal{B}_2}, \dots$ be subnets of $(x_\alpha)_\alpha$ such that $(y^{(n+1)}_\beta)_{\beta \in \mathcal{B}_{n+1}}$ is also a subnet of $(y^{(n)}_\beta)_{\beta \in \mathcal{B}_n}$ for every $n \in \mathbb{N}$. Then, there is a subnet $(z_\gamma)_{\gamma \in \mathcal{C}}$ of $(x_\alpha)_\alpha$ with the following property: For every $n \in \mathbb{N}$ there is a $\gamma_n \in \mathcal{C}$ such that $(z_\gamma)_\gamma$ restricted to $\{\gamma \in \mathcal{C} \mid \gamma_n \preceq_{\mathcal{C}} \gamma\}$ is a subnet of $(y^{(n)}_\beta)_{\beta \in \mathcal{B}_n}$

Informally speaking, $(z_\gamma)_\gamma$ is a subnet of each $(y^{(n)}_\beta)_{\beta \in \mathcal{B}_n}$ when we ignore the first elements of $(z_\gamma)_\gamma$ and hence a "diagonal subnet".

Proof: Recall that, by the definition of a subnet, for each $n \in \mathbb{N}$ there is a function $T_n \colon \mathcal{B}_n \to \mathcal{A}$ with $y_\beta = x_{T_n(\beta)}$ for every $\beta \in \mathcal{B}_n$. Define $\mathcal{C} := \mathbb{N} \times \bigtimes_{n \in \mathbb{N}} \mathcal{B}_n$ and let $\preceq_{\mathcal{C}}$ denote the product ordering on $\mathcal{C}$. That is, we have $(N, (\beta_n)_n) \preceq_{\mathcal{C}} (N', (\beta'_n)_n)$ if and only if $N \le N'$ and $\beta_n \preceq_n \beta'_n$ for every $n \in \mathbb{N}$, where $\preceq_n$ denotes the directed order on $\mathcal{B}_n$. It is easy to show that $(\mathcal{C}, \preceq_{\mathcal{C}})$ is a directed set. We define $T \colon \mathcal{C} \to \mathcal{A}$ by $T((N, (\beta_n)_n)) := T_N(\beta_N)$. By construction the net $(z_\gamma)_{\gamma \in \mathcal{C}}$ with $z_\gamma := x_{T(\gamma)}$ is a subnet of $(x_\alpha)_\alpha$ and has the desired property. □

2.2 Initial topology

Sometimes it is desirable to define a topology which makes a certain set of functions on X continuous. This is the idea behind the initial topology.

Definition 2.10 (Initial topology)
Let X and I be non-empty sets and for every $i \in I$ let $f_i\colon X \to Y_i$ be a function from X to some topological space (Y_i, τ_i). Let $\mathcal{F}$ denote the set $\{ f_i \mid i \in I \}$. The initial topology on X induced by $\mathcal{F}$ *is defined as the coarsest topology on X for which all functions of $\mathcal{F}$ are continuous.*

Let τ denote the initial topology on X induced by $\mathcal{F}$. The topology τ is uniquely determined by the following property (cf. [vQ79, Satz 3.13]): A function $g\colon Z \to X$ from a topological space (Z, τ_Z) to X is continuous with respect to τ if and only if $f_i \circ g$ is continuous for every $i \in I$.

A typical example of an initial topology is the product topology. Given a set $\{ (Z_i, \tau_i) \mid i \in I \}$ of topological spaces, the product topology on $Z := \prod_{i \in I} Z_i$ can be defined as the initial topology induced by all canonical projections $\pi_i\colon Z \to (Z_i, \tau_i)$ with $i \in I$.

Another example of an initial topology is the weak topology for Borel measures on a Polish space X. This topology can be defined as the initial topology induced by the functions $\mu \to \int f \,\mathrm{d}\mu$ with $f \in \mathcal{C}_b(X)$. See section 2.4 for more details.

Since a function is continuous if and only if the preimages of open sets are open, it is clear that $f_i^{-1}(O_i)$ must be open in (X, τ) for every $i \in I$ and every open subset O_i of (Y, τ_i) and that the set $\{ f_i^{-1}(O_i) \mid i \in I, O_i \in \tau_i \}$ is a topological subbase for τ. This implies that convergence in the initial topology can be characterized by the following property.

Lemma 2.11
A net $(x_\alpha)_\alpha$ in X converges to a point $x \in X$ with respect to the initial topology induced by $\mathcal{F}$ if and only if $f(x_\alpha)$ converges to $f(x)$ for every $f \in \mathcal{F}$.

Assume now that all the spaces (Y_i, τ_i) coincide with some space (Y, τ_Y). Recall that we say that $\mathcal{F}$ *separates points* in X (or that $\mathcal{F}$ is *point separating* in X) if for all $x, y \in X$ with $x \neq y$ there is a function $f \in \mathcal{F}$ with $f(x) \neq f(y)$. In this case we have the following useful lemma.

Lemma 2.12
Let $\mathcal{F}$ be a set of functions from X to a Hausdorff space (Y, τ_Y). If $\mathcal{F}$ separates points in X, then the initial topology induced by $\mathcal{F}$ is also a Hausdorff topology.

Proof: Let x, y be two different points in X. Then there is a $f \in \mathcal{F}$ with $f_i(x) \neq f_i(y)$. Because Y is a Hausdorff space, there are disjoint open environments O_x and O_y of $f(x)$ and $f(y)$, respectively. Hence, $f^{-1}(O_x)$ and $f^{-1}(O_y)$ are disjoint open environments of x and y, respectively. □

2.3 A note on Polish spaces

A topological space is called *Polish* (or a *Polish space*) if it is separable and metrizable with a complete metric. We say that a metric space (X, r) is *Polish* if it is separable and the metric r is complete. Nearly all topological and metric spaces which appear in this thesis will be Polish, though for some of them this property is not apparent and we will put some effort into showing that they are Polish.

Polish spaces have nice hereditary properties. Countable products of Polish spaces are also Polish and so are closed subsets of Polish spaces. Thus, the space $\mathbb{R}^{\mathbb{N}}$ equipped with the product topology and all its closed subsets are Polish. The following proposition from [Eng89, Corollary 4.3.25] shows that these are in fact the only Polish spaces (up to homeomorphisms).

Proposition 2.13
A topological space is Polish if and only if it is homeomorphic to a closed subset of $\mathbb{R}^{\mathbb{N}}$.

Thus, one can think of the set of all closed subspaces of $\mathbb{R}^{\mathbb{N}}$ as being the "set of all Polish spaces", even though the latter set does not exist in the axiomatic set theory of Zermelo and Fraenkel.

In this thesis we will be concerned with sets of Polish metric spaces and to be formally correct we require these spaces to be subsets of $\mathbb{R}^{\mathbb{N}}$. However, in light of Proposition 2.13, this is not a restriction and our theory stays valid for general Polish metric spaces.

2.4 The weak topology and the Prokhorov metric

Let (X,r) be a Polish metric space. By $\mathcal{M}_f(X)$ we denote the set of all finite Borel measures on X equipped with the weak topology. The *weak topology* on $\mathcal{M}_f(X)$ is defined as the initial topology with respect to all functions $\mu \mapsto \int f \,\mathrm{d}\mu$ with $f \in \mathcal{C}_b(X)$, where $\mathcal{C}_b(X)$ denotes the set of bounded and continuous functions from X to $\mathbb{R}$. Therefore, a net $(\mu_\alpha)_\alpha$ of finite Borel measures on X converges to a finite Borel measure μ in the weak topology if and only if

$$\int f \,\mathrm{d}\mu_\alpha \to \int f \,\mathrm{d}\mu$$

for every test function $f \in \mathcal{C}_b(X)$. We say that μ_α *converges weakly* to μ and denote this weak convergence by $\mu_\alpha \xrightarrow{w} \mu$ or $\text{w-lim}_\alpha \mu_\alpha = \mu$.

It is well known that the set $\mathcal{M}_f(X)$ equipped with the weak topology is a Polish space and that the *Prokhorov metric* d_{P} is a complete metric for this topology (cf. for example [Pro56]). The Prokhorov distance $d_{\mathrm{P}}(\mu,\eta)$ between two finite measures $\mu,\eta \in \mathcal{M}_f(X)$ is defined as the infimum over all $\varepsilon > 0$ such that

$$\mu(A) \leq \eta(B(A,\varepsilon)) + \varepsilon \quad \text{and} \quad \eta(A) \leq \mu(B(A,\varepsilon)) + \varepsilon \tag{E2.2}$$

for all closed sets $A \subset X$, where $B(A,\varepsilon) = \bigcup_{a\in A} B(a,\varepsilon)$ and $B(a,\varepsilon)$ is the open ball of radius ε around a. To emphasize that we are using the Prokhorov metric for measures on a specific metric space (X,r), we sometimes write d_{P}^X or $d_{\mathrm{P}}^{(X,r)}$ instead of d_{P}.

The subsequent lemma shows that weak convergence of finite measures is equivalent to weak convergence of the normalized measures and convergence of the masses. To formalize this statement we need to introduce some notation. The *mass* of a finite Borel measure $\mu \in \mathcal{M}_f(X)$ is denoted by

$$\mathfrak{m}(\mu) := \mu(X).$$

For a tuple $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m) \in \mathcal{M}_f(X)^m$ of finite measures we define

$$\mathfrak{m}(\boldsymbol{\mu}) := (\mathfrak{m}(\mu_1), \dots, \mathfrak{m}(\mu_m)).$$

Furthermore, for every $K \geq 0$ we define the sets

$$\mathcal{M}_{\leq K}(X) := \{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \leq K\}$$

and

$$\mathcal{M}_K(X) := \{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) = K\}.$$

In particular, $\mathcal{M}_1(X)$ denotes the set of probability measures on X. Note that for two finite measures $\mu, \eta \in \mathcal{M}_K(X)$

$$\mu(A) \leq \eta(B(A,\varepsilon)) + \varepsilon \text{ for all closed } A \subset X$$

already implies

$$\eta(A) \leq \mu(B(A,\varepsilon)) + \varepsilon \text{ for all closed } A \subset X$$

by [EK86, Lemma 3.1.1]. Thus, for measures with the same mass one only needs to check one of the properties of (E2.2) to compute the Prokhorov distance.

The *normalization* of a measure $\mu \in \mathcal{M}_f(X)$ is defined as

$$\overline{\mu} := \begin{cases} \frac{\mu}{\mathfrak{m}(\mu)} & \mu \neq o \\ o & \mu = o. \end{cases}$$

Here, o denotes the *null measure*, which is 0 on all sets. For a tuple of measures $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m) \in \mathcal{M}_f(X)^m$ we define $\overline{\boldsymbol{\mu}} = (\overline{\mu_1}, \dots, \overline{\mu_m})$.

The following lemma follows directly from the definition of weak convergence and the fact that $\mu = \mathfrak{m}(\mu) \cdot \overline{\mu}$.

Lemma 2.14
Let X be a Polish space and let $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_f(X)$. The following are equivalent:

(a) $\mu_n \xrightarrow{w} \mu$.

(b) *(i)* $\mathfrak{m}(\mu_n) \to \mathfrak{m}(\mu)$ *and*

(ii) $\mu = o$ *or* $\overline{\mu_n} \xrightarrow{w} \overline{\mu}$.

In particular, the map $\mu \to \overline{\mu}$ is everywhere continuous except at o.

Recall that a set $\mathcal{F} \subset \mathcal{M}_f(X)$ is called *tight* if for every $\varepsilon > 0$ there is a compact set $C \subset X$ such that $\mu(\complement C) < \varepsilon$ for every $\mu \in \mathcal{F}$. We say that a single measure $\mu \in \mathcal{M}_f(X)$ is tight if the set $\{\mu\}$ is tight. Finite measures on Polish spaces are always tight. It is well known that for *probability* measures tightness is equivalent to relative compactness. However, for *finite* measures we also need to ensure that the masses of the measures are bounded. This is part of the original theorem of Prokhorov in [Pro56, Theorem 1.12].

Proposition 2.15 (Prokhorov's theorem)
Let X be a Polish space and $\Gamma \subset \mathcal{M}_f(X)$. The set Γ is relatively compact in the weak topology if and only if Γ is tight and the set $\{\mathfrak{m}(\mu) \mid \mu \in \Gamma\}$ is bounded in $\mathbb{R}$.

Observe that $\{\mathfrak{m}(\mu) \mid \mu \in \Gamma\}$ is bounded if and only if Γ is bounded in the Prokhorov metric since

$$|\mathfrak{m}(\mu) - \mathfrak{m}(\eta)| \leq d_{\mathrm{P}}(\mu, \eta) \leq \max(\mathfrak{m}(\mu), \mathfrak{m}(\eta))$$

for all $\mu, \eta \in \mathcal{M}_f(X)$.

2.5 Separating and convergence determining sets

In this section we introduce some concepts related to weak convergence and discuss their relationships with each other. It should be noted that we distinguish between the properties "separating" and "point separating".

Definition 2.16 (Separating and convergence determining sets)
Let $\mathcal{F} \subset B(X)$ be a set of bounded Borel measurable functions on a topological space X. We say that $\mathcal{F}$ is separating *if for any two Borel probability measures $\mathbb{P}, \mathbb{Q}$ on X*

$$\int f \,\mathrm{d}\mathbb{P} = \int f \,\mathrm{d}\mathbb{Q} \quad \forall f \in \mathcal{F}$$

implies that $\mathbb{P} = \mathbb{Q}$.

If moreover $\mathcal{F} \subset \mathcal{C}_b(X)$, then we say that $\mathcal{F}$ is convergence determining *if for Borel probability measures $\mu, \mu_1, \mu_2, \dots$ on X*

$$\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu \quad \forall f \in \mathcal{F}$$

implies weak convergence $\mu_n \xrightarrow{w} \mu$.

Clearly, convergence determining sets are always separating.

Convergence determining sets are useful because we only need to check convergence with respect to a smaller subset of $\mathcal{C}_b(X)$ to establish weak convergence. The following result due to Le Cam provides conditions for a set $\mathcal{F} \subset \mathcal{C}_b(X)$ to be convergence determining. We cite the version from [HJ77, Lemma 4.1].

Proposition 2.17 (Le Cam's theorem)
Let X be a completely regular Hausdorff space (e. g. a metric space) and let $\mathcal{F} \subset \mathcal{C}_b(X)$ be closed under multiplication. If $\mathcal{F}$ induces the topology of X, then $\mathcal{F}$ is convergence determining for Radon probability measures on X.

Recall that a topological space X is called *completely regular* if for any closed set $C \subset X$ and any point $x \in X \setminus C$ there is a continuous function $f\colon X \to \mathbb{R}$ such that $f(x) = 0$ and $f(C) = \{1\}$. In this case we say that f separates C and x. Hence, in a completely regular space we can separate points from closed sets by continuous functions. If (X, r) is a metric space, then the function $f(\cdot) := 1 - \frac{r(\cdot, C)}{r(x, C)}$ is continuous and separates C from x. Therefore, every metric space is completely regular.

By [Wil70, Theorem 14.12] a topological space (X, τ) is completely regular if and only if its topology τ coincides with the initial topology induced by $\mathcal{C}_b(X)$. Therefore, complete regularity is a natural requirement in Proposition 2.17, because otherwise no subset $\mathcal{F} \subset \mathcal{C}_b(X)$ can induce the topology of X.

Another way of proving that a set is convergence determining is to show that it "strongly separates points".

Definition 2.18 (Strongly separating points)
Let X be a topological space and let $\mathcal{F}$ be a set of real-valued functions on X. We say that $\mathcal{F}$ strongly separates points *if for every $x \in X$ and every neighborhood O of x there are finitely many functions $f_1, \dots, f_n$ in $\mathcal{F}$ such that*

$$\inf_{y \notin O} \max_{1 \le i \le n} |f_i(y) - f_i(x)| > 0.$$

Obviously, if X is a Hausdorff space and $\mathcal{F}$ strongly separates points, then $\mathcal{F}$ also separates points in X.

Ethier and Kurtz proved in [EK86, Theorem 4.4.5] that every algebra of functions $\mathcal{F} \subset \mathcal{C}_b(X)$ which strongly separates points is also convergence determining. However, their result is in fact a special case of Proposition 2.17. This follows from the next lemma, which is based on [BK10, Lemma 4].

Lemma 2.19
Let X be a Hausdorff space and let $\mathcal{F} \subset \mathcal{C}(X)$ be a set of continuous functions on X. Then $\mathcal{F}$ strongly separates points if and only if $\mathcal{F}$ induces the topology of X.

In the remainder of this section we provide some corollaries which will be useful later on. The results are well-known and have been proven before with various methods. Here, we deduce them directly from Proposition 2.17

The following statement can also be found (in a more general version) in [EK86, Proposition 3.4.6].

Corollary 2.20
Let $(X_n)_n$ be a sequence of Polish spaces and let $X = \prod_{n=1}^{\infty} X_n$ be equipped with the product topology. The set $\mathcal{F} \subset \mathcal{C}_b(X)$ defined by

$$\mathcal{F} := \Big\{ \prod_{n=1}^{N} f_n \;\Big|\; N \in \mathbb{N}, f_n \in \mathcal{C}_b(X_n) \Big\}$$

is convergence determining for $\mathcal{M}_1(X)$.

Proof: Each $\mathcal{C}_b(X_n)$ induces the topology on X_n since X_n is completely regular as a metrizable space. Hence, the set $\mathcal{F}$ induces the product topology on X (i. e. the topology of pointwise convergence). Moreover, $\mathcal{F}$ is closed under multiplication and X is metrizable as a countable product of metrizable spaces. We can now apply Proposition 2.17 to obtain the result. □

Corollary 2.21 (Fdd convergence implies weak convergence)
Let $(X_n)_n$ *be a sequence of Polish spaces and let* $X = \prod_{n=1}^{\infty} X_n$ *be equipped with the product topology. Moreover, let* $\mu, \mu_1, \mu_2, \dots$ *be probability measures on* X. *Then,* μ_n *converges weakly to* μ *if and only if the finite dimensional distributions of* μ_n *converge weakly to those of* μ.

Proof: The set $\mathcal{F}$ from Corollary 2.20 is convergence determining for $\mathcal{M}_1(X)$. The functions of $\mathcal{F}$ depend only on finitely many coordinates and thus weak convergence of the finite dimensional distributions yields

$$\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$$

for every $f \in \mathcal{F}$. □

The following statement can also be found in [Bil68, Theorem 3.2].

Corollary 2.22
Let $m \in \mathbb{N}$ *and* $X_1, \dots, X_m$ *be Polish spaces. Let* $\mu^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots$ *be probability measures on* X_i *for each* $i \in \{1, \dots, m\}$. *The following are equivalent:*

(a) $\mu_n^{(i)} \xrightarrow{w} \mu^{(i)}$ *for each* $i \in \{1, \dots, m\}$.

(b) $\bigotimes_{i=1}^m \mu_n^{(i)} \xrightarrow{w} \bigotimes_{i=1}^m \mu^{(i)}$ *as probability measures on the product space* $\prod_{i=1}^m X_i$.

In particular, if all spaces coincide to some Polish space X, *then a sequence* $(\mu_n)_n \subset \mathcal{M}_1(X)$ *converges weakly to* $\mu \in \mathcal{M}_1(X)$ *if and only if* $\mu_n^{\otimes m}$ *converges weakly to* $\mu^{\otimes m}$.

Proof: Follows directly from Corollary 2.20. □

2.6 Push-forward operators

Let μ be a measure on a metric space (X, r) and let g be a Borel measurable function from X to another metric space (Y, d). It is well known that the push-forward $\mu \circ g^{-1}$ is a measure on the space Y. This construction allows us to compare measures which are defined on different metric spaces by pushing them forward to a common space using "nice" functions. In this thesis we will mainly deal with two-level measures $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$, i. e. finite measures on the set of finite measures on X. To compare two-level measures with different underlying spaces, say $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ and $\lambda \in$

$\mathcal{M}_f(\mathcal{M}_f(Y))$, we need a way to push-forward ν to a two-level measure on Y using a function $g\colon X \to Y$ between the underlying spaces. To achieve this we define a push-forward operator $g_*\colon \mathcal{M}_f(X) \to \mathcal{M}_f(Y)$ by $g_*\mu = \mu \circ g^{-1}$ and then use the push-forward of this operator, i. e. $\nu \circ (g_*)^{-1} \in \mathcal{M}_f(\mathcal{M}_f(Y))$.

Definition 2.23 (One-level and two-level push-forward)
Let (X,r) and (Y,d) be Polish metric spaces and g be a Borel measurable function from X to Y. We define the (one-level) push-forward operator of g *by*

$$\begin{aligned} g_*\colon \mathcal{M}_f(X) &\to \mathcal{M}_f(Y) \\ \mu &\mapsto g_*\mu = \mu \circ g^{-1} \end{aligned} \tag{E2.3}$$

and the two-level push-forward operator g_{**} of g *by*

$$\begin{aligned} g_{**}\colon \mathcal{M}_f(\mathcal{M}_f(X)) &\to \mathcal{M}_f(\mathcal{M}_f(Y)) \\ \nu &\mapsto g_{**}\nu := \nu \circ (g_*)^{-1}. \end{aligned} \tag{E2.4}$$

In this thesis the function g will usually be an isometry between X and Y. In this case the "structure" of the push-forward measure $g_*\mu$ is the same as of the original measure $\mu \in \mathcal{M}_f(X)$. The same is true for the two-level push-forward measure $g_{**}\nu$ with $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.

Let $\varphi\colon Y \to \mathbb{R}$ be measurable and let $\mu \in \mathcal{M}_f(X)$ and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. For the push-forward measures $g_*\mu$ and $g_{**}\nu$ we have the following obvious transformation formulas (assuming that the integrals exist):

$$\int \varphi \,\mathrm{d}(g_*\mu) = \int \varphi \circ g \,\mathrm{d}\mu \tag{E2.5}$$

and

$$\begin{aligned} \int_{\mathcal{M}_f(Y)} \int_Y \varphi \,\mathrm{d}\mu \,\mathrm{d}(g_{**}\nu)(\mu) &= \int_{\mathcal{M}_f(X)} \int_Y \varphi \,\mathrm{d}(g_*\mu) \,\mathrm{d}\nu(\mu) \\ &= \int_{\mathcal{M}_f(X)} \int_X \varphi \circ g \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu). \end{aligned} \tag{E2.6}$$

The following lemma summarizes some useful properties of the one-level and two-level push-forward operator.

Lemma 2.24 (Properties of push-forward operators)
Let $(X,d_X),(Y,d_Y)$ be Polish metric spaces, let $h,g,g_1,g_2,\dots$ be measurable functions from X to Y. Then we have:

(a) If g is continuous, then g_* and g_{**} defined as in (E2.3) *and* (E2.4)*, respectively, are continuous.*

(b) If g_n converges pointwise to g, then g_{n*} and g_{n**} converge pointwise to g_* and g_{**}, respectively. That is, $g_{n*}\mu$ converges weakly to $g_*\mu$ and $g_{n**}\nu$ converges weakly to $g_{**}\nu$ for all $\mu \in \mathcal{M}_f(X)$ and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.

(c) *Let* $\mu \in \mathcal{M}_f(X)$ *and* $\varepsilon > 0$. *Define* $M_\varepsilon := \{ x \in X \mid d_Y(g(x), h(x)) < \varepsilon \}$ *and* $\delta := \mu(\complement M_\varepsilon)$. *Then we have*

$$d_\mathrm{P}(g_*\mu, h_*\mu) \le \max(\varepsilon, \delta).$$

(d) *If* $(Z, \mathcal{Z})$ *is another measurable space and if* $f : Y \to Z$ *is measurable, then*

$$(g \circ f)_* = g_* \circ f_* \tag{E2.7}$$

and

$$(g \circ f)_{**} = g_{**} \circ f_{**} \tag{E2.8}$$

i. e. we have $(g \circ f)_*\mu = g_*(f_*\mu)$ *and* $(g \circ f)_{**}\nu = g_{**}(f_{**}\nu)$ *for all* $\mu \in \mathcal{M}_f(X)$ *and* $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.

Proof: The proofs of (a) and (b) are straightforward with the transformation formulas (E2.5) and (E2.6). To show (c), let $A \subset X$ be a closed set and let $m := \max(\varepsilon, \delta)$. Then we have

$$\begin{aligned} g_*\mu(A) &= \mu(g^{-1}(A)) \\ &= \mu(g^{-1}(A) \cap M_\varepsilon) + \mu(g^{-1}(A) \cap \complement M_\varepsilon) \\ &\le \mu(h^{-1}(B(A, \varepsilon))) + \delta \\ &\le h_*\mu(B(A, m)) + m \end{aligned}$$

and in the same way we can show that

$$h_*\mu(A) \le g_*\mu(B(A, m)) + m.$$

This holds for every closed set $A \subset X$. From the definition of the Prokhorov metric d_P we see that

$$d_\mathrm{P}(g_*\mu, h_*\mu) \le m = \max(\varepsilon, \delta).$$

To prove assertion (d), let $\mu \in \mathcal{M}_f(X)$ and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. By using the definition of the push-forward operator we get

$$(g \circ f)_*\mu = \mu \circ (g \circ f)^{-1} = \mu \circ f^{-1} \circ g^{-1} = g_*(\mu \circ f^{-1}) = g_*(f_*\mu),$$

which yields (E2.7). From this we obtain

$$\begin{aligned} (g \circ f)_{**}\nu &= \nu \circ ((g \circ f)_*)^{-1} = \nu \circ (g_* \circ f_*)^{-1} \\ &= \nu \circ f_*^{-1} \circ g_*^{-1} = g_{**}(\nu \circ f_*^{-1}) \\ &= g_{**}(f_{**}\nu), \end{aligned}$$

which yields (E2.8). □

Chapter 3

Metric measure spaces

This chapter is a summary of metric measure spaces with *finite* measures. The basis of this chapter is an article by Greven, Pfaffelhuber and Winter ([GPW09]) in which the authors study metric measure spaces equipped with *probability* measures. Parts of the results of [GPW09] have already been generalized to finite measures by Glöde in [Glö12]. Also note that Greven, Pfaffelhuber and Winter mention finite measures in [GPW09, Remarks 3.3 and 7.2] in the context of their compactness and tightness theorems.

The results of this chapter are not new, yet to the best of our knowledge they have not been comprehensively treated in a single publication. We will provide full proofs (with few exceptions) because we want to present the most important tools and techniques before we generalize the results further in chapter 4.

Outline of this chapter: In section 3.1 we provide the basic definition of metric measure spaces (mm spaces) and prove the reconstruction theorem, which shows how mm spaces can be reconstructed from the distances between randomly sampled points. In section 3.2 we define mm-monomials, a class of test functions on the set $\mathbb{M}$ of mm spaces, and show that they separate points in $\mathbb{M}$. Then we study the Gromov-weak topology on $\mathbb{M}$, which is defined as the topology induced by the mm-monomials. The subsequent section 3.3 deals with the Gromov-Prokhorov metric d_{GP} on $\mathbb{M}$. We prove embedding properties for convergent sequences and show that $(\mathbb{M}, d_{\mathrm{GP}})$ is a Polish metric space. In section 3.4 we define some invariant characteristics of mm spaces, namely the distance distribution and the modulus of mass distribution, and study their continuity properties with respect to the Gromov-weak topology and the Gromov-Prokhorov metric. In section 3.5 we characterize compact sets with respect to the metric d_{GP}. One of the compactness criteria is given in terms of the distance distribution and the modulus of mass distribution, thus linking compactness in the Gromov-Prokhorov metric with the Gromov-weak topology. With this link we are able to prove in section 3.6 that the Gromov-weak topology is in fact metrized by the Gromov-Prokhorov metric d_{GP}. Finally, in section 3.7 we deal with distributions of random metric measure spaces. We provide tightness results and show that the mm-monomials are convergence determining.

3.1 The reconstruction theorem for mm spaces

In this section we define the notion of metric measure spaces and show how a metric measure space $[X, r, \mu]$ can be reconstructed from its mass $\mathfrak{m}(\mu)$ and from the distances between random points sampled by the normalization $\overline{\mu}$ (recall the definition of $\mathfrak{m}(\mu)$ and $\overline{\mu}$ from section 2.4).

While there are various definitions of metric measure spaces to be found in the mathematical literature, all of these definitions have in common that they denote isomorphism classes of triples (X, r, μ) where (X, r) is a Polish metric space and μ a Borel measure on X. The definitions differ mostly in the measure μ. Some authors assume that μ is a probability measure (e. g. [GPW09]), some that μ is a finite measure (e. g. [Glö12]) and others allow even infinite measures (e. g. [Gro99, Vil09]). In this thesis we restrict the definition to finite measures.

Definition 3.1 (mm space and mm-isomorphism)

(a) *A triple* (X, r, μ) *is called a* metric measure triple (mm triple) *if* $X \subset \mathbb{R}^{\mathbb{N}}$ *is non-empty,* (X, r) *is a Polish metric space and* $\mu \in \mathcal{M}_f(X)$*.*

If additionally μ *is a probability measure, then we call* (X, r, μ) *a* metric probability measure triple (mpm triple).

(b) *Two mm triples* (X, r, μ) *and* (Y, d, η) *are called* mm-isomorphic *(or* measure-preserving isometric*) if there exists a measurable function* $f: X \to Y$ *such that* $\eta = f_*\mu$ *and* f *is isometric on the set* $\operatorname{supp}\mu$ *(but not necessarily on the whole space* X*). The function* f *is called an* mm-isomorphism*. To denote that both spaces are mm-isomorphic, we write* $(X, r, \mu) \cong (Y, d, \eta)$ *or* $(X, r, \mu) \cong_f (Y, d, \eta)$ *if we want to emphasize that* f *is an mm-isomorphism.*

(c) *The relation of being mm-isomorphic is an equivalence relation on the set of metric measure triples. The equivalence class of a metric measure triple* (X, r, μ) *is called a* metric measure space (mm space) *and denoted by* $[X, r, \mu]$*. The set of all mm spaces (i. e. the set of equivalence classes of metric measure triples) is denoted by* $\mathbb{M}$*.*

If (X, r, μ) *is a metric probability measure triple, we call its equivalence class* $[X, r, \mu]$ *a* metric probability measure space (mpm space)*. The set of all mpm spaces (i. e. the set of equivalence classes of metric probability measure triple) is denoted by* $\mathbb{M}_1$*.*

Generic elements of $\mathbb{M}$ will be denoted by $\mathcal{X} = [X, r, \mu]$, $\mathcal{X}_n = [X_n, r_n, \mu_n]$ or $\mathcal{Y} = [Y, d, \eta]$.

Remarks 3.2

(a) Recall from section 2.3 that every Polish space is homeomorphic to a subset of $\mathbb{R}^{\mathbb{N}}$. Thus the assumption $X \subset \mathbb{R}^{\mathbb{N}}$ in Definition 3.1 is not a restriction and the theory of metric measure spaces stays valid even for arbitrary Polish metric spaces.

(b) An mm space is mm-isomorphic to the support of its measure in the following sense: Let $[X,r,\mu]$ be an arbitrary mm space with $S := \operatorname{supp}\mu \neq \emptyset$. Then $[X,r,\mu] = [S,r_S,\mu_S]$ where r_S denotes the restriction of the metric r to $S \times S$ and μ_S denotes the restriction of μ to S. Therefore, we may assume without loss of generality that $X = \operatorname{supp}\mu$ and we will do so in many of the proofs of this chapter.

(c) In the definition of mm triples/spaces we do not allow the underlying metric space X to be empty because this would break the symmetry in the definition of an mm-isomorphism. This is due to the fact that there is no function from a non-empty set to an empty set. To be precise, let (X,r) be some non-empty Polish metric space. One would expect the mm triples $(\emptyset,0,o)$ and (X,r,o) to be mm-isomorphic, since both measure have the same (empty) structure. Indeed, the empty function $f\colon \emptyset \to X$ is measure-preserving and isometric on the empty set $\operatorname{supp} o$. However, there is no function from X to $\emptyset$. Thus, there is a measure-preserving isometry from $(\emptyset,0,o)$ to (X,r,o), but none from (X,r,o) to $(\emptyset,0,o)$.

Item (b) of the previous remark shows that the definition of metric measure spaces puts the emphasis on the measure μ and on its support rather than on the full space (X,r). The structure of the space outside of $\operatorname{supp}\mu$ is not important as long as (X,r) is Polish.

The remainder of this section is devoted to the reconstruction theorem, which shows how mm spaces can be reconstructed from the distances between randomly sampled points. We start with the Glivenko-Cantelli theorem, which states how a probability measure can be reconstructed from an i. i. d. sample.

For a sequence $\boldsymbol{x} = (x_i)_i$ in a metric space X and $n \in \mathbb{N}$ we define the *n-th empirical distribution* $\Xi_n(\boldsymbol{x}) \in \mathcal{M}_1(X)$ by

$$\Xi_n(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

and the *(infinite) empirical distribution* $\Xi_\infty(\boldsymbol{x}) \in \mathcal{M}_1(X) \cup \{\, o \,\}$ by

$$\Xi_\infty(\boldsymbol{x}) := \begin{cases} \text{w-lim}_n \, \Xi_n(\boldsymbol{x}), & \text{if the weak limit exists} \\ o, & \text{else.} \end{cases}$$

Recall that o denotes the null measure, which is 0 on all sets. A sequence $\boldsymbol{x} = (x_i)_i$ is said to be *equidistributed* with respect to a probability measure $\mu \in \mathcal{M}_1(X)$, if the empirical measures $\Xi_n(\boldsymbol{x})$ converge weakly to μ. The Glivenko-Cantelli theorem shows that typical i. i. d. samples are equidistributed with respect to their sampling measure. The following version of the theorem is due to Dudley (see [Dud02, Theorem 11.4.1]).

Proposition 3.3 (Glivenko-Cantelli theorem)
Let μ be a probability measure on a separable metric space X and let $\boldsymbol{x} = (x_i)_i$ be an i. i. d. sequence sampled by μ. Then $\Xi_n(\boldsymbol{x})$ almost surely converges weakly to μ.

An immediate consequence of the Glivenko-Cantelli theorem is that i. i. d. sequences must be dense in the support of the sampling measure.

Corollary 3.4 (i. i. d. sequences are dense)
Let μ be a probability measure on a separable metric space X and let $(x_i)_i$ be an i. i. d. sequence sampled by μ. Then $(x_i)_i$ is almost surely dense in $\operatorname{supp}\mu$.

Proof: Assume that $\boldsymbol{x} = (x_i)_i$ is not dense in $\operatorname{supp}\mu$. Then there are $y \in \operatorname{supp}\mu$ and $\varepsilon > 0$ such that $B(y,\varepsilon)$ contains none of the points $(x_i)_i$. Hence we have

$$\liminf_{n\to\infty} \Xi_n(\boldsymbol{x})(B(y,\varepsilon)) = 0 < \mu(B(y,\varepsilon)).$$

It follows from the Portmanteau theorem that $\Xi_n(\boldsymbol{x})$ does not converge weakly to μ and the probability of this is event is 0 by the Glivenko-Cantelli theorem. □

By means of the Glivenko-Cantelli theorem we can reconstruct a metric probability measure space $[X,r,\mu] \in \mathbb{M}_1$ from the distances between points sampled by μ. Imagine that $(x_i)_i$ is a realization of an infinite i. i. d. sequence sampled by μ and that the only thing we know about $[X,r,\mu]$ are the mutual distances $(r(x_i,x_j))_{i,j\in\mathbb{N}}$ between these points. We can now define a metric d' on $\mathbb{N}$ by setting $d'(i,j) := r(x_i,x_j)$. Since the sequence $(x_i)_i$ is typically dense in the support of μ, the completion (Y,d) of the metric space $(\mathbb{N},d')$ is isometric to (X,r). By the Glivenko-Cantelli theorem we can reconstruct μ from $(x_i)_i$ by taking weak limits of the empirical distributions. Thus, the weak limit $\eta := \text{w-lim}\,\frac{1}{n}\sum_{i=1}^{n}\delta_i$ exists and has the same structure as μ (but on another isometric space). It follows that the mpm triples (Y,d,η) and (X,r,μ) are mm-isomorphic.

This shows that the infinite distance matrix $(r(x_i,x_j))_{i,j\in\mathbb{N}}$ carries enough information to determine the mpm space $[X,r,\mu]$. However, the matrix $(r(x_i,x_j))_{i,j\in\mathbb{N}}$ is symmetric and its diagonal values are all 0. The relevant information is therefore contained in the upper half of the matrix. We introduce such upper half distance matrices in the next definition.

Definition 3.5 (Sets of distance matrices)
For every integer $m \in \mathbb{N}_{\geq 2}$ we define the set

$$\mathbb{D}_m := \left\{ (r_{ij})_{1\leq i<j\leq m} \in \mathbb{R}_+^{\binom{m}{2}} \;\middle|\; r_{ij} \leq r_{ik} + r_{kj} \;\; \forall i,j,k \in [m] \right\},$$

where $[m]$ denotes the set $\{1,\dots,m\}$. Moreover, we define

$$\mathbb{D}_{\mathbb{N}} := \left\{ (r_{ij})_{1\leq i<j} \in \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \;\middle|\; r_{ij} \leq r_{ik} + r_{kj} \;\; \forall i,j,k \in \mathbb{N} \right\}.$$

Elements of these sets are called distance matrices. *We equip these sets with the subspace topology inherited from the product topologies on $\mathbb{R}_+^{\binom{m}{2}}$ and $\mathbb{R}_+^{\binom{\mathbb{N}}{2}}$, respectively.*

Sometimes we tacitly regard the upper half matrices of $\mathbb{D}_m$ and $\mathbb{D}_{\mathbb{N}}$ as full matrices and use the entries r_{ii} or r_{ji} with $i<j$. It should be clear that in this case we assume $r_{ii}=0$ and $r_{ji}=r_{ij}$.

To get the distance matrix from a sample of points we introduce the so-called distance maps.

Definition 3.6 (Distance maps)
For a metric space (X,r) and $m\in\mathbb{N}_{\geq 2}$ we define the following distance maps*:*

$$\begin{aligned} R^X_m\colon X^m &\to \mathbb{D}_m \\ \boldsymbol{x} &\mapsto (r(x_i,x_j))_{1\leq i<j\leq m} \end{aligned}$$

and

$$\begin{aligned} R^X_{\mathbb{N}}\colon X^{\mathbb{N}} &\to \mathbb{D}_{\mathbb{N}} \\ \boldsymbol{x} &\mapsto (r(x_i,x_j))_{1\leq i<j}. \end{aligned}$$

For convenience we often suppress the super- and subscript in the distance maps above and simply write R or R^X instead of R^X_m and $R^X_{\mathbb{N}}$ when the space and dimension are clear from the context.

We are now able to formalize the reconstruction approach from above. The proof relies on an idea due to Vershik, which can be found in [Gro99, section $3\frac{1}{2}$.7]. Note that in order to reconstruct a metric measure space $[X,r,\mu]\in\mathbb{M}$ with *finite* measure μ we have to decompose μ into its mass $\mathfrak{m}(\mu)$ and its normalization $\overline{\mu}$.

Proposition 3.7 (Reconstruction theorem for mm spaces)
Let $[X,r,\mu]$ and $[Y,d,\eta]$ be metric measure spaces. The following are equivalent:

(a) $[X,r,\mu]=[Y,d,\eta]$ *(i. e. (X,r,μ) and (Y,d,η) are mm-isomorphic),*

(b) $\mathfrak{m}(\mu)\cdot R^X_*\overline{\mu}^{\otimes n}=\mathfrak{m}(\eta)\cdot R^Y_*\overline{\eta}^{\otimes n}$ *for every $n\in\mathbb{N}_{\geq 2}$,*

(c) $\mathfrak{m}(\mu)\cdot R^X_*\overline{\mu}^{\otimes\infty}=\mathfrak{m}(\eta)\cdot R^Y_*\overline{\eta}^{\otimes\infty}$.

Proof: It follows from all three assertions that $\mathfrak{m}(\mu)$ and $\mathfrak{m}(\eta)$ must be equal. Moreover, the proposition is obviously true for $\mu=\eta=o$. Therefore, we may assume without loss of generality that μ and η are probability measures with $X=\operatorname{supp}\mu$ and $Y=\operatorname{supp}\eta$ (cf. Remark 3.2 for the latter assumption).

(a) $\Rightarrow$ (b): Let $f\colon X\to Y$ be an isometry with $\eta=f_*\mu$. For each $n\in\mathbb{N}_{\geq 2}$ define

$$\begin{aligned} f_n\colon X^n &\to Y^n \\ (x_1,\ldots,x_n) &\mapsto (f(x_1),\ldots,f(x_n)). \end{aligned} \tag{E3.1}$$

Because f is isometric we have $R^Y\circ f_n=R^X$. Together with (E2.7) this yields

$$R^Y_*\eta^{\otimes n}=R^Y_*(f_*\mu)^{\otimes n}=R^Y_*(f_{n*}(\mu^{\otimes n}))=(R^Y\circ f_n)_*\mu^{\otimes n}=R^X_*\mu^{\otimes n}.$$

(b) ⇒ (c): Follows by the Kolmogorov extension theorem since $R_*^X\mu^{\otimes\infty}$ and $R_*^Y\eta^{\otimes\infty}$ are the projective limits of $(R_*^X\mu^{\otimes n})_n$ and $(R_*^Y\eta^{\otimes n})_n$, respectively.

(c) ⇒ (a): Let $\boldsymbol{\xi}=(\xi_i)_i\sim\mu^{\otimes\infty}$ be a random i. i. d. sequence in X. The Glivenko-Cantelli theorem (Proposition 3.3) states that $\Xi_m(\boldsymbol{\xi})$ almost surely converges to μ. Moreover, the sequence $\boldsymbol{\xi}$ is almost surely dense in $\operatorname{supp}\mu=X$ by Corollary 3.4. The analog statements are true for a random sequence in Y with law $\eta^{\otimes\infty}$. Together with $R_*^X\mu^{\otimes\infty}=R_*^Y\eta^{\otimes\infty}$ this implies that there are dense sequences $\boldsymbol{x}=(x_i)_i\subset X$ and $\boldsymbol{y}=(y_i)_i\subset Y$ which satisfy

$$r(x_i,x_j)=d(y_i,y_j) \text{ for all } i,j\in\mathbb{N} \tag{E3.2}$$

and

$$\begin{aligned}\mu&=\operatorname*{w-lim}_m\Xi_m(\boldsymbol{x}),\\ \eta&=\operatorname*{w-lim}_m\Xi_m(\boldsymbol{y}).\end{aligned}$$

Define a function f' by $f'(x_i)=y_i$ for all $i\in\mathbb{N}$. Then f' is isometric and can be extended to an isometric function $f\colon X\to Y$. The isometry properties (E3.2) yield

$$\Xi_n(\boldsymbol{y})=f_*\Xi_n(\boldsymbol{x})$$

for each $n\in\mathbb{N}$. Recall that the push-forward operator f_* is continuous with respect to weak convergence (cf. Lemma 2.24). Hence, we obtain

$$\eta=f_*\mu$$

by letting n go to infinity. □

The objects that we use in the previous proposition to determine a metric measure space are called distance matrix distribution.

Definition 3.8
Let $[X,r,\mu]$ be a metric measure space.

- *(a) For each $n\in\mathbb{N}_{\geq 2}$ the measure $\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes n}$ is called the n*-point distance matrix distribution of $[X,r,\mu]$.
- *(b) The measure $\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty}$ is called the* distance matrix distribution of $[X,r,\mu]$.

Proposition 3.7 states that a metric measure space is uniquely determined by its distance matrix distribution. This suggests to define a topology on $\mathbb{M}$ by means of weak convergence of the distance matrix distributions, i. e. a topology in which $[X_n,r_n,\mu_n]\to[X,r,\mu]$ if and only if

$$\mathfrak{m}(\mu_n)\cdot R_*\overline{\mu_n}^{\otimes\infty}\xrightarrow{w}\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty}$$

or equivalently (cf. Corollary 2.21) if and only if

$$\mathfrak{m}(\mu_n)\cdot R_*\overline{\mu_n}^{\otimes m}\xrightarrow{w}\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes m}$$

for each $m\in\mathbb{N}_{\geq 2}$. We formalize this approach in the next section.

3.2 Mm-monomials and the Gromov-weak topology

In this section we define a class of test functions on $\mathbb{M}$ and study the induced topology. The test functions are designed in such a way that the induced topology formalizes the weak convergence of distance matrix distributions. Moreover, it turns out that this topology also generalizes weak convergence to measures which are defined on different metric spaces.

Definition 3.9 (Mm-monomials)
A function $\Phi\colon \mathbb{M}\to\mathbb{R}$ *is called an* mm-monomial *if it has the form*

$$\Phi([X,r,\mu]) = \psi(\mathfrak{m}(\mu)) \int \varphi\circ R \,\mathrm{d}\overline{\mu}^{\otimes n}, \tag{E3.3}$$

where $n\in\mathbb{N}_{\geq 2}$, $\varphi\in\mathcal{C}_b(\mathbb{D}_n)$ *and* $\psi\in\mathcal{C}_b(\mathbb{R}_+)$ *with* $\psi(0)=0$. *The* degree *of* Φ *is defined as the smallest integer* $n\in\mathbb{N}_{\geq 2}$ *such that* Φ *can be written in the form* (E3.3).

The set of all mm-monomials will be denoted by $\mathcal{T}^{(1)}$.

Roughly speaking, we split μ into its mass $\mathfrak{m}(\mu)$ and its normalization $\overline{\mu}$. Then we sample finitely many points $x_1,\ldots,x_n$ with $\overline{\mu}$ and evaluate the distance between these points. Since the metric space spanned by the points $x_1,\ldots,x_n$ is determined by the mutual distances (up to isometry), this is a way of looking at the structure of sampled finite subspaces of X.

Lemma 3.10
Every mm-monomial $\Phi\in\mathcal{T}^{(1)}$ *is well-defined, i. e. we have* $\Phi([X,r,\mu])=\Phi([Y,d,\eta])$ *if the two mm triples* (X,r,μ) *and* (Y,d,η) *are mm-isomorphic.*

Proof: Let (X,r,μ) and (Y,d,η) be two mm-isomorphic mm triples and let $f\colon X\to Y$ be isometric on $\operatorname{supp}\mu$ with $\eta = f_*\mu$. For every $n\in\mathbb{N}_{\geq 2}$ define f_n as in (E3.1). Moreover, let $\Phi\in\mathcal{T}^{(1)}$ be as in (E3.3). Then we obtain

$$\begin{aligned}
\Phi([Y,d,\eta]) &= \psi(\mathfrak{m}(\eta)) \int \varphi\circ R^Y \,\mathrm{d}\overline{\eta}^{\otimes n}\\
&= \psi(\mathfrak{m}(f_*\mu)) \int \varphi\circ R^Y \,\mathrm{d}(f_*\overline{\mu})^{\otimes n}\\
&= \psi(\mathfrak{m}(\mu)) \int \varphi\circ (R^Y\circ f_n) \,\mathrm{d}\overline{\mu}^{\otimes n}\\
&= \psi(\mathfrak{m}(\mu)) \int \varphi\circ R^X \,\mathrm{d}\overline{\mu}^{\otimes n}\\
&= \Phi([X,r,\mu]).
\end{aligned}$$ □

Remarks 3.11
(a) It is easy to see that if Φ_1,Φ_2 are mm-monomials of degree n_1 and n_2, respectively, then $\Phi_1\cdot\Phi_2$ is an mm-monomial of degree n_1+n_2. Thus $\mathcal{T}^{(1)}$ is closed under multiplication. However, $\Phi_1+\Phi_2$ is in general not an mm-monomial anymore, so $\mathcal{T}^{(1)}$ is *not* closed under addition.

(b) We call elements of $\mathcal{T}^{(1)}$ mm-*monomials*, because we think of them as monomials in the measure μ. Like the set of monomials (as functions on $\mathbb{R}$), the set $\mathcal{T}^{(1)}$ is closed under multiplication but not under addition. For similar reasons we call the elements of the linear span of $\mathcal{T}^{(1)}$ *mm-polynomials*. However, the set $\mathcal{T}^{(1)}$ is sufficient for our needs and we will not use the algebra of mm-polynomials.

It follows easily from the results of the last section that the a metric measure space is uniquely determined by the values of the mm-monomials.

Proposition 3.12
$\mathcal{T}^{(1)}$ separates points in $\mathbb{M}$.

Proof: Let $[X,r,\mu]$ and $[Y,d,\eta]$ be two mm spaces with $\Phi([X,r,\mu]) = \Phi([Y,d,\eta])$ for every $\Phi \in \mathcal{T}^{(1)}$. It follows that $\mathfrak{m}(\mu) = \mathfrak{m}(\eta)$ (by choosing mm-monomials with $\varphi \equiv 1$). Observe that an mm-monomial as in (E3.3) can be rewritten in the following form:

$$\Phi([X,r,\mu]) = \psi(\mathfrak{m}(\mu)) \int \varphi \circ R \, \mathrm{d}\overline{\mu}^{\otimes n} = \psi(\mathfrak{m}(\mu)) \int \varphi \, \mathrm{d}\big(R_* \overline{\mu}^{\otimes n}\big) .$$

By ranging over all φ we get

$$\mathfrak{m}(\mu) \cdot R_* \overline{\mu}^{\otimes n} = \mathfrak{m}(\eta) \cdot R_* \overline{\eta}^{\otimes n}$$

for every $n \in \mathbb{N}_{\geq 2}$. By Proposition 3.7 this implies that both spaces are equal. □

We use the mm-monomials to define a topology on $\mathbb{M}$. Because $\mathcal{T}^{(1)}$ separates points, the induced topology is Hausdorff.

Definition 3.13
The Gromov-weak topology τ_{Gw} *on* $\mathbb{M}$ *is defined as the initial topology induced by* $\mathcal{T}^{(1)}$.

Remark 3.14 (Gromov-weak convergence generalizes weak convergence)
The mm-monomials are designed in such a way that Gromov-weak convergence generalizes weak convergence. That is, if a sequence $(\mu_n)_n \subset \mathcal{M}_f(X)$ of finite measures converges weakly to μ, then $\Phi([X,r,\mu_k]) \to \Phi([X,r,\mu])$ for every $\Phi \in \mathcal{T}^{(1)}$. This follows simply by the fact that $\mathfrak{m}(\mu_k) \to \mathfrak{m}(\mu)$ and $\overline{\mu_k}^{\otimes n} \xrightarrow{w} \overline{\mu_k}^{\otimes n}$ for every $n \in \mathbb{N}$ (cf. Lemma 2.14 and Corollary 2.22).

This is also the reason for the restriction $\psi(0) = 0$ in the definition of mm-monomials. Recall that the normalization $\mu \mapsto \overline{\mu}$ is discontinuous at $\mu = o$. To smooth out this discontinuity we need $\psi(0) = 0$. Without this condition $\mu_n \xrightarrow{w} o$ would in general *not* imply $\Phi([X,r,\mu_n]) \to \Phi([X,r,o]) = 0$.

In the next lemma we characterize Gromov-weak convergence through test functions *without* decomposing μ into its mass $\mathfrak{m}(\mu)$ and normalization $\overline{\mu}$. Recall the definition of nets from section 2.1.

Lemma 3.15 (Alternative test functions for τ_{Gw})
A net $(\mathcal{X}_\alpha)_{\alpha\in\mathcal{A}}$ of mm spaces converges Gromov-weakly to an mm space $\mathcal{X}$ if and only if $\tilde\Phi(\mathcal{X}_\alpha)\to\tilde\Phi(\mathcal{X})$ for every function $\tilde\Phi\colon \mathbb{M}\to\mathbb{R}$ of the form

$$\tilde\Phi([X,r,\mu]) = \int \varphi\circ R \,\mathrm{d}\mu^{\otimes m} \tag{E3.4}$$

with $m\in\mathbb{N}_{\geq 2}$ and $\varphi\in\mathcal{C}_b(\mathbb{D}_m)$.

Proof: Let $\mathcal{X}=[X,r,\mu]$ and $\mathcal{X}_\alpha=[X_\alpha,r_\alpha,\mu_\alpha]$ for every $\alpha\in\mathcal{A}$. First assume that $\mathcal{X}_\alpha$ converges Gromov-weakly to $\mathcal{X}$. It follows that $\mathfrak{m}(\mu_\alpha)$ converges to $\mathfrak{m}(\mu)$ (by convergence of all mm-monomials with $\varphi\equiv 1$). Hence, for $M>\mathfrak{m}(\mu)$ we eventually have $\mathfrak{m}(\mu_\alpha)<M$. Let $\tilde\Phi$ be as in (E3.4) and let $\psi\in\mathcal{C}_b(\mathbb{R}_+)$ be such that $\psi(x)=x^m$ for every $x\leq M$. Then, eventually

$$\tilde\Phi(\mathcal{X}_\alpha) = \int \varphi\circ R\,\mathrm{d}\mu_\alpha^{\otimes m} = \psi(\mathfrak{m}(\mu_\alpha))\int\varphi\circ R\,\mathrm{d}\overline{\mu_\alpha}^{\otimes m}.$$

The right hand side of this equation is an mm-monomial and converges to

$$\psi(\mathfrak{m}(\mu))\int\varphi\circ R\,\mathrm{d}\overline{\mu}^{\otimes m} = \int\varphi\circ R\,\mathrm{d}\mu^{\otimes m} = \tilde\Phi(\mathcal{X}).$$

This holds for every $\tilde\Phi$ of the form (E3.4).

The other direction of the proof follows in a similar manner by showing first that $\mathfrak{m}(\mu_\alpha)$ converges to $\mathfrak{m}(\mu)$ (by choosing $\tilde\Phi$ as in (E3.4) with $\varphi\equiv 1$) and then writing mm-monomials $\Phi\in\mathcal{T}^{(1)}$ as the product of two converging nets by

$$\Phi(\mathcal{X}_\alpha) = \psi(\mathfrak{m}(\mu_\alpha))\int\varphi\circ R\,\mathrm{d}\overline{\mu_\alpha}^{\otimes m} = \frac{\psi(\mathfrak{m}(\mu_\alpha))}{\mathfrak{m}(\mu_\alpha)^m}\cdot\int\varphi\circ R\,\mathrm{d}\mu_\alpha^{\otimes m}$$

(excluding the trivial case where $\mathfrak{m}(\mu)=0$). □

The preceding lemma shows that the Gromov-weak topology τ_{Gw} is also induced by all functions $\tilde\Phi$ of the form (E3.4). However, unlike the mm-monomials, these functions are not bounded. Hence, they do not necessarily determine probability measures on $\mathbb{M}$, nor are they useful for defining generators of $\mathbb{M}$-valued Markov processes.

In Proposition 3.7 we proved that a metric measure space $[X,r,\mu]$ is uniquely determined by the distance matrix distribution

$$\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty}\in\mathcal{M}_f(\mathbb{D}_\mathbb{N}).$$

This suggests to pull back the weak topology on $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$ to define a topology on $\mathbb{M}$. The following proposition shows that this topology coincides with the Gromov-weak topology. Before we provide the statement recall that an injective function $\iota\colon Z_1\to Z_2$ between two topological spaces Z_1,Z_2 is called *bicontinuous* if both ι and its inverse ι^{-1} are continuous. We tacitly assume that the inverse function ι^{-1} is defined on the image $\iota(Z_1)$.

Proposition 3.16 ($\mathbb{M}$ is embedded in $\mathcal{M}_f(\mathbb{D}_{\mathbb{N}})$)
The function

$$\begin{aligned}\iota\colon (\mathbb{M},\tau_{Gw}) &\to \mathcal{M}_f(\mathbb{D}_{\mathbb{N}})\\ [X,r,\mu] &\mapsto \mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty}\end{aligned}$$

is an embedding, i. e. it is injective and bicontinuous.

Proof: The function ι is injective by Proposition 3.7. To show continuity of ι, let $([X_\alpha,r_\alpha,\mu_\alpha])_{\alpha\in\mathcal{A}}$ be a net of mm spaces which converges Gromov-weakly to an mm space $[X,r,\mu]$. By choosing mm-monomials $\Phi\in\mathcal{T}^{(1)}$ with $\varphi\equiv 1$ we see that

$$\mathfrak{m}(\mu_\alpha)\to\mathfrak{m}(\mu). \tag{E3.5}$$

In the case $\mathfrak{m}(\mu)=0$, this yields

$$\mathfrak{m}(\mu_\alpha)\cdot R_*\overline{\mu_\alpha}^{\otimes\infty}\xrightarrow{w} o=\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty},$$

where o denotes the null measure which we defined in section 2.4. If $\mathfrak{m}(\mu)>0$, then (E3.5) and the convergence of test functions of $\mathcal{T}^{(1)}$ implies that

$$R_*\overline{\mu_\alpha}^{\otimes m}\xrightarrow{w} R_*\overline{\mu}^{\otimes m}$$

for every $m\in\mathbb{N}_{\geq 2}$. Since weak convergence of finite dimensional distributions implies weak convergence of the projective limits (on countable product spaces, cf. Corollary 2.21), we get

$$R_*\overline{\mu_\alpha}^{\otimes\infty}\xrightarrow{w} R_*\overline{\mu}^{\otimes\infty}.$$

This shows that ι is continuous.

Because $\mathcal{M}_f(\mathbb{D}_{\mathbb{N}})$ is metrizable, the inverse function ι^{-1} is continuous if and only if it is sequentially continuous. Let $[X,r,\mu]\in\mathbb{M}$ and let $([X_n,r_n,\mu_n])_n$ be a sequence of mm spaces with

$$\mathfrak{m}(\mu_n)\cdot R_*\overline{\mu_n}^{\otimes\infty}\xrightarrow{w}\mathfrak{m}(\mu)\cdot R_*\overline{\mu}^{\otimes\infty}.$$

It follows that

$$\mathfrak{m}(\mu_n)\to\mathfrak{m}(\mu). \tag{E3.6}$$

If $\mathfrak{m}(\mu)=0$, we have $\Phi([X_n,r_n,\mu_n])\to\Phi([X,r,\mu])$ for every $\Phi\in\mathcal{T}^{(1)}$ (because of the restriction $\psi(0)=0$). If $\mathfrak{m}(\mu)>0$, we have

$$R_*\overline{\mu_n}^{\otimes\infty}\xrightarrow{w} R_*\overline{\mu}^{\otimes\infty}$$

and therefore weak convergence of the marginals

$$R_*\overline{\mu_n}^{\otimes m}\xrightarrow{w} R_*\overline{\mu}^{\otimes m}$$

with $m\in\mathbb{N}_{\geq 2}$. This together with (E3.6) yields convergence for all test functions in $\mathcal{T}^{(1)}$. We have shown that ι^{-1} is continuous and this completes the proof. □

Proposition 3.16 states that $(\mathbb{M}, \tau_{Gw})$ is homeomorphic to a subset of the Polish space $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$. It follows immediately that $(\mathbb{M}, \tau_{Gw})$ is separable and metrizable (as a subset of a separable and metrizable space).

Corollary 3.17
$(\mathbb{M}, \tau_{Gw})$ *is separable and metrizable.*

In particular this implies that $(\mathbb{M}, \tau_{Gw})$ is first-countable and that its topology and the continuity of functions on $\mathbb{M}$ can be characterized by sequences.

It is possible to metrize $(\mathbb{M}, \tau_{Gw})$ by pulling back the Prokhorov metric from the set $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$. But the following example shows that $\mathbb{M}$ is not closed in $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$. Hence the pull-back of the Prokhorov metric is not complete. In the subsequent section we will define a different metric which is complete and also metrizes the Gromov-weak topology.

Example 3.18 ($\mathbb{M}$ is not closed in $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$)
There are many elements of $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$ which do not stem from a metric measure space. One of them is $\delta_{\mathbb{1}_\infty}$, where $\mathbb{1}_\infty \in \mathbb{D}_\mathbb{N}$ is the distance matrix with all entries equal to 1. Assume that there is an mpm space $[X, r, \mu]$ with $R_* \mu^{\otimes\infty} = \delta_{\mathbb{1}_\infty}$ (it is obvious that μ must be a probability measure). It follows that

$$\mu^{\otimes\infty}(\{(x_i)_i \in X^\mathbb{N} \mid r(x_i, x_j) = 1 \ \forall i \neq j\}) = 1.$$

The measure μ must be defined on an infinite discrete space and has no atoms. Such a measure does not exists. Therefore, $\delta_{\mathbb{1}_\infty}$ cannot be the distance matrix distribution of an metric measure space.

However, the measure $\delta_{\mathbb{1}_\infty}$ is the weak limit of distance matrix distribution of metric measure spaces. For each $n \in \mathbb{N}$ define a metric measure space $\mathcal{Y}_n := [Y_n, d_n, \eta_n]$ by

$$\begin{aligned} Y_n &:= \{1, \ldots, n\}, \\ d_n(i,j) &:= \mathbb{1}_{i \neq j}, \\ \eta_n &:= \frac{1}{n} \sum_{i=1}^{n} \delta_i, \end{aligned}$$

i. e. η_n is a uniform distribution on n discrete points. Then we have

$$R_* \eta_n^{\otimes m} = \delta_{\mathbb{1}_m} + O\left(\frac{1}{n}\right) \xrightarrow{w} \delta_{\mathbb{1}_m}$$

for every $m \in \mathbb{N}$, where $\mathbb{1}_m$ denotes the distance matrix in $\mathbb{D}_m$ with all entries equal to 1. Hence, the finite dimensional distributions of $R_* \eta_n^{\otimes\infty}$ converge to those of $\delta_{\mathbb{1}_\infty}$ and we obtain $R_* \eta_n^{\otimes\infty} \xrightarrow{w} \delta_{\mathbb{1}_\infty}$ (see Corollary 2.21). This shows that $(\mathbb{M}, \tau_{Gw})$ is not closed in $\mathcal{M}_f(\mathbb{D}_\mathbb{N})$.

3.3 The Gromov-Prokhorov metric and its topology

The aim of this section is to introduce a complete metric on $\mathbb{M}$ which metrizes the Gromov-weak topology. We will put some effort into studying this metric before we are able to prove in section 3.6 that both topologies coincide.

We will not prove the statements in this section because the proofs in [GPW09, section 5] for mpm spaces are literally the same as the proofs for mm spaces.

As stated in the last section, the Gromov-weak topology generalizes weak convergence to measures which are defined on different sets. This suggests to generalize a complete metric for the weak topology, e. g. the Prokhorov metric d_P, to metrize this topology. To compare measures which are defined on different metric spaces, we use the same method as the Gromov-Hausdorff metric. We embed the involved metric spaces into a common metric space Z and push the involved measures forward to this space. This allows us to use the Prokhorov metric on Z to gauge the distance between the measures.

Definition 3.19 (Gromov-Prokhorov metric and topology)
Let $\mathcal{X} = [X,r,\mu]$ and $\mathcal{Y} = [Y,d,\eta]$ be mm spaces. We define the Gromov-Prokhorov distance *between $\mathcal{X}$ and $\mathcal{Y}$ by*

$$d_\mathrm{GP}(\mathcal{X},\mathcal{Y}) := \inf_{Z,\iota_X,\iota_Y} d_\mathrm{P}^Z(\iota_{X*}\mu, \iota_{Y*}\eta), \tag{E3.7}$$

where the infimum ranges over all isometric embeddings $\iota_X : X \to Z$, $\iota_Y : Y \to Z$ into a common Polish metric space (Z,r_Z) with $Z \subset \mathbb{R}^\mathbb{N}$ and where d_P^Z denotes the Prokhorov metric for measures on Z. The topology induced by this metric is called the Gromov-Prokhorov topology *and denoted by τ_{GP}.*

It is easy to see that the Gromov-Prokhorov metric is well-defined, i. e. that the distance between two metric measure spaces does not depend on the representatives of the mm spaces. Moreover, convergence with respect to d_GP obviously generalizes weak convergence. That is, if $\mu_n \xrightarrow{w} \mu$ on a Polish metric space (X,r), then $d_\mathrm{GP}([X,r,\mu_n],[X,r,\mu])$ converges to 0.

Remark 3.20
We need the condition $Z \subset \mathbb{R}^\mathbb{N}$ in Definition 3.19 to ensure that the infimum is taken over a well-defined set. This is not a restriction since every Polish space can be embedded in $\mathbb{R}^\mathbb{N}$ by section 2.3.

Finding a "good" space Z and embeddings ι_X, ι_Y to compute $d_\mathrm{GP}(\mathcal{X},\mathcal{Y})$ can be a challenging task. If X and Y have a similar structure, there might be a natural way to overlap X and Y in such a way that the Prokhorov distance in (E3.7) is small. However, if X and Y are very different, it might be better to choose Z as the disjoint union $X \sqcup Y$. Then, the problem of finding an appropriate space Z and embeddings ι_X, ι_Y reduces to finding a metric r' on $X \sqcup Y$ such that it extends the old metrics r and d and connects the spaces X and Y in an optimal way. The next lemma shows that this procedure gives the same distance as the original definition.

Lemma 3.21 (Alternative characterization of Gromov-Prokhorov metric)
For all mm spaces $\mathcal{X}=[X,r,\mu]$ and $\mathcal{Y}=[Y,d,\eta]$ we have

$$d_{\mathrm{GP}}(\mathcal{X},\mathcal{Y})=\inf_{r'} d_{\mathrm{P}}^{X\sqcup Y,r'}(\mu,\eta),$$

where the infimum ranges over all metrics r' on the disjoint union $X\sqcup Y$ which extend the metrics r and d.

The following two embedding lemmas show that we can embed convergent (or Cauchy) sequences into a single ambient metric space in which we have weak convergence of the push-forward measures. Loosely speaking, metric measure spaces converge with respect to the Gromov-Prokhorov metric if and only if there is a common Polish metric space on which their measures converge weakly.

Lemma 3.22 (Embedding of convergent sequences of mm spaces)
Let $(\mathcal{X}_n)_n$ be a sequence of mm spaces with $\mathcal{X}_n=[X_n,r_n,\mu_n]$ which converges to an mm space $\mathcal{X}=[X,r,\mu]$ in the Gromov-Prokhorov topology. Then there is a Polish metric space (Z,r_Z) and isometric embeddings $\iota,\iota_1,\iota_2,\dots$ of $X,X_1,X_2,\dots$, respectively, into Z such that

$$d_{\mathrm{P}}^{Z}(\iota_{n*}\mu_n,\iota_*\mu)\to 0.$$

Lemma 3.23 (Embedding of sequences of mm spaces)
Let $(\varepsilon_n)_n$ be a sequence of positive real numbers and let $(\mathcal{X}_n)_n$ be a sequence of mm spaces with $\mathcal{X}_n=[X_n,r_n,\mu_n]$ and

$$d_{\mathrm{GP}}(\mathcal{X}_n,\mathcal{X}_{n+1})<\varepsilon_n$$

for every $n\in\mathbb{N}$. Then there is a Polish metric space (Z,r_Z) and isometric embeddings $\iota_1,\iota_2,\dots$ of $X_1,X_2,\dots$, respectively, into Z such that

$$d_{\mathrm{P}}^{Z}(\iota_{n*}\mu_n,(\iota_{(n+1)*}\mu_{n+1})<\varepsilon_n$$

for every $n\in\mathbb{N}$.

It follows easily from the preceding embedding lemma that the metric d_{GP} is complete. Separability follows in a similar way as for the Prokhorov metric. The countable set of all finite metric measure spaces with rational values and rational distances is dense in $(\mathbb{M},d_{\mathrm{GP}})$. We therefore have the following result (see [GPW09, section 5] for a comprehensive proof).

Proposition 3.24
$(\mathbb{M},d_{\mathrm{GP}})$ *is a Polish metric space.*

Since $\mathbb{M}_1$ is obviously a closed subset of $(\mathbb{M},d_{\mathrm{GP}})$, it follows directly that $(\mathbb{M}_1,d_{\mathrm{GP}})$ is also a Polish metric space.

We close this section by showing that the topology τ_{GP} is finer than τ_{Gw}. Note that eventually both topologies will turn out to be the same (see section 3.6). But for the time being we are only able to prove one direction.

Lemma 3.25 (τ_{GP} is finer than τ_{Gw})
Every mm-monomial $\Phi \in \mathcal{T}^{(1)}$ is continuous with respect to the Gromov-Prokhorov topology. Therefore, Gromov-Prokhorov convergence implies Gromov-weak convergence and τ_{GP} is finer than τ_{Gw}.

Proof: Let Φ be an mm-monomial and let $([X_n, r_n, \mu_n])_n$ converge to $[X, r, \mu]$ with respect to the Gromov-Prokhorov metric. By Lemma 3.22 we can embed all involved metric spaces in a common metric space (Z, r_Z) in which we have weak convergence of the push-forwards of the measures. The function

$$\begin{aligned} \mathcal{M}_f(Z) &\to \mathbb{R} \\ \mu &\mapsto \Phi([Z, r_Z, \mu]) \end{aligned}$$

is continuous since $\mu \mapsto \mathfrak{m}(\mu)$ and $\mu \mapsto \mu^{\otimes n}$ are continuous (cf. Lemma 2.14 and Corollary 2.22). Hence we have

$$\Phi([X_n, r_n, \mu_n]) = \Phi([Z, r_Z, \iota_{n*}\mu_n]) \to \Phi([Z, r_Z, \iota_*\mu]) = \Phi([X, r, \mu]). \qquad \square$$

3.4 Distance distribution and modulus of mass distribution

In this section we define the distance distribution and the modulus of mass distribution, which have been introduced for probability measures in [GPW09]. The definitions can easily be generalized to finite measures. We keep the names distance *distribution* and modulus of mass *distribution*, even though in our case we do not necessarily deal with probability measures.

Both the distance distribution and the modulus of mass distribution are used to characterize compact sets in $(\mathbb{M}, d_{\mathrm{GP}})$ in section 3.5. To motivate their definition, we look at typical "strongly divergent" sequences, i. e. sequences without any convergent subsequence, in the following examples. Note that the last two examples are taken from [GPW09].

Example 3.26
Define $\mathcal{Z}_n := [Z, r_Z, \zeta_n]$ with $Z := \{1\}$, $r_Z \equiv 0$ and $\zeta_n := n\delta_1$. It is easy to see, that $d_{\mathrm{P}}^Z(\zeta_i, \zeta_j) = |i-j|$ and hence $d_{\mathrm{GP}}(\mathcal{Z}_i, \mathcal{Z}_j) = |i-j|$ for all $i, j \in \mathbb{N}$. Therefore, the sequence $(\mathcal{Z}_n)_n$ does not have a convergent subsequence. The problem here is obviously that the mass of the measures goes to infinity.

Example 3.27
For each $n \in \mathbb{N}$ define a metric measure space $\mathcal{X}_n := [X_n, r_n, \mu_n]$ with

$$\begin{aligned} X_n &:= \{0, 1\}, \\ r_n(i, j) &:= n\mathbb{1}_{i \neq j}, \\ \mu_n &:= \frac{1}{2}(\delta_0 + \delta_1). \end{aligned}$$

The measure μ_n is a uniform distribution between two points with mutual distance n. Since this distance goes to infinity, we cannot embed the metric spaces (X_n, r_n) into a common metric space such that the push-forwards of the measures converge. This shows that $(\mathcal{X}_n)_n$ diverges and so does every subsequence.

Example 3.28
In Example 3.18 we defined metric measure spaces $\mathcal{Y}_n := [Y_n, d_n, \eta_n]$ with

$$\begin{aligned} Y_n &:= \{1, \ldots, n\}, \\ d_n(i,j) &:= \mathbb{1}_{i \neq j}, \\ \eta_n &:= \frac{1}{n} \sum_{i=1}^{n} \delta_i. \end{aligned}$$

The measure η_n is a uniform distribution on n points with mutual distance 1. As n increases the mass is distributed among more and more discrete points. Hence, a potential limit would be a uniform distribution on an infinite discrete space. But such an object does not exist. A formal proof for the divergence of the sequence $(\mathcal{Y}_n)_n$ follows from Example 3.18. There we showed that the distance matrix distribution of $\mathcal{Y}_n$ converges to $\delta_{\mathbb{1}_\infty}$ and that $\delta_{\mathbb{1}_\infty}$ is not the distance matrix distribution of a metric measure space. Therefore, $(\mathcal{Y}_n)_n$ does not converge, nor does any of its subsequences.

The divergence of the sequence $(\mathcal{X}_n)_n$ of example 3.27 is related to the diameter of the support of μ_n, but also to the distribution of mass on the support. To see the latter relationship, observe that the diameter of the support of $\mu'_n := (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_1$ diverges, but that $[X_n, r_n, \mu'_n]$ converges with respect to d_{GP}. To describe the type of divergence of $(\mathcal{X}_n)_n$ we look at the distance between two randomly sampled points. The distribution of this distance is given by $r_{n*}\mu_n^{\otimes 2} = \frac{1}{2}(\delta_0 + \delta_n)$ and is also divergent. We call $r_{n*}\mu_n^{\otimes 2}$ the distance distribution of the measure μ_n.

Definition 3.29 (Distance distribution)
Let μ be a finite Borel measure on a Polish metric space (X, r). The distance distribution $w(\mu) \in \mathcal{M}_f(\mathbb{R}_+)$ of μ *is defined by* $w(\mu) := r_*\mu^{\otimes 2}$.

We will prove in Corollary 3.36 that the function $[X, r, \mu] \mapsto w(\mu)$ is continuous. Hence, the convergence of the distance distributions is a necessary condition for the convergence of a sequence of metric measure spaces.

The divergence of the sequence $(\mathcal{Y}_n)_n$ of Example 3.28 is related to the fact that the mass of η_n is scattered among many discrete points. Heuristically, we can regard those points as being "thin", since there is only a small portion of the mass of η_n in their vicinity. As n increases, the points of Y_n become thinner and thinner, but since they are discrete, η_n cannot converge to a continuous measure. To capture such behavior we introduce the modulus of mass distribution in the next definition. We will return to Example 3.28 at the end of this section.

Definition 3.30 (Modulus of mass distribution)
Let μ be a finite Borel measure on a Polish metric space (X,r). The modulus of mass distribution of μ *is the function defined by*

$$\begin{aligned} V_\cdot(\mu)\colon \mathbb{R}_+ &\to \mathbb{R}_+ \\ \delta &\mapsto V_\delta(\mu) := \inf\{\,\varepsilon > 0 \mid \mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\}) \le \varepsilon\,\}. \end{aligned}$$

The following lemma summarizes some useful properties of the modulus of mass distribution.

Lemma 3.31 (Properties of the modulus of mass distribution)
Let μ be a finite Borel measure on a Polish metric space (X,r).

(a) The function $\delta \mapsto V_\delta(\mu)$ is non-decreasing and bounded by the total mass $\mathfrak{m}(\mu)$. Moreover, we have $\lim_{\delta\searrow 0} V_\delta(\mu) = 0$.

(b) For $\varepsilon,\delta > 0$ we have $V_\delta(\mu) < \varepsilon$ if and only if $\mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\}) < \varepsilon$.

(c) Let $V_\delta(\mu) < \varepsilon$ with $\varepsilon,\delta > 0$. Then there is a finite set $A \subset X$ with $|A| \le \max(1, \frac{\mathfrak{m}(\mu)}{\delta})$ such that $\mu(\complement B(A,\varepsilon)) < \varepsilon$.

Note that the last assertion implies that if we have a set Γ of finite measures (possibly defined on different metric spaces) with $\sup_{\mu\in\Gamma} V_\delta(\mu) < \varepsilon$, then we can cover most of their supports (up to a set of mass ε) by *the same number* of ε-balls.

Proof (of Lemma 3.31): Assertion (a) was proved for probability measures in Lemma 6.5 of [GPW09]. The same proof holds true for finite measures.

That $V_\delta(\mu) < \varepsilon$ implies $\mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\}) < \varepsilon$ was proved in [GPW09, Lemma 6.4]. To prove the other direction, assume that $V_\delta(\mu) \ge \varepsilon$. Notice that we have

$$\mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\}) = \int \mathbb{1}_{[0,\delta]}(\mu(B(x,\varepsilon)))\,\mathrm{d}\mu(x). \tag{E3.8}$$

The function

$$\varepsilon \mapsto \mathbb{1}_{[0,\delta]}(\mu(B(x,\varepsilon)))$$

is left-continuous for every $x \in X$. It follows from (E3.8) and the dominated convergence theorem that the function

$$\varepsilon \mapsto \mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\})$$

is also left-continuous. From the definition of $V_\delta(\mu)$ we see that $V_\delta(\mu) \ge \varepsilon$ implies

$$\mu(\{\,x \in X \mid \mu(B(x,\varepsilon')) \le \delta\,\}) > \varepsilon' \text{ for every } \varepsilon' < \varepsilon.$$

This together with the left-continuity yields

$$\mu(\{\,x \in X \mid \mu(B(x,\varepsilon)) \le \delta\,\}) = \lim_{\varepsilon'\nearrow\varepsilon} \mu(\{\,x \in X \mid \mu(B(x,\varepsilon')) \le \delta\,\} \ge \varepsilon.$$

Assertion (c) was proved for probability measures in [GPW09, Lemma 6.9]. We extend this proof to finite measures. In case $\mathfrak{m}(\mu) \le \varepsilon$ there must be an $x \in X$ with $\mu(\complement B(x,\varepsilon)) < \varepsilon$ and we are done. Otherwise we define $D := \{ x \in X \mid \mu(B(x,\varepsilon)) > \delta \}$. Because $V_\delta(\mu)$ is less than ε, we have $\mu(\complement D) < \varepsilon < \mathfrak{m}(\mu)$ and D is not empty. By [BBI01, page 278] there exists an ε-separated discrete subset A of D that is maximal. That is, we have $r(x_1,x_2) \ge \varepsilon$ for $x_1, x_2 \in A$ with $x_1 \ne x_2$ and adding a further point of D would destroy this property. It follows from the maximality that $D \subset B(A,\varepsilon)$ and therefore $\mu(\complement B(A,\varepsilon)) \le \mu(\complement D) < \varepsilon$. Moreover, we see that

$$\mathfrak{m}(\mu) \ge \mu(B(A,\varepsilon)) = \sum_{x \in A} \mu(B(x,\varepsilon)) \ge |A|\delta,$$

which yields the claim. □

It is easy to see that the distance distribution and the modulus of mass distribution are invariant under mm-isomorphisms. That is, if (X,r,μ) and (Y,d,η) are representatives of the same metric measure space, then $w(\mu) = w(\eta)$ and $V_\delta(\mu) = V_\delta(\eta)$ for all $\delta \ge 0$ (cf. [GPW09, Remark 2.10]). In the remainder of this section we study the continuity properties of the distance distribution and the modulus of mass distribution with respect to the Gromov-weak and the Gromov-Prokhorov topology. For this purpose we introduce the random distance distribution which is closely related to $w(\cdot)$ and $V_\delta(\cdot)$.

Definition 3.32 (Random distance distribution)
Let μ be a finite Borel measure on a Polish metric space (X,r). We define

$$\begin{aligned} r_x \colon X &\to \mathbb{R}_+ \\ x' &\mapsto r(x,x') \end{aligned}$$

for each $x \in X$ and

$$\begin{aligned} \hat{r} \colon X &\to \mathcal{M}_f(\mathbb{R}_+) \\ x &\mapsto (r_x)_*\mu. \end{aligned}$$

The random distance distribution $W(\mu) \in \mathcal{M}_f(\mathcal{M}_f(\mathbb{R}_+))$ of μ *is defined as*

$$W(\mu) := \hat{r}_*\mu.$$

If μ is a probability measure, then $\hat{r}(x) = (r_x)_*\mu$ is the distribution of the distance between x and a randomly sampled point x'. Therefore, $\hat{r}(x)$ can be seen as the *local* distance distribution in x. With the random distance distribution $W(\mu)$ we first sample a point x and then look at the local distance distribution in x.

The following lemma shows that the distance distribution and the modulus of mass distribution can be expressed in terms of the random distance distribution.

Lemma 3.33 ($w(\mu)$ and $V_\delta(\mu)$ in terms of $W(\mu)$)
Let μ be a finite Borel measure on a Polish space (X,r). Then we have

$$w(\mu) = \int \eta \, \mathrm{d}W(\mu)(\eta)$$

and

$$V_\delta(\mu) = \inf\{\, \varepsilon > 0 \mid W(\mu)(\{\, \eta \in \mathcal{M}_f(\mathbb{R}_+) \mid \eta([0,\varepsilon)) \le \delta \,\}) \le \varepsilon \,\}$$

for all $\delta \ge 0$.

Proof: The first equation follows by simple transformations:

$$\int \eta \, \mathrm{d}W(\mu)(\eta) = \int \eta \, \mathrm{d}(\hat{r}_*\mu)(\eta) = \int \hat{r}(x) \, \mathrm{d}\mu(x) = \int r_{x*}\mu \, \mathrm{d}\mu(x) = r_*\mu^{\otimes 2} = w(\mu).$$

To see the second equality, observe that for $\delta, \varepsilon > 0$ we have

$$\begin{aligned} &W(\mu)(\{\, \eta \in \mathcal{M}_f(\mathbb{R}_+) \mid \eta([0,\varepsilon)) \le \delta \,\}) \\ &\quad = \mu(\{\, x \in X \mid \hat{r}(x)([0,\varepsilon)) \le \delta \,\}) \\ &\quad = \mu(\{\, x \in X \mid \mu(B(x,\varepsilon)) \le \delta \,\}). \end{aligned}$$

The claim follows directly from the definition of $V_\delta(\mu)$. □

Remark 3.34 (Random distance distribution and its moment measures)
$W(\mu)$ is an element of $\mathcal{M}_f(\mathcal{M}_f(\mathbb{R}_+))$. We call such measures two-level measures. Two-level measures like this will be treated extensively in section 4.2. In anticipation of this section we provide a short overview over the tools which we need here:

The m-th moment measures of $W(\mu)$ is the measure on $\mathbb{R}_+^m$ defined by

$$\mathfrak{M}^m_{W(\mu)} := \int \eta^{\otimes m} \, \mathrm{d}W(\mu)(\eta).$$

Observe that with $K := \mathfrak{m}(\mu)$ we actually have

$$W(\mu) \in \mathcal{M}_K(\mathcal{M}_K(\mathbb{R}_+))$$

(recall that $\mathcal{M}_K(\ldots)$ denotes the set of all measures with mass K). This follows directly from Definition 3.32 since we have $\mathfrak{m}(W(\mu)) = \mathfrak{m}(\hat{r}_*\mu) = \mathfrak{m}(\mu)$ and $\mathfrak{m}(\hat{r}(x)) = \mathfrak{m}((r_x)_*\mu) = \mathfrak{m}(\mu)$ for all $x \in X$. Therefore, the mass of the m-th moment measures of $W(\mu)$ is equal to K^{m+1}. It will follow from Proposition 4.10 that weak convergence of the random distance distribution is equivalent to weak convergence of all moment measures. That is, if $\mu, \mu_1, \mu_2, \ldots$ are finite Borel measures (possibly defined on different Polish spaces) we have

$$W(\mu_n) \xrightarrow{w} W(\mu) \text{ if and only if } \mathfrak{M}^k_{W(\mu_n)} \xrightarrow{w} \mathfrak{M}^k_{W(\mu)} \text{ for every } k \in \mathbb{N}. \tag{E3.9}$$

With the help of the preceding remark we are able to show that the random distance distribution is continuous in the Gromov-weak topology.

Lemma 3.35 (Random distance distribution is continuous)
The function

$$\begin{aligned} \mathbb{M} &\to \mathcal{M}_f(\mathcal{M}_f(\mathbb{R}_+)) \\ [X,r,\mu] &\mapsto W(\mu) \end{aligned}$$

is continuous with respect to both the Gromov-weak and the Gromov-Prokhorov topology.

Proof: Since the Gromov-Prokhorov topology is finer than the Gromov-weak topology by Lemma 3.25, it suffices to prove continuity for the latter topology. Let $([X_n,r_n,\mu_n])_n$ be a sequence of metric measure spaces which converges Gromov-weakly to $[X,r,\mu] \in \mathbb{M}$. This implies that the masses $(\mathfrak{m}(\mu_n))_n$ are bounded from above by some constant $K > 0$ and that $W(\mu_n) \in \mathcal{M}_{\leq K}(\mathcal{M}_{\leq K}(\mathbb{R}_+))$ for every $n \in \mathbb{N}$. To prove that $W(\mu_n)$ converges to $W(\mu)$ if suffices to show that all moment measures converge (cf. Remark 3.34). Fix $m \in \mathbb{N}$ and $\varphi \in \mathcal{C}_b(\mathbb{R}_+^m)$. The integral of φ with respect to the m-th moment measure is

$$\begin{aligned} \int\int \varphi \,\mathrm{d}\eta^{\otimes m}\,\mathrm{d}W(\mu_n)(\eta) &= \int\int \varphi \,\mathrm{d}(r_{x*}\mu_n)^{\otimes m}\,\mathrm{d}\mu_n(x) \\ &= \int\int \varphi(r_x(x_1),\dots,r_x(x_m))\,\mathrm{d}\mu_n^{\otimes m}(x_1,\dots,x_m)\,\mathrm{d}\mu_n(x) \\ &= \int \varphi(r(x,x_1),\dots,r(x,x_m))\,\mathrm{d}\mu_n^{\otimes m+1}(x,x_1,\dots,x_m). \end{aligned}$$

This is a test function of the form (E3.4). Since we have Gromov-weak convergence, this converges to

$$\int \varphi(r(x,x_1),\dots,r(x,x_m))\,\mathrm{d}\mu^{\otimes m+1}(x,x_1,\dots,x_m) = \int\int \varphi\,\mathrm{d}\eta^{\otimes m}\,\mathrm{d}W(\mu)(\eta)$$

by Lemma 3.15. Because m and φ are arbitrary, we have shown that all moment measures converge and the proof is complete. □

Since the distance distribution and the modulus of mass distribution can be expressed in terms of the random distance distribution, we can use the previous lemma to obtain continuity results for both of them.

Corollary 3.36 (Distance distribution is continuous)
The function

$$\begin{aligned} \mathbb{M} &\to \mathcal{M}_f(\mathbb{R}_+) \\ [X,r,\mu] &\mapsto w(\mu) \end{aligned}$$

is continuous with respect to both the Gromov-weak and the Gromov-Prokhorov topology.

Proof: It follows from Lemma 3.33 that $w(\mu)$ is the first moment measure of $W(\mu)$. The random distance distribution is continuous by Lemma 3.35 and thus the claim follows directly from (E3.9). □

Lemma 3.37 (Modulus of mass distribution is upper semi-continous) *Let $\delta > 0$ be fixed. The function*

$$\begin{aligned} \mathbb{M} &\to \mathbb{R}_+ \\ [X,r,\mu] &\mapsto V_\delta(\mu) \end{aligned}$$

is upper semi-continuous with respect to both the Gromov-weak and the Gromov-Prokhorov topology.

Proof: The proof for metric probability measure spaces can be found in [GPW09, Proposition 6.6 (iv)]. It remains valid even if we replace metric probability measure spaces by metric measure spaces with *finite* measures. □

Recall that the upper semi-continuity of $V_\delta(\mu)$ means that

$$\limsup_{n\to\infty} V_\delta(\mu_n) \le V_\delta(\mu) \tag{E3.10}$$

when $[X_n, r_n, \mu_n]$ converges to $[X, r, \mu]$. Since the modulus of mass distribution is increasing and vanishes for $\delta \to 0$ (cf. Lemma 3.31) we obtain

$$\lim_{\delta\searrow 0} \sup_{n\in\mathbb{N}} V_\delta(\mu_n) \to 0. \tag{E3.11}$$

In other words, the moduli of mass distribution of the sequence $(\mu_n)_n$ are uniformly continuous in $\delta = 0$.

Let us now come back to the metric measure spaces $\mathcal{Y}_n = [Y_n, d_n, \eta_n]$ of Example 3.28. Observe that for fixed $\delta > 0$ we have

$$V_\delta(\eta_n) = \begin{cases} 0, & n < \frac{1}{\delta} \\ 1, & n \ge \frac{1}{\delta}. \end{cases}$$

This yields $\lim_{\delta\searrow 0} \sup_{n\in\mathbb{N}} V_\delta(\eta_n) = 1$ and the same holds for every subsequence of $(\eta_n)_n$. Hence, $(\eta_n)_n$ violates (E3.11) and $(\mathcal{Y}_n)_n$ cannot have a convergent subsequence. This shows that $(\mathcal{Y}_n)_n$ is "strongly divergent" as we claimed in the beginning of this section.

The inequality (E3.10) shows that we can deduce an upper bound for $V_\delta(\mu_n)$ from the limit mm space because V_δ is *upper* semi-continuous. Sometimes, however, it is necessary to deduce an upper bound for the limit in terms of the approximating sequence. For this reason we provide a *lower* semi-continuous alternative to the modulus of mass distribution in the following lemma. We will apply this lemma in the example in chapter 5.

Lemma 3.38
For fixed $\varepsilon, \delta > 0$ the function

$$\begin{aligned} \mathbb{M} &\to \mathbb{R}_+ \\ [X,r,\mu] &\mapsto \mu(\{\, x \in X \mid \mu(\overline{B}(x,\varepsilon)) < \delta \,\}) \end{aligned}$$

is lower semi-continuous with respect to both the Gromov-weak and the Gromov-Prokhorov topology.

Proof: The proof of [GPW09, Proposition 6.6 (iv)] actually shows that

$$\begin{aligned} \mathbb{M} &\to \mathbb{R}_+ \\ [X,r,\mu] &\mapsto \mu(\{\, x \in X \mid \mu(B(x,\varepsilon)) \leq \delta \,\}) \end{aligned}$$

is upper semi-continuous by using the Portmanteau theorem for closed sets. The proof of our claim is similar, but uses the Portmanteau theorem for open sets instead. □

3.5 Compact sets

Now we characterize relatively compact subsets of $(\mathbb{M}, d_{\mathrm{GP}})$. The proof is rather long, that is why we omit it here and refer the interested reader to the proof of [GPW09, Proposition 7.1]. The proof in [GPW09] for mpm spaces easily generalizes to mm space when we additionally require that the mass of the measures is uniformly bounded (cf. [GPW09, Remark 7.2]).

Since d_{GP}-convergence generalizes weak convergence, it is not surprising that a characterization of compactness resembles the Prokhorov theorem. That is, a set $\Gamma \subset \mathbb{M}$ is relatively compact if and only if the masses of the measures are uniformly bounded and the measures are in some way "tight". Tightness in this context means that given an arbitrary small ε we have that for every mm space $\mathcal{X} = [X,r,\mu] \in \Gamma$ there is a compact set $C_{\mathcal{X},\varepsilon} \subset X$ with $\mu(\complement C_{\mathcal{X},\varepsilon}) < \varepsilon$ and that the sets $\{\, C_{\mathcal{X},\varepsilon} \mid \mathcal{X} \in \Gamma \,\}$ are "uniformly compact", i. e. relatively compact in the Gromov-Hausdorff topology. This is the first characterization of relative compactness, see property (d) of the next proposition.

The set $\{\, C_{\mathcal{X},\varepsilon} \mid \mathcal{X} \in \Gamma \,\}$ is relatively compact in the Gromov-Hausdorff topology if and only if it is uniformly totally bounded (see Proposition A.5), meaning that for every ε there is a natural number N such that $\operatorname{diam} C_{\mathcal{X},\varepsilon} \leq N$ and $C_{\mathcal{X},\varepsilon}$ can be covered by at most N balls of radius ε. This yields another characterization of relative compactness of Γ, see property (c) of Proposition 3.39.

We also provide a condition for relative compactness in terms of the distance distribution $w(\mu)$ and the modulus of mass distribution $V_\delta(\mu)$. Since $w(\mu)$ and $V_\delta(\mu)$ are continuous with respect to the Gromov-weak topology, this links compactness in the Gromov-Prokhorov topology with the Gromov-weak topology. We will use this link in the subsequent section to show that both topologies coincide.

Proposition 3.39 (Characterization of relatively compact subsets of $\mathbb{M}$)
Let $\Gamma \subset \mathbb{M}$ be a set of mm spaces. The following are equivalent:

(a) Γ *is relatively compact in the Gromov-Prokhorov topology* τ_{GP}.

(b) $\{\mathfrak{m}(\mu) \mid [X,r,\mu] \in \Gamma\}$ *is bounded in* $\mathbb{R}_+$ *and*

- $\sup_{[X,r,\mu]\in\Gamma} V_\delta(\mu) \to 0$ *for* $\delta \searrow 0$,
- $\{w(\mu) \mid [X,r,\mu] \in \Gamma\}$ *is tight in* $\mathcal{M}_f(\mathbb{R}_+)$.

(c) $\{\mathfrak{m}(\mu) \mid [X,r,\mu] \in \Gamma\}$ *is bounded in* $\mathbb{R}_+$ *and for every* $\varepsilon > 0$ *there is an* $N_\varepsilon \in \mathbb{N}$ *such that for every* $\mathcal{X} = [X,r,\mu] \in \Gamma$ *there exists a measurable subset* $X_{\mathcal{X},\varepsilon} \subset X$ *with*

- $\mu\left(\complement X_{\mathcal{X},\varepsilon}\right) < \varepsilon$,
- $X_{\mathcal{X},\varepsilon}$ *can be covered by at most* N_ε *balls of radius* ε,
- *the diameter of* $X_{\mathcal{X},\varepsilon}$ *is at most* N_ε.

(d) $\{\mathfrak{m}(\mu) \mid [X,r,\mu] \in \Gamma\}$ *is bounded in* $\mathbb{R}_+$ *and for every* $\varepsilon > 0$ *and* $\mathcal{X} = [X,r,\mu] \in \Gamma$ *there is a compact subset* $C_{\mathcal{X},\varepsilon} \subset X$ *such that*

- $\mu\left(\complement C_{\mathcal{X},\varepsilon}\right) < \varepsilon$,
- $\mathcal{C}_\varepsilon := \{C_{\mathcal{X},\varepsilon} \mid \mathcal{X} \in \Gamma\}$ *is relatively compact in the Gromov-Hausdorff topology.*

3.6 Equivalence of Gromov-weak and Gromov-Prokhorov topology

We are now in a position to verify that the Gromov-weak topology is metrized by the Gromov-Prokhorov metric. The proof is almost the same as the proof for mpm spaces in [GPW09, Theorem 5]. However, because this is one of the main results of this chapter, we provide the proof in full detail.

Proposition 3.40
The Gromov-weak topology τ_{Gw} *and the Gromov-Prokhorov topology* τ_{GP} *coincide.*

In particular, this means that every statement we made about one of the two topologies (e. g. embedding lemmas, compactness criteria, Polishness) hold for both topologies.

Proof: We already proved in Lemma 3.25 that the Gromov-Prokhorov topology is finer than the Gromov-weak topology. To prove the other direction, it suffices to show that every τ_{Gw}-convergent sequence

$$[X_n, r_n, \mu_n] \xrightarrow{\tau_{Gw}} [X, r, \mu] \tag{E3.12}$$

also converges with respect to the Gromov-Prokhorov topology τ_{GP} (recall from Corollary 3.17 that τ_{GP} is metrizable, thus we may use sequences instead of nets). We will show that $([X_n, r_n, \mu_n])_n$ is relatively compact with respect to τ_{GP}. Since $\mathcal{T}^{(1)}$ separates points and τ_{GP}-convergence implies τ_{Gw}-convergence, it follows that every subsequence has a further convergent subsequence with limit $[X,r,\mu]$. Consequently, the sequence itself must converge to $[X,r,\mu]$ in the topology τ_{GP}.

To show relative compactness of the sequence we use the compactness criterion (b) of Proposition 3.39. By Lemma 3.36 both $\mathfrak{m}(\mu_n)$ and $w(\mu_n)$ are convergent. Thus $(\mathfrak{m}(\mu_n))_n$ is bounded and $(w(\mu_n))_n$ is tight. Moreover, by Lemma 3.37 we have

$$\limsup_n V_\delta(\mu_n) \leq V_\delta(\mu)$$

for every $\delta > 0$. Since $V_\delta(\eta) \to 0$ for $\delta \searrow 0$ and for every finite measure η (cf. Lemma 3.31), this yields $\sup_n V_\delta(\mu_n) \to 0$ for $\delta \searrow 0$ and the assumptions of Proposition 3.39 are satisfied. □

3.7 Distributions on $\mathbb{M}$

With Le Cam's theorem we can easily deduce that the mm-monomials are convergence determining for random metric measure spaces.

Proposition 3.41
The set $\mathcal{T}^{(1)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M})$.

Proof: $\mathcal{T}^{(1)}$ is closed under multiplication and induces the topology of $\mathbb{M}$. Therefore, we can apply Proposition 2.17 to obtain the result. □

Tightness conditions for distributions on $\mathbb{M}$ can directly be derived from the compactness criterion 3.39. We omit the proof and refer to [GPW09, Theorem 3, Remark 3.2 and 3.3].

Proposition 3.42 (Tightness criterion for $\mathcal{M}_1(\mathbb{M})$)
A set $\mathcal{P} \subset \mathcal{M}_1(\mathbb{M})$ is tight if and only if for every $\varepsilon > 0$ there are $\delta > 0$ and $c > 0$ such that for every $P \in \mathcal{P}$ we have

(a) $P(\mathfrak{m}(\mu) \geq c) < \varepsilon$,

(b) $P(V_\delta(\mu) \geq \varepsilon) < \varepsilon$,

(c) $P(w(\mu)([c,\infty]) \leq \varepsilon) < \varepsilon$,

where μ denotes the measure of a random mm space $[X,r,\mu]$ with law P.

Chapter 4

Metric two-level measure spaces

In this chapter we generalize the results of chapter 3. Instead of metric measure spaces we will consider metric *two-level* measure spaces $[X,r,\nu]^{(2)}$, where $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ is a two-level measure on the Polish metric space (X,r). For didactic reasons we first consider the easier case $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ and afterwards the more general case $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.

Outline of this chapter: In section 4.1 we provide the basic definitions of metric two-level measure spaces and discuss the problems that occur when generalizing the theory of metric measure spaces. Section 4.2 deals with two-level measures $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ on a fixed Polish space X. We introduce important tools such as the moment measures and the approximation of two-level measures via density functions. In section 4.3 we consider metric two-level *probability* measure spaces (m2pm spaces), which are isomorphism classes of triples (X,r,ν) with $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$. We provide a reconstruction theorem for m2pm spaces and a point separating class of test functions on the set $\mathbb{M}_1^{(2)}$ of m2pm spaces. Furthermore, we introduce the two-level Gromov-weak topology on $\mathbb{M}_1^{(2)}$ as the topology which is induced by these test functions. We define the two-level Gromov-Prokhorov metric on $\mathbb{M}_1^{(2)}$ and show that this metric is complete. Then, we characterize compact subsets with respect to this metric and prove that the two-level Gromov-weak topology is metrized by the two-level Gromov-Prokhorov metric and thus Polish. Finally, we study distributions on $\mathbb{M}_1^{(2)}$. In section 4.4 we extend our arguments to general metric two-level measure spaces (m2m spaces), i. e. isomorphism classes of triples (X,r,ν) with $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. We provide a point separating class of test functions on the set $\mathbb{M}^{(2)}$ of m2m spaces and study the induced topology, which we also call the two-level Gromov-weak topology. Moreover, we extend the two-level Gromov-Prokhorov metric from $\mathbb{M}_1^{(2)}$ to $\mathbb{M}^{(2)}$. We characterize compact nets and compact subsets in $\mathbb{M}^{(2)}$ and eventually prove that the two-level Gromov-Prokhorov metric on $\mathbb{M}^{(2)}$ metrizes the two-level Gromov-weak topology. In the end we also study properties of distributions on $\mathbb{M}^{(2)}$. We characterize tight subsets and show that the test functions which induce the two-level Gromov-weak topology are convergence determining.

4.1 Motivation and basic definitions

In this section we motivate and provide the definition of metric two-level measure spaces (hereinafter abbreviated as m2m spaces). At the end of the section we also explain how we generalize the theory of mm spaces to m2m spaces and discuss the obstacles which arise in the process.

Let us first explain what we mean by a two-level measure.

Definition 4.1 (Two-level measures)
Let X be a Polish metric space. A measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ is called a two-level measure on X. *If additionally $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$, then we say that ν is a* two-level probability measure on X.

We call these measures two-level measures because we think of them as describing a hierarchical structure on X consisting of two different levels. A two-level probability measure ν on X is the distribution of a random probability measure $\mu \in \mathcal{M}_1(X)$. The measure μ represents a group (= distribution) of particles in X and this group is the *first level* of the hierarchical structure. In this sense the measure ν is a distribution of groups and this distribution of groups is the *second level* of the hierarchy.

We want to generalize the theory of metric measure spaces to be able to model biological processes with hierarchical structures, e. g. host-parasite systems, in which the parasite population is distributed between different hosts. For this reason we consider metric two-level measure triples $(X,r,\nu)^{(2)}$, where (X,r) is a Polish metric space and ν a two-level measure on X. In a host-parasite system X would be the parasite population, r would represent the genealogical tree of this population and ν would be a sampling measure on the hosts, which in turn are represented by measures on X (e. g. with ν we can sample hosts $\mu \in \mathcal{M}_f(X)$ and with μ we can sample parasites inside of the host). Because we are only interested in the structure of $(X,r,\nu)^{(2)}$ and not in its particular representation, we go over to isomorphism classes. The definition of an isomorphism should preserve the hierarchical structure of the measure ν, that is why we use the two-level push-forward (cf. section 2.6). We regard two metric two-level measure triples $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ as being isomorphic if there is a function $f\colon X \to Y$ with $\lambda = f_{**}\nu$ such that f is isometric on the "effective support of ν in X". By effective support we mean the smallest closed subset $A \subset X$ such that $\operatorname{supp}\mu \subset A$ for ν-almost every $\mu \in \mathcal{M}_f(X)$. It turns out that this set A coincides with the support $\operatorname{supp}\mathfrak{M}_\nu$ of the first moment measure $\mathfrak{M}_\nu := \int \mu \,\mathrm{d}\nu(\mu)$ (we will prove this in Corollary 4.8).

Definition 4.2 (m2m space and m2m-isomorphism)

(a) *A triple $(X,r,\nu)^{(2)}$ is called a* metric two-level measure triple (m2m triple) *if $X \subset \mathbb{R}^{\mathbb{N}}$ is non-empty, (X,r) is a Polish metric space and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.*

If moreover $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$, then we call $(X,r,\nu)^{(2)}$ a metric two-level probability measure triple (m2pm triple).

(b) *Two m2m triples $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ are called* m2m-isomorphic *if there exists a measurable function $f\colon X \to Y$ such that $\lambda = f_{**}\nu$ and f is isometric on* *the set* $\operatorname{supp}\mathfrak{M}_\nu$ *(but not necessarily on the whole space* X*). The function* f *is called an* m2m-isomorphism. *To denote that both spaces are m2m-isomorphic, we write* $(X,r,\nu)^{(2)} \cong (Y,d,\lambda)^{(2)}$ *or* $(X,r,\nu)^{(2)} \cong_f (Y,d,\lambda)^{(2)}$ *if we want to emphasize that* f *is an m2m-isomorphism.*

(c) *The relation of being m2m-isomorphic is an equivalence relation on the set of m2m triples. The equivalence class of an m2m triple* $(X,r,\mu)^{(2)}$ *is called a* metric two-level measure space (m2m space) *and is denoted by* $[X,r,\nu]^{(2)}$*. The set of all m2m spaces (i. e. the set of equivalence classes of m2m triples) is denoted by* $\mathbb{M}^{(2)}$*.*

If $(X,r,\nu)^{(2)}$ *is an m2pm triple, then we call its equivalence class* $[X,r,\nu]^{(2)}$ *a* metric two-level probability measure space (m2pm space)*. The set of all m2pm spaces (i. e. the set of equivalence classes of m2pm triples) is denoted by* $\mathbb{M}_1^{(2)}$*.*

Generic elements of $\mathbb{M}^{(2)}$ will be denoted by $\mathcal{X} = [X,r,\nu]^{(2)}$, $\mathcal{X}_n = [X_n,r_n,\nu_n]^{(2)}$ or $\mathcal{Y} = [Y,d,\lambda]^{(2)}$.

Remarks 4.3

(a) Recall from section 2.3 that every Polish space is homeomorphic to a subset of $\mathbb{R}^{\mathbb{N}}$. Thus, the assumption $X \subset \mathbb{R}^{\mathbb{N}}$ in Definition 4.2 is not a restriction and the theory developed in this thesis stays valid even for arbitrary Polish metric spaces.

(b) Let $[X,r,\nu]^{(2)}$ be an m2m space with $S := \operatorname{supp}\mathfrak{M}_\nu \neq \emptyset$ and let r_S denote the restriction of r to $S \times S$ and ν_S denote the restriction of ν to $\mathcal{M}_f(S)$. Because the support of ν is a subset of $\{\,\mu \in \mathcal{M}_f(X) \mid \operatorname{supp}\mu \subset S\,\}$ (see Corollary 4.8), we have $[X,r,\nu]^{(2)} = [S,r_S,\nu_S]^{(2)}$. This shows that an m2m space is always m2m-isomorphic to the support of its first moment measure. Therefore, we may assume without loss of generality that $X = \operatorname{supp}\mathfrak{M}_\nu$ and we will do so in many of the proofs of this chapter.

(c) In the definition of m2m triples/spaces we prohibit X from being empty because this would break the symmetry in the definition of m2m-isomorphism. See (c) of Remarks 3.2 for details.

We will first treat the special case of m2pm spaces and only afterwards the more general case of m2m spaces. We decided to do this for the following reasons:

- M2pm spaces are a very important special case. They are useful for modeling population processes with fixed population size and resampling dynamics such as the two-level Fleming-Viot process in [Daw18]. Thus, they deserve a thorough treatment on their own.

- The theory of m2pm spaces is much more accessible and partly a straight-forward generalization of the theory of mm spaces. Readers familiar with metric measure spaces will easily understand the generalization to m2pm spaces.

- The theory for general m2m spaces on the other hand is more complicated. Several new problems arise which do not exists in the m2pm case. By treating the latter case first, we hope that these problems and their solutions become more apparent.

In the remainder of this section we explain how we proceed in section 4.3 (m2pm spaces) and section 4.4 (m2m spaces) and discuss the differences of the two cases. We start with section 4.3:

There, we will provide a reconstruction theorem for m2pm spaces, which states that an m2pm space $[X,r,\nu]^{(2)} \in \mathbb{M}_1^{(2)}$ is uniquely determined by its distance array distribution $R_*\mathfrak{M}_\nu^{\infty,\infty}$, where $\mathfrak{M}_\nu^{\infty,\infty}$ is the infinite mixed moment measure defined by

$$\mathfrak{M}_\nu^{\infty,\infty} := \int \int \bigotimes_{i=1}^{\infty} \mu_i^{\otimes\infty} \,\mathrm{d}\nu^{\otimes\infty}(\mu_1,\mu_2,\dots)$$

and where R maps an infinite matrix $(x_{ij})_{ij} \in X^{\mathbb{N}\times\mathbb{N}}$ to its distance array

$$R((x_{ij})_{ij}) := \big(r(x_{ij},x_{kl})\big)_{i,j,k,l\in\mathbb{N}}. \tag{E4.1}$$

Heuristically, the infinite mixed moment measure samples infinitely many measures $(\mu_i)_i \subset \mathcal{M}_1(X)$ with distribution ν and then samples with each of these measures μ_i infinitely many points $(x_{ij})_j \subset X$. Then, the distance array distribution is the distribution of the random distance array in (E4.1).

We introduce the two-level Gromov-weak topology τ'_{2Gw} on $\mathbb{M}_1^{(2)}$ as the initial topology with respect to a point separating class $\mathcal{T}_1^{(2)}$ of test functions on $\mathbb{M}_1^{(2)}$. The definition of these test functions is based on the idea of sampling finite subspaces of the underlying metric space X. It turns out that τ'_{2Gw}-convergence is the same as weak convergence of the associated distance array distributions, i. e. $(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ is homeomorphic to a subset of $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$, where $\mathbb{D}_{\mathbb{N}\times\mathbb{N}}$ is the set of distance arrays as in (E4.1). It follows that the topology τ'_{2Gw} is metrizable. Hence, we can use sequences to prove continuity of functions and compactness of subsets.

We also define a complete metric $d_{2\mathrm{GP}}$ on $\mathbb{M}_1^{(2)}$, the so-called two-level Gromov-Prokhorov metric, which is a straight-forward generalization of the one-level Gromov-Prokhorov metric. The induced topology on $\mathbb{M}_1^{(2)}$ is denoted by τ'_{2GP}. We characterize compactness with respect to τ'_{2GP} in terms of the distance distribution and the modulus of mass distribution *of the first moment measure* $\mathfrak{M}_\nu$, i. e. in terms of $w(\mathfrak{M}_\nu)$ and $V_\delta(\mathfrak{M}_\nu)$. Finally, we prove that τ'_{2Gw} and τ'_{2GP} coincide by showing that τ'_{2Gw}-convergent sequences are relatively compact with respect to τ'_{2GP}. Therefore, $d_{2\mathrm{GP}}$ is a complete metric for the topology τ'_{2Gw} and $(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ is a Polish space. Moreover, because τ'_{2Gw} is induced by $\mathcal{T}_1^{(2)}$, we can apply Le Cam's theorem (see Proposition 2.17) to show that $\mathcal{T}_1^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}_1^{(2)})$.

In section 4.4 we are concerned with general m2m spaces $[X,r,\nu]^{(2)}$ with measures $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. It turns out that in this case we cannot proceed in the same way as in section 4.3. One of the reasons is that there is no reconstruction theorem for general m2m spaces, meaning that an m2m space $[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)}$ cannot be identified with a distance array distribution (i. e. with a finite measure on distance arrays). The problem is that $\nu(\{o\})$ may be positive and that we cannot sample points from X with the null measure o. However, we can still define a point separating class $\mathcal{T}^{(2)}$ of test functions on $\mathbb{M}^{(2)}$ based on the idea of sampling finite subspaces of the underlying metric space X. We define the two-level Gromov-weak topology τ_{2Gw} on $\mathbb{M}^{(2)}$ as the initial topology induced by $\mathcal{T}^{(2)}$. Since we cannot embed $\mathbb{M}^{(2)}$ as a subset of distance matrix distributions, there is no easy way of proving that $(\mathbb{M}^{(2)}, \tau_{2Gw})$ is metrizable. For this reason we will use nets instead of sequences in order to prove continuity of functions and compactness of subsets.

The two-level Gromov-Prokhorov metric $d_{2\mathrm{GP}}$, which we first define on $\mathbb{M}_1^{(2)}$ in section 4.3, can easily be extended to a complete metric on $\mathbb{M}^{(2)}$. The induced topology on $\mathbb{M}^{(2)}$ is denoted by τ_{2GP}. Then, we characterize compact *nets* with respect to τ_{2GP} in terms of $w(\mathfrak{M}_{f_K\cdot\nu})$ and $V_\delta(\mathfrak{M}_{f_K\cdot\nu})$, where the measures $f_K\cdot\nu$ with $K>0$ are a certain class of approximations of ν. These approximations are necessary because the first moment measure $\mathfrak{M}_\nu$ of a two-level measure ν can be infinite and because the distance distribution and the modulus of mass distribution are only meaningful for finite measures.

Finally, we prove that τ_{2Gw} and τ_{2GP} coincide by showing that τ_{2Gw}-convergent nets are compact nets with respect to τ_{2GP}. It follows that $d_{2\mathrm{GP}}$ is a complete metric for τ_{2Gw} and that $(\mathbb{M}^{(2)}, \tau_{2Gw})$ is a Polish space. Furthermore, we can use Le Cam's theorem to prove that $\mathcal{T}^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}^{(2)})$.

Recall from Example 2.7 that the elements of a compact net do not necessarily form a relatively compact set. Thus, to be able to prove the equality of τ_{2Gw} and τ_{2GP} we really need a characterization of compact nets (instead of a characterization of compact subsets).

Note that we call both topologies τ'_{2Gw} on $\mathbb{M}_1^{(2)}$ and τ_{2Gw} on $\mathbb{M}^{(2)}$ the two-level Gromov-weak topology even though they are induced by different classes of test functions. This is due to the fact that τ'_{2Gw} is just the subspace topology of τ_{2Gw}, as we show in section 4.4. The same holds for the two-level Gromov-Prokhorov topology τ'_{2GP} on $\mathbb{M}_1^{(2)}$ and τ_{2GP} on $\mathbb{M}^{(2)}$.

4.2 Two-level measures and moment measures

In this section we study two-level measures $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ on a fixed Polish space X. We define the moment measures of ν and discuss the interrelationship between ν and its moment measures. At the end of the section we also provide a tightness criterion for (sets of) two-level measures.

Two-level measures are closely related to random measures, because the normalization $\overline{\nu}$ of a two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ is a distribution of a random finite measure on X. Hence, it is not surprising that we use tools from the theory of random measures. One of these tools are the moment measures.

Definition 4.4 (Moment measures)
Let $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ be a two-level measure on a Polish space X and let $m \in \mathbb{N}$ and

$\boldsymbol{n} \in \mathbb{N}^m$.

(a) *The* m-th moment measure $\mathfrak{M}_\nu^m \in \mathcal{M}(X^m)$ of ν *is defined by*

$$\mathfrak{M}_\nu^m(\,\cdot\,) := \int \mu^{\otimes m}(\,\cdot\,)\,\mathrm{d}\nu(\mu).$$

The first moment measure is also called the intensity measure of ν *and will be denoted by* $\mathfrak{M}_\nu$.

(b) *The* $\boldsymbol{n}$-mixed moment measure $\mathfrak{M}_\nu^{\boldsymbol{n}} \in \mathcal{M}(X^{|\boldsymbol{n}|})$ of ν *is defined by*

$$\mathfrak{M}_\nu^{\boldsymbol{n}}(\,\cdot\,) := \int \boldsymbol{\mu}^{\otimes \boldsymbol{n}}(\,\cdot\,)\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}),$$

where $\boldsymbol{\mu}^{\otimes \boldsymbol{n}}$ *denotes the measure* $\boldsymbol{\mu}^{\otimes \boldsymbol{n}} := \bigotimes_{i=1}^m \mu_i^{\otimes n_i} \in \mathcal{M}_f(X^{|\boldsymbol{n}|})$.

(c) *We say that* ν *is* of k-th order *if the k-th moment measure* $\mathfrak{M}_\nu^k$ *is a finite measure. We say that* ν *is* of infinite order *if all moment measures* $(\mathfrak{M}_\nu^k)_{k\in\mathbb{N}}$ *are finite measures.*

By definition we have

$$\mathfrak{M}_\nu^{\boldsymbol{n}} = \bigotimes_{i=1}^m \mathfrak{M}_\nu^{n_i},$$

i. e. $\mathfrak{M}_\nu^{\boldsymbol{n}}$ is a mix of different moment measures. That is why we call $\mathfrak{M}_\nu^{\boldsymbol{n}}$ a mixed moment measure.

If ν is a two-level probability measure, the m-th moment measures $\mathfrak{M}_\nu^m$ describes a random experiment in which we first sample a measure $\mu \in \mathcal{M}_1(X)$ with ν and then sample m points in X with distribution μ. The $\boldsymbol{n}$-mixed moment measure corresponds to first sampling m measures $\mu_1, \ldots, \mu_m \in \mathcal{M}_1(X)$ and then for each $i \in \{1, \ldots, m\}$ sampling n_i points in X with distribution μ_i.

Obviously, every two-level probability measure is of infinite order. For these measures we can also define the infinite (mixed) moment measure, which is the projective limit of the (mixed) moment measures.

Definition 4.5 (Infinite moment measures)
Let $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ *be a two-level probability measure on a Polish space* X.

(a) *The* infinite moment measure $\mathfrak{M}_\nu^\infty \in \mathcal{M}_1(X^{\mathbb{N}})$ of ν *is defined by*

$$\mathfrak{M}_\nu^\infty(\,\cdot\,) := \int \mu^{\otimes\infty}(\,\cdot\,)\,\mathrm{d}\nu(\mu),$$

(b) *The* infinite mixed moment measure $\mathfrak{M}_\nu^{\infty,\infty} \in \mathcal{M}_1(X^{\mathbb{N}\times\mathbb{N}})$ of ν *is defined by*

$$\mathfrak{M}_\nu^{\infty,\infty}(\,\cdot\,) := \int \boldsymbol{\mu}^{\otimes\infty}(\,\cdot\,)\,\mathrm{d}\nu^{\otimes\infty}(\boldsymbol{\mu}),$$

where $\boldsymbol{\mu}^{\otimes\infty}$ *denotes the measure* $\boldsymbol{\mu}^{\otimes\infty} := \bigotimes_{i=1}^\infty \mu_i^{\otimes\infty} \in \mathcal{M}_1(X^{\mathbb{N}\times\mathbb{N}})$.

Informally speaking, the mass of a two-level measure is distributed between the two levels and the first moment measure can be seen as a projection of this mass to a one-level measure.

Remark 4.6 (First moment measure as a projection)
The first moment measure is a projection in the following sense: $\mathcal{M}_1(X)$ can be identified with $\{\,\delta_\mu \mid \mu \in \mathcal{M}_1(X)\,\}$, which is a closed subspace of $\mathcal{M}_1(\mathcal{M}_1(X))$. Then, the map

$$\begin{aligned} P\colon \mathcal{M}_1(\mathcal{M}_1(X)) &\to \mathcal{M}_1(\mathcal{M}_1(X)) \\ \nu &\mapsto \delta_{\mathfrak{M}_\nu} \end{aligned}$$

is a projection onto this subspace, i. e. we have $P \circ P = P$.

However, for general two-level measure this is not possible anymore since the first moment measure of a two-level measure is not necessarily finite.

Recall that the support $\operatorname{supp}\mu$ of a Borel measure μ is defined as the smallest closed subset A of X with $\mu(\complement A) = 0$. Equivalently, it is the set of all $x \in X$ with $\mu(B(x,\varepsilon)) > 0$ for every $\varepsilon > 0$. The following two results show that a two-level measure ν is effectively a finite measure on $\mathcal{M}_f(\operatorname{supp}\mathfrak{M}_\nu)$, i. e. we can restrict X to $\operatorname{supp}\mathfrak{M}_\nu$ without losing information about ν.

Lemma 4.7 (First moment measure is a supporting measure)
Let X be a Polish space and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. The first moment measure $\mathfrak{M}_\nu$ is a supporting measure *of ν in the sense that for any non-negative measurable function $f\colon X \to \mathbb{R}$ we have*

$$\int f \,\mathrm{d}\mathfrak{M}_\nu = 0 \iff \int f \,\mathrm{d}\mu = 0 \text{ for } \nu\text{-almost every } \mu \in \mathcal{M}_f(X).$$

Proof: By the definition of the moment measures we have

$$\int f \,\mathrm{d}\mathfrak{M}_\nu = \int\!\!\int f \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu).$$

Now if $\int f \,\mathrm{d}\mathfrak{M}_\nu = 0$, then for every $n \in \mathbb{N}$ the set of all $\mu \in \mathcal{M}_f(X)$ with $\int f \,\mathrm{d}\mu > \frac{1}{n}$ must have ν-measure zero. Consequently, $\int f \,\mathrm{d}\mu = 0$ for ν-almost every $\mu \in \mathcal{M}_f(X)$.

On the other hand, if $\nu(\complement A) = 0$ for $A := \{\,\mu \in \mathcal{M}_f(X) \mid \int f \,\mathrm{d}\mu = 0\,\}$, then

$$\int_X f \,\mathrm{d}\mathfrak{M}_\nu = \int_{\mathcal{M}_f(X)} \int_X f \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu) = \int_A \int_X f \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu) = 0. \qquad \square$$

Corollary 4.8
Let X be a Polish space and $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. Then $A := \operatorname{supp}\mathfrak{M}_\nu$ is the smallest closed subset of X with the property

$$\operatorname{supp}\mu \subset A \text{ for } \nu\text{-almost every } \mu \in \mathcal{M}_f(X). \tag{E4.2}$$

Proof: Let A be any closed subset of X. Using Lemma 4.7 we see that the following assertions are all equivalent:

$$\begin{aligned} &\mathfrak{M}_\nu(\complement A)=0 \\ \iff &\int \mathbb{1}_{(\complement A)}\,\mathrm{d}\mathfrak{M}_\nu=0 \\ \iff &\int \mathbb{1}_{(\complement A)}\,\mathrm{d}\mu=0 \text{ for } \nu\text{-almost every } \mu\in\mathcal{M}_f(X) \\ \iff &\operatorname{supp}\mu\subset A \text{ for } \nu\text{-almost every } \mu\in\mathcal{M}_f(X). \end{aligned}$$

This yields the claim, since the set $\operatorname{supp}\mathfrak{M}_\nu$ is defined as the smallest closed set $A\subset X$ with $\mathfrak{M}_\nu(\complement A)=0$. □

Now the question arises when a two-level measure is determined by its moment measures. This problem is analogously to the classical moment problem for distributions on $\mathbb{R}$ (see for example [ST43] for an extensive treatment about the classical moment problem). The following example shows that the moment measures may not contain enough information about the associated two-level measure. Let $(\varepsilon_n)_n$ be a sequence of positive real numbers with $\sum_{n=1}^{\infty}\varepsilon_n=1$ and define

$$\nu:=\sum_{n=1}^{\infty}\varepsilon_n\delta_{(\varepsilon_n^{-1}\delta_1)}.$$

Then we have $\mathfrak{M}_\nu^k=\infty\cdot\delta_1$ for every $k\in\mathbb{N}$. This holds true for every sequence $(\varepsilon_n)_n$ with $\sum\varepsilon_n=1$. Hence, different two-level measures may have the same moment measures.

The following proposition shows that a two-level measures is determined by its moment measures provided that the mass of the moment measures grows slow enough. It carries over the Carleman condition for distributions on $\mathbb{R}_+$ (see for example [ST43, Theorem 1.10]) to two-level measures.

Proposition 4.9 (Carleman condition for two-level measures)
Let $\nu\in\mathcal{M}_f(\mathcal{M}_f(X))$ be a two-level measure on a Polish space X. If

$$\sum_{k=1}^{\infty}\mathfrak{m}(\mathfrak{M}_\nu^k)^{-\frac{1}{2k}}=\infty, \tag{E4.3}$$

then ν is determined by its moment measures $(\mathfrak{M}_\nu^k)_{k\in\mathbb{N}}$, i. e. if λ is a two-level measure on X with $\mathfrak{M}_\lambda^k=\mathfrak{M}_\nu^k$ for all $k\in\mathbb{N}$, then $\lambda=\nu$.

Proof: A version of this proposition with $\nu\in\mathcal{M}_1(\mathcal{M}_f(X))$ can be found in [Daw93, Theorem 3.2.9] (based on [Zes83, Theorem 2.1]). This extends directly to measures $\nu\in\mathcal{M}_f(\mathcal{M}_f(X))$, since ν is just a multiple of the measure $\overline{\nu}\in\mathcal{M}_1(\mathcal{M}_f(X))$ and multiplication by a factor does not change the divergence of the series in (E4.3). □

Observe that

$$\mathfrak{m}(\mathfrak{M}^k_\nu) = \int \mathfrak{m}(\mu)^k \,\mathrm{d}\nu(\mu) = \int m^k \,\mathrm{d}(\mathfrak{m}_*\nu)(m), \tag{E4.4}$$

where $\mathfrak{m}_*\nu$ is the push-forward of ν under the function $\mathfrak{m}\colon \mu \in \mathcal{M}_f(X) \mapsto \mathfrak{m}(\mu) \in \mathbb{R}_+$. We call the measure $\mathfrak{m}_*\nu \in \mathcal{M}_f(\mathbb{R}_+)$ the *mass distribution of* ν. The mass distribution should not be confused with the modulus of mass distribution, which is, despite the similar name, not related. By (E4.4) there is a one-to-one correspondence between the mass of the moment measures and the moments of the mass distribution.

It is not surprising that there is a strong relationship between weak convergence of two-level measures and weak convergence of the associated moment measures. The following proposition states under which conditions one implies the other.

Proposition 4.10
Let $(\nu_n)_n \subset \mathcal{M}_f(\mathcal{M}_f(X))$ be a sequence of two-level measures on a Polish space X, such that each ν_n is of infinite order.

(a) If for each $k \in \mathbb{N}$ the k-th moment measures $(\mathfrak{M}^k_{\nu_n})_n$ converge weakly to a finite measure μ_k on X^k and if

$$\sum_{k=1}^{\infty} \mathfrak{m}(\mu_k)^{-\frac{1}{2k}} = \infty,$$

then ν_n converges weakly to a two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ with moment measures $\mathfrak{M}^k_\nu = \mu_k$ for each $k \in \mathbb{N}$.

(b) If ν_n converges weakly to some two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ and if

$$\sup_{n\in\mathbb{N}} \mathfrak{m}(\mathfrak{M}^k_{\nu_n}) < \infty$$

for each $k \in \mathbb{N}$, then ν is of infinite order and the moment measures of ν_n converge weakly to the moment measures of ν.

Proof: See [Daw93, Theorem 3.2.9] and [Zes83, Theorem 2.2] for versions of this proposition with $\nu, \nu_n \in \mathcal{M}_1(\mathcal{M}_f(X))$. This extends directly to measures $\nu, \nu_n \in \mathcal{M}_f(\mathcal{M}_f(X))$ (cf. Lemma 2.14 and the proof of Proposition 4.9). □

Note that for two-level probability measures all of the conditions in Propositions 4.9 and 4.10 are fulfilled. Thus a two-level probability measure is always determined by its moment measures and weak convergence is equivalent to weak convergence of the moment measures. However, in this thesis we will also be concerned with general two-level measures which do not necessarily satisfy these conditions.

Because the moment measures are important for our proofs (for example to sample points from the underlying space X) we need to approximate two-level measures $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ by measures ν_K from $\mathcal{M}_f(\mathcal{M}_{\le K}(X))$ to ensure that the moment measures are finite and do not grow to fast. The simplest choice for ν_K would be the restriction of ν to $\mathcal{M}_{\le K}(X)$, i. e. $\nu_K(\cdot) = \nu(\cdot \cap \mathcal{M}_{\le K}(X))$. However, this rough cut-off may lead to discontinuities if $\nu(\mathcal{M}_K(X))$ is greater than 0. Therefore it is better to "cut off" ν with a continuous density f_K as defined below.

Let $\{ g_K \in \mathcal{C}_b(\mathbb{R}_+) \mid K > 0 \}$ be a set of functions having the following properties:

(a) $0 \le g_K \le 1$ for every K,

(b) $g_K(x) = 0$ for $x \ge K$,

(c) $g_K \to 1$ for $K \to \infty$ uniformly on every bounded interval.

For example we may use

$$g_K(x) := \begin{cases} 1 & 0 \le x \le \frac{K}{2} \\ 2 - \frac{2x}{K} & \frac{K}{2} < x \le K \\ 0 & K < x. \end{cases}$$

However, the particular choice of the functions is not important as long as the properties (a), (b) and (c) are satisfied and as long as the choice is fixed for the remainder of this thesis.

Furthermore, define $f_K := g_K \circ \mathfrak{m}$. For a two-level measure $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$ we denote by $f_K \cdot \nu$ the measure that has density f_K with respect to ν. It is the unique measure which satisfies

$$(f_K \cdot \nu)(A) = \int_A f_K(\mu)\, \mathrm{d}\nu(\mu) = \int_A g_K(\mathfrak{m}(\mu))\, \mathrm{d}\nu(\mu) \tag{E4.5}$$

for every measurable set $A \subset \mathcal{M}_f(X)$. The measures $\{ f_K \cdot \nu \mid K > 0 \}$ will serve as approximations of ν. The next lemma describes the relationship between ν and $f_K \cdot \nu$.

Lemma 4.11
Let $\nu, \nu_1, \nu_2, \ldots \in \mathcal{M}_f(\mathcal{M}_f(X))$ be two-level measures on a Polish space X. Then we have:

(a) $f_K \cdot \nu \xrightarrow{w} \nu$ for $K \to \infty$ and ν is uniquely determined by the measures $(f_K \cdot \nu)_{K>0}$, i. e. if λ is a two-level measure on X with $f_K \cdot \lambda = f_K \cdot \nu$ for all $K > 0$, then $\lambda = \nu$.

(b) ν_n converges weakly to ν if and only if
- *the set of mass distributions $(\mathfrak{m}_* \nu_n)_n$ is relatively compact and*
- *$f_K \cdot \nu_n$ converges weakly to $f_K \cdot \nu$ for every $K > 0$.*

Proof: By the dominated convergence theorem we have

$$\int h \, \mathrm{d}(f_K \cdot \nu) = \int h f_K \, \mathrm{d}\nu = \int h(\mu)\, g_K(\mathfrak{m}(\mu))\, \mathrm{d}\nu(\mu) \xrightarrow{K \to \infty} \int h(\mu)\, \mathrm{d}\nu(\mu)$$

for every $h \in \mathcal{C}_b(\mathcal{M}_f(X))$ and this shows assertion (a).

To prove assertion (b), assume first that ν_n converges weakly to ν. Since the density functions $(f_K)_{K>0}$ are continuous and bounded on $\mathcal{M}_f(X)$, it follows immediately that $f_K \cdot \nu_n$ converges weakly to $f_K \cdot \nu$ for every $K > 0$. Moreover, $\mathfrak{m}_*\nu_n$ converges weakly to $\mathfrak{m}_*\nu$ by assertion (a) of Lemma 2.24. Hence $(\mathfrak{m}_*\nu_n)_n$ is relatively compact.

To prove the other direction, assume that $(\mathfrak{m}_*\nu_n)_n$ is relatively compact and that $f_K \cdot \nu_n$ converges weakly to $f_K \cdot \nu$ for every $K > 0$. Let $\varphi \in \mathcal{C}_b(\mathcal{M}_f(X))$ and $\varepsilon > 0$ be arbitrary. By the triangle inequality we have

$$\begin{aligned}\left|\int \varphi \,\mathrm{d}\nu_n - \int \varphi \,\mathrm{d}\nu\right| &\leq \left|\int \varphi \,\mathrm{d}\nu_n - \int \varphi \,\mathrm{d}(f_K \cdot \nu_n)\right| + \left|\int \varphi \,\mathrm{d}(f_K \cdot \nu_n) - \int \varphi \,\mathrm{d}(f_K \cdot \nu)\right| \\ &\quad + \left|\int \varphi \,\mathrm{d}(f_K \cdot \nu) - \int \varphi \,\mathrm{d}\nu\right| \\ &\leq \|\varphi\|_\infty \int |1 - f_K| \,\mathrm{d}\nu_n + \left|\int \varphi \,\mathrm{d}(f_K \cdot \nu_n) - \int \varphi \,\mathrm{d}(f_K \cdot \nu)\right| \\ &\quad + \|\varphi\|_\infty \int |1 - f_K| \,\mathrm{d}\nu,\end{aligned} \tag{E4.6}$$

where $\|\varphi\|_\infty$ denotes the supremum of φ. Because $(\mathfrak{m}_*\nu_n)_n$ is relatively compact, $\mathfrak{m}(\nu)$ and $(\mathfrak{m}(\nu_n))_n$ are bounded by some constant $M > 0$ and there is a $b > 0$ with

$$\nu\left(\complement \mathcal{M}_{\leq b}(X)\right) = \mathfrak{m}_*\nu\left(\complement[0,b]\right) < \varepsilon$$

and

$$\nu_n\left(\complement \mathcal{M}_{\leq b}(X)\right) = \mathfrak{m}_*\nu_n\left(\complement[0,b]\right) < \varepsilon$$

for every $n \in \mathbb{N}$. Since g_K converges to 1 uniformly on bounded sets, there exists a $K > 0$ such that $|1 - g_K(x)| < \varepsilon$ for $x \in [0,b]$. It follows that

$$\begin{aligned}\int |1 - f_K| \,\mathrm{d}\nu &= \int_{\mathcal{M}_{\leq b}(X)} |1 - f_K| \,\mathrm{d}\nu + \int_{\complement \mathcal{M}_{\leq b}(X)} |1 - f_K| \,\mathrm{d}\nu \\ &\leq \varepsilon\,\mathfrak{m}(\nu) + \nu(\complement \mathcal{M}_{\leq b}(X)) \\ &\leq \varepsilon(M+1)\end{aligned}$$

and the same upper bound holds if we replace ν by ν_n. Finally, since $f_K \cdot \nu_n$ converges weakly to $f_K \cdot \nu$, there is an $n_\varepsilon \in \mathbb{N}$ such that

$$\left|\int \varphi \,\mathrm{d}(f_K \cdot \nu_n) - \int \varphi \,\mathrm{d}(f_K \cdot \nu)\right| < \varepsilon$$

for all $n \geq n_\varepsilon$. Thus, the right hand side of (E4.6) is less than $\varepsilon\,(2\|\varphi\|_\infty(M+1)+1)$ for all $n \geq n_\varepsilon$. Because ε and φ are arbitrary, this shows that $\int \varphi\nu_n$ converges to $\int \varphi\nu$ for every $\varphi \in \mathcal{C}_b(\mathcal{M}_f(X))$. □

Observe that every measure $\lambda \in \mathcal{M}_f(\mathcal{M}_{\leq K}(X))$ satisfies

$$\mathfrak{m}(\mathfrak{M}^m_\lambda) \leq K^m \mathfrak{m}(\lambda)$$

for every $m \in \mathbb{N}$ and thus

$$\sum_{m=1}^{\infty} \mathfrak{m}(\mathfrak{M}_{\lambda}^{m})^{-\frac{1}{2m}} \geq \sum_{m=1}^{\infty} K^{-\frac{1}{2}} \mathfrak{m}(\lambda)^{-\frac{1}{2m}} = \infty.$$

It follows from Proposition 4.9 and 4.10 that every measure $\lambda \in \mathcal{M}_f(\mathcal{M}_{\leq K}(X))$ is uniquely determined by its moment measures and weak convergence of measures of $\mathcal{M}_f(\mathcal{M}_{\leq K}(X))$ is equivalent to weak convergence of the associated moment measures. Since $f_K \cdot \nu \in \mathcal{M}_f(\mathcal{M}_{\leq K}(X))$ for every two-level measure ν we can rephrase Lemma 4.11 in terms of moment measures.

Corollary 4.12
Let $\nu, \nu_1, \nu_2, \ldots \in \mathcal{M}_f(\mathcal{M}_f(X))$ be two-level measures on a Polish space X. Then we have:

(a) ν is uniquely determined by the moment measures $\{\mathfrak{M}_{f_K \cdot \nu}^{m} \mid K > 0,\ m \in \mathbb{N}\}$, i. e. if λ is a two-level measure on X such that $\mathfrak{M}_{f_K \cdot \lambda}^{m} = \mathfrak{M}_{f_K \cdot \nu}^{m}$ for all $K > 0$ and $m \in \mathbb{N}$, then $\lambda = \nu$.

(b) ν_n converges weakly to ν if and only if

- *$(\mathfrak{m}_* \nu_n)_n$ is relatively compact and*
- *$\mathfrak{M}_{f_K \cdot \nu_n}^{m}$ converges weakly to $\mathfrak{M}_{f_K \cdot \nu}^{m}$ for every $K > 0$ and every $m \in \mathbb{N}$.*

We close this section with a characterization of tightness of two-level measures in terms of the mass distributions and compact sets of the underlying space X. This result is an extension of a characterization of tightness of two-level *probability* measures in [Kle08, Exercise 13.4.1].

Lemma 4.13 (Tightness of two-level measures)
Let $\Gamma \subset \mathcal{M}_f(\mathcal{M}_f(X))$ be a set of two-level measures on a Polish space X. The following are equivalent:

(a) Γ is tight.

(b) $\{\mathfrak{m}_* \nu \mid \nu \in \Gamma\}$ is tight in $\mathcal{M}_f(\mathbb{R}_+)$ and for all $\varepsilon, \delta > 0$ there is a compact set $K_{\varepsilon,\delta} \subset X$ such that

$$\nu\left(\complement\{\mu \in \mathcal{M}_f(X) \mid \mu\left(\complement K_{\varepsilon,\delta}\right) \leq \delta\}\right) < \varepsilon$$

for every $\nu \in \Gamma$.

(c) For every $\varepsilon > 0$ and $\delta > 0$ there exists a $C_{\varepsilon,\delta} > 0$ and a compact set $K_{\varepsilon,\delta} \subset X$, such that

$$\nu\left(\complement\{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \leq C_{\varepsilon,\delta},\ \mu\left(\complement K_{\varepsilon,\delta}\right) \leq \delta\}\right) < \varepsilon$$

for every $\nu \in \Gamma$.

Remark 4.14

(a) The lemma above holds true if we restrict the statements in (b) and (c) to the special case $\delta = \varepsilon$.

(b) By Prokhorov's theorem (Theorem 2.15) a set $\Gamma \subset \mathcal{M}_f(\mathcal{M}_f(X))$ is relatively compact if and only if $\sup_{\nu \in \Gamma} \mathfrak{m}(\nu) < \infty$ and one of the conditions of the lemma above are satisfied.

Proof (of Lemma 4.13): (a) $\Rightarrow$ (b): Fix ε and δ. Because Γ is tight, there is a compact set $\mathcal{K}_\varepsilon \subset \mathcal{M}_f(X)$ with $\nu(\complement \mathcal{K}_\varepsilon) < \varepsilon$ for every $\nu \in \Gamma$. By Prokhorov's theorem (Proposition 2.15) there is a constant $C_{\varepsilon,\delta} > 0$ and a compact set $K_{\varepsilon,\delta} \subset X$ such that

$$\mathcal{K}_\varepsilon \subset \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon,\delta},\ \mu(\complement K_{\varepsilon,\delta}) \le \delta \,\}.$$

Consequently, we have for every $\nu \in \Gamma$

$$\nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mu(\complement K_{\varepsilon,\delta}) \le \delta \,\}) \le \nu(\complement \mathcal{K}_\varepsilon) < \varepsilon$$

and

$$\mathfrak{m}_* \nu(\complement [0, C_{\varepsilon,\delta}])) = \nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon,\delta} \,\}) \le \nu(\complement \mathcal{K}_\varepsilon) < \varepsilon.$$

Since ε is arbitrary, the latter inequality implies that $\{\, \mathfrak{m}_\nu \mid \nu \in \Gamma \,\}$ is tight.

(b) $\Rightarrow$ (c): Fix $\delta, \varepsilon > 0$. By assumption there is a $C > 0$ and a compact set $K \subset X$ with

$$\nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C \,\}) = \mathfrak{m}_* \nu(\complement [0, C]) < \frac{\varepsilon}{2}$$

and

$$\nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mu(\complement K) \le \delta \,\}) < \frac{\varepsilon}{2}$$

for every $\nu \in \Gamma$. We obtain

$$\begin{aligned} &\nu\big(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C,\ \mu(\complement K) \le \delta \,\}\big) \\ &\quad \le \nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C \,\}) \\ &\quad\quad + \nu(\complement \{\, \mu \in \mathcal{M}_f(X) \mid \mu(\complement K) \le \delta \,\}) \\ &\quad < \varepsilon. \end{aligned}$$

(c) $\Rightarrow$ (a): Fix $\varepsilon > 0$ and define $\varepsilon_n = \delta_n := (\frac{1}{2})^n \varepsilon$. Moreover, define

$$\mathcal{K}_\varepsilon := \bigcap_{n \in \mathbb{N}} \{\, \mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon_n,\delta_n},\ \mu(\complement K_{\varepsilon_n,\delta_n}) \le \delta_n \,\},$$

where $C_{\varepsilon_n,\delta_n}$ and $K_{\varepsilon_n,\delta_n}$ are as in assertion (c). Then we have for every $\nu \in \Gamma$

$$\begin{aligned}\nu(\complement\mathcal{K}_\varepsilon) &= \nu\left(\bigcup_{n\in\mathbb{N}} \complement\{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon_n,\delta_n},\ \mu(\complement K_{\varepsilon_n,\delta_n}) \le \delta_n)\}\right) \\ &\le \sum_{n\in\mathbb{N}} \nu(\complement\{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon_n,\delta_n},\ \mu(\complement K_{\varepsilon_n,\delta_n}) \le \delta_n)\}) \\ &\le \sum_{n\in\mathbb{N}} \varepsilon_n = \varepsilon.\end{aligned}$$

It is easy to see that $\mathcal{K}_\varepsilon$ is tight and bounded in $\mathcal{M}_f(X)$ and therefore relatively compact. By the Portmanteau theorem the set

$$\{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \le C_{\varepsilon_n,\delta_n},\ \mu(\complement K_{\varepsilon_n,\delta_n}) \le \delta_n\}$$

is closed for every n. It follows that $\mathcal{K}_\varepsilon$ is also closed (as an intersection of closed sets) and hence compact. □

4.3 Metric two-level probability measure spaces

In this section we study metric two-level probability measure spaces (m2pm spaces). Recall that an m2pm space is an isomorphism class $[X, r, \nu]^{(2)}$, where (X, r) is a Polish metric space and $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ a two-level probability measure on X, and that $\mathbb{M}_1^{(2)}$ denotes the set of all m2pm spaces.

4.3.1 The reconstruction theorem for m2pm spaces

We first provide a reconstruction theorem for m2pm spaces similar to the reconstruction theorem for mm spaces. We show how an m2pm space $[X, r, \nu]^{(2)} \in \mathbb{M}_1^{(2)}$ can be rebuild from distances between randomly sampled points in X. The points from X are sampled according to the infinite mixed moment measures.

The Glivenko-Cantelli theorem states that a probability measure μ can be reconstructed from an infinite i. i. d. sample. A similar result holds true if μ is a *random* probability measure, as can be seen from the following proposition, which is based on [Ald85, page 14]. Recall from page 25 that $\Xi_n(\boldsymbol{x})$ and $\Xi_\infty(\boldsymbol{x})$ denote the n-th empirical distribution and the infinite empirical distribution of a sequence $\boldsymbol{x}$.

Proposition 4.15 (Glivenko-Cantelli for random probability measures) *Let $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ be a two-level probability measure on a Polish space X and let $\boldsymbol{x} = (x_i)_i \in X^{\mathbb{N}}$ be a random sequence with law $\mathfrak{M}_\nu^\infty$. Then the weak limit*

$$\mu := \underset{n\to\infty}{\text{w-lim}}\, \Xi_n(\boldsymbol{x})$$

exists almost surely and the random probability measure μ has law ν.

Proof: Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let $\boldsymbol{x}\colon \Omega \to X^{\mathbb{N}}$ be an infinite random sequence with law $\mathfrak{M}_\nu^\infty$. The sequence $\boldsymbol{x}$ is exchangeable. Thus, by de Finetti's theorem (see [Ald85, Theorem 3.1]) there is a random probability measure $\mu\colon \Omega \to \mathcal{M}_1(X)$, $\omega \mapsto \mu_\omega$ such that $\mu_\cdot^{\otimes\infty}$ is a regular conditional distribution of $\boldsymbol{x}$ given μ. Using the disintegration theorem (cf. [Ald85, Equation (2.4)]) and the Glivenko-Cantelli theorem (cf. Proposition 3.3) we see that

$$\mathbb{P}(\Xi_\infty(\boldsymbol{x}) = \mu|\mu)(\omega) = \int \mathbb{1}_{\mu_\omega}(\Xi_\infty(\boldsymbol{x}))\, \mathrm{d}\mu_\omega^{\otimes\infty}(\boldsymbol{x}) = 1$$

for $\mathbb{P}$-almost every $\omega \in \Omega$. Therefore we have

$$\mathbb{P}(\Xi_\infty(\boldsymbol{x}) = \mu) = 1$$

and the weak limit $\text{w-lim}_n \Xi_n(\boldsymbol{x})$ exists almost surely and has law $\mathfrak{L}(\mu)$. Observe that all moment measure of $\mathfrak{L}(\mu)$ and ν are equal (because the law of $\boldsymbol{x}$ is $\mathfrak{M}_\nu^\infty$). Since two-level probability measures are uniquely determined by their moment measures (see Proposition 4.9), we have $\mathfrak{L}(\mu) = \nu$ and this completes the proof. □

Remark 4.16
Proposition 4.15 implies that

$$\int F(\mu)\, \mathrm{d}\nu(\mu) = \int \int F(\Xi_\infty(\boldsymbol{x}))\, \mathrm{d}\mu^{\otimes\infty}(\boldsymbol{x})\, \mathrm{d}\nu(\mu)$$

for every bounded Borel function $F \in B(\mathcal{M}_1(X))$.

The previous proposition shows that if we sample a measure μ with a two-level probability measure ν and then sample an i. i. d. sequence $\boldsymbol{x}$ with μ, then we can rebuild the measure μ directly from the sample $\boldsymbol{x}$. Our goal is now to reconstruct ν from a sample in X. To this end, we need not one sampled measure μ, but an i. i. d. sequence of measures $(\mu_i)_i$. If we sample an infinite i. i. d. sample $(x_{ij})_j$ with each measure μ_i, then, from all of the sampled points we can reconstruct first the measures μ_i and then the two-level measure ν. The distribution of the matrix $(x_{ij})_{ij} \in X^{\mathbb{N}\times\mathbb{N}}$ is now given by the infinite mixed moment measure $\mathfrak{M}_\nu^{\infty,\infty}$.

Proposition 4.17 (Reconstruction of two-level probability measures)
Let $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ be a two-level probability measure on a non-empty Polish space X and let $(x_{ij})_{i,j\in\mathbb{N}}$ be a random infinite matrix with distribution $\mathfrak{M}_\nu^{\infty,\infty}$. Then almost surely we have:

(a) For every $i \in \mathbb{N}$ the weak limit

$$\mu_i := \underset{n}{\text{w-lim}}\, \Xi_n((x_{ij})_j) = \underset{n}{\text{w-lim}}\, \frac{1}{n}\sum_{j=1}^{n} \delta_{x_{ij}}$$

exists and the random probability measure μ_i has law ν.

(b) *The two-level measure* $\Xi_n((\mu_i)_i) = \frac{1}{n}\sum_{i=1}^n \delta_{\mu_i}$ *converges weakly to* ν.

(c) *The sequence* $(x_{ij})_i$ *is dense in* $\operatorname{supp}\mathfrak{M}_\nu$ *for every* $j \in \mathbb{N}$.

Proof: Assertion (a) follows from Proposition 4.15 since the random sequence $(x_{ij})_j$ has law $\mathfrak{M}_\nu^\infty$ for every $i \in \mathbb{N}$. The sequence of random measures $(\mu_i)_i$ is therefore an i. i. d. sequence in $\mathcal{M}_1(X)$ sampled by ν. By applying the Glivenko-Cantelli theorem we obtain assertion (b). To see assertion (c), observe that $(x_{ij})_i$ is an i. i. d. sequence in X sampled by the measure $\mathfrak{M}_\nu$ and that an i. i. d. sequence is almost surely dense in the support of the sampling measure by Corollary 3.4. □

The previous proposition implies that we can reconstruct the two-level measure ν *and* the metric space (X,r) even if we only know the distances $r(x_{ij},x_{kl})$ between the points $(x_{ij})_{ij}$ sampled by the infinite mixed moment measure. We exploit this fact in the reconstruction theorem for m2pm spaces. In order to formulate this theorem we need to introduce generalizations of distance matrices and distance maps suitable for two-level sampling. The distances between the points x_{ij} are now contained in the upper half of the *array* $(r(x_{ij},x_{kl}))_{i,j,k,l\in\mathbb{N}}$.

Definition 4.18 (Sets of distance arrays)
For every $m \in \mathbb{N}$ *and* $\boldsymbol{n} \in \mathbb{N}^m$ *with* $|\boldsymbol{n}| \geq 2$ *we define*

$$\mathbb{D}_{\boldsymbol{n}} := \left\{ \left(r_{ij,kl}\right)_{\substack{(i,j),(k,l)\in[\boldsymbol{n}] \\ (i,j)<(k,l)}} \in \mathbb{R}_+^{\binom{|\boldsymbol{n}|}{2}} \;\middle|\; r_{ij,kl} \leq r_{ij,gh} + r_{gh,kl} \;\forall (i,j),(k,l),(g,h) \in [\boldsymbol{n}] \right\},$$

where $[\boldsymbol{n}] := \{ (i,j) \mid i \in [m], j \in [n_i] \}$ *and where* $\mathbb{N}\times\mathbb{N}$ *is ordered lexicographically. Furthermore, we define*

$$\mathbb{D}_{\mathbb{N}\times\mathbb{N}} := \left\{ \left(r_{ij,kl}\right)_{\substack{(i,j),(k,l)\in\mathbb{N}^2 \\ (i,j)<(k,l)}} \in \mathbb{R}_+^{\binom{\mathbb{N}\times\mathbb{N}}{2}} \;\middle|\; r_{ij,kl} \leq r_{ij,gh} + r_{gh,kl} \;\forall (i,j),(k,l),(g,h) \in \mathbb{N}\times\mathbb{N} \right\}.$$

Elements of these sets are called distance arrays. *We equip these set with the subspace topology inherited from the product topologies on* $\mathbb{R}_+^{\binom{|\boldsymbol{n}|}{2}}$ *and* $\mathbb{R}_+^{\binom{\mathbb{N}\times\mathbb{N}}{2}}$*, respectively.*

Occasionally we regard the upper half distance arrays from the previous definition as full arrays and use the entries $r_{ij,ij}$ or $r_{kl,ij}$ with $(i,j)<(k,l)$. It should be clear that in this case we assume $r_{ij,ij} = 0$ and $r_{kl,ij} = r_{ij,kl}$.

Definition 4.19 (Distance maps)
For a metric space (X,r)*,* $m \in \mathbb{N}$ *and* $\boldsymbol{n} = (n_1,\dots,n_m) \in \mathbb{N}^m$ *with* $|\boldsymbol{n}| \geq 2$ *we define the following* distance maps*:*

$$\begin{aligned} R_{\boldsymbol{n}}^X : X^{\boldsymbol{n}} &\to \mathbb{D}_{\boldsymbol{n}} \\ \boldsymbol{x} &\mapsto (r(x_{ij},x_{kl}))_{(i,j),(k,l)\in[\boldsymbol{n}]} \end{aligned}$$

and

$$\begin{aligned} R^X_{\mathbb{N}\times\mathbb{N}}\colon X^{\mathbb{N}\times\mathbb{N}} &\to \mathbb{D}_{\mathbb{N}\times\mathbb{N}} \\ \boldsymbol{x} &\mapsto (r(x_{ij},x_{kl}))_{(i,j),(k,l)\in\mathbb{N}^2} \end{aligned}$$

For convenience we often suppress the super- and subscript in the distance maps above and simply write R or R^X instead of $R^X_{\boldsymbol{n}}$ and $R^X_{\mathbb{N}\times\mathbb{N}}$ when the space and dimension are clear from the context.

It is not surprising that the reconstruction theorem for m2pm spaces resembles the reconstruction theorem for mm spaces (see Proposition 3.7). The difference is that now we use the mixed moment measures to sample points from the underlying metric space X of an m2pm space $[X,r,\nu]^{(2)}$. To rebuild the m2pm space from the distances of the sampled points, we need to reconstruct the first and the second level of ν simultaneously.

Theorem 4.20 (Reconstruction theorem for m2pm spaces)
Let $[X,r,\nu]^{(2)}$ and $[Y,d,\lambda]^{(2)}$ be m2pm spaces. The following are equivalent:

(a) $[X,r,\nu]^{(2)} = [Y,d,\lambda]^{(2)}$ (i. e. $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ are m2m-isomorphic).

(b) $R^X_*\mathfrak{M}^{\boldsymbol{n}}_\nu = R^Y_*\mathfrak{M}^{\boldsymbol{n}}_\lambda$ for every $m\in\mathbb{N}$ and $\boldsymbol{n}\in\mathbb{N}^m$ with $|\boldsymbol{n}|\geq 2$.

(c) $R^X_*\mathfrak{M}^{\infty,\infty}_\nu = R^Y_*\mathfrak{M}^{\infty,\infty}_\lambda$.

Proof: Without loss of generality we may assume that $X=\operatorname{supp}\mathfrak{M}_\nu$ and $Y=\operatorname{supp}\mathfrak{M}_\lambda$ (cf. Remark 4.3).

(a) $\Rightarrow$ (b): Let $f\colon X\to Y$ be an isometry with $\lambda = f_{**}\nu$ and let $m\in\mathbb{N}$ and $\boldsymbol{n}\in\mathbb{N}^m$ with $|\boldsymbol{n}|\geq 2$. Define $f_{\boldsymbol{n}}\colon X^{|\boldsymbol{n}|}\to Y^{|\boldsymbol{n}|}$ by $f_{\boldsymbol{n}}(x_{11},\dots,x_{mn_m}) = (f(x_{11}),\dots,f(x_{mn_m}))$. Because f is isometric, we have

$$R^X = R^Y\circ f_{\boldsymbol{n}}.$$

Observe that to prove $R^X_*\mathfrak{M}^{\boldsymbol{n}}_\nu = R^Y_*\mathfrak{M}^{\boldsymbol{n}}_\lambda$ it suffices to show

$$\int\varphi\,\mathrm{d}(R^X_*\mathfrak{M}^{\boldsymbol{n}}_\nu) = \int\varphi\,\mathrm{d}(R^Y_*\mathfrak{M}^{\boldsymbol{n}}_\lambda) \tag{E4.7}$$

for every $\varphi\in\mathcal{C}_b(\mathbb{D}_{\boldsymbol{n}})$. Using the transformation formula (E2.5), the definition of $\mathfrak{M}^{\boldsymbol{n}}_\nu$ and $\lambda=f_{**}\nu$ we see that the right hand side of (E4.7) equals

$$\int\varphi\,\mathrm{d}\left(R^Y_*\mathfrak{M}^{\boldsymbol{n}}_\lambda\right) = \int\int\varphi\circ R^Y\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}\lambda^{\otimes m}(\boldsymbol{\mu}) = \int\int\varphi\circ R^Y\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}(f_{**}\nu)^{\otimes m}(\boldsymbol{\mu}). \tag{E4.8}$$

Recall that we defined $\boldsymbol{\mu}^{\otimes\boldsymbol{n}} := \bigotimes_{i=1}^m\mu_i^{\otimes n_i}$. This and the transformation formula (E2.6) imply that the right hand side of (E4.8) is equal to

$$\int\int\varphi\circ R^Y\,\mathrm{d}\left(\bigotimes_{i=1}^m(f_*\mu_i)^{\otimes n_i}\right)\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) = \int\int\varphi\circ R^Y\,\mathrm{d}f_{\boldsymbol{n}*}(\boldsymbol{\mu}^{\otimes\boldsymbol{n}})\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}).$$

By using the transformation formula (E2.5) again and $R^X = R^Y \circ f_{\boldsymbol{n}}$ we obtain

$$\int\int \varphi \circ R^Y \circ f_{\boldsymbol{n}} \,\mathrm{d}\boldsymbol{\mu}^{\otimes \boldsymbol{n}} \,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) = \int\int \varphi \circ R^X \,\mathrm{d}\boldsymbol{\mu}^{\otimes \boldsymbol{n}} \,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) = \int \varphi \,\mathrm{d}(R^X_* \mathfrak{M}^{\boldsymbol{n}}_\nu).$$

Therefore, we have proved the equality (E4.7) for every $\varphi \in \mathcal{C}_b(\mathbb{D}_{\boldsymbol{n}})$ and this yields the claim.

(b) ⇒ (c): Follows by the Kolmogorov extension theorem since the distributions $R^X_* \mathfrak{M}^{\infty,\infty}_\nu$ and $R^Y_* \mathfrak{M}^{\infty,\infty}_\lambda$ are the projective limits of $(R^X_* \mathfrak{M}^{\boldsymbol{n}}_\nu)_{\boldsymbol{n}}$ and $(R^Y_* \mathfrak{M}^{\boldsymbol{n}}_\lambda)_{\boldsymbol{n}}$, respectively.

(c) ⇒ (a): Assertion (c) together with Proposition 4.17 implies that there are dense matrices $\boldsymbol{x} = (x_{ij})_{ij} \in X^{\mathbb{N}\times\mathbb{N}}$ and $\boldsymbol{y} = (y_{ij})_{ij} \in Y^{\mathbb{N}\times\mathbb{N}}$ with

$$R^X(\boldsymbol{x}) = R^Y(\boldsymbol{y}) \tag{E4.9}$$

such that

$$\nu = \operatorname*{w-lim}_n \Xi_n((\mu_i)_i) \quad \text{with} \quad \mu_i := \operatorname*{w-lim}_n \Xi_n((x_{ij})_j)$$

and

$$\lambda = \operatorname*{w-lim}_n \Xi_n((\eta_i)_i) \quad \text{with} \quad \eta_i := \operatorname*{w-lim}_n \Xi_n((y_{ij})_j).$$

Equation (E4.9) yields

$$r(x_{ij}, x_{kl}) = d(y_{ij}, y_{kl}) \text{ for all } i,j,k,l \in \mathbb{N}.$$

If we a define a function f' by $f'(x_{ij}) := y_{ij}$ for all $i,j \in \mathbb{N}$, then f' is isometric and can be extended continuously to an isometry $f\colon X \to Y$. By construction we have

$$\Xi_n((y_{ij})_j) = f_* \Xi_n((x_{ij})_j)$$

for all $i,n \in \mathbb{N}$. Since the push-forward operator f_* is continuous by Lemma 2.24 we obtain $\eta_i = f_* \mu_i$ for each $i \in \mathbb{N}$ by taking the limit $n \to \infty$. It follows that

$$\Xi_n((\eta_i)_i) = f_{**} \Xi_n((\mu_i)_i)$$

for every $n \in \mathbb{N}$. By taking the limit $n \to \infty$ again, we get $\lambda = f_{**}\nu$ and this completes the proof. □

The measures $R_* \mathfrak{M}^{\boldsymbol{n}}_\nu$ and $R_* \mathfrak{M}^{\infty,\infty}_\nu$ in the previous theorem are analogs of the distance matrix distributions of chapter 3. We call them distance array distributions.

Definition 4.21
Let $[X,r,\nu]^{(2)}$ be an m2pm space.

(a) For each $m \in \mathbb{N}$ and $\boldsymbol{n} \in \mathbb{N}^m$ with $|\boldsymbol{n}| \geq 2$ the measure $R_* \mathfrak{M}^{\boldsymbol{n}}_\nu$ is called the $\boldsymbol{n}$-point distance array distribution *of the m2pm space $[X,r,\nu]^{(2)}$.*

(b) *The measure* $R_*\mathfrak{M}_\nu^{\infty,\infty}$ *is called the* distance array distribution of the m2pm space $[X,r,\nu]^{(2)}$.

Theorem 4.20 shows that an m2pm space $[X,r,\nu]^{(2)}$ can be identified with its distance array distribution $R_*\mathfrak{M}_\nu^{\infty,\infty} \in \mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$. We can now pull back the weak topology of $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$ to obtain a topology on $\mathbb{M}_1^{(2)}$ in which $[X_n,r_n,\nu_n]^{(2)} \to [X,r,\nu]^{(2)}$ if and only if

$$R_*\mathfrak{M}_{\nu_n}^{\infty,\infty} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\infty,\infty} \tag{E4.10}$$

or equivalently if and only if

$$R_*\mathfrak{M}_{\nu_n}^{\boldsymbol{n}} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\boldsymbol{n}} \tag{E4.11}$$

for all $m \in \mathbb{N}$ and $\boldsymbol{n} \in \mathbb{N}^m$ with $|\boldsymbol{n}| \geq 2$. We will elaborate on this approach in the next section.

4.3.2 M2pm-polynomials and the two-level Gromov-weak topology

Now we define a class of test functions on $\mathbb{M}_1^{(2)}$, the so-called m2pm-polynomials. We show that they separate points in $\mathbb{M}_1^{(2)}$ and study the induced topology. This topology formalizes the weak convergence of distance array distributions given in (E4.10).

Definition 4.22 (M2pm-polynomials)
A function $\Phi\colon \mathbb{M}_1^{(2)} \to \mathbb{R}$ *is called an* m2pm-polynomial *if it is of the form*

$$\Phi([X,r,\nu]^{(2)}) = \int\int \varphi\circ R \,\mathrm{d}\boldsymbol{\mu}^{\otimes \boldsymbol{n}}\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}), \tag{E4.12}$$

where $m \in \mathbb{N}$, $\boldsymbol{n} \in \mathbb{N}^m$ *with* $|\boldsymbol{n}| \geq 2$ *and* $\varphi \in \mathcal{C}_b(\mathbb{D}_{\boldsymbol{n}})$.
The set of all m2pm-polynomials is denoted by $\mathcal{T}_1^{(2)}$.

Recall that we defined $\boldsymbol{\mu}^{\otimes\boldsymbol{n}} := \bigotimes_{i=1}^m \mu_i^{\otimes n_i} \in \mathcal{M}_1(X^{|\boldsymbol{n}|})$ in section 4.2.

Remarks 4.23

(a) For $i \in \{1,2\}$ let $m_i \in \mathbb{N}$, $\boldsymbol{n}^{(i)} = (n_1^{(i)},\dots,n_{m_i}^{(i)}) \in \mathbb{N}^{m_i}$ and let Φ_i be an m2pm-polynomial of the form

$$\Phi_i([X,r,\nu]^{(2)}) = \int\int \varphi\circ R\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}^{(i)}}\,\mathrm{d}\nu^{\otimes m_i}(\boldsymbol{\mu}).$$

Then $\Phi_1+\Phi_2$ can easily be written as an m2pm-polynomial of the form (E4.12) with $m = m_1+m_2$ and $\boldsymbol{n} = (n_1^{(1)},\dots,n_{m_1}^{(1)},n_1^{(2)},\dots,n_{m_2}^{(2)})$ and the same is true for $\Phi_1\cdot\Phi_2$. Hence, the set $\mathcal{T}_1^{(2)}$ is closed under addition and under multiplication.

(b) We think of elements of $\mathcal{T}_1^{(2)}$ as polynomials of the two-level measure ν. Similar to the set of polynomials (as functions on $\mathbb{R}$), the set $\mathcal{T}_1^{(2)}$ is an algebra, i. e. it is closed under addition and under multiplication. See also Remarks 3.11 for the naming convention about monomials and polynomials.

Heuristically, with an m2pm-polynomial we first sample m measures $\mu_1,\dots,\mu_m$, then sample n_i points from X with each measure μ_i and then evaluate the distance between the sampled points. This suggests that we can express an m2pm-polynomial in terms of the mixed moment measure $\mathfrak{M}_\nu^{\boldsymbol{n}}$. Indeed we have

$$\Phi([X,r,\nu]^{(2)}) = \int\int \varphi\circ R \,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) = \int \varphi\circ R\,\mathrm{d}\mathfrak{M}_\nu^{\boldsymbol{n}} = \int \varphi\,\mathrm{d}\left(R_*\mathfrak{M}_\nu^{\boldsymbol{n}}\right). \quad \text{(E4.13)}$$

With this equation and the reconstruction theorem for m2pm spaces (see Theorem 4.20) it is easy to see that m2pm-polynomials are well-defined, i. e. that their values do not depend on the choice of the representatives of the m2pm spaces.

Lemma 4.24
Every m2pm-polynomial $\Phi\in\mathcal{T}_1^{(2)}$ is well-defined.

The following theorem shows that an m2pm spaces is uniquely determined by the values of the m2pm-polynomials.

Theorem 4.25
$\mathcal{T}_1^{(2)}$ separates points in $\mathbb{M}_1^{(2)}$.

Proof: We know from the reconstruction theorem for m2pm spaces (see Theorem 4.20) that an m2pm space $[X,r,\nu]^{(2)}$ is uniquely determined by the measures $R_*\mathfrak{M}_\nu^{\boldsymbol{n}}$ with $m\in\mathbb{N}$ and $\boldsymbol{n}\in\mathbb{N}^m$. It follows from (E4.13) that these measures are determined by the values $\Phi([X,r,\nu]^{(2)})$ with $\Phi\in\mathcal{T}_1^{(2)}$. □

We use $\mathcal{T}_1^{(2)}$ to induce a topology on $\mathbb{M}_1^{(2)}$. It follows from the preceding theorem that this topology is Hausdorff.

Definition 4.26
The two-level Gromov-weak topology τ'_{2Gw} *on $\mathbb{M}_1^{(2)}$ is defined as the initial topology induced by $\mathcal{T}_1^{(2)}$.*

Remark 4.27 (τ'_{2Gw}-convergence generalizes weak convergence)
The topology τ'_{2Gw} generalizes weak convergence of two-level probability measures. That is, if $\nu,\nu_1,\nu_2,\dots$ are two-level probability measures on the same Polish metric space X such that $\nu_n \xrightarrow{w} \nu$, then the m2pm spaces $[X,r,\nu_n]^{(2)}$ converges two-level Gromov-weakly to $[X,r,\nu]^{(2)}$. This follows from the fact that for every m2pm-polynomial $\Phi\in\mathcal{T}_1^{(2)}$ the function

$$\begin{aligned}&\mathcal{M}_1(\mathcal{M}_1(X))\to\mathbb{R}\\ &\qquad\nu\mapsto\Phi([X,r,\nu]^{(2)}) = \int\int\varphi\circ R\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu})\end{aligned}$$

is continuous as a concatenation of continuous functions (recall from Corollary 2.22 that $\nu\mapsto\nu^{\otimes m}$ and $\boldsymbol{\mu}\mapsto\boldsymbol{\mu}^{\otimes\boldsymbol{n}}$ are continuous).

The reconstruction theorem for m2pm spaces states that an m2pm space $[X,r,\nu]^{(2)}$ can be identified with its distance array distribution $R_*\mathfrak{M}_\nu^{\infty,\infty} \in \mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$. In the following theorem we show that this identification is in fact an embedding of $(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ in $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$. This shows that the topology τ'_{2Gw} formalizes the convergence approach in (E4.10) and (E4.11).

Theorem 4.28 ($\mathbb{M}_1^{(2)}$ is embedded in $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$)
The function

$$\begin{aligned}\iota\colon (\mathbb{M}_1^{(2)},\tau'_{2Gw}) &\to \mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})\\ [X,r,\nu]^{(2)} &\mapsto R_*\mathfrak{M}_\nu^{\infty,\infty}\end{aligned}$$

is an embedding, i. e. it is injective and bicontinuous.

Proof: The function ι is injective by Theorem 4.20. To show that it is continuous, let $([X_\alpha,r_\alpha,\nu_\alpha]^{(2)})_{\alpha\in\mathcal{A}}$ be a net of m2pm spaces which converges two-level Gromov-weakly to an m2pm space $[X,r,\nu]^{(2)}$. The convergence of all test functions $\Phi \in \mathcal{T}_1^{(2)}$ together with (E4.13) yields

$$R_*\mathfrak{M}_{\nu_\alpha}^{\boldsymbol{n}} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\boldsymbol{n}}$$

for every $m \in \mathbb{N}$ and $\boldsymbol{n} \in \mathbb{N}^m$. Since weak convergence of finite dimensional distributions implies weak convergence of the projective limits (on countable product spaces, cf. Corollary 2.21), we obtain

$$R_*\mathfrak{M}_{\nu_\alpha}^{\infty,\infty} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\infty,\infty}.$$

This shows that ι is continuous.

Because $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$ is metrizable, the inverse function ι^{-1} is continuous if and only if it is sequentially continuous. Let $[X,r,\nu]^{(2)}$ be an m2pm space and let $([X_k,r_k,\nu_k]^{(2)})_k$ be a sequence of m2pm spaces with

$$R_*\mathfrak{M}_{\nu_k}^{\infty,\infty} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\infty,\infty}.$$

Then we also have weak convergence of the marginal distributions

$$R_*\mathfrak{M}_{\nu_k}^{\boldsymbol{n}} \xrightarrow{w} R_*\mathfrak{M}_\nu^{\boldsymbol{n}}$$

for every $m \in \mathbb{N}$ and $\boldsymbol{n} \in \mathbb{N}^m$. This and (E4.13) yields convergence of all m2pm-polynomials and thus τ'_{2Gw}-convergence. We have shown that ι^{-1} is continuous and this completes the proof. □

Since $(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ is homeomorphic to a subset of the Polish space $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$, it is also separable and metrizable (as a subset of a separable and metrizable space).

Corollary 4.29
$(\mathbb{M}_1^{(2)},\tau'_{2Gw})$ *is separable and metrizable.*

Therefore, sequences are sufficient to characterize continuity and compactness in the space $(\mathbb{M}_1^{(2)}, \tau'_{2Gw})$.

Recall from Example 3.18 that we can embed $(\mathbb{M}, \tau_{Gw})$ in $\mathcal{M}_f(\mathbb{D}_N)$, but that the embedding of $(\mathbb{M}, \tau_{Gw})$ is not closed. With the same arguments we can show that $(\mathbb{M}_1^{(2)}, \tau'_{2Gw})$ is not closed in $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$ (use $R_*\mathfrak{M}^{\infty,\infty}_{\nu_n}$ of $[X_n, r_n, \nu_n]^{(2)}$ with $\nu_n := \delta_{\mu_n}$ as the approximating sequence for $\delta_{\mathbb{1}_\infty}$, where X_n, r_n, μ_n are defined as in Example 3.18). Therefore, the Prokhorov metric on $\mathcal{M}_1(\mathbb{D}_{\mathbb{N}\times\mathbb{N}})$ can be pulled back to metrize τ'_{2Gw}, but this metric is not complete. For this reason we provide a complete metric for τ'_{2Gw}, the so-called two-level Gromov-Prokhorov metric, in the next subsection.

4.3.3 The two-level Gromov-Prokhorov metric and its topology

Now we generalize the Gromov-Prokhorov metric d_{GP} of section 3.3 and define the so-called *two-level* Gromov-Prokhorov metric $d_{2\mathrm{GP}}$ on $\mathbb{M}_1^{(2)}$. Generalizing d_{GP} is straight-forward. We simply replace the one-level push-forward with the two-level push-forward. Not surprisingly, all the results of this section are analogs of the statements of section 3.3. In particular we will show that $(\mathbb{M}, d_{2\mathrm{GP}})$ is a Polish metric space. Moreover, the topology induced by the metric $d_{2\mathrm{GP}}$ will turn out to be the same as the two-level Gromov-weak topology. We will prove this fact in section 4.3.6.

Definition 4.30 (Two-level Gromov-Prokhorov metric and topology)
Let $\mathcal{X} = [X, r, \nu]^{(2)}$ and $\mathcal{Y} = [Y, d, \lambda]^{(2)}$ be two m2pm spaces. We define the two-level Gromov-Prokhorov distance $d_{2\mathrm{GP}}(\mathcal{X}, \mathcal{Y})$ *between $\mathcal{X}$ and $\mathcal{Y}$ by*

$$d_{2\mathrm{GP}}(\mathcal{X}, \mathcal{Y}) := \inf_{Z, \iota_X, \iota_Y} d_{\mathrm{P}}^{\mathcal{M}_1(Z)}(\iota_{X**}\nu, \iota_{Y**}\lambda), \tag{E4.14}$$

where the infimum ranges over all isometric embeddings $\iota_X : X \to Z$, $\iota_Y : Y \to Z$ into a common Polish metric space (Z, r_Z) with $Z \subset \mathbb{R}^{\mathbb{N}}$ and where $d_{\mathrm{P}}^{\mathcal{M}_1(Z)}$ denotes the Prokhorov metric for measures on *the space $\mathcal{M}_1(Z)$.*

The induced topology on $\mathbb{M}_1^{(2)}$ is called the two-level Gromov-Prokhorov topology *and denoted by τ'_{2GP}.*

Remarks 4.31

(a) The condition $Z \subset \mathbb{R}^{\mathbb{N}}$ in Definition 4.30 is not a restriction since every Polish space can be embedded in $\mathbb{R}^{\mathbb{N}}$ (cf. section 2.3). We only need this condition to ensure that the infimum is taken over a well-defined set.

(b) Note that the two-level Gromov-Prokhorov metric can easily be extended to $\mathbb{M}^{(2)}$, because the Prokhorov metric is defined for finite measures. The results of this section still hold for this extension. Most of the proofs stay literally the same, but some require small modifications (for instance: to prove that the extension is a metric on $\mathbb{M}^{(2)}$, we first need to introduce a point separating class of functions on $\mathbb{M}^{(2)}$). We spell out the details in section 4.4.3.

It is easy to see that the definition of $d_{2\mathrm{GP}}$ does not depend on the representatives of the equivalence classes. Hence, the two-level Gromov-Prokhorov distance is well-defined. Moreover, convergence with respect to $d_{2\mathrm{GP}}$ generalizes weak convergence in the following sense: If $\nu, \nu_1, \nu_2, \dots$ are two-level measures on (X,r) with $\nu_n \xrightarrow{w} \nu$, then $d_{2\mathrm{GP}}([X,r,\nu_n]^{(2)}, [X,r,\nu]^{(2)}) \to 0$.

We show in Theorem 4.36 that $d_{2\mathrm{GP}}$ indeed is a metric and even complete, but before we are able to prove this we need to provide some preparatory lemmas. The next lemma shows that we can always use the disjoint union $Z = X \sqcup Y$ to compute the two-level Gromov-Prokhorov distance in (E4.14).

Lemma 4.32 (Alternative characterization of $d_{2\mathrm{GP}}$)
For all m2pm spaces $\mathcal{X} = [X,r,\nu]^{(2)}$ and $\mathcal{Y} = [Y,d,\lambda]^{(2)}$ we have

$$d_{2\mathrm{GP}}(\mathcal{X},\mathcal{Y}) = \inf_{r'} d_{\mathrm{P}}^{\mathcal{M}_1(X \sqcup Y, r')}(\nu,\lambda),$$

where the infimum ranges over all complete metrics r' on the disjoint union $X \sqcup Y$ which extend the metrics r and d.

In the previous lemma we regard ν and λ as elements of the same space $\mathcal{M}_1(\mathcal{M}_1(X \sqcup Y))$ without using the canonical embeddings $\iota_X : X \to X \sqcup Y$ and $\iota_Y : Y \to X \sqcup Y$. We will continue to use this abbreviated notation in the remainder of this section whenever it seems appropriate (i. e. when we are embedding spaces into their disjoint union). In the same manner we will regard points $x \in X$ and $y \in Y$ as elements of $X \sqcup Y$ and write $r'(x,y)$ instead of $r'(\iota_X(x), \iota_Y(y))$ when r' is a metric on $X \sqcup Y$.

Proof (of Lemma 4.32): By the definition of $d_{2\mathrm{GP}}$ we always have

$$d_{2\mathrm{GP}}(\mathcal{X},\mathcal{Y}) \le \inf_{r'} d_{\mathrm{P}}^{\mathcal{M}_1(X \sqcup Y, r')}(\nu,\lambda).$$

To prove the other direction, let $d_{2\mathrm{GP}}(\mathcal{X},\mathcal{Y}) < \varepsilon$. Then there is a Polish metric space (Z, r_Z) and isometric embeddings $\iota_X : X \to Z$, $\iota_Y : Y \to Z$ such that

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z,r_Z)}(\iota_{X**}\nu, \iota_{Y**}\lambda) < \varepsilon.$$

For $\delta > 0$ define r'_δ as the metric on $X \sqcup Y$ which extends r and d and satisfies

$$r'_\delta(x,y) = \delta + r_Z(\iota_X(x), \iota_Y(y))$$

for every $x \in X$ and $y \in Y$. The metric space $(X \sqcup Y, r'_\delta)$ is complete and separable and we have

$$d_{\mathrm{P}}^{\mathcal{M}_1(X \sqcup Y, r'_\delta)}(\nu,\lambda) < \varepsilon + \delta.$$

Ranging over all possible ε and δ yields the claim. □

By applying the previous lemma iteratively we obtain the following two embedding lemmas. The first lemma implies that we can embed a Cauchy sequence of m2pm spaces into a single metric space in which the two-level push-forwards of the measures form a Cauchy sequence. This fact will be used later to show that $d_{2\mathrm{GP}}$ is complete.

Lemma 4.33 (Embedding of sequences of m2pm spaces)
Let $(\varepsilon_n)_n$ be a sequence of positive real numbers and let $(\mathcal{X}_n)_n$ be a sequence of m2pm spaces with $\mathcal{X}_n = [X_n, r_n, \nu_n]^{(2)}$ and

$$d_{2\mathrm{GP}}(\mathcal{X}_n, \mathcal{X}_{n+1}) < \varepsilon_n$$

for every $n \in \mathbb{N}$. Then there exists a Polish metric space (Z, r_Z) and isometric embeddings $\iota_1, \iota_2, \dots$ of $X_1, X_2, \dots$, respectively, into Z such that

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z,r_Z)}(\iota_{n**}\nu_n, \iota_{(n+1)**}\nu_{n+1}) < \varepsilon_n$$

for all $n \in \mathbb{N}$.

Proof: We define $Z_n := \bigsqcup_{k=1}^n X_k$ for every $n \in \mathbb{N}$ and $Z_\infty := \bigsqcup_{k=1}^\infty X_k$. We will inductively define metrics d_n on Z_n using Lemma 4.32. By this lemma there exists a metric d_2 on $Z_2 = X_1 \sqcup X_2$ such that (Z_2, d_2) is complete and separable and

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z_2,d_2)}(\nu_1, \nu_2) < \varepsilon_1.$$

This metric extends r_1 and r_2 and therefore we have $[Z_2, d_2, \nu_2]^{(2)} = [X_2, r_2, \nu_2]^{(2)}$ and

$$d_{2\mathrm{GP}}([Z_2, d_2, \nu_2]^{(2)}, \mathcal{X}_3) = d_{2\mathrm{GP}}(\mathcal{X}_2, \mathcal{X}_3) < \varepsilon_2.$$

By using Lemma 4.32 again we find a metric d_3 on $Z_3 = Z_2 \sqcup X_3$ such that (Z_3, d_3) is complete and separable and

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z_3,d_3)}(\nu_2, \nu_3) < \varepsilon_2.$$

This procedure can be continued ad infinitum and in this way we obtain a metric d_∞ on Z_∞. The metric space (Z_∞, d_∞) is separable but not necessarily complete. For this reason we define (Z, r_Z) as the completion of (Z_∞, d_∞). It is easily seen that this completion has the desired properties with ι_n being the canonical embedding of X_n into $Z_\infty \subset Z$ for every $n \in \mathbb{N}$. □

A similar statement holds for convergent sequences of m2pm spaces. Roughly speaking, m2pm spaces converge with respect to the two-level Gromov-Prokhorov metric if and only if there is a common metric space on which their two-level measures converge weakly.

Lemma 4.34 (Embedding of convergent sequences of m2pm spaces)
Let $(\mathcal{X}_n)_n$ be a sequence of m2pm spaces with $\mathcal{X}_n = [X_n, r_n, \nu_n]^{(2)}$ which converges to an m2pm space $\mathcal{X} = [X, r, \nu]^{(2)}$ in the two-level Gromov-Prokhorov topology. Then there is a Polish metric space (Z, r_Z) and isometric embeddings $\iota, \iota_1, \iota_2, \dots$ of $X, X_1, X_2, \dots$, respectively, into Z such that

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z,r_Z)}(\iota_{n**}\nu_n, \iota_{**}\nu) \to 0.$$

Proof: The proof is similar to the proof of Lemma 4.33, but this time we always route through the space X. Let $(\varepsilon_n)_n$ be a sequence of positive reals converging to 0 with

$$d_{2\mathrm{GP}}(\mathcal{X}_i, \mathcal{X}) < \varepsilon_n$$

for every $n \in \mathbb{N}$. We define $Z_n := X \sqcup (\bigsqcup_{k=1}^n X_k)$ for every $n \in \mathbb{N}$ and $Z_\infty := X \sqcup (\bigsqcup_{k=1}^\infty X_k)$. By the same arguments as in the proof of Lemma 4.33 we find metrics d_n such that (Z_n, d_n) is complete and separable for every $n \in \mathbb{N}$ and

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z_n,d_n)}(\nu_n, \nu) < \varepsilon_n.$$

This way we obtain a metric d_∞ on Z_∞. The completion of the metric space (Z_∞, d_∞) has the properties that were claimed in the lemma. □

With the preceding embedding lemma it is easy to show that τ'_{2GP}-convergence implies τ'_{2Gw}-convergence. We show in section 4.3.6 that actually both topologies coincide. But at this point we are only able to prove one direction.

Lemma 4.35 (τ'_{2GP} is finer than τ'_{2Gw})
Every m2pm-monomial $\Phi \in \mathcal{T}_1^{(2)}$ is continuous with respect to the two-level Gromov-Prokhorov topology. Therefore, two-level Gromov-Prokhorov convergence implies two-level Gromov-weak convergence and τ'_{2GP} is finer than τ'_{2Gw}.

Proof: Let $\mathcal{X} = [X, r, \nu]^{(2)}$, $\mathcal{X}_1 = [X_1, r_1, \nu_1]^{(2)}$, ... be m2pm spaces with $d_{2\mathrm{GP}}(\mathcal{X}_n, \mathcal{X}) \to 0$. By Lemma 4.34 we may assume without loss of generality that all metric spaces coincide, i. e. $(Z, r_Z) = (X, r) = (X_1, r_1) = \ldots$, and that we have $d_{\mathrm{P}}^{\mathcal{M}_1(Z,r_Z)}(\nu_n, \nu) \to 0$. For every $\Phi \in \mathcal{T}_1^{(2)}$ the function $\lambda \mapsto \Phi([Z, r_Z, \lambda]^{(2)})$ from $\mathcal{M}_1(\mathcal{M}_1(X))$ to $\mathbb{R}$ is continuous with respect to weak convergence (see Remark 4.27) and thus we obtain $\Phi(\mathcal{X}_n) \to \Phi(\mathcal{X})$. □

Finally, we are in a position to prove that the two-level Gromov-Prokhorov distance is a complete metric for $\mathbb{M}_1^{(2)}$.

Theorem 4.36
$(\mathbb{M}_1^{(2)}, d_{2\mathrm{GP}})$ is a Polish metric space.

Proof: It is obvious that $d_{2\mathrm{GP}}$ is non-negative, symmetric and satisfies $d_{2\mathrm{GP}}(\mathcal{X}, \mathcal{X}) = 0$ for every m2pm space $\mathcal{X}$. Now let $\mathcal{X}, \mathcal{Y}$ be m2pm spaces with $d_{2\mathrm{GP}}(\mathcal{X}, \mathcal{Y}) = 0$. With Lemma 4.35 we get $\Phi(\mathcal{X}) = \Phi(\mathcal{Y})$ for every $\Phi \in \mathcal{T}_1^{(2)}$ (by using the constant sequence $\mathcal{X}_n = \mathcal{X}$). Since $\mathcal{T}_1^{(2)}$ separates points, this implies $\mathcal{X} = \mathcal{Y}$.

To prove the triangle inequality, let $\mathcal{X}_i = [X_i, r_i, \nu_i]^{(2)} \in \mathbb{M}_1^{(2)}$ for $i \in \{1, 2, 3\}$. Let $\varepsilon_1, \varepsilon_2 > 0$ such that

$$d_{2\mathrm{GP}}(\mathcal{X}_1, \mathcal{X}_2) < \varepsilon_1 \quad \text{and} \quad d_{2\mathrm{GP}}(\mathcal{X}_2, \mathcal{X}_3) < \varepsilon_2.$$

It follows from the proof of Lemma 4.33 that there is a complete metric d_3 on $Z_3 := X_1 \sqcup X_2 \sqcup X_3$ which extends r_1, r_2 and r_3 such that

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z_3,d_3)}(\nu_1, \nu_2) < \varepsilon_1 \quad \text{and} \quad d_{\mathrm{P}}^{\mathcal{M}_1(Z_3,d_3)}(\nu_2, \nu_3) < \varepsilon_2.$$

The triangle inequality for d_{P} yields

$$d_{\mathrm{P}}^{\mathcal{M}_1(Z_3,d_3)}(\nu_1,\nu_3) < \varepsilon_1 + \varepsilon_2.$$

It follows from the definition of $d_{2\mathrm{GP}}$ that

$$d_{2\mathrm{GP}}(\mathcal{X}_1,\mathcal{X}_3) \le d_{\mathrm{P}}^{\mathcal{M}_1(Z_3,d_3)}(\nu_1,\nu_3) < \varepsilon_1 + \varepsilon_2.$$

By taking the infimum over all possible ε_1 and ε_2 we obtain the triangle inequality for $d_{2\mathrm{GP}}$.

To prove that $d_{2\mathrm{GP}}$ is complete, let $\mathcal{X}_n = [X_n, r_n, \nu_n]^{(2)}$ be a Cauchy sequence with respect to $d_{2\mathrm{GP}}$. By Lemma 4.33 we can embed the metric spaces $((X_n, r_n))_n$ into a common Polish metric space (Z, r_Z) using isometries $(\iota_n)_n$. The two-level push-forward measures $(\iota_{n**}\nu_n)_n$ form a Cauchy sequence in $\mathcal{M}_1(\mathcal{M}_1(Z))$ and thus converge weakly to some $\nu \in \mathcal{M}_1(\mathcal{M}_1(Z))$. It follows that $(\mathcal{X}_n)_n$ converges to the m2pm space $[Z, r_Z, \nu]^{(2)}$ with respect to the two-level Gromov-Prokhorov metric.

To show that $(\mathbb{M}^{(2)}, d_{2\mathrm{GP}})$ is separable, we define $\mathbb{S}_1$ as the set of all m2pm spaces $[S, d, \lambda]^{(2)}$ such that $|S| < \infty$, the metric d takes only rational values and λ is of the form

$$\lambda = \sum_{i=1}^{M} a_i \delta_{\left(\sum_{j=1}^{N_i} b_{ij}\delta_{x_{ij}}\right)} \tag{E4.15}$$

with $M, N_1, \ldots, N_M \in \mathbb{N}$, $x_{ij} \in S$ $a_i, b_{ij} \in \mathbb{Q}_+$ and

$$\sum_{i=1}^{M} a_i = 1, \qquad \sum_{j=1}^{N_i} b_{ij} = 1$$

for all i, j. That is, $(S, d, \lambda)^{(2)}$ is a finite m2pm triple with only rational distances and λ is an atomic probability measure on the set of atomic probability measures with only rational values. The set $\mathbb{S}_1$ is obviously countable. To prove density, let $[X, r, \nu]^{(2)}$ be an arbitrary m2pm space and let $\varepsilon > 0$. Because the set of measures of the form (E4.15) is dense in $\mathcal{M}_1(\mathcal{M}_1(X))$, there is a $\lambda \in \mathcal{M}_1(\mathcal{M}_1(X))$ of this form with $d_{\mathrm{P}}(\lambda, \nu) < \frac{\varepsilon}{2}$. Then, $S := \operatorname{supp}\mathfrak{M}_\lambda$ is finite and $d_{2\mathrm{GP}}([X, r, \nu]^{(2)}, [S, d, \lambda]^{(2)}) < \frac{\varepsilon}{2}$. The last step is to approximate the metric r by a rational version d such that $|d(x, y) - r(x, y)| < \frac{\varepsilon}{2}$ for all $x, y \in S$. The m2pm space $[S, d, \lambda]^{(2)}$ is in $\mathbb{S}_1$ and we have $d_{2\mathrm{GP}}([X, r, \nu]^{(2)}, [S, d, \lambda]^{(2)}) < \varepsilon$. □

4.3.4 Distance distribution and modulus of mass distribution

In this section we examine how the distance distribution and the modulus of mass distribution can be used to describe convergence (or divergence) behavior of sequences in $\mathbb{M}_1^{(2)}$.

At the beginning of section 3.4 we provided examples of strongly divergent sequences of metric measure spaces (where strongly divergent means that they have no convergent subsequence). We adapt these examples to obtain strongly divergent sequences of m2pm spaces. Observe that now the problems which prevent convergence can arise on *both levels* of the two-level measures.

Example 4.37
In Example 3.27 we defined metric measure spaces $\mathcal{X}_n := [X_n, r_n, \mu_n]$ with

$$\begin{aligned} X_n &:= \{0,1\}, \\ r_n(i,j) &:= n\mathbb{1}_{i \neq j}, \\ \mu_n &:= \frac{1}{2}(\delta_0 + \delta_1) \end{aligned}$$

and argued why this sequence cannot have a convergent subsequence. With the same arguments as before it follows that neither $([X_n, r_n, \delta_{\mu_n}]^{(2)})_n$ nor $([X_n, r_n, \nu_n]^{(2)})_n$ with $\nu_n := \frac{1}{2}(\delta_{\delta_0} + \delta_{\delta_1})$ can have a convergent subsequence.

We showed in section 3.4 that the divergence of the mm spaces $([X_n, r_n, \mu_n])_n$ can be captured by the divergence of their distance distributions $w(\mu_n) = \frac{1}{2}(\delta_0 + \delta_n)$. Recall that the distance distribution of a finite measure μ on a Polish metric space (X, r) is defined as $w(\mu) := r_*\mu^{\otimes 2}$. A two-level measure ν on X is a finite measure on the Polish metric space $(\mathcal{M}_f(X), d_{\mathrm{P}})$, where d_{P} denotes the Prokhorov metric. Therefore, the distance distribution of ν is $w(\nu) := d_{\mathrm{P}*}\nu^{\otimes 2}$. However, $w(\nu)$ only captures the spread of the mass on the upper level of ν. For instance, in Example 4.37 we have $w(\delta_{\mu_n}) = 0$ and $w(\nu_n) = \frac{1}{2}(\delta_0 + \delta_n)$, even though the associated m2pm spaces clearly diverge in a similar way. The remedy is to project both levels of the two-level measures to a single level, i. e. to look at the first moment measures

$$\mathfrak{M}_{\nu_n} = \frac{1}{2}(\delta_0 + \delta_1) \text{ and } \mathfrak{M}_{\delta_{\mu_n}} = \mu_n = \frac{1}{2}(\delta_0 + \delta_1).$$

Then, we apply the distance distribution to obtain

$$w(\mathfrak{M}_{\nu_n}) = w(\mathfrak{M}_{\delta_{\mu_n}}) = \frac{1}{2}(\delta_0 + \delta_n).$$

This shows that we can capture the divergence of the sequences $([X_n, r_n, \delta_{\mu_n}]^{(2)})_n$ and $([X_n, r_n, \nu_n]^{(2)})_n$ by using the distance distribution of the first moment measures.

Example 4.38
In Example 3.28 we defined metric measure spaces $\mathcal{Y}_n := [Y_n, d_n, \eta_n]$ with

$$\begin{aligned} Y_n &:= \{1, \dots, n\}, \\ d_n(i,j) &:= \mathbb{1}_{i \neq j}, \\ \eta_n &:= \frac{1}{n}\sum_{i=1}^{n} \delta_i \end{aligned}$$

and showed that $(\mathcal{Y}_n)_n$ is strongly divergent. In a similar manner one can show that both sequences $([Y_n, d_n, \delta_{\eta_n}]^{(2)})_n$ and $([Y_n, d_n, \lambda_n]^{(2)})_n$ with $\lambda_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\delta_i}$ do not have a convergent subsequence.

With the same reasoning as above one can show that applying the modulus of mass distribution directly to the two-level measures δ_{η_n} and λ_n is insufficient for describing the divergence of the sequences $([Y_n, d_n, \delta_{\eta_n}]^{(2)})_n$ and $([Y_n, d_n, \lambda_n]^{(2)})_n$. The solution is again to use the first moment measures. Then, it is not very difficult to see that

$$V_\delta(\mathfrak{M}_{\delta_{\eta_n}}) = V_\delta(\mathfrak{M}_{\lambda_n}) = \begin{cases} 0, & n < \frac{1}{\delta} \\ 1, & n \geq \frac{1}{\delta}. \end{cases}$$

Thus, the divergence of the sequences can be described in terms of the modulus of mass distribution of the first moment measures.

In the remainder of this section we study the continuity behavior of the functions $[X, r, \nu]^{(2)} \mapsto w(\mathfrak{M}_\nu)$ and $[X, r, \nu]^{(2)} \mapsto V_\delta(\mathfrak{M}_\nu)$.

Lemma 4.39
The function

$$\begin{aligned} \mathbb{M}_1^{(2)} &\to \mathbb{M}_1 \\ [X, r, \nu]^{(2)} &\mapsto [X, r, \mathfrak{M}_\nu] \end{aligned}$$

is continuous with respect to both the two-level Gromov-weak topology τ'_{2Gw} and the two-level Gromov-Prokhorov topology τ'_{2GP}.

Proof: Because τ'_{2GP} is finer than τ'_{2Gw} (see Lemma 4.35), it suffices to prove continuity only for τ'_{2Gw}. Let $([X_n, r_n, \nu_n]^{(2)})_n$ be a sequence of m2pm spaces which converges to $[X, r, \nu]^{(2)}$ in the two-level Gromov-weak topology and let $\Phi \in \mathcal{T}^{(1)}$ be an mm-monomial of degree m as in (E3.3). Then we see that

$$\begin{aligned} \Phi([X_n, r_n, \mathfrak{M}_{\nu_n}]) &= \psi(\mathfrak{m}(\mathfrak{M}_{\nu_n})) \int \varphi \circ R \, \mathrm{d}\mathfrak{M}_{\nu_n}^{\otimes m} \\ &= \psi(1) \int \int \varphi \circ R \, \mathrm{d}\boldsymbol{\mu}^{\otimes(1,1,\dots,1)} \, \mathrm{d}\nu_n^{\otimes m}(\boldsymbol{\mu}). \end{aligned}$$

The right hand side is an m2pm-polynomial and thus converges to

$$\begin{aligned} &\psi(1) \int \int \varphi \circ R \, \mathrm{d}\boldsymbol{\mu}^{\otimes(1,1,\dots,1)} \, \mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) \\ &= \Phi([X, r, \mathfrak{M}_\nu]) \end{aligned}$$

Since $\Phi \in \mathcal{T}^{(1)}$ is arbitrary, this proves that $[X_n, r_n, \mathfrak{M}_{\nu_n}]$ converges Gromov-weakly to $[X, r, \mathfrak{M}_\nu]$. □

The previous lemma allows us to reuse the results of section 3.4 to obtain continuity properties of $w(\mathfrak{M}_\nu)$ and $V_\delta(\mathfrak{M}_\nu)$.

Corollary 4.40
The function

$$\begin{aligned} \mathbb{M}_1^{(2)} &\to \mathcal{M}_1(\mathbb{R}_+) \\ [X, r, \nu]^{(2)} &\mapsto w(\mathfrak{M}_\nu) \end{aligned}$$

is continuous with respect to both the two-level Gromov-weak topology τ'_{2Gw} *and the two-level Gromov-Prokhorov topology* τ'_{2GP}.

Proof: If follows from Lemma 4.39 and Lemma 3.36 that

$$[X,r,\nu]^{(2)} \mapsto [X,r,\mathfrak{M}_\nu] \mapsto w(\mathfrak{M}_\nu)$$

is continuous as a concatenation of continuous functions. □

Corollary 4.41
Let $\delta > 0$ *be fixed. The function*

$$\mathbb{M}_1^{(2)} \to \mathbb{R}_+$$
$$[X,r,\nu]^{(2)} \mapsto V_\delta(\mathfrak{M}_\nu)$$

is upper semi-continuous with respect to both the two-level Gromov-weak topology τ'_{2Gw} *and the two-level Gromov-Prokhorov topology* τ'_{2GP}.

Proof: It follows from Lemma 4.39 and Lemma 3.37 that for every $\delta > 0$

$$[X,r,\nu]^{(2)} \mapsto [X,r,\mathfrak{M}_\nu] \mapsto V_\delta(\mathfrak{M}_\nu)$$

is upper semi-continuous as a concatenation of a continuous and an upper semi-continuous functions. □

4.3.5 Compact sets

In this section we generalize the results of Proposition 3.39 to m2pm spaces and characterize relatively compact subsets of $(\mathbb{M}_1^{(2)}, d_{2\mathrm{GP}})$.

From the examples at the beginning of section 4.3.4 it is easy to imagine how we can generalize the compactness criterion (b) of Proposition 3.39 to the m2pm case: We simply replace the one-level measures with the the first moment measures of the two-level measures. This is the content of assertion (b) of the following theorem.

The other compactness criteria of the next theorem can be seen as analogs of tightness (but for two-level measures defined on different sets). It follows from Lemma 4.13 that a set Γ of two-level probability measures on the same Polish space X is tight if and only if for every $\varepsilon > 0$ there is a compact set $K \subset X$ such that

$$\nu(\complement\{\mu \in \mathcal{M}_1(X) \mid \mu(\complement K)) \le \varepsilon\} < \varepsilon$$

for all $\nu \in \Gamma$. If now Γ is a set of m2pm spaces, i. e. we consider two-level probability measures on different sets, we can regard Γ as being "tight" if for every $\varepsilon > 0$ and every $\mathcal{X} = [X,r,\nu]^{(2)} \in \Gamma$ there is a compact set $K_\mathcal{X} \subset X$ with

$$\nu(\complement\{\mu \in \mathcal{M}_1(X) \mid \mu(\complement K_\mathcal{X})) \le \varepsilon\} < \varepsilon$$

such that the sets $\{K_\mathcal{X} \mid \mathcal{X} \subset \Gamma\}$ are "uniformly compact", meaning that they are relatively compact in the Gromov-Hausdorff topology. This yields compactness criterion (d) of the following theorem.

Theorem 4.42 (Characterization of relatively compact subsets of $\mathbb{M}_1^{(2)}$)
Let $\Gamma \subset \mathbb{M}_1^{(2)}$ be a set of m2pm spaces. The following are equivalent:

(a) Γ *is relatively compact in the two-level Gromov-Prokhorov topology* τ'_{2GP}.

(b) The following two conditions are fulfilled:
- $\sup_{[X,r,\nu]^{(2)}\in\Gamma} V_\delta(\mathfrak{M}_\nu) \to 0$ *for* $\delta \searrow 0$,
- $\{\, w(\mathfrak{M}_\nu) \mid [X,r,\nu]^{(2)} \in \Gamma \,\}$ *is tight in* $\mathcal{M}_1(\mathbb{R}_+)$.

(c) For all $\varepsilon > 0$ *there is an* $N_\varepsilon \in \mathbb{N}$ *such that for every* $\mathcal{X} = [X,r,\nu]^{(2)} \in \Gamma$ *there exists a measurable subset* $X_{\mathcal{X},\varepsilon} \subset X$ *with*
- $\nu(\complement\{\, \mu \in \mathcal{M}_1(X) \mid \mu(\complement X_{\mathcal{X},\varepsilon}) \le \varepsilon \,\}) < \varepsilon$,
- $X_{\mathcal{X},\varepsilon}$ *can be covered by at most* N_ε *balls of radius* ε,
- *the diameter of* $X_{\mathcal{X},\varepsilon}$ *is at most* N_ε.

(d) For all $\varepsilon > 0$ *and* $\mathcal{X} = [X,r,\nu]^{(2)} \in \Gamma$ *there is a compact set* $C_{\mathcal{X},\varepsilon} \subset X$ *such that*
- $\nu(\complement\{\, \mu \in \mathcal{M}_1(X) \mid \mu(\complement C_{\mathcal{X},\varepsilon}) \le \varepsilon \,\}) < \varepsilon$,
- $\mathcal{C}_\varepsilon := \{\, C_{\mathcal{X},\varepsilon} \mid \mathcal{X} \in \Gamma \,\}$ *is relatively compact in the Gromov-Hausdorff topology.*

The proof of the previous theorem is rather long and cumbersome. We omit it here and refer to the more general Theorem 4.68, which provides similar compactness criteria for general m2m spaces.

4.3.6 Equivalence of two-level Gromov-weak and two-level Gromov-Prokhorov topology on $\mathbb{M}_1^{(2)}$

We now prove that the two-level Gromov-weak topology τ'_{2Gw} on $\mathbb{M}_1^{(2)}$ is metrized by the two-level Gromov-Prokhorov metric $d_{2\mathrm{GP}}$. The proof is similar to the one of Proposition 3.40.

Theorem 4.43
The two-level Gromov-weak topology τ'_{2Gw} *and the two-level Gromov-Prokhorov topology* τ'_{2GP} *on* $\mathbb{M}_1^{(2)}$ *coincide.*

Proof: We have already proved in Lemma 4.35 that τ'_{2GP} is finer than τ'_{2Gw}. To show the other direction, we prove that every τ_{2Gw}-convergent sequence

$$[X_n, r_n, \nu_n]^{(2)} \xrightarrow{\tau_{2Gw}} [X, r, \nu]^{(2)}$$

also converges with respect to the two-level Gromov-Prokhorov topology τ_{2GP} (recall from Corollary 4.29 that τ_{2Gw} is metrizable, thus we may use sequences instead of nets).

We will show that $([X_n, r_n, \nu_n]^{(2)})_n$ is relatively compact with respect to τ_{2GP}. Since $\mathcal{T}_1^{(2)}$ separates points and τ_{2GP}-convergence implies τ_{2Gw}-convergence, it follows that every subsequence has a further convergent subsequence with limit $[X,r,\nu]^{(2)}$. Consequently, the sequence itself must converge to $[X,r,\nu]^{(2)}$ in the topology τ_{2GP}.

To show relative compactness of the sequence we use the compactness criterion (b) of Theorem 4.42. By Corollary 4.40 the measures $w(\mathfrak{M}_{\nu_n})$ are convergent, hence tight. Moreover, by Corollary 4.41 we have

$$\limsup_n V_\delta(\mathfrak{M}_{\nu_n}) \le V_\delta(\mathfrak{M}_\nu)$$

for every $\delta > 0$. Since $V_\delta(\eta) \to 0$ for $\delta \searrow 0$ and every finite measure η (by Lemma 3.31), this yields $\sup_n V_\delta(\mathfrak{M}_{\nu_n}) \to 0$ for $\delta \searrow 0$ and the assumptions of Theorem 4.42 are satisfied. □

4.3.7 Distributions on $\mathbb{M}_1^{(2)}$

Now we provide some results about random m2pm spaces. First we deduce from Le Cam's theorem that m2pm-polynomials are convergence determining.

Proposition 4.44
The set $\mathcal{T}_1^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}_1^{(2)})$.

Proof: $\mathcal{T}_1^{(2)}$ is closed under multiplication and induces the topology of $\mathbb{M}_1^{(2)}$. It follows from Proposition 2.17 that $\mathcal{T}_1^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}_1^{(2)})$. □

In the following proposition we characterize tight subsets of $\mathcal{M}_1(\mathbb{M}_1^{(2)})$. We omit the proof because it uses standard arguments and the result follows from the more general Proposition 4.71.

Proposition 4.45 (Tightness criterion for $\mathcal{M}_1(\mathbb{M}_1^{(2)})$)
A set $\mathcal{P} \subset \mathcal{M}_1(\mathbb{M}_1^{(2)})$ is tight if and only if for every $\varepsilon > 0$ there are $\delta > 0$ and $c > 0$ such that for every $P \in \mathcal{P}$ we have

(a) $P(w(\mathfrak{M}_\nu)([c,\infty)) \ge \varepsilon) < \varepsilon$ and

(b) $P(V_\delta(\mathfrak{M}_\nu) \ge \varepsilon) < \varepsilon$,

where ν denotes the two-level measure of a random m2pm space $[X,r,\nu]^{(2)}$ with law P.

The following version of the preceding proposition is particularly useful for the applications in chapter 5. We write $\eta \le \mu$ for two finite measure $\eta,\mu \in \mathcal{M}_f(X)$ if $\eta(A) \le \mu(A)$ for every measurable $A \subset X$.

Corollary 4.46
A set $\mathcal{P} \subset \mathcal{M}_1(\mathbb{M}_1^{(2)})$ is tight if the following two conditions hold:

(a) There is a finite Borel measure μ on $\mathbb{R}_+$, such that

$$P[w(\mathfrak{M}_\nu)] := \int w(\mathfrak{M}_\nu)\, \mathrm{d}P([X,r,\nu]^{(2)}) \le \mu$$

for every $P \in \mathcal{P}$.

(b) $\lim_{\delta \searrow 0} \sup_{P \in \mathcal{P}} P[\mathfrak{M}_\nu(\{ x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta \})] = 0$ *for every* $\varepsilon > 0$.

Remark 4.47
In case $\mathcal{P}$ is a sequence $(P_n)_n$, we can replace $\sup_{P \in \mathcal{P}}$ in property (b) by $\limsup_{n \to \infty}$.

Proof (of Corollary 4.46): We will show that $\mathcal{P}$ satisfies both properties of Proposition 4.45. Let $\varepsilon > 0$. Then, there exists a $c > 0$ such that $\mu([c,\infty)) < \varepsilon^2$. By Markov's inequality we get

$$P(w(\mathfrak{M}_\nu)([c,\infty)) \geq \varepsilon) \leq \frac{P[w(\mathfrak{M}_\nu)([c,\infty))]}{\varepsilon} \leq \frac{\mu([c,\infty))}{\varepsilon} < \varepsilon$$

for every $P \in \mathcal{P}$. Moreover, by condition (b) there is a $\delta > 0$ such that

$$P[\mathfrak{M}_\nu(\{ x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\frac{\varepsilon}{2})) < 2\delta \})] < \varepsilon^2$$

for all $P \in \mathcal{P}$. With Lemma 3.31 and with Markov's inequality we obtain

$$\begin{aligned} P(V_\delta(\mathfrak{M}_\nu) \geq \varepsilon) &= P(\mathfrak{M}_\nu(\{ x \in X \mid \mathfrak{M}_\nu(B(x,\varepsilon)) \leq \delta \}) \geq \varepsilon) \\ &\leq \varepsilon^{-1} P[\mathfrak{M}_\nu(\{ x \in X \mid \mathfrak{M}_\nu(B(x,\varepsilon)) \leq \delta \})] \\ &\leq \varepsilon^{-1} P[\mathfrak{M}_\nu(\{ x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\frac{\varepsilon}{2})) < 2\delta \})] \\ &< \varepsilon \end{aligned}$$

for all $P \in \mathcal{P}$. Thus, $\mathcal{P}$ is tight by Proposition 4.45. □

4.4 Metric two-level measure spaces

Now we study general m2m spaces $[X,r,\nu]^{(2)}$ with $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$. One of the main difficulties is that there is no reconstruction theorem for m2m spaces. That is, we cannot identify an m2m spaces with its distance array distribution as we did in Theorem 4.20 for m2pm spaces. The main problem is that $\nu(\{ o \})$ may be positive and that we cannot sample points from the underlying metric space X with the null measures o.

4.4.1 Reconstruction of two-level measures

In this subsection we generalize Proposition 4.17 and show how a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_f(X))$ can be reconstructed from sampled points in X provided that $\nu(\{ o \}) = 0$.

Let us first extend Proposition 4.15 to two-level measures $\nu \in \mathcal{M}_1(\mathcal{M}_f(X))$ by decomposing finite measures $\mu \in \mathcal{M}_f(X)$ into their total mass $\mathfrak{m}(\mu)$ and their normalized probability measure $\overline{\mu}$.

Lemma 4.48 (Glivenko-Cantelli for random measures)
Let X be a Polish space and let $\nu \in \mathcal{M}_1(\mathcal{M}_f(X))$ with $\nu(\{o\}) = 0$. Let $(m, \boldsymbol{x}) \in \mathbb{R}_+ \times X^{\mathbb{N}}$ be random with law $\int \delta_{\mathfrak{m}(\mu)} \otimes \overline{\mu}^{\otimes\infty} \,\mathrm{d}\nu(\mu)$. Then the weak limit

$$\mu := m \cdot \operatorname*{w-lim}_{n\to\infty} \Xi_n(\boldsymbol{x})$$

exists almost surely and the random measure μ has law ν.

Proof: Define the normalizing map

$$\begin{aligned} \mathfrak{n} \colon \mathcal{M}_f(X) &\to \mathcal{M}_1(X) \cup \{o\} \\ \mu &\mapsto \overline{\mu} \end{aligned}$$

(recall the convention $\overline{o} = o$) and the infinite empirical distribution

$$\begin{aligned} \Xi_\infty \colon X^{\mathbb{N}} &\to \mathcal{M}_1(X) \cup \{o\} \\ \boldsymbol{x} &\mapsto \begin{cases} \text{w-lim}_n \Xi_n(\boldsymbol{x}), & \text{if the weak limit exists} \\ o, & \text{else.} \end{cases} \end{aligned}$$

The random sequence $\boldsymbol{x}$ has law

$$\int \overline{\mu}^{\otimes\infty} \,\mathrm{d}\nu(\mu) = \int \mu^{\otimes\infty} \,\mathrm{d}\mathfrak{n}_*\nu(\mu) \in \mathcal{M}_1(X^{\mathbb{N}}).$$

Thus, we can apply Proposition 4.15 to show that the weak limit $\text{w-lim}_n \Xi_n(\boldsymbol{x})$ exists almost surely.

To prove the second claim, we define

$$\begin{aligned} \pi \colon \mathcal{M}_f(X) &\to (\mathbb{R}_{>0} \times \mathcal{M}_1(X)) \cup \{(0, o)\} \\ \mu &\mapsto (\mathfrak{m}(\mu), \overline{\mu}). \end{aligned}$$

The function π is bijective with $\pi^{-1}(m, \mu) = m \cdot \mu$. Therefore, $m \cdot \Xi_\infty(\boldsymbol{x})$ has law ν if and only if $(m, \Xi_\infty(\boldsymbol{x}))$ has law $\pi_*\nu = \int \delta_{\mathfrak{m}(\mu)} \otimes \delta_{\overline{\mu}} \,\mathrm{d}\nu(\mu)$. From the construction of $(m, \Xi_\infty(\boldsymbol{x}))$ we know that its law is $\int \delta_{\mathfrak{m}(\mu)} \otimes \Xi_{\infty *}(\overline{\mu}^{\otimes\infty}) \,\mathrm{d}\nu(\mu)$. Hence, our claim is that

$$\int \delta_{\mathfrak{m}(\mu)} \otimes \delta_{\overline{\mu}} \,\mathrm{d}\nu(\mu) = \int \delta_{\mathfrak{m}(\mu)} \otimes \Xi_{\infty *}(\overline{\mu}^{\otimes\infty}) \,\mathrm{d}\nu(\mu).$$

To prove that both measures coincide, it suffices to show that the integrals over all test functions of the form $(m, \mu) \mapsto \mathbb{1}_A(m) F(\mu)$ with $A \in \mathcal{B}(\mathbb{R}_+)$, $F \in B(\mathcal{M}_1(X))$ coincide (cf. for example [EK86, Proposition 3.4.6]). That is, we have to show

$$\int \mathbb{1}_A(\mathfrak{m}(\mu)) F(\overline{\mu}) \,\mathrm{d}\nu(\mu) = \int \int \mathbb{1}_A(\mathfrak{m}(\mu)) F(\Xi_\infty(\boldsymbol{x})) \,\mathrm{d}\overline{\mu}^{\otimes\infty}(\boldsymbol{x}) \,\mathrm{d}\nu(\mu). \tag{E4.16}$$

Both sides vanish if $\nu(\mathcal{M}_A(X)) = 0$ with $\mathcal{M}_A(X) := \{\mu \in \mathcal{M}_f(X) \mid \mathfrak{m}(\mu) \in A\}$. Otherwise, we define $\nu_A(\,\cdot\,) := \nu(\,\cdot\, \cap \mathcal{M}_A(X))$. By using Remark 4.16 with $\mathfrak{n}_*\overline{\nu_A}$ instead of ν we see that

$$\begin{aligned}\int F(\overline{\mu})\,\mathrm{d}\overline{\nu_A}(\mu) &= \int F(\mu)\,\mathrm{d}\mathfrak{n}_*\overline{\nu_A}(\mu)\\ &= \int\int F(\Xi_\infty(\boldsymbol{x}))\,\mathrm{d}\mu^{\otimes\infty}(\boldsymbol{x})\,\mathrm{d}\mathfrak{n}_*\overline{\nu_A}(\mu)\\ &= \int\int F(\Xi_\infty(\boldsymbol{x}))\,\mathrm{d}\overline{\mu}^{\otimes\infty}(\boldsymbol{x})\,\mathrm{d}\overline{\nu_A}(\mu)\end{aligned}$$

Hence, we obtain

$$\begin{aligned}\int \mathbb{1}_A(\mathfrak{m}(\mu))F(\overline{\mu})\,\mathrm{d}\nu(\mu) &= \int F(\overline{\mu})\,\mathrm{d}\nu_A(\mu)\\ &= \mathfrak{m}(\nu_A)\int F(\overline{\mu})\,\mathrm{d}\overline{\nu_A}(\mu)\\ &= \mathfrak{m}(\nu_A)\int\int F(\Xi_\infty(\boldsymbol{x}))\,\mathrm{d}\overline{\mu}^{\otimes\infty}(\boldsymbol{x})\,\mathrm{d}\overline{\nu_A}(\mu)\\ &= \int\int \mathbb{1}_A(\mathfrak{m}(\mu))F(\Xi_\infty(\boldsymbol{x}))\,\mathrm{d}\overline{\mu}^{\otimes\infty}(\boldsymbol{x})\,\mathrm{d}\nu(\mu)\end{aligned}$$

and this establishes formula (E4.16). □

It follows that we can reconstruct a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_f(X))$ from a sample of masses and points in X.

Proposition 4.49 (Reconstruction of two-level measures)
Let X be a non-empty Polish space and let $\nu \in \mathcal{M}_1(\mathcal{M}_f(X))$ with $\nu(\{o\}) = 0$. Moreover, let $((m_i)_i, (x_{ij})_{ij}) \in \mathbb{R}^{\mathbb{N}} \times X^{\mathbb{N}\times\mathbb{N}}$ be random with law

$$\int \bigotimes_{i=1}^{\infty} \delta_{\mathfrak{m}(\mu_i)} \otimes \overline{\boldsymbol{\mu}}^{\otimes\infty}\,\mathrm{d}\nu^{\otimes\infty}(\boldsymbol{\mu}).$$

Then, the following assertions hold almost surely:

(a) For every $i \in \mathbb{N}$ the weak limit

$$\mu_i := m_i \cdot \operatorname*{w-lim}_{n\to\infty} \Xi_n((x_{ij})_j) = m_i \cdot \operatorname*{w-lim}_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \delta_{x_{ij}}$$

exists and the random measure μ_i has law ν.

(b) The two-level measure $\Xi_n((\mu_i)_i) = \frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_i}$ converges weakly to ν.

(c) The sequence $(x_{ij})_i$ is dense in $\operatorname{supp}\mathfrak{M}_\nu$ for every $j \in \mathbb{N}$.

Proof: If we fix $i \in \mathbb{N}$, then $(m_i, (x_{ij})_j)$ has law $\int \delta_{\mathfrak{m}(\mu)} \otimes \overline{\mu}^{\otimes\infty}\,\mathrm{d}\nu(\mu)$ and assertion (a) follows directly from Lemma 4.48.

Assertion (b) follows from the Glivenko-Cantelli theorem (Proposition 3.3), since $(\mu_i)_i$ is an i. i. d. sequence of random measures, each with law ν.

To see assertion (c), observe that $(x_{ij})_i$ is an i. i. d. sequence of random variables in X, each with law $\eta(\cdot) := \int \overline{\mu}(\cdot)\, \mathrm{d}\nu(\mu)$. By Corollary 3.4 the sequence is almost surely dense in $\operatorname{supp}\eta = \operatorname{supp}\mathfrak{M}_\nu$. □

The last proposition implies that we can even reconstruct both the measure ν *and* the space (X,r) if we only know the masses $(m_i)_i$ of the sampled measures and the distances $r(x_{ij},x_{kl})$ between all points x_{ij} and x_{kl} with $i,j,k,l \in \mathbb{N}$. We exploit this fact in the next section to show that a certain class of test functions on $\mathbb{M}^{(2)}$ separates points.

4.4.2 M2m-monomials and the two-level Gromov-weak topology

We now define a point separating class of test functions on $\mathbb{M}^{(2)}$, the so-called m2m-monomials, and study the induced topology. The m2m-monomials are based on the idea of sampling finite subspaces of an m2m space $[X,r,\nu]^{(2)}$. Similar to the m2pm-polynomials, we sample finitely many measures $\mu_1,\cdots,\mu_m \in \mathcal{M}_f(X)$ with ν and then sample points from X with each of these measures and evaluate the distances between the sampled points (or rather evaluate the finite metric space spanned by them). Since we are now using finite measures, we have to decompose the measures ν and $\mu_1,\ldots,\mu_m$ into their mass and their normalization to obtain bounded test functions.

Note that for m2m spaces we need to introduce three different types of test functions. We explain the reasons for this after the definition.

Definition 4.50 (M2m-monomials)
We say that a function $\Phi\colon \mathbb{M}^{(2)} \to \mathbb{R}$ is an m2m-monomial *if it has one of the following forms:*

$$\Phi([X,r,\nu]^{(2)}) = \chi(\mathfrak{m}(\nu)), \tag{TF1}$$

$$\Phi([X,r,\nu]^{(2)}) = \chi(\mathfrak{m}(\nu)) \int \psi(\mathfrak{m}(\boldsymbol{\mu}))\, \mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu}), \tag{TF2}$$

$$\Phi([X,r,\nu]^{(2)}) = \chi(\mathfrak{m}(\nu)) \int \psi(\mathfrak{m}(\boldsymbol{\mu})) \int \varphi\circ R(\boldsymbol{x})\, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}}(\boldsymbol{x})\, \mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu}), \tag{TF3}$$

where $m \in \mathbb{N}$, $\boldsymbol{n} \in \mathbb{N}^m$ with $|\boldsymbol{n}| \geq 2$, $\chi \in \mathcal{C}_b(\mathbb{R}_+)$, $\psi \in \mathcal{C}_b(\mathbb{R}_+^m)$, $\varphi \in \mathcal{C}_b(\mathbb{D}_{|\boldsymbol{n}|})$ with $\chi(0) = 0$ and $\psi(\boldsymbol{a}) = 0$ whenever any of the components of the vector $\boldsymbol{a} \in \mathbb{R}_+^m$ is 0.

The set of all m2m-monomials is denoted by $\mathcal{T}^{(2)}$.

Recall from section 4.2 that $\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} := \bigotimes_{i=1}^m \overline{\mu_i}^{\otimes n_i} \in \mathcal{M}_1(X^{|\boldsymbol{n}|})$.

We will prove in Theorem 4.53 that $\mathcal{T}^{(2)}$ separates points in $\mathbb{M}^{(2)}$. The three types of m2m-monomials serve different purposes for determining an m2m space $[X,r,\nu]^{(2)}$. Test functions of the form (TF1) simply determine the mass $\mathfrak{m}(\nu)$, whereas test functions of the form (TF2) determine the normalized mass distribution $\mathfrak{m}_*\overline{\nu}$. The space (X,r) (more precisely, the support $\operatorname{supp}\mathfrak{M}_\nu$ equipped with the restriction of r) and the structure of $\overline{\nu}$ are determined by the test functions of type (TF3).

Remarks 4.51

(a) Note that $\mathcal{T}^{(2)}$ is closed under multiplication. We show this for two m2m-monomials of type (TF3). For $i \in \{1,2\}$ let Φ_i be an m2m-monomial with

$$\Phi_i([X,r,\nu]^{(2)}) = \chi_i(\mathfrak{m}(\nu)) \int \psi_i(\mathfrak{m}(\boldsymbol{\mu})) \int \varphi_i \circ R(\boldsymbol{x}) \, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n_i}}(\boldsymbol{x}) \, \mathrm{d}\overline{\nu}^{\otimes m_i}(\boldsymbol{\mu})$$

with $m_i, \boldsymbol{n_i}, \chi_i, \psi_i, \varphi_i$ as in (TF3). Then $(\Phi_1 \cdot \Phi_2)([X,r,\nu]^{(2)})$ is equal to

$$\chi_1(\mathfrak{m}(\nu))\chi_2(\mathfrak{m}(\nu)) \int \psi_1(\mathfrak{m}(\mu_1))\psi_2(\mathfrak{m}(\mu_2)) \int \varphi_1 \circ R(\boldsymbol{x_1}) \; \varphi_2 \circ R(\boldsymbol{x_2}) \, \mathrm{d}(\overline{\boldsymbol{\mu_1}}^{\otimes \boldsymbol{n_1}} \otimes \overline{\boldsymbol{\mu_2}}^{\otimes \boldsymbol{n_2}})(\boldsymbol{x_1}, \boldsymbol{x_2}) \, \mathrm{d}\overline{\nu}^{\otimes (m_1+m_2)}(\boldsymbol{\mu_1}, \boldsymbol{\mu_2})$$

It is easy to see that this can be written as an m2m-monomial of the form (TF3) with $m = m_1 + m_2$ and $\boldsymbol{n} = (\boldsymbol{n_1}, \boldsymbol{n_2})$. With similar arguments it follows that $\Phi_1 \cdot \Phi_2 \in \mathcal{T}^{(2)}$ even if Φ_1, Φ_2 are of type (TF1) or (TF2).

However, $\Phi_1 + \Phi_2$ is in general not an m2m-monomial anymore, hence $\mathcal{T}^{(2)}$ is *not* closed under addition.

(b) We think of elements of $\mathcal{T}^{(2)}$ as being monomials of the two-level measure ν, hence the name m2m-monomial. Elements of the linear span of $\mathcal{T}^{(2)}$ are called *m2m-polynomials*. However, the set $\mathcal{T}^{(2)}$ is sufficient for our needs and we will not use the algebra of m2m-polynomials. See also Remarks 3.11 for the naming convention about monomials and polynomials

(c) With some abuse of notation we are able to write all three types of m2m-monomials in the form (TF3). If we define $\int \ldots \mathrm{d}\lambda^{\otimes 0} = 1$ for every measure λ and allow $m \in \mathbb{N}_0$ and $\boldsymbol{n} \in \mathbb{N}_0^m$, then (TF1) and (TF2) can also be written as (TF3) (with $m = 0$ or with $\boldsymbol{n} = (0, \ldots, 0)$, respectively).

However, this notation is somewhat questionable, since there is no measure such that the integral over every function is 1 and since the natural choice for $\lambda^{\otimes 0}$ is the null measure o, which would imply $\int \ldots \mathrm{d}\lambda^{\otimes 0} = 0$. Moreover, the three types of m2m-monomials determine different parts of an m2m space and will be used separately in proofs. For these reasons we prefer to distinguish between the m2m-monomials of the form (TF1), (TF2) and (TF3).

Lemma 4.52
Every m2m-monomial $\Phi \in \mathcal{T}^{(2)}$ is well-defined, that is we have

$$\Phi([X,r,\nu]^{(2)}) = \Phi([Y,d,\lambda]^{(2)})$$

if the two m2m triples $(X,r,\nu)^{(2)}$ and $(Y,d,\lambda)^{(2)}$ are m2m-isomorphic.

Proof: Let $\mathcal{X} = (X,r,\nu)^{(2)}$ and $\mathcal{Y} = (Y,d,\lambda)^{(2)}$ be isomporhic m2m triples. That is, there is a function $f\colon X \to Y$ such that f is isometric on $\operatorname{supp} \mathfrak{M}_\nu$ and $\lambda = f_{**}\nu$. For every $n\in\mathbb{N}$ we define $f_n\colon X^n\to Y^n$ by $f_n(x_1,\dots,x_n)=(f(x_1),\dots,f(x_n))$. Because of the isometric properties of f, we have $R^Y(f_n(\boldsymbol{x}))=R^X(\boldsymbol{x})$ for every $\boldsymbol{x}\in(\operatorname{supp}\mathfrak{M}_\nu)^n$. Using the transformation formulas (E2.5), (E2.6) and the fact that f_* and f_{**} do not change the mass of the measures, we conclude that for every Φ as in (TF3)

$$\begin{aligned}\Phi([Y,d,\lambda]^{(2)})&=\Phi([Y,d,f_{**}\nu]^{(2)})\\&=\chi(\mathfrak{m}(f_{**}\nu))\int\psi(\mathfrak{m}(\boldsymbol{\mu}))\int\varphi\circ R^Y(\boldsymbol{x})\,\mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}(\boldsymbol{x})\,\mathrm{d}\overline{f_{**}\nu}^{\otimes m}(\boldsymbol{\mu})\\&=\chi(\mathfrak{m}(\nu))\int\psi(\mathfrak{m}(f_{m*}\boldsymbol{\mu}))\int\varphi\circ R^Y(\boldsymbol{x})\,\mathrm{d}\overline{f_{m*}\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}(\boldsymbol{x})\,\mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu})\\&=\chi(\mathfrak{m}(\nu))\int\psi(\mathfrak{m}(\boldsymbol{\mu}))\int\varphi\circ R^Y(f_{|\boldsymbol{n}|}(\boldsymbol{x}))\,\mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}(\boldsymbol{x})\,\mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu})\\&=\chi(\mathfrak{m}(\nu))\int\psi(\mathfrak{m}(\boldsymbol{\mu}))\int\varphi\circ R^X(\boldsymbol{x})\,\mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}(\boldsymbol{x})\,\mathrm{d}\overline{\nu}^{\otimes m}(\boldsymbol{\mu})\\&=\Phi([X,r,\nu]^{(2)}).\end{aligned}$$

Equality for m2m-monomials of type (TF1) and (TF2) follows in the same way. □

Theorem 4.53
$\mathcal{T}^{(2)}$ separates points in $\mathbb{M}^{(2)}$.

Proof: Let $\mathcal{X}=[X,r,\nu]^{(2)}$ and $\mathcal{Y}=[Y,d,\lambda]^{(2)}$ be m2m spaces with $\Phi(\mathcal{X})=\Phi(\mathcal{Y})$ for every $\Phi\in\mathcal{T}^{(2)}$. We exclude the trivial case $\nu=\lambda=o$. Because $\Phi(\mathcal{X})=\Phi(\mathcal{Y})$ for every m2m-monomial of type (TF1) and (TF2), we conclude that $\mathfrak{m}(\nu)=\mathfrak{m}(\lambda)$ and $\mathfrak{m}_*\nu=\mathfrak{m}_*\lambda$. In particular we have $\nu(\{o\})=\lambda(\{o\})$. Without loss of generality we may assume that ν and λ are *probability* measures with $\nu(\{o\})=\lambda(\{o\})=0$. Now, from the equality for all $\Phi\in\mathcal{T}^{(2)}$ of type (TF3) it follows that

$$\int\bigotimes_{i=1}^m\delta_{\mathfrak{m}(\mu_i)}\otimes R^X_*\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu})=\int\bigotimes_{i=1}^m\delta_{\mathfrak{m}(\mu_i)}\otimes R^Y_*\overline{\boldsymbol{\mu}}^{\otimes\boldsymbol{n}}\,\mathrm{d}\lambda^{\otimes m}(\boldsymbol{\mu})$$

for all $m\in\mathbb{N}$ and $\boldsymbol{n}\in\mathbb{N}^m$ with $|\boldsymbol{n}|\geq 2$. By taking the projective limit we obtain

$$\int\bigotimes_{i=1}^\infty\delta_{\mathfrak{m}(\mu_i)}\otimes R^X_*\overline{\boldsymbol{\mu}}^{\otimes\infty}\,\mathrm{d}\nu^{\otimes\infty}(\boldsymbol{\mu})=\int\bigotimes_{i=1}^\infty\delta_{\mathfrak{m}(\mu_i)}\otimes R^Y_*\overline{\boldsymbol{\mu}}^{\otimes\infty}\,\mathrm{d}\lambda^{\otimes\infty}(\boldsymbol{\mu}).$$

Together with Proposition 4.49 this implies that there are $\boldsymbol{m}\in\mathbb{R}_+^{\mathbb{N}}$ and dense matrices $\boldsymbol{x}=(x_{ij})_{ij}\in X^{\mathbb{N}\times\mathbb{N}}$, $\boldsymbol{y}=(y_{ij})_{ij}\in Y^{\mathbb{N}\times\mathbb{N}}$ with

$$(\boldsymbol{m},R^X(\boldsymbol{x}))=(\boldsymbol{m},R^Y(\boldsymbol{y}))\tag{E4.17}$$

such that

$$\nu=\operatorname*{w-lim}_n\Xi_n((\mu_i)_i)\quad\text{with}\quad\mu_i:=m_i\cdot\operatorname*{w-lim}_n\Xi_n((x_{ij})_j)$$

and

$$\lambda = \text{w-}\lim_n \Xi_n((\eta_i)_i) \quad \text{with} \quad \eta_i := m_i \cdot \text{w-}\lim_n \Xi_n((y_{ij})_j).$$

Equation (E4.17) yields

$$r(x_{ij}, x_{kl}) = d(y_{ij}, y_{kl}) \text{ for all } i,j,k,l \in \mathbb{N}.$$

If we define a function f' by $f'(x_{ij}) := y_{ij}$ for all $i,j \in \mathbb{N}$, then f' is isometric and can be extended continuously to an isometric function $f\colon \operatorname{supp}\mathfrak{M}_\nu \to Y$. By construction we have

$$\Xi_n((y_{ij})_j) = f_*\Xi_n((x_{ij})_j)$$

for all $i,n \in \mathbb{N}$. Since the push-forward operator f_* is continuous by Lemma 2.24 we obtain $\eta_i = f_*\mu_i$ for each $i \in \mathbb{N}$ by taking the limit $n \to \infty$. It follows that

$$\Xi_n((\eta_i)_i) = f_{**}\Xi_n((\mu_i)_i)$$

for every $n \in \mathbb{N}$. By taking the limit $n \to \infty$ again, we get $\lambda = f_{**}\nu$ and this completes the proof. □

Remark 4.54 (There is no reconstruction theorem for m2m spaces)
In the preceding proof we reconstruct an m2m space $[X,r,\nu]^{(2)}$ with $\nu(\{o\}) = 0$ and $\mathfrak{m}(\nu) = 1$ from the measure

$$\int \bigotimes_{i=1}^{\infty} \delta_{\mathfrak{m}(\mu_i)} \otimes R_*\overline{\boldsymbol{\mu}}^{\otimes\infty} \, \mathrm{d}\nu^{\otimes\infty}(\boldsymbol{\mu})$$

in $\mathcal{M}_1(\mathbb{R}_+^{\mathbb{N}} \times \mathbb{D}_{\mathbb{N}\times\mathbb{N}})$. One might be tempted to think that a general m2m space $[X,r,\nu]^{(2)}$ is determined by the measure

$$\mathfrak{m}(\nu) \cdot \int \bigotimes_{i=1}^{\infty} \delta_{\mathfrak{m}(\mu_i)} \otimes R_*\overline{\boldsymbol{\mu}}^{\otimes\infty} \, \mathrm{d}\overline{\nu}^{\otimes\infty}(\boldsymbol{\mu}) \tag{E4.18}$$

in $\mathcal{M}_f(\mathbb{R}_+^{\mathbb{N}} \times \mathbb{D}_{\mathbb{N}\times\mathbb{N}})$. However, the association between an m2m space and the measure in (E4.18) is not unique. For instance, the measure in (E4.18) is equal to o for every m2m space $[\{x\}, 0, c\delta_o]^{(2)}$ with $c \geq 0$. The problem here is that $\nu(\{o\})$ may be positive and that we cannot sample points from X with the null measure. For this reason there is no reconstruction theorem for general m2m spaces, i. e. we cannot identify an m2m space with a measure on distance array distributions like we did in Theorem 4.20 for m2pm spaces.

We use the m2m-monomials to define a topology on $\mathbb{M}^{(2)}$. Since $\mathcal{T}^{(2)}$ separates points, the induced topology is a Hausdorff topology.

Definition 4.55
The two-level Gromov-weak topology τ_{2Gw} *on* $\mathbb{M}^{(2)}$ *is defined as the initial topology induced by* $\mathcal{T}^{(2)}$.

It is easy to see that the restriction of τ_{2Gw} to $\mathbb{M}_1^{(2)}$ coincides with the topology τ'_{2Gw} of Definition 4.26. That is why we call both τ'_{2Gw} and τ_{2Gw} the two-level Gromov-weak topology.

Remark 4.56 (Weak convergence implies $\boldsymbol{\tau_{2Gw}}$ convergence)
Note that weak convergence of two-level measures implies two-level Gromov-weak convergence of the corresponding m2m spaces. To see this, let (X,r) be a fixed Polish metric space and let $\Phi \in \mathcal{T}^{(2)}$ be arbitrary. The function

$$\begin{aligned} \mathcal{M}_f(\mathcal{M}_f(X)) &\to \mathbb{R} \\ \nu &\mapsto \Phi([X,r,\nu]^{(2)}) \end{aligned}$$

is a composition of functions which all are continuous (recall from Corollary 2.22 that $\overline{\nu} \mapsto \overline{\nu}^{\otimes m}$ and $\overline{\mu} \mapsto \overline{\mu}^{\otimes \boldsymbol{n}}$ are continuous) except for the normalizing functions $\nu \mapsto \overline{\nu}$ and $\boldsymbol{\mu} \mapsto \overline{\boldsymbol{\mu}}$. However, these discontinuities are smoothed by the constraints on χ and ψ in the definition of the m2m-monomials (recall that we require $\chi(0) = 0$ and $\psi(\boldsymbol{a}) = 0$ whenever any of the components of the vector $\boldsymbol{a}$ is 0). Therefore, the function $\nu \mapsto \Phi([X,r,\nu]^{(2)})$ is continuous for every $\Phi \in \mathcal{T}^{(2)}$ and $\nu_n \xrightarrow{w} \nu$ implies that $([X,r,\nu_n]^{(2)})_n$ converges two-level Gromov-weakly to $[X,r,\nu]^{(2)}$.

The notion of m2m spaces is in fact a generalization of the notion of mm spaces. Every metric measure space $[X,r,\mu] \in \mathbb{M}$ is associated to a metric two-level measure space $[X,r,\delta_\mu]^{(2)} \in \mathbb{M}^{(2)}$. In this way we can embed the topological space $(\mathbb{M},\tau_{Gw})$ in $(\mathbb{M}^{(2)},\tau_{2Gw})$, as the following lemma shows.

Lemma 4.57 (m2m spaces are a generalization of mm spaces)
The function

$$\begin{aligned} (\mathbb{M},\tau_{Gw}) &\to (\mathbb{M}^{(2)},\tau_{2Gw}) \\ [X,r,\mu] &\mapsto [X,r,\delta_\mu]^{(2)} \end{aligned}$$

is an embedding, i. e. it is injective and bicontinuous.

Proof: Observe that every mm-monomial $\Phi \in \mathcal{T}^{(1)}$ can be rewritten as an m2m-monomial by

$$\Phi([X,r,\mu]) = \psi(\mathfrak{m}(\mu)) \int \varphi \circ R \,\mathrm{d}\overline{\mu}^{\otimes n} = \int \psi(\mathfrak{m}(\eta)) \int \varphi \circ R \,\mathrm{d}\overline{\eta}^{\otimes n} \,\mathrm{d}\delta_\mu(\eta).$$

Conversely, every m2m-monomial $\Phi \in \mathcal{T}^{(2)}$ of the form (TF3) can be written as an mm-monomial by

$$\begin{aligned} \Phi([X,r,\delta_\mu]^{(2)}) &= \chi(\mathfrak{m}(\delta_\mu)) \int \psi(\mathfrak{m}(\boldsymbol{\eta})) \int \varphi \circ R \,\mathrm{d}\overline{\boldsymbol{\eta}}^{\otimes \boldsymbol{n}} \,\mathrm{d}\delta_\mu^{\otimes m}(\boldsymbol{\eta}) \\ &= \tilde{\psi}(\mathfrak{m}(\mu)) \int \varphi \circ R \,\mathrm{d}\overline{\mu}^{\otimes |\boldsymbol{n}|}, \end{aligned}$$

where $\tilde{\psi}(x) = \chi(1)\psi(x,\dots,x)$, and a similar statement holds true for m2m-monomials of the form (TF2). Because the monomials separate points and induce the topologies on $\mathbb{M}$ and $\mathbb{M}^{(2)}$, this implies injectivity and bicontinuity of the function given in the statement. □

Since there is no reconstruction theorem for general m2m spaces (cf. Remark 4.54), we cannot embed $\mathbb{M}^{(2)}$ in a space of measures like we did with $\mathbb{M}_1^{(2)}$ in Theorem 4.28. Thus, there is no easy way of proving that τ_{2Gw} is metrizable. We therefore use nets instead of sequences in order to prove continuity of functions or compactness of sets.

Sometimes it is convenient to work with simpler test functions than (TF3). For this reason we provide equivalent conditions for two-level Gromov-weak convergence in the following lemma.

Lemma 4.58 (Alternative test functions for τ_{2Gw})
Let $(\mathcal{X}_\alpha)_{\alpha\in\mathcal{A}}$ be a net of m2m spaces with $\mathcal{X}_\alpha = [X_\alpha, r_\alpha, \nu_\alpha]^{(2)}$ and let $\mathcal{X} = [X, r, \nu]^{(2)}$ be another m2m space. $(\mathcal{X}_\alpha)_\alpha$ converges two-level Gromov-weakly to $\mathcal{X}$ if and only if the following two conditions hold:

(a) $\mathfrak{m}_*\nu_\alpha$ converges weakly to $\mathfrak{m}_*\nu$ in $\mathcal{M}_f(\mathbb{R}_+)$

(b) $\tilde{\Phi}(\mathcal{X}_\alpha)$ converges to $\tilde{\Phi}(\mathcal{X})$ for every $\tilde{\Phi}\colon \mathbb{M}^{(2)} \to \mathbb{R}$ of the form

$$\tilde{\Phi}([X,r,\nu]^{(2)}) = \int \psi(\mathfrak{m}(\boldsymbol{\mu})) \int \varphi \circ R \,\mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} \,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) \tag{TF4}$$

with $m \in \mathbb{N}$, $\boldsymbol{n} \in \mathbb{N}^m$ with $|\boldsymbol{n}| \geq 2$, $\varphi \in \mathcal{C}_b(\mathbb{D}_{\boldsymbol{n}})$, and $\psi \in \mathcal{C}_b(\mathbb{R}_+^m)$ with $\psi(\boldsymbol{a}) = 0$ whenever any of the components of the vector $\boldsymbol{a} \in \mathbb{R}_+^m$ is 0.

Note that in (TF4) we use ν instead of the normalized version $\overline{\nu}$.

Proof: (a) is equivalent to convergence of all m2m-monomials of type (TF1) and (TF2). (b) follows from the convergence of functions of the form (TF3) by choosing the function $\chi \in \mathcal{C}_b(\mathbb{R}_+)$ such that $\chi(x) = x^m$ for $x \in [0, \mathfrak{m}(\nu)+1]$. The other direction is obvious. □

Remark 4.59
The previous lemma shows that the functions $[X,r,\nu]^{(2)} \mapsto \mathfrak{m}(\nu)$ and $[X,r,\nu]^{(2)} \mapsto \mathfrak{m}_*\nu$ are continuous on $\mathbb{M}^{(2)}$ in the two-level Gromov-weak topology.

4.4.3 The two-level Gromov-Prokhorov metric and its topology

Since the Prokhorov metric d_{P} is defined for finite measures, we can easily extend the metric $d_{2\mathrm{GP}}$ of Definition 4.30 to $\mathbb{M}^{(2)}$. We call this extension again the two-level Gromov-Prokhorov metric.

Definition 4.60 (Two-level Gromov-Prokhorov metric and topology)
Let $\mathcal{X} = [X,r,\nu]^{(2)}$ and $\mathcal{Y} = [Y,d,\lambda]^{(2)}$ be two m2m spaces. We define the two-level Gromov-Prokhorov distance *$d_{2\mathrm{GP}}(\mathcal{X},\mathcal{Y})$ between $\mathcal{X}$ and $\mathcal{Y}$ by*

$$d_{2\mathrm{GP}}(\mathcal{X},\mathcal{Y}) := \inf_{Z,\iota_X,\iota_Y} d_{\mathrm{P}}^{\mathcal{M}_f(Z)}(\iota_{X**}\nu, \iota_{Y**}\lambda), \tag{E4.19}$$

where the infimum ranges over all isometric embeddings $\iota_X : X \to Z$, $\iota_Y : Y \to Z$ *into a common Polish metric space* (Z, r_Z) *with* $Z \subset \mathbb{R}^{\mathbb{N}}$ *and where* $d_{\mathrm{P}}^{\mathcal{M}_f(Z)}$ *denotes the Prokhorov metric for measures on the space* $\mathcal{M}_f(Z)$.

The induced topology on $\mathbb{M}^{(2)}$ *is called the* two-level Gromov-Prokhorov topology *and denoted by* τ_{2GP}.

All results of section 4.3.3 remain true for the extension of $d_{2\mathrm{GP}}$ to $\mathbb{M}^{(2)}$. Most of the proofs are literally the same. We simply need to replace $\mathcal{M}_1(\dots)$ and $\mathcal{M}_1(\mathcal{M}_1(\dots))$ by $\mathcal{M}_f(\dots)$ and $\mathcal{M}_f(\mathcal{M}_f(\dots))$, respectively. This holds in particular for the alternative characterization of $d_{2\mathrm{GP}}$ in Lemma 4.32 and the embedding Lemmas 4.33 and 4.34.

Lemma 4.61 (τ_{2GP} is finer than τ_{2Gw})
Every m2m-monomial $\Phi \in \mathcal{T}^{(2)}$ *is continuous with respect to the two-level Gromov-Prokhorov topology. Hence, two-level Gromov-Prokhorov convergence implies two-level Gromov-weak convergence and* τ_{2GP} *is finer than* τ_{2Gw}.

Proof: The proof is the same as the proof of Lemma 4.35 but uses Remark 4.56 instead of Remark 4.27. □

Note that we will show in section 4.4.6 that in fact both topologies τ_{2GP} and τ_{2Gw} coincide. But for the time being we are only able to prove one direction.

Theorem 4.62
$(\mathbb{M}^{(2)}, d_{2\mathrm{GP}})$ *is a Polish metric space.*

Proof: There are only two differences to the proof of Theorem 4.36. The first one is that we need to use m2m-monomials instead of m2pm-polynomials to show the positive definiteness of the metric: Let $\mathcal{X}, \mathcal{Y}$ be m2m spaces with $d_{2\mathrm{GP}}(\mathcal{X}, \mathcal{Y}) = 0$. With Lemma 4.61 we get $\Phi(\mathcal{X}) = \Phi(\mathcal{Y})$ for every m2m-monomial $\Phi \in \mathcal{T}^{(2)}$ (by using the constant sequence $\mathcal{X}_n = \mathcal{X}$). Since $\mathcal{T}^{(2)}$ separates points in $\mathbb{M}^{(2)}$, this implies $\mathcal{X} = \mathcal{Y}$.

The second difference to the proof of Theorem 4.36 is that we need to find a different countable dense subset $\mathbb{S} \subset \mathbb{M}^{(2)}$. Define $\mathbb{S}$ as the set of all m2m spaces $[S, d, \lambda]^{(2)}$ such that $|S| < \infty$, the metric d takes only rational values and

$$\lambda = \sum_{i=1}^{M} a_i \delta_{\left(\sum_{j=1}^{N_i} b_{ij} \delta_{x_{ij}}\right)} \tag{E4.20}$$

with $M, N_1, \dots, N_M \in \mathbb{N}$, $x_{ij} \in S$, $a_i, b_{ij} \in \mathbb{Q}_+$ for all i, j (the difference to the set $\mathbb{S}_1$ from the proof of Theorem 4.36 is that the values a_i and b_{ij} do not need to sum up to 1 anymore). That is, $(S, d, \lambda)^{(2)}$ is a finite m2m triple with only rational distances and λ is a finite atomic measure on finite atomic measures with only rational values. The set $\mathbb{S}$ is obviously countable. To prove density, let $[X, r, \nu]^{(2)}$ be an arbitrary m2m space and let $\varepsilon > 0$. Because the set of measures of the form (E4.20) is dense in $\mathcal{M}_f(\mathcal{M}_f(X))$, there is a $\lambda \in \mathcal{M}_f(\mathcal{M}_f(X))$ of this form with $d_{\mathrm{P}}(\lambda, \nu) < \frac{\varepsilon}{2}$. Then, $S := \operatorname{supp} \mathfrak{M}_\lambda$ is finite and $d_{2\mathrm{GP}}([X, r, \nu]^{(2)}, [S, d, \lambda]^{(2)}) < \frac{\varepsilon}{2}$. The last step is to approximate the metric r by a rational version d such that $|d(x,y) - r(x,y)| < \frac{\varepsilon}{2}$ for all $x, y \in S$. Then, $[S, d, \lambda]^{(2)}$ is in $\mathbb{S}$ and we have $d_{2\mathrm{GP}}([X, r, \nu]^{(2)}, [S, d, \lambda]^{(2)}) < \varepsilon$. □

4.4.4 Distance distribution and modulus of mass distribution

Now we provide continuity results for the distance distribution and the modulus of mass distribution for m2m spaces. Recall that we showed in section 4.3.4 how the convergence (or rather divergence) of a sequence $([X_n,r_n,\nu_n]^{(2)})_n$ of m2pm spaces can be described in terms of $w(\mathfrak{M}_{\nu_n})$ and $V_\delta(\mathfrak{M}_{\nu_n})$. However, the first moment measure $\mathfrak{M}_\nu$ of a general m2m space $[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)}$ may be infinite. Because the distance distribution and the modulus of mass distribution are only meaningful for finite measures, $w(\mathfrak{M}_\nu)$ and $V_\delta(\mathfrak{M}_\nu)$ are not very useful in the general setting. The remedy is to approximate the measure ν in such a way that the moment measures are finite. We introduced such an approximation on page 56 in section 4.2. There we defined density functions $(f_K)_{K\geq 0}$ such that the moment measures of $f_K \cdot \nu$ are finite and $f_K \cdot \nu \xrightarrow{w} \nu$ for $K \to \infty$. The following lemma summarizes the continuity behavior of this approximation.

Lemma 4.63 (Properties of the approximation $f_K \cdot \nu$)
The following statements hold true in the two-level Gromov-weak topology τ_{2Gw} and in the two-level Gromov-Prokhorov topology τ_{2GP}:

(a) *The function*

$$\begin{aligned} \mathbb{M}^{(2)} &\to \mathbb{M}^{(2)} \\ [X,r,\nu]^{(2)} &\mapsto [X,r,f_K\cdot\nu]^{(2)} \end{aligned}$$

is continuous for every $K>0$.

(b) *The function*

$$\begin{aligned} \mathbb{M}^{(2)} &\to \mathbb{M} \\ [X,r,\nu]^{(2)} &\mapsto [X,r,\mathfrak{M}_{f_K\cdot\nu}] \end{aligned}$$

is continuous for every $K>0$, where $\mathbb{M}$ denotes the set of metric measure spaces equipped with the Gromov-weak topology.

(c) $[X,r,f_K\cdot\nu]^{(2)} \to [X,r,\nu]^{(2)}$ *for $K\to\infty$ and for every m2m space* $[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)}$.

Proof: (a) with respect to τ_{2Gw}: Fix $K>0$ and let the net $([X_\alpha,r_\alpha,\nu_\alpha]^{(2)})_{\alpha\in\mathcal{A}}$ converge to $[X,r,\nu]^{(2)}$ in the two-level Gromov-weak topology. We use Lemma 4.58 to show that $[X_\alpha,r_\alpha,f_K\cdot\nu_\alpha]^{(2)}$ converges to $[X,r,f_K\cdot\nu]^{(2)}$. Since $\mathfrak{m}_*\nu_\alpha \xrightarrow{w} \mathfrak{m}_*\nu$ (cf. Remark 4.59) and g_K is continuous and bounded, we have for every $h\in\mathcal{C}_b(\mathbb{R}_+)$

$$\int h\,\mathrm{d}\mathfrak{m}_*(f_K\cdot\nu_\alpha) = \int h(\mathfrak{m}(\mu))g_K(\mathfrak{m}(\mu))\,\mathrm{d}\nu_\alpha(\mu) = \int h(z)g_K(z)\,\mathrm{d}\mathfrak{m}_*\nu_\alpha(z)$$

and this converges to

$$\int h(z)g_K(z)\,\mathrm{d}\mathfrak{m}_*\nu(z) = \int h\,\mathrm{d}\mathfrak{m}_*(f_K\cdot\nu).$$

It follows that $\mathfrak{m}_*(f_K \cdot \nu_\alpha)$ converges weakly to $\mathfrak{m}_*(f_K \cdot \nu)$. Now let $\tilde{\Phi}$ be as in (TF4). Then

$$\begin{aligned}\tilde{\Phi}([X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}) &= \int \psi(\mathfrak{m}(\boldsymbol{\mu})) \int \varphi \circ R \, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} \, \mathrm{d}(f_K \cdot \nu_\alpha)^{\otimes m}(\boldsymbol{\mu}) \\ &= \int \psi(\mathfrak{m}(\boldsymbol{\mu})) \prod_{i=1}^m g_K(\mathfrak{m}(\mu_i)) \int \varphi \circ R \, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} \, \mathrm{d}\nu_\alpha^{\otimes m}(\boldsymbol{\mu})\end{aligned}$$

and this converges to

$$\int \psi(\mathfrak{m}(\boldsymbol{\mu})) \prod_{i=1}^m g_K(\mathfrak{m}(\mu_i)) \int \varphi \circ R \, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} \, \mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) = \tilde{\Phi}([X, r, f_K \cdot \nu]^{(2)})$$

because of the τ_{2Gw}-convergence of the net $([X_\alpha, r_\alpha, \nu_\alpha]^{(2)})_\alpha$.

Therefore, both conditions of Lemma 4.58 are satisfied and $[X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}$ converges to $[X, r, f_K \cdot \nu]^{(2)}$.

(a) with respect to τ_{2GP}: Let $([X_n, r_n, \nu_n]^{(2)})_n$ be a τ_{2GP}-convergent sequence of m2m spaces with limit $[X, r, \nu]^{(2)}$. By Lemma 4.34 we can embed all the metric spaces isometrically into a common Polish metric space (Z, r_Z) such that the measures $(\nu_n)_n$ converge weakly to ν in $\mathcal{M}_f(\mathcal{M}_f(Z))$. Recall from Lemma 4.11 that the function $\nu \to f_K \cdot \nu$ is weakly continuous on $\mathcal{M}_f(\mathcal{M}_f(Z))$. Therefore, $f_K \cdot \nu_n$ converges weakly to $f_K \cdot \nu$ in $\mathcal{M}_f(\mathcal{M}_f(Z))$. Since weak convergence implies convergence of the corresponding m2m spaces in the two-level Gromov-Prokhorov topology (see Remark 4.56), we see that $[X_n, r_n, f_K \cdot \nu_n]^{(2)} \to [X, r, f_K \cdot \nu]^{(2)}$ with respect to τ_{2GP}.

(b): Recall from Lemma 4.61 that two-level Gromov-Prokhorov topology τ_{2GP} is finer than the two-level Gromov-weak topology τ_{2Gw}. Thus, it suffices to show continuity only with respect to τ_{2Gw}. Let $K > 0$ and let the net $([X_\alpha, r_\alpha, \nu_\alpha]^{(2)})_{\alpha \in \mathcal{A}}$ converge to $[X, r, \nu]^{(2)}$ in the two-level Gromov-weak topology. By assertion (a), $[X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}$ converges two-level Gromov-weakly to $[X, r, f_K \cdot \nu]^{(2)}$. We want to show that the *metric measure spaces* $[X_\alpha, r_\alpha, \mathfrak{M}_{f_K \cdot \nu_\alpha}]$ converge Gromov-weakly to $[X, r, \mathfrak{M}_{f_K \cdot \nu}]$. By Lemma 3.15 we need to show that $\tilde{\Phi}([X_\alpha, r_\alpha, \mathfrak{M}_{f_K \cdot \nu_\alpha}])$ converges to $\tilde{\Phi}([X, r, \mathfrak{M}_{f_K \cdot \nu}])$ for every test function $\tilde{\Phi} \colon \mathbb{M} \to \mathbb{R}_+$ of the form

$$\tilde{\Phi}([X, r, \mu]) = \int \varphi \circ R \, \mathrm{d}\mu^{\otimes m}$$

with $m \in \mathbb{N}_{\geq 2}$ and $\varphi \in \mathcal{C}_b(\mathbb{D}_m)$. Fix such a $\tilde{\Phi}$ and let $\boldsymbol{n} := (1, \ldots, 1) \in \mathbb{N}^m$. With the definition of the first moment measure $\mathfrak{M}_{f_K \cdot \nu}$ we obtain

$$\tilde{\Phi}([X_\alpha, r_\alpha, \mathfrak{M}_{f_K \cdot \nu_\alpha}]) = \int \varphi \circ R \, \mathrm{d}(\mathfrak{M}_{f_K \cdot \nu_\alpha})^{\otimes m} = \int \int \varphi \circ R \, \mathrm{d}\boldsymbol{\mu}^{\otimes \boldsymbol{n}} \, \mathrm{d}(f_K \cdot \nu_\alpha)^{\otimes m}(\boldsymbol{\mu}).$$

If we choose a bounded function $\psi \in \mathcal{C}_b(\mathbb{R}_+^m)$ such that $\psi(x_1, \ldots x_m) = \prod_{i=1}^m x_i$ on $[0, K]^m$, the right hand side of the last equation can be written as

$$\Phi([X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}) := \int \psi(\mathfrak{m}(\boldsymbol{\mu})) \int \varphi \circ R \, \mathrm{d}\overline{\boldsymbol{\mu}}^{\otimes \boldsymbol{n}} \, \mathrm{d}(f_K \cdot \nu_\alpha)^{\otimes m}(\boldsymbol{\mu}).$$

Φ is a test function as in (TF4). Since $[X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}$ converges two-level Gromov-weakly, we also have convergence of all test functions of this form (by Lemma 4.58) and thus

$$\tilde{\Phi}([X_\alpha, r_\alpha, \mathfrak{M}_{f_K\cdot\nu_\alpha}]) = \Phi([X_\alpha, r_\alpha, f_K \cdot \nu_\alpha]^{(2)}) \to \Phi([X, r, f_K \cdot \nu]^{(2)}) = \tilde{\Phi}([X, r, \mathfrak{M}_{f_K\cdot\nu}]).$$

(c): $f_K \cdot \nu$ converges weakly to ν for $K \to \infty$ by Lemma 4.11. Thus, $[X, r, f_K \cdot \nu]^{(2)}$ converges to $[X, r, \nu]^{(2)}$ in the two-level Gromov-Prokhorov metric. Since two-level Gromov-Prokhorov convergence implies two-level Gromov-weak convergence (cf. Lemma 4.61), assertion (c) is true for both topologies. □

We can now apply the distance distribution and the modulus of mass distribution to the first moment measure $\mathfrak{M}_{f_K\cdot\nu}$ of the approximation $f_K \cdot \nu$. The following corollaries show that $w(\mathfrak{M}_{f_K\cdot\nu})$ and $V_\delta(\mathfrak{M}_{f_K\cdot\nu})$ are (semi-) continuous. The results follow immediately by combining assertion (b) of Lemma 4.63 with Corollary 3.36 and Lemma 3.37.

Corollary 4.64
For every $K > 0$ the function

$$\begin{aligned} \mathbb{M}^{(2)} &\to \mathcal{M}_f(\mathbb{R}_+) \\ [X, r, \nu]^{(2)} &\mapsto w(\mathfrak{M}_{f_K\cdot\nu}) \end{aligned}$$

is continuous with respect to both the two-level Gromov-weak topology τ_{2Gw} and the two-level Gromov-Prokhorov topology τ_{2GP}.

Corollary 4.65
For all $\delta, K > 0$ the function

$$\begin{aligned} \mathbb{M}^{(2)} &\to \mathbb{R}_+ \\ [X, r, \nu]^{(2)} &\mapsto V_\delta(\mathfrak{M}_{f_K\cdot\nu}) \end{aligned}$$

is upper semi-continuous with respect to both the two-level Gromov-weak topology τ_{2Gw} and the two-level Gromov-Prokhorov topology τ_{2GP}.

Therefore, if a net $([X_\alpha, r_\alpha, \nu_\alpha]^{(2)})_{\alpha\in\mathcal{A}}$ of m2m spaces converges two-level Gromov-weakly to $[X, r, \nu]^{(2)}$, then

$$\limsup_{\alpha} V_\delta(\mathfrak{M}_{f_K\cdot\nu_\alpha}) \le V_\delta(\mathfrak{M}_{f_K\cdot\nu}).$$

Because $V_\delta(\mathfrak{M}_{f_K\cdot\nu}) \to 0$ for $\delta \searrow 0$ (see Lemma 3.31), this implies that for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$V_\delta(\mathfrak{M}_{f_K\cdot\nu_\alpha}) < \varepsilon \tag{E4.21}$$

eventually. We will use this property in the next section to characterize compact subsets of $\mathbb{M}^{(2)}$.

4.4.5 Compact sets

In this section we examine compactness in $(\mathbb{M}^{(2)}, d_{2\mathrm{GP}})$. The main results of this section are Proposition 4.67 and Theorem 4.68 in which we characterize compact nets and relative compact subsets in $\mathbb{M}^{(2)}$. It might seem odd to look at nets in a metric space. However, in the proof of Theorem 4.69 we need to show that every τ_{2Gw}-convergent net is also τ_{2GP}-convergent. Thus it is useful for us to have a better understanding about τ_{2GP}-compact nets.

But first we introduce a useful class of compact subsets of $\mathbb{M}^{(2)}$ whose union is dense. For every $N \in \mathbb{N}$ we define $\mathbb{A}_N \subset \mathbb{M}^{(2)}$ as the set of all m2m spaces $[X, r, \nu]^{(2)}$ such that

- $\operatorname{supp} \mathfrak{M}_\nu$ consists of at most N points,
- the diameter of $\operatorname{supp} \mathfrak{M}_\nu$ is at most N,
- $\nu \in \mathcal{M}_{\leq N}(\mathcal{M}_{\leq N}(X))$.

Observe that the union $\bigcup_{N\in\mathbb{N}} \mathbb{A}_N$ is dense in $(\mathbb{M}^{(2)}, d_{2\mathrm{GP}})$ since it contains the dense set $\mathbb{S}$ from the proof of Theorem 4.62.

Lemma 4.66
The set $\mathbb{A}_N$ is compact in the two-level Gromov-Prokhorov topology for every $N \in \mathbb{N}$.

Proof: We prove that $\mathbb{A}_N$ is sequentially compact. Let $([X_n, r_n, \nu_n]^{(2)})_n$ be a sequence in $\mathbb{A}_N$. Without loss of generality we assume $X_n = \operatorname{supp} \mathfrak{M}_{\nu_n}$ for all $n \in \mathbb{N}$. The finite metric spaces $((X_n, r_n))_n$ are determined by the number of points and the mutual distances between the points. All of these are bounded by N. By Proposition A.5 the sequence $((X_n, r_n))_n$ is relatively compact in the Gromov-Hausdorff topology and there is a subsequence which converges to some compact metric space (X, r). For the sake of convenience we denote this subsequence again by $((X_n, r_n))_n$. Observe that the limit metric spaces satisfies $|X| \leq N$ and $\operatorname{diam} X \leq N$. By Lemma A.3 there is a compact metric space (Z, r_Z) and isometric embeddings $\iota, \iota_1, \iota_2, \ldots$ of $X, X_1, X_2, \ldots$ into Z such that

$$d_H^Z(\iota_n(X_n), \iota(X)) \to 0,$$

where d_H^Z denotes the Hausdorff metric on Z. Because Z is compact, both $\mathcal{M}_{\leq N}(Z)$ and $\mathcal{M}_{\leq N}(\mathcal{M}_{\leq N}(Z))$ are compact too. Therefore, $((\iota_{n**}\nu_n))_n$ has a subsequence which converges weakly to some measure $\nu \in \mathcal{M}_{\leq N}(\mathcal{M}_{\leq N}(Z))$. Then, the corresponding subsequence of $([X_n, r_n, \nu_n]^{(2)})_n$ converges to $[Z, r_Z, \nu]^{(2)} = [X, r, \nu]^{(2)} \in \mathbb{A}_N$. □

Proposition 4.67 (Characterization of compact nets in $\mathbb{M}^{(2)}$)
Let $(\mathcal{A}, \preceq)$ be a directed set and let $(\mathcal{X}_\alpha)_{\alpha\in\mathcal{A}}$ be a net in $\mathbb{M}^{(2)}$ with $\mathcal{X}_\alpha = [X_\alpha, r_\alpha, \nu_\alpha]^{(2)}$. The following are equivalent:

(a) $(\mathcal{X}_\alpha)_\alpha$ is a compact net with respect to the two-level Gromov-Prokhorov topology.

(b) For every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $d_{2\mathrm{GP}}(\mathcal{X}_\alpha, \mathbb{A}_N) < \varepsilon$ eventually.

(c) $(\mathfrak{m}_*\nu_\alpha)_\alpha$ is a compact net in $\mathcal{M}_f(\mathbb{R}_+)$ and for every $K > 0$ we have:

- *For every $\varepsilon > 0$ there is a $\delta > 0$ such that $V_\delta(\mathfrak{M}_{f_K \cdot \nu_\alpha}) < \varepsilon$ eventually,*
- *$(w(\mathfrak{M}_{f_K \cdot \nu_\alpha}))_\alpha$ is a compact net in $\mathcal{M}_f(\mathbb{R}_+)$.*

(d) $(\mathfrak{m}_* \nu_\alpha)_\alpha$ is a compact net and for every $\varepsilon > 0$ there is an $N_\varepsilon \in \mathbb{N}$ such that for every α there is a set $X_{\alpha,\varepsilon} \subset X_\alpha$ such that eventually

- *$\nu_\alpha(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement X_{\alpha,\varepsilon}) \le \varepsilon \}) < \varepsilon$,*
- *$X_{\alpha,\varepsilon}$ can be covered by at most N_ε balls of radius ε,*
- *the diameter of $X_{\alpha,\varepsilon}$ is at most N_ε.*

Proof: (a) $\Rightarrow$ (b): We prove this assertion by contradiction and assume that (b) does not hold. It follows that for each $N \in \mathbb{N}$ there is a subnet $(\mathcal{X}_{T_N(\beta)})_{\beta \in \mathcal{B}_N}$ with $d_{2\mathrm{GP}}(\mathcal{X}_{T_N(\beta)}, \mathbb{A}_N) \ge \varepsilon$. By a diagonal argument (cf. Lemma 2.9) we can construct a subnet $(\mathcal{X}_{T(\gamma)})_{\gamma \in \mathcal{C}}$ such that for every $N \in \mathbb{N}$ we eventually have $d_{2\mathrm{GP}}(\mathcal{X}_{T(\gamma)}, \mathbb{A}_N) \ge \varepsilon$. The net $(\mathcal{X}_{T(\gamma)})_\gamma$ is compact as a subnet of a compact net. Hence, it has a convergent subnet. Let $\mathcal{X}$ be the limit of this subnet. Observe that the set $\bigcup_{N \in \mathbb{N}} \mathbb{A}_N$ is dense in $\mathbb{M}^{(2)}$ since it contains the dense set $\mathbb{S}$ from the proof of Theorem 4.62. Thus, there is a natural number N_0 with $d_{2\mathrm{GP}}(\mathcal{X}, \mathbb{A}_{N_0}) < \varepsilon$. Consequently, the subnet converging to $\mathcal{X}$ is eventually ε-close to $\mathbb{A}_{N_0}$. But this contradicts the construction of the net $(\mathcal{X}_{T(\gamma)})_\gamma$.

(b) $\Rightarrow$ (a): Observe that assertion (b) also holds for any subnet of $(\mathcal{X}_\alpha)_\alpha$. Thus, it is enough to show that any net which satisfies (b) has a convergent subnet. Since $(\mathcal{X}_\alpha)_\alpha$ is eventually ε-close to a compact set $\mathbb{A}_N$, there is a subnet which is eventually ε-close to a convergent net in $\mathbb{A}_N$. Consequently, this subnet is eventually contained in a ball of radius 2ε. By using this argument repeatedly with $\varepsilon = 1/n$, we can construct subnets $(\mathcal{X}_{T_n(\beta)})_{\beta \in \mathcal{B}_n}$ for each $n \in \mathbb{N}$ such that $(\mathcal{X}_{T_n(\beta)})_{\beta \in \mathcal{B}_n}$ is a subnet of $(\mathcal{X}_{T_{n-1}(\beta)})_{\beta \in \mathcal{B}_{n-1}}$ and eventually contained in a ball of radius $1/n$. By a diagonal argument (cf. Lemma 2.9) we can construct a subnet which is a Cauchy net. That is, for every $\varepsilon > 0$ the subnet is eventually contained in a ball of radius ε. Since $(\mathbb{M}^{(2)}, d_{2\mathrm{GP}})$ is complete, this subnet net is convergent.

(a) $\Rightarrow$ (c): Note that both functions $[X,r,\nu]^{(2)} \mapsto \mathfrak{m}_* \nu$ and $[X,r,\nu]^{(2)} \mapsto w(\mathfrak{M}_{f_K \cdot \nu})$ are continuous (see Remark 4.59 and Corollary 4.64) and that the continuous image of a compact net is again compact. Thus $(\mathfrak{m}_* \nu_\alpha)_\alpha$ and $(w(\mathfrak{M}_{f_K \cdot \nu_\alpha}))_\alpha$ are both compact.

To prove the last property, assume that it does not hold, i. e. there are $K, \varepsilon > 0$ such that for every $\delta > 0$ the value of $V_\delta(\mathfrak{M}_{f_K \cdot \nu_\alpha})$ is frequently greater or equal than ε. Inductively we can construct subnets $(\mathcal{X}_{T_n(\beta)})_{\beta \in \mathcal{B}_n}$ for each $n \in \mathbb{N}$ such that $(\mathcal{X}_{T_n(\beta)})_{\beta \in \mathcal{B}_n}$ is a subnet of $(\mathcal{X}_{T_{n-1}(\beta)})_{\beta \in \mathcal{B}_{n-1}}$ and $V_{1/n}(\mathfrak{M}_{f_K \cdot \nu_{T_n(\beta)}}) \ge \varepsilon$ eventually. By a diagonal argument (cf. Lemma 2.9) we can construct a subnet $(\mathcal{X}_{T(\gamma)})_{\gamma \in \mathcal{C}}$ such that $V_\delta(\mathfrak{M}_{f_K \cdot \nu_\gamma}) \ge \varepsilon$ eventually for every $\delta > 0$. By (E4.21) this subnet cannot have a convergent subnet in contradiction to (a).

(c) $\Rightarrow$ (d): Let $1 > \varepsilon > 0$. The net $(\mathfrak{m}_* \nu_\alpha)_\alpha$ is compact and thus tight. Therefore, there is a K such that eventually $\mathfrak{m}(\nu_\alpha) - \mathfrak{m}(f_K \cdot \nu_\alpha) < \varepsilon/3$. Then, assertion (d) is a consequence of the following two claims, which we will prove at the end of this proof.

Claim 1: There is a positive integer N_1 and for every $\alpha \in \mathcal{A}$ there is a bounded set $C_\alpha \subset X_\alpha$ with $\operatorname{diam} C_\alpha \le N_1$ such that eventually

$$f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement C_\alpha) \le \frac{\varepsilon}{3} \} \Big) \le \frac{\varepsilon}{3}. \tag{E4.22}$$

Claim 2: There is a positive integer N_2 and for every $\alpha \in \mathcal{A}$ there is a finite set $A_\alpha \subset X_\alpha$ with $|A_\alpha| \le N_2$ such that eventually

$$f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement B(A_\alpha, \varepsilon)) < \frac{\varepsilon}{3} \} \Big) < \frac{\varepsilon}{3}. \tag{E4.23}$$

With these two claims we define $N_\varepsilon = \max(N_1, N_2)$ and the sets $X_{\alpha,\varepsilon} = C_\alpha \cap B(A_\alpha, \varepsilon)$. Then eventually the set $X_{\alpha,\varepsilon}$ has diameter of at most N_ε and can be covered by at most N_ε balls of radius ε. Moreover, eventually we have

$$\begin{aligned}
&\nu_\alpha \left(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement X_{\alpha,\varepsilon}) \le \varepsilon \} \right) \\
&\quad < \big(\mathfrak{m}(\nu_\alpha) - \mathfrak{m}(f_K \cdot \nu_\alpha) \big) + f_K \cdot \nu_\alpha \big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement X_{\alpha,\varepsilon}) \le \varepsilon \} \big) \\
&\quad \le \frac{\varepsilon}{3} + f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement C_\alpha) \le \frac{\varepsilon}{3} \} \Big) \\
&\qquad + f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mu(\complement B(A_\alpha, \varepsilon) < \frac{\varepsilon}{3} \} \Big) \\
&\quad < \varepsilon,
\end{aligned}$$

and the assertion is proved.

Proof of Claim 1: The net $(w(\mathfrak{M}_{f_K \cdot \nu_\alpha}))_\alpha$ is compact and thus tight. Therefore, there is an $a > 0$ such that eventually

$$w(\mathfrak{M}_{f_K \cdot \nu_\alpha})([a, \infty)) < \frac{1}{2} \left(\frac{\varepsilon}{3} \right)^4.$$

Let N_1 be a positive integer with $2a \le N_1$. We now construct bounded sets $C_\alpha \subset X_\alpha$ with $\operatorname{diam} C_\alpha < 2a \le N_1$ such that eventually (E4.22) holds. In the case

$$f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mathfrak{m}(\mu) \le \frac{\varepsilon}{3} \} \Big) \le \frac{\varepsilon}{3},$$

this is satisfied for $C_\alpha := B(x, a)$ for any $x \in X_\alpha$ and we are done. If, on the other hand,

$$f_K \cdot \nu_\alpha \Big(\complement \{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mathfrak{m}(\mu) \le \frac{\varepsilon}{3} \} \Big) > \frac{\varepsilon}{3}, \tag{E4.24}$$

then we define $C_\alpha := \{ x \in X_\alpha \mid \mathfrak{M}_{f_K \cdot \nu_\alpha}(\complement B(x, a)) < \frac{1}{2} \left(\frac{\varepsilon}{2} \right)^2 \}$. We show $\operatorname{diam} C_\alpha < 2a$ by contradiction: Assume there are $x, y \in C_\alpha$ with $r_\alpha(x, y) \ge 2a$. It follows that $B(x, a) \cap B(y, a) = \emptyset$ and therefore

$$\begin{aligned}
\left(\frac{\varepsilon}{2} \right)^2 &> \mathfrak{M}_{f_K \cdot \nu_\alpha}(\complement B(x, a)) + \mathfrak{M}_{f_K \cdot \nu_\alpha}(\complement B(y, a)) \\
&\ge \mathfrak{M}_{f_K \cdot \nu_\alpha}\big(\complement (B(x, a) \cap B(y, a)) \big) \\
&= \mathfrak{M}_{f_K \cdot \nu_\alpha}(X_\alpha) = \int \mu(X_\alpha) \, \mathrm{d}(f_K \cdot \nu_\alpha)(\mu) \\
&\ge \frac{\varepsilon}{3} \Big(f_K \cdot \nu_\alpha \Big) \Big(\{ \mu \in \mathcal{M}_f(X_\alpha) \mid \mathfrak{m}(\mu) > \frac{\varepsilon}{3} \} \Big).
\end{aligned}$$

This contradicts (E4.24), so the diameter of C_α must be less than $2a$.

Furthermore, we eventually have

$$\begin{aligned}\frac{1}{2}\Big(\frac{\varepsilon}{3}\Big)^4 &> w(\mathfrak{M}_{f_K\cdot\nu_\alpha})([a,\infty)) \\ &= (\mathfrak{M}_{f_K\cdot\nu_\alpha})^{\otimes 2}(\{(x,y)\in X_\alpha^2 \mid y\notin B(x,a)\}) \\ &\geq (\mathfrak{M}_{f_K\cdot\nu_\alpha})^{\otimes 2}(\{(x,y)\in X_\alpha^2 \mid x\notin C_\alpha, y\notin B(x,a)\}) \\ &\geq \frac{1}{2}\Big(\frac{\varepsilon}{3}\Big)^2\mathfrak{M}_{f_K\cdot\nu_\alpha}(\complement C_\alpha).\end{aligned}$$

In the last inequality we used the fact that

$$\mathfrak{M}_{f_K\cdot\nu_\alpha}(\complement B(x,a)) \geq \frac{1}{2}\Big(\frac{\varepsilon}{2}\Big)^2$$

for $x\notin C_\alpha$ by the very definition of C_α. We conclude that eventually we have

$$\begin{aligned}\Big(\frac{\varepsilon}{2}\Big)^2 &> \mathfrak{M}_{f_K\cdot\nu_\alpha}(\complement C_\alpha) = \int \mu(\complement C_\alpha)\,\mathrm{d}(f_K\cdot\nu_\alpha)(\mu) \\ &\geq \frac{\varepsilon}{3}\Big(f_K\cdot\nu_\alpha\Big)\Big(\{\mu\in\mathcal{M}_f(X_\alpha)\mid \mu(\complement C_\alpha) > \frac{\varepsilon}{3}\}\Big)\end{aligned}$$

and this yields (E4.22).

Proof of Claim 2: Because $(\mathfrak{m}_*\nu_\alpha)_\alpha$ is a compact net, $\mathfrak{m}(\nu_\alpha)$ is eventually bounded by some positive real number m. By assumption there are $\delta>0$ and $\alpha_0\in\mathcal{A}$ such that $V_\delta(\mathfrak{M}_{f_K\cdot\nu_\alpha}) < \frac{\varepsilon^2}{9}$ for every $\alpha\succeq\alpha_0$. Let N_2 be the largest positive integer with $N_2\leq\max\big(1,\frac{Km}{\delta}\big)$. By Lemma 3.31 we can find for every $\alpha\succeq\alpha_0$ a finite set $A_\alpha\subset X_\alpha$ with $|A_\alpha|\leq N_2$ such that

$$\begin{aligned}\frac{\varepsilon^2}{9} &> \mathfrak{M}_{f_K\cdot\nu_\alpha}\Big(\complement B(A_\alpha,\frac{\varepsilon^2}{9})\Big) \\ &\geq \mathfrak{M}_{f_K\cdot\nu_\alpha}(\complement B(A_\alpha,\varepsilon)) \\ &= \int \mu(\complement B(A_\alpha,\varepsilon))\,\mathrm{d}(f_K\cdot\nu_\alpha)(\mu) \\ &\geq \frac{\varepsilon}{3}\Big(f_K\cdot\nu_\alpha\Big)\Big(\{\mu\in\mathcal{M}_f(X_\alpha)\mid\mu(\complement B(A_\alpha,\varepsilon))\geq\frac{\varepsilon}{3}\}\Big).\end{aligned}$$

This leads to the desired inequality (E4.23).

(d) ⇒ (b): Let $\varepsilon>0$ be arbitrary. By assumption there are $N_\varepsilon\in\mathbb{N}$ and $\alpha_0\in\mathcal{A}$ such that for all $\alpha\succeq\alpha_0$ there are subsets $X_{\alpha,\varepsilon}\subset X_\alpha$ with $\operatorname{diam} X_{\alpha,\varepsilon}\leq N_\varepsilon$ such that

$$\nu_\alpha(\complement\{\mu\in\mathcal{M}_f(X_\alpha)\mid\mu(\complement X_{\alpha,\varepsilon})\leq\varepsilon\}) < \varepsilon \tag{E4.25}$$

and $X_{\alpha,\varepsilon}$ can be covered by $N_\alpha\leq N_\varepsilon$ balls $B(x_1^{(\alpha)},\varepsilon),\dots,B(x_{N_\alpha}^{(\alpha)},\varepsilon)$. For every $\alpha\succeq\alpha_0$ we define a function $F_\alpha\colon X_\alpha\to X_\alpha$ by

$$F_\alpha(x) = \begin{cases} x_1^{(\alpha)}, & \text{if } x\in B(x_1^{(\alpha)},\varepsilon) \text{ or } x\notin\bigcup_{j=1}^{N_\alpha}B(x_j^{(\alpha)},\varepsilon) \\ x_i^{(\alpha)}, & \text{if } x\in B(x_i^{(\alpha)},\varepsilon)\setminus\bigcup_{j=1}^{i-1}B(x_j^{(\alpha)},\varepsilon) \text{ for } i\in\{2,\dots,N_\alpha\}.\end{cases}$$

By assertion (c) of Lemma 2.24 we have $d_P(\mu, F_{\alpha*}\mu) \le \varepsilon$ for every $\mu \in \mathcal{M}_f(X_\alpha)$ with $\mu(\complement X_{\alpha,\varepsilon}) \le \varepsilon$. The definition of the Prokhorov metric and inequality (E4.25) yield $d_P(\nu_\alpha, F_{\alpha**}\nu_\alpha) \le \varepsilon$. Because $(\mathfrak{m}_*\nu_\alpha)_\alpha$ is compact, $\mathfrak{m}(\nu_\alpha)$ is eventually bounded from above by some positive number m and $(\mathfrak{m}_*\nu_\alpha)_\alpha$ is tight. Therefore, there are $K > 0$ and $\alpha_1 \succeq \alpha_0$ such that for all $\alpha \succeq \alpha_1$ we have $d_P(\nu'_\alpha, \nu_\alpha) < \varepsilon$, where ν'_α is the restriction of ν_α to $\mathcal{M}_{\le K}(X_\alpha)$. With the triangle inequality we obtain $d_P(\nu_\alpha, F_{\alpha**}\nu'_\alpha) \le 2\varepsilon$. It follows that

$$d_{2GP}([X_\alpha, r_\alpha, \nu_\alpha]^{(2)}, [X_\alpha, r_\alpha, F_{\alpha**}\nu'_\alpha]^{(2)}) \le 2\varepsilon.$$

Observe that with $N := \max(m, K, N_\varepsilon)$ we eventually have

$$[X_\alpha, r_\alpha, F_{\alpha**}\nu'_\alpha]^{(2)} \in \mathbb{A}_N.$$

Since $\varepsilon > 0$ is arbitrary, this proves the claim. □

Theorem 4.68 (Characterization of relatively compact subsets of $\mathbb{M}^{(2)}$)
Let $\Gamma \subset \mathbb{M}^{(2)}$ be a set of m2m spaces. The following are equivalent:

(a) Γ is relatively compact in the two-level Gromov-Prokhorov topology τ_{2Gw}.

(b) For every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d_{2GP}(\mathcal{X}, \mathbb{A}_N) < \varepsilon$ for every $\mathcal{X} \in \Gamma$.

(c) $\{\mathfrak{m}_*\nu \mid [X,r,\nu]^{(2)} \in \Gamma\}$ is relatively compact in $\mathcal{M}_f(\mathbb{R}_+)$ and for every $K > 0$ we have

- $\sup_{[X,r,\nu]^{(2)} \in \Gamma} V_\delta(\mathfrak{M}_{f_K \cdot \nu}) \to 0$ *for* $\delta \searrow 0$,
- $\{w(\mathfrak{M}_{f_K \cdot \nu}) \mid [X,r,\nu]^{(2)} \in \Gamma\}$ *is tight in* $\mathcal{M}_f(\mathbb{R}_+)$.

(d) $\{\mathfrak{m}_*\nu \mid [X,r,\nu]^{(2)} \in \Gamma\}$ is relatively compact in $\mathcal{M}_f(\mathbb{R}_+)$ and for every $\varepsilon > 0$ there is an $N_\varepsilon \in \mathbb{N}$ such that for every $\mathcal{X} = [X,r,\nu]^{(2)} \in \Gamma$ there exists a measurable subset $X_{\mathcal{X},\varepsilon} \subset X$ with

- $\nu(\complement\{\mu \in \mathcal{M}_f(X) \mid \mu(\complement X_{\mathcal{X},\varepsilon}) \le \varepsilon\}) < \varepsilon$,
- $X_{\mathcal{X},\varepsilon}$ *can be covered by at most* N_ε *balls of radius* ε,
- *the diameter of* $X_{\mathcal{X},\varepsilon}$ *is at most* N_ε.

(e) $\{\mathfrak{m}_*\nu \mid [X,r,\nu]^{(2)} \in \Gamma\}$ is relatively compact in $\mathcal{M}_f(\mathbb{R}_+)$ and for every $\varepsilon > 0$ and $\mathcal{X} = [X,r,\nu]^{(2)} \in \Gamma$ there is a compact subset $C_{\mathcal{X},\varepsilon} \subset X$ such that

- $\nu(\complement\{\mu \in \mathcal{M}_f(X) \mid \mu(\complement C_{\mathcal{X},\varepsilon}) \le \varepsilon\}) < \varepsilon$,
- $\mathcal{C}_\varepsilon := \{C_{\mathcal{X},\varepsilon} \mid \mathcal{X} \in \Gamma\}$ *is relatively compact in the Gromov-Hausdorff topology.*

Note that in assertion (c) is suffices to have the property only for a diverging sequence $(K_n)_n \nearrow \infty$.

Proof (of Theorem 4.68): The equivalence of assertions (a) to (d) follows easily from Proposition 4.67 and the fact that a set is relatively compact if and only if every net in it is compact (cf. Lemma 2.8). In the remainder of this proof we show that assertion (d) and (e) are equivalent.

Note that a set $\mathcal{C}$ of compact metric spaces is relatively compact in the Gromov-Hausdorff-topology if and only if the diameter of the elements of $\mathcal{C}$ is uniformly bounded and for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that every element of $\mathcal{C}$ can be covered by at most N balls of radius ε (see Proposition A.5). Therefore, assertion (e) readily implies assertion (d).

To prove the other direction, let $\varepsilon > 0$. For $n \in \mathbb{N}$ let $\varepsilon_n := \frac{\varepsilon}{2} \cdot \left(\frac{1}{2}\right)^n$ and let N_{ε_n} and $X_{\mathcal{X},\varepsilon_n}$ be as in assertion (d). Without loss of generality we may assume that every $X_{\mathcal{X},\varepsilon_n}$ is closed. For every $\mathcal{X} = [X, r, \nu]^{(2)} \in \Gamma$ and every $n \in \mathbb{N}$ there is a compact set $\mathcal{K}_{\mathcal{X},\varepsilon_n} \subset \mathcal{M}_f(X)$ with $\nu(\complement \mathcal{K}_{\mathcal{X},\varepsilon_n}) < \varepsilon_n$. Because $\mathcal{K}_{\mathcal{X},\varepsilon_n}$ is tight, we have

$$\mathcal{K}_{\mathcal{X},\varepsilon_n} \subset \{ \mu \in \mathcal{M}_f(X) \mid \mu(\complement C_{\mathcal{X},\varepsilon_n}) < \varepsilon_n \}$$

for some compact set $C_{\mathcal{X},\varepsilon_n} \subset X$ and thus

$$\nu(\complement \{ \mu \in \mathcal{M}_f(X) \mid \mu(\complement C_{\mathcal{X},\varepsilon_n}) \leq \varepsilon_n \}) \leq \nu(\complement \mathcal{K}_{\mathcal{X},\varepsilon_n}) < \varepsilon_n.$$

Define $C_{\mathcal{X},\varepsilon} := \bigcap_{n\in\mathbb{N}} (X_{\mathcal{X},\varepsilon_n} \cap C_{\mathcal{X},\varepsilon_n})$. The set $C_{\mathcal{X},\varepsilon}$ is compact as a closed subset of a compact set. Observe that for every $\mu \in \mathcal{M}_f(X)$ with

$$\mu(\complement X_{\mathcal{X},\varepsilon_n}) \leq \varepsilon_n$$

and

$$\mu(\complement C_{\mathcal{X},\varepsilon_n}) \leq \varepsilon_n$$

for every $n \in \mathbb{N}$ we have

$$\begin{aligned} \mu(\complement C_{\mathcal{X},\varepsilon}) &= \mu\left(\complement \bigcap_{n\in\mathbb{N}} (X_{\mathcal{X},\varepsilon_n} \cap C_{\mathcal{X},\varepsilon_n})\right) \\ &= \mu\left(\bigcup_{n\in\mathbb{N}} (\complement X_{\mathcal{X},\varepsilon_n} \cup \complement C_{\mathcal{X},\varepsilon_n})\right) \\ &\leq \sum_{n\in\mathbb{N}} \mu(\complement X_{\mathcal{X},\varepsilon_n}) + \mu(\complement C_{\mathcal{X},\varepsilon_n}) \\ &\leq \sum_{n\in\mathbb{N}} 2\varepsilon_n = \varepsilon. \end{aligned}$$

Hence, we obtain

$$\begin{aligned}
&\nu\left(\complement\{\mu\in\mathcal{M}_f(X)\mid\mu(\complement C_{\mathcal{X},\varepsilon})\le\varepsilon\}\right)\\
&\le\nu\left(\complement\bigcap_{n\in\mathbb{N}}\{\mu\in\mathcal{M}_f(X)\mid\mu(\complement X_{\mathcal{X},\varepsilon_n})\le\varepsilon_n,\ \mu(\complement C_{\mathcal{X},\varepsilon_n})\le\varepsilon_n\}\right)\\
&\le\sum_{n\in\mathbb{N}}\Big(\nu\left(\complement\{\mu\in\mathcal{M}_f(X)\mid\mu(\complement X_{\mathcal{X},\varepsilon_n})\le\varepsilon_n\}\right)\\
&\quad+\nu\left(\complement\{\mu\in\mathcal{M}_f(X)\mid\mu(\complement C_{\mathcal{X},\varepsilon_n})\le\varepsilon_n\}\right)\Big)\\
&<2\sum_{n\in\mathbb{N}}\varepsilon_n=\varepsilon.
\end{aligned}$$

The set $\mathcal{C}_\varepsilon=\{C_{\mathcal{X},\varepsilon}\mid\mathcal{X}\in\Gamma\}$ is relatively compact in the Gromov-Hausdorff-topology because the diameter of each $C_{\mathcal{X},\varepsilon}$ is bounded by N_{ε_1} and for every $\delta>0$ there is a natural number N such that each $C_{\mathcal{X},\varepsilon}$ can be covered by at most N balls of radius δ (cf. the remark at the beginning of this proof). Thus, assertion (e) is fulfilled and the proof is complete. □

4.4.6 Equivalence of two-level Gromov-weak and two-level Gromov-Prokhorov topology on $\mathbb{M}^{(2)}$

We are finally able to prove the fact that the two-level Gromov-weak topology and the two-level Gromov-Prokhorov topology on $\mathbb{M}^{(2)}$ coincide. Our proof relies mostly on the compactness criteria of Proposition 4.67.

Theorem 4.69
The two-level Gromov-weak topology τ_{2Gw} and the two-level Gromov-Prokhorov topology τ_{2GP} on $\mathbb{M}^{(2)}$ coincide.

Note that now every statement we made about one of the two topologies (e. g. embedding lemmas, compactness criteria, Polishness) also holds true for the other topology.

Proof: We have already proved in Lemma 4.61 that τ_{2GP} is finer than τ_{2Gw}. To show the other direction we prove that every τ_{2Gw}-convergent net $(\mathcal{X}_\alpha)_{\alpha\in\mathcal{A}}$ with $\mathcal{X}_\alpha\xrightarrow{\tau_{2Gw}}\mathcal{X}$ also converges in the two-level Gromov-Prokhorov topology τ_{2GP} (recall that we do not know yet whether τ_{2Gw} is metrizable, thus we have to use nets instead of sequences). We will do so by proving that $(\mathcal{X}_\alpha)_\alpha$ is a compact net with respect to τ_{2GP}. Because $\mathcal{T}^{(2)}$ separates points (cf. Theorem 4.53), this implies that every subnet has a converging subnet with the same limit $\mathcal{X}$. Hence, $(\mathcal{X}_\alpha)_\alpha$ itself converges to $\mathcal{X}$ in the d_{2GP}-metric.

To prove compactness, we use property (c) of Proposition 4.67. For every $\alpha\in\mathcal{A}$ let $\mathcal{X}_\alpha=[X_\alpha,r_\alpha,\nu_\alpha]^{(2)}$. Recall that the functions $[X,r,\nu]^{(2)}\mapsto\mathfrak{m}_*\nu$ and $[X,r,\nu]^{(2)}\mapsto w(\mathfrak{M}_{f_K\cdot\nu})$ are continuous with respect to τ_{2Gw} by Remark 4.59 and Corollary 4.64. Therefore, both $(\mathfrak{m}_*\nu_\alpha)_\alpha$ and $(w(\mathfrak{M}_{f_K\cdot\nu_\alpha}))_\alpha$ are convergent and thus compact. Furthermore, by (E4.21) for every $\varepsilon>0$ there is a $\delta>0$ such that eventually $V_\delta(\mathfrak{M}_{f_K\cdot\nu_\alpha})<\varepsilon$. Thus, the assumptions of Proposition 4.67 are satisfied and $(\mathcal{X}_\alpha)_\alpha$ is a compact net in the two-level Gromov-Prokhorov topology. □

4.4.7 Distributions on $\mathbb{M}^{(2)}$

In this section we provide results about random m2m spaces. With Le Cam's theorem it is easy to deduce that the m2m-monomials are convergence determining.

Proposition 4.70
The set $\mathcal{T}^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}^{(2)})$.

Proof: The set $\mathcal{T}^{(2)}$ separates points, is closed under multiplication and induces the topology of $\mathbb{M}^{(2)}$. Therefore, it is convergence determining for $\mathcal{M}_1(\mathbb{M}^{(2)})$ by Proposition 2.17. □

This means that a sequence $(P_n)_n$ in $\mathcal{M}_1(\mathbb{M}^{(2)})$ converges weakly to $P \in \mathcal{M}_1(\mathbb{M}^{(2)})$ if and only if $P_n[\Phi]$ converges to $P[\Phi]$ for every $\Phi \in \mathcal{T}^{(2)}$. Here, $P[\Phi]$ denotes the expectation $\int \Phi(\mathcal{X})\,\mathrm{d}P(\mathcal{X})$.

We now provide a characterization of tight subsets of $\mathcal{M}_1(\mathbb{M}^{(2)})$. Since tightness is defined in terms of compact sets, it is not surprising that we use Theorem 4.68 to find conditions for tightness.

Proposition 4.71 (Tightness criterion for $\mathcal{M}_1(\mathbb{M}^{(2)})$)
A set $\mathcal{P} \subset \mathcal{M}_1(\mathbb{M}^{(2)})$ is tight if and only if for every $\varepsilon > 0$ and $K > 0$ there exist $\delta > 0$ and $c > 0$ such that for every $P \in \mathcal{P}$ each of the following conditions hold:

(a) $P(\mathfrak{m}(\nu) \geq c) < \varepsilon$.

(b) $P(\mathfrak{m}_*\nu([c,\infty)) \geq \varepsilon) < \varepsilon$.

(c) $P(V_\delta(\mathfrak{M}_{f_K \cdot \nu}) \geq \varepsilon) < \varepsilon$.

(d) $P(w(\mathfrak{M}_{f_K \cdot \nu})([c,\infty)) \geq \varepsilon) < \varepsilon$.

Here, ν denotes the two-level measure of a random m2m space $[X,r,\nu]^{(2)}$ with law P.

Proof: Let $\varepsilon, K > 0$. If $\mathcal{P}$ is tight, then there is a compact set $C \subset \mathbb{M}^{(2)}$ such that $P(\complement C) < \varepsilon$ for every $P \in \mathcal{P}$. By property (c) of Theorem 4.68 we can choose $\delta > 0$ and $c > \sup\{\, \mathfrak{m}(\nu) \mid [X,r,\nu]^{(2)} \in C \,\}$ such that for every $[X,r,\nu]^{(2)} \in C$ we have

$$\begin{aligned} &\mathfrak{m}_*\nu([c,\infty)) < \varepsilon, \\ &V_\delta(\mathfrak{M}_{f_K \cdot \nu}) < \varepsilon, \\ &w(\mathfrak{M}_{f_K \cdot \nu})([c,\infty)) < \varepsilon \end{aligned}$$

and the claim follows immediately.

To prove the other direction, let $\varepsilon > 0$. We are going to construct a relatively compact set $C \subset \mathbb{M}^{(2)}$ such that $P(\complement C) < \varepsilon$ for every $P \in \mathcal{P}$. First, define $\varepsilon_n := \frac{\varepsilon}{4} \cdot 2^{-n}$ and $K_n := n$ for every $n \in \mathbb{N}$. By assumption there are $c, \delta_n, c_n > 0$ such that

$$\begin{aligned} &P(\mathfrak{m}(\nu) \geq c) < \tfrac{\varepsilon}{4}, \\ &P(\mathfrak{m}_*\nu([c_n,\infty)) \geq \varepsilon_n) < \varepsilon_n, \\ &P(V_{\delta_n}(\mathfrak{M}_{f_{K_n} \cdot \nu}) \geq \varepsilon_n) < \varepsilon_n, \\ &P(w(\mathfrak{M}_{f_{K_n} \cdot \nu})([c_n,\infty)) \geq \varepsilon_n) < \varepsilon_n \end{aligned} \tag{E4.26}$$

for every $P \in \mathcal{P}$ and $n \in \mathbb{N}$. Let C_n be the set of all $[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)}$ with

$$\begin{aligned}&\mathfrak{m}_*\nu([c_n,\infty)) < \varepsilon_n,\\ &V_{\delta_n}(\mathfrak{M}_{f_{K_n}\cdot\nu}) < \varepsilon_n,\\ &w(\mathfrak{M}_{f_{K_n}\cdot\nu})([c_n,\infty)) < \varepsilon_n.\end{aligned}$$

By (E4.26) we have $P(\complement C_n) < 3\varepsilon_n$ for every $P \in \mathcal{P}$. With

$$C := \{\,[X,r,\nu]^{(2)} \in \mathbb{M}^{(2)} \mid \mathfrak{m}(\nu) < c\,\} \cap \bigcap_{n\in\mathbb{N}} C_n,$$

we get

$$P(\complement C) < \frac{\varepsilon}{4} + \sum_{n\in\mathbb{N}} 3\varepsilon_n = \varepsilon$$

for every $P \in \mathcal{P}$. Moreover, C satisfies the compactness criterion given in property (c) of Theorem 4.68 and is therefore relatively compact. □

Chapter 5

The nested Kingman coalescent measure tree

In this chapter we introduce the nested Kingman coalescent measure tree, which is a random m2pm space modeling the genealogy of a nested Kingman coalescent. The nested Kingman coalescent measure tree is defined as a weak limit of random finite m2pm spaces. To show convergence, we will apply the tightness criterion of section 4.3.7.

Outline of this chapter: We start by providing the definition of the nested Kingman coalescent in section 5.1. In section 5.2 we introduce the nested Kingman coalescent measure tree and state the main result of this chapter, which is the existence of the nested Kingman coalescent measure tree. The proof for this result is given in section 5.3.

5.1 The nested Kingman coalescent

Nested coalescents were introduced in [Bla16, BDLSJ18] to jointly model the species and the gene coalescents of a population of multiple species. The nested Kingman coalescent is a special case of the model developed in these publications (cf. also [BRSSJ18] for further research about the nested Kingman coalescent). In this section we provide a definition of the nested Kingman coalescent first for a finite set $I \subset \mathbb{N}^2$ of individuals and then for infinitely many individuals (i. e. $I = \mathbb{N}^2$). We use $\mathbb{N}^2$ to encode individuals because we think of $(i,j) \in \mathbb{N}^2$ as the j-th individual of the i-th species.

For a non-empty set I let $\mathcal{E}(I) \subset I^2$ denote the set of equivalence relations on I equipped with the discrete topology. The equivalence classes of an equivalence relation are called *blocks*. We say that a pair $(\mathcal{R}_1,\mathcal{R}_2) \in \mathcal{E}(I) \times \mathcal{E}(I)$ is *nested* (or that $\mathcal{R}_2$ is nested in $\mathcal{R}_1$) if $\mathcal{R}_1 \supset \mathcal{R}_2$. Note that $(\mathcal{R}_1,\mathcal{R}_2)$ is nested if and only if for every block π_2 of $\mathcal{R}_2$ there is a block π_1 of $\mathcal{R}_1$ with $\pi_2 \subset \pi_1$. Let

$$\mathcal{N}(I) := \{ (\mathcal{R}_1,\mathcal{R}_2) \in \mathcal{E}(I)^2 \mid \mathcal{R}_2 \subset \mathcal{R}_1 \}$$

denote the set of nested equivalence relations equipped with the discrete topology.

Moreover, we define the following equivalence relations on $\mathbb{N}^2$, which will be the initial states of the nested Kingman coalescent:

$$G_0 := \{ (\boldsymbol{x},\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{N}^2 \},$$
$$S_0 := \{ ((i,j),(i,k)) \mid i,j,k \in \mathbb{N} \}.$$

G_0 is the equivalence relation with only singleton blocks and S_0 is the equivalence relation whose blocks are the different species of the population.

Definition 5.1 (Finite nested Kingman coalescent)
Let I be a finite subset of $\mathbb{N}^2$ and $\gamma_s,\gamma_g > 0$. Let

$$\mathcal{R}^{(I)} = (\mathcal{R}^{(I)}(t))_{t\geq 0} = (\mathcal{R}_s^{(I)}(t), \mathcal{R}_g^{(I)}(t))_{t\geq 0}$$

be a continuous-time Markov process with values in $\mathcal{N}(I)$. We call $\mathcal{R}^{(I)}$ the finite nested Kingman coalescent on I with rates (γ_s,γ_g) *if it has the following properties:*

- *The initial state is $\mathcal{R}_s^{(I)}(0) = S_0 \cap I^2$ and $\mathcal{R}_g^{(I)}(0) = G_0 \cap I^2$, i. e. for $\boldsymbol{x} = (x_1,x_2) \in I^2$ and $\boldsymbol{y} = (y_1,y_2) \in I^2$ we have*

$$(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{R}_s^{(I)}(0) \Leftrightarrow x_1 = y_1,$$
$$(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{R}_g^{(I)}(0) \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}.$$

- *The* species coalescent $\mathcal{R}_s^{(I)} = (\mathcal{R}_s^{(I)}(t))_{t\geq 0}$ *behaves like a Kingman coalescent with rate γ_s, i. e. any two blocks in $\mathcal{R}_s^{(I)}(t)$ merge at rate γ_s*

- *The* gene coalescent $\mathcal{R}_g^{(I)} = (\mathcal{R}_s^{(I)}(t))_{t\geq 0}$ *behaves in the following way: any two blocks π_1,π_2 of $\mathcal{R}_g^{(I)}$ such that $\pi_1 \cup \pi_2$ is contained in a single block of $\mathcal{R}_s^{(I)}(t)$ merge at rate γ_g. Other blocks cannot merge.*

The definition of the finite nested Kingman coalescent describes the behavior of a Markov process with only finitely many states. Thus, it is clear that such a process exists and is unique in distribution. Figure 5.1 shows a realization of a finite nested Kingman coalescent.

We now define the nested Kingman coalescent for an infinite set of individuals (that is, with $I = \mathbb{N}^2$). For the existence of this process we refer to the construction of (more general) nested coalescents in [BDLSJ18, section 5].

Definition 5.2 (Nested Kingman coalescent)
Let $\gamma_s,\gamma_g > 0$. The nested Kingman coalescent with rates (γ_s,γ_g) *is a continuous-time Markov process $\mathcal{R} = (\mathcal{R}(t))_{t\geq 0} = (\mathcal{R}_s(t), \mathcal{R}_g(t))_{t\geq 0}$ with values in $\mathcal{N}(\mathbb{N}^2)$ such that for any finite $I \subset \mathbb{N}^2$ the restriction of $\mathcal{R}$ to $\mathcal{N}(I)$ is a finite nested Kingman coalescent on I with rates (γ_s,γ_g).*

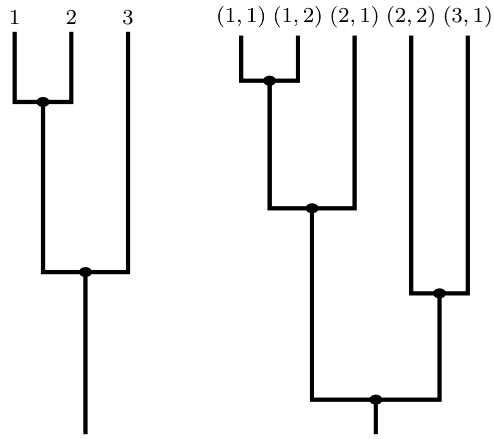

Figure 5.1: A realization of a finite nested Kingman coalescent on the set $I = \{(1,1),(1,2),(2,1),(2,2),(3,1)\}$. The species coalescent on the left starts with the equivalence classes $\{(1,1),(1,2)\},\{(2,1),(2,2)\}$ and $\{(3,1)\}$ (here abbreviated by 1, 2 and 3). Its branch points are speciation events. The tree on the right side depicts the gene coalescent. Notice that merging events in the gene coalescent can only happen after the associated species have merged in the species coalescent.

It follows immediately from the definition that the initial states of the species and the gene coalescent are

$$\mathcal{R}_s(0) = S_0 \quad \text{and} \quad \mathcal{R}_g(0) = G_0.$$

It is well-known that the standard Kingman coalescent immediately comes down from infinity, meaning that after any positive time the coalescent almost surely has only finitely many blocks left, even if we start with infinitely many blocks. The same is true for the nested Kingman coalescent as stated in the next lemma. A proof can be found in [BDLSJ18, section 6].

Lemma 5.3 (Nested Kingman coalescent comes down from infinity)
The nested Kingman coalescent immediately comes down from infinity. That is, if $\mathcal{R} = (\mathcal{R}_s(t), \mathcal{R}_g(t))_{t\geq 0}$ is a nested Kingman coalescent, then for every $t > 0$ both $\mathcal{R}_s(t)$ and $\mathcal{R}_g(t)$ almost surely consist of only finitely many blocks.

5.2 The nested Kingman coalescent measure tree

In this section we define a random m2pm space called the nested Kingman coalescent measure tree. Roughly speaking, it is the genealogical tree of the gene coalescent of a nested Kingman coalescent equipped with a two-level probability measure that represents uniform sampling of species on the second level and uniform sampling of individuals in a single species on the first level.

Let $\mathcal{R} = (\mathcal{R}_s, \mathcal{R}_g)$ be a nested Kingman coalescent and $\mathbb{P}$ its law. For $\boldsymbol{x} = (x_1, x_2) \in \mathbb{N}^2$ and $\boldsymbol{y} = (y_1, y_2) \in \mathbb{N}^2$ we define the coalescence time of $\boldsymbol{x}$ and $\boldsymbol{y}$ by

$$r_g(\boldsymbol{x},\boldsymbol{y}) := \inf\{\, t \geq 0 \mid (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{R}_g(t) \,\}.$$

The law of $r_g(\boldsymbol{x},\boldsymbol{y})$ is

$$r_g(\boldsymbol{x},\boldsymbol{y}) \sim \begin{cases} \mathrm{Exp}(\gamma_g) * \mathrm{Exp}(\gamma_s), & \text{if } x_1 \neq y_1 \\ \mathrm{Exp}(\gamma_g), & \text{if } x_1 = y_1 \text{ and } x_2 \neq y_2 \\ \delta_0, & \text{if } \boldsymbol{x} = \boldsymbol{y}, \end{cases} \tag{E5.1}$$

where $\mathrm{Exp}(\gamma)$ denotes the exponential distribution with parameter $\gamma > 0$ and $\mu * \eta$ denotes the convolution of two distributions μ and η. The function r_g satisfies the triangle inequality and is thus a (random) metric on $\mathbb{N}^2$ (in fact it is even an ultrametric). Let (Z, r) denote the completion of the metric space $(\mathbb{N}^2, r_g)$.

Remark 5.4 (Partial exchangeability of nested Kingman coalescent)
The distances between individuals of the nested Kingman coalescent are partially exchangeable in the following sense. Let $\widetilde{P}$ be the set of finite permutations p on $\mathbb{N}^2$ with the property that $\pi_1 p(i,j) = \pi_1 p(i,k)$ for all $i,j,k \in \mathbb{N}$, where π_1 denotes the projection to the first component of a vector (i. e. the first components of $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ coincide whenever the first components of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}^2$ coincide). Then, for every $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n} \in \mathbb{N}^2$ the law of $R(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})$ is equal to the law of $R(p(\boldsymbol{x_1}), \ldots, p(\boldsymbol{x_n}))$ for every permutation $p \in \widetilde{P}$. In other words, the law of $R(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})$ is invariant under $\widetilde{P}$.

For all $M, N \in \mathbb{N}$ we define the two-level measure $\nu_{M,N} \in \mathcal{M}_1(\mathcal{M}_1(Z))$ by

$$\nu_{M,N} := \frac{1}{M} \sum_{i=1}^{M} \delta_{\left(\frac{1}{N} \sum_{j=1}^{N} \delta_{(i,j)}\right)}. \tag{E5.2}$$

$\nu_{M,N}$ samples uniformly one of the first M species and then samples uniformly one of the first N individuals in that species. Let $H_{M,N}$ be the function that maps a realization of the nested Kingman coalescent to the m2pm space $[Z, r, \nu_{M,N}]^{(2)}$ and define

$$\mathbb{Q}_{M,N} := H_{M,N*}\mathbb{P} \in \mathcal{M}_1(\mathbb{M}_1^{(2)}).$$

Theorem 5.5 (Nested Kingman coalescent measure tree)

(a) The sequence $(\mathbb{Q}_{M,N})_N$ is weakly convergent for every $M \in \mathbb{N}$. We denote its limit by $\mathbb{Q}_M$.

(b) The sequence $(\mathbb{Q}_M)_M$ is weakly convergent.

Therefore, the limit

$$\mathbb{Q} := \underset{M\to\infty}{\text{w-lim}}\, \underset{N\to\infty}{\text{w-lim}}\, \mathbb{Q}_{M,N} = \underset{M\to\infty}{\text{w-lim}}\, \mathbb{Q}_M$$

exists. We call any random variable with values in $\mathbb{M}_1^{(2)}$ and law $\mathbb{Q}$ a *nested Kingman coalescent measure tree.*

Remark 5.6 (Generalization to nested Λ-coalescent)
Recall that the Λ-coalescent is a generalization of the Kingman coalescent that allows multiple mergers and in which the merging rates are described by a finite measure $\Lambda \in \mathcal{M}_f([0,1])$. In a similar manner we can generalize the nested Kingman coalescent to a *nested* (Λ_s,Λ_g)*-coalescent*, where the species coalescent behaves like a Λ_s-coalescent and the gene coalescent behaves like a Λ_g-coalescent (inside of single species blocks). The proof of Theorem 5.5 is valid even for the nested (Λ_s,Λ_g)-coalescent as long as the nested coalescent immediately comes down from infinity. The latter condition is true if and only if both Λ_s and Λ_g are such that the corresponding Λ-coalescents immediately come down from infinity (cf. [BRSSJ18]).

It is an open question whether these conditions can be relaxed to the more general dust free property (cf. [GPW09, Theorem 4] in which the authors construct a Λ-coalescent measure tree for all Λ-coalescents which satisfy the dust free property).

5.3 Proof of Theorem 5.5

The proofs of both statements of Theorem 5.5 use the same kind of argument. First we show that the sequence under consideration has at most one limit point. Then we show that the sequence is relatively compact, i. e. every subsequence has a convergent subsequence. Consequently, because the limit point is unique, we may conclude that the original sequence is convergent.

Two main tools when working with coalescents are exchangeability and relative frequencies of blocks. We already showed in Remark 5.4 that the nested Kingman coalescent is partially exchangeable. Let us now define the relative frequencies of blocks.

Definition 5.7
For all $i,l \in \mathbb{N}$ *and* $t \in \mathbb{R}_+$ *we define*

$$\mathfrak{f}_{i,l}(t) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[0,t]}(r_g((i,1),(l,n))).$$

$\mathfrak{f}_{i,l}(t)$ *is the relative frequency of the block of* $(i,1)$ *w. r. t. species* l *at time* t.

It is a standard fact from coalescent theory that in a Kingman coalescent the relative frequencies of blocks exist. By definition the gene coalescent restricted to a single species $i \in \mathbb{N}$ is a Kingman coalescent. Thus, the relative frequency $\mathfrak{f}_{i,i}(t)$ exists for every $t \geq 0$. Moreover, the Kingman coalescent almost surely has proper frequencies (cf. [Pit99, Theorem 8]), implying that

$$\mathbb{P}(\mathfrak{f}_{i,i}(t) = 0) = 0 \tag{E5.3}$$

for every $t > 0$.

For $\mathfrak{f}_{i,l}(t)$ with $i \neq l$ the situation is a little different since the species i and l have to merge first. Let $\tau_s(i,l)$ denote the coalescent time of the species i and l in the species tree. Clearly, we have $\mathfrak{f}_{i,l}(t) = 0$ for $t < \tau_s(i,l)$. For $t \geq \tau_s(i,l)$ the gene coalescent restricted to $\{(l,j) \mid j \in \mathbb{N}\} \cup \{(i,1)\}$ behaves like a Kingman coalescent. Thus, the relative frequency $\mathfrak{f}_{i,l}(t)$ also exists for $t \geq \tau_s(i,l)$.

Because the nested Kingman coalescent starts with singleton blocks at time $t = 0$, we have $\mathfrak{f}_{i,l}(0) = 0$ for all $i, l \in \mathbb{N}$. Moreover, almost surely the function $t \mapsto \mathfrak{f}_{i,l}(t)$ is non-decreasing (the relative frequency increases when blocks merge) and converges to 1 (eventually all blocks have merged to a single block).

In the next lemma we state an analog of equation (E5.3) for the relative frequencies $\mathfrak{f}_{i,l}(t)$ with $i \neq l$. Though it may happen that $\mathfrak{f}_{i,l}(t) = 0$ *for some* $l \neq i$, it may not happen *for all* $l \neq i$. This follows from the partial exchangeability explained in Remark 5.4.

Lemma 5.8
For every $t > 0$ *and every* $i \in \mathbb{N}$

$$\mathbb{P}(\mathfrak{f}_{i,l}(t) = 0 \text{ for all } l \neq i) = 0. \tag{E5.4}$$

Proof: We define for every $n \in \mathbb{N}$ and $t' \in \mathbb{R}_+$ the infinite vector $\boldsymbol{\mathfrak{f}}_n(t') \in [0,1]^{\mathbb{N}}$ by

$$\boldsymbol{\mathfrak{f}}_n(t') := (\mathfrak{f}_{n,l}(t'))_{l \neq n} = (\mathfrak{f}_{n,1}(t'), \mathfrak{f}_{n,2}(t'), \dots, \mathfrak{f}_{n,n-1}(t'), \mathfrak{f}_{n,n+1}(t'), \dots).$$

We fix $t > 0$ and $i \in \mathbb{N}$. Observe that equation (E5.4) is equivalent to $\mathbb{P}(\boldsymbol{\mathfrak{f}}_i(t) = \mathbf{0}) = 0$ and that the sequence of vectors $(\boldsymbol{\mathfrak{f}}_n(t))_{n \in \mathbb{N}}$ is exchangeable (cf. Remark 5.4). By de Finetti's theorem there is a random probability measure Ξ on $[0,1]^{\mathbb{N}}$ such that $\Xi^{\otimes\infty}$ is a regular conditional distribution of $(\boldsymbol{\mathfrak{f}}_n(t))_n$ given $\sigma(\Xi)$ (cf. [Ald85, Theorem 3.1]).

We prove the claim by contradiction. Assume that

$$0 < \mathbb{P}(\boldsymbol{\mathfrak{f}}_i(t) = \mathbf{0}) = \int \Xi(\mathbf{0}) \,\mathrm{d}\mathbb{P}.$$

This is true if and only if $\mathbb{P}(\Xi(\mathbf{0}) > 0) > 0$ and in this case

$$\begin{aligned} &\mathbb{P}(\boldsymbol{\mathfrak{f}}_n(t) = \mathbf{0} \text{ for infinitely many } n \in \mathbb{N}) \\ &\quad = \mathbb{P}(\boldsymbol{\mathfrak{f}}_n(t) = \mathbf{0} \text{ for infinitely many } n \in \mathbb{N} \mid \Xi(\mathbf{0}) > 0) \cdot \mathbb{P}(\Xi(\mathbf{0}) > 0) \\ &\quad = \mathbb{P}(\Xi(\mathbf{0}) > 0) \\ &\quad > 0. \end{aligned} \tag{E5.5}$$

However, if $\boldsymbol{\mathfrak{f}}_n(t) = \mathbf{0}$ for infinitely many $n \in \mathbb{N}$, then there is an increasing sequence of positive integers $(n_k)_k$ with $\boldsymbol{\mathfrak{f}}_{n_k}(t) = \mathbf{0}$. Observe that $\boldsymbol{\mathfrak{f}}_{n_k}(t) = \mathbf{0}$ implies that at time t the block of $(n_k, 1)$ has not merged with a block of another species. Therefore, each $(n_l, 1)$ is in a different block and $\mathcal{R}_g(t)$ contains an infinite number of blocks. But by Lemma 5.3 we know that almost surely $\mathcal{R}_g(t)$ contains only finitely many blocks. This is a contradiction to (E5.5). Consequently, we must have $\mathbb{P}(\boldsymbol{\mathfrak{f}}_i(t) = \mathbf{0}) = 0$. □

Before we start to prove Theorem 5.5, observe that the first moment measure of $\nu_{M,N}$ from (E5.2) is

$$\mathfrak{M}_{\nu_{M,N}} = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{N} \sum_{j=1}^{N} \delta_{(i,j)},$$

i. e. $\mathfrak{M}_{\nu_{M,N}}$ is a uniform distribution on the set $\{1,\ldots,M\}\times\{1,\ldots,N\}$.

5.3.1 Weak convergence of $(\mathbb{Q}_{M,N})_N$ for fixed M

Let $M\in\mathbb{N}$ be fixed.

Uniqueness: Recall that the m2pm-polynomials $\Phi\in\mathcal{T}_1^{(2)}$ are of the form

$$\Phi([X,r,\nu]^{(2)}) = \int\int \varphi\circ R\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}\nu^{\otimes m}(\boldsymbol{\mu}) \tag{E5.6}$$

with $m\in\mathbb{N}$, $\boldsymbol{n}=(n_1,\ldots,n_m)\in\mathbb{N}^m$ with $|\boldsymbol{n}|\geq 2$ and $\varphi\in\mathcal{C}_b(\mathbb{D}_{\boldsymbol{n}})$ and that $\mathcal{T}_1^{(2)}$ is convergence determining for $\mathcal{M}_1(\mathbb{M}_1^{(2)})$ (see Proposition 4.44). We will explain why for each such Φ the limit $\lim_{N\to\infty}\mathbb{Q}_{M,N}[\Phi]$ exists. Thus, the sequence $(\mathbb{Q}_{M,N})_N$ has at most one limit point.

Fix a m2pm-polynomial $\Phi\in\mathcal{T}_1^{(2)}$ as in (E5.6). Then, we have

$$\begin{aligned}\mathbb{Q}_{M,N}[\Phi] &= \int \Phi\,\mathrm{d}\mathbb{Q}_{M,N}\\ &= \int \Phi([Z,r,\nu_{M,N}]^{(2)})\,\mathrm{d}\mathbb{P}\\ &= \int\int\int \varphi\circ R\,\mathrm{d}\boldsymbol{\mu}^{\otimes\boldsymbol{n}}\,\mathrm{d}(\nu_{M,N})^{\otimes m}(\boldsymbol{\mu})\,\mathrm{d}\mathbb{P}.\end{aligned}$$

One can show that this converges to

$$\int \frac{1}{M^m}\sum_{i_1,\ldots,i_m=1}^{M} \varphi\Big(R\big((i_1,1),\ldots,(i_1,n_1),(i_2,n_1+1),\ldots,(i_2,n_1+n_2),(i_3,n_1+n_2+1),\ldots,(i_m,|\boldsymbol{n}|)\big)\Big)\,\mathrm{d}\mathbb{P} \tag{E5.7}$$

for $N\to\infty$ using the partial exchangeability of the distances under $\mathbb{P}$ (cf. Remark 5.4). However, writing down a formal proof for general m and $\boldsymbol{n}$ is cumbersome and we would have to introduce a lot of notation. For this reason we omit the proof. The reader may easily verify our claim for small m and $\boldsymbol{n}$ to understand what is going on here. Heuristically, Φ corresponds to sampling m species, then sampling $n_1,\ldots,n_m$ individuals in these species and then evaluating the (genetic) distances between the individuals. We sample with the two-level measure $\nu_{M,N}$, which means we uniformly sample from the first M species and in each of these species we uniformly sample from the first N individuals. Since M is finite, it is possible that some species are sampled more than once. But for $N\to\infty$ the probability to sample a single individual more than once goes to 0.

Relative compactness: We use Corollary 4.46. Thus, we have to show the following:

(a) there is a finite Borel measure μ_0 on $\mathbb{R}_+$ with $\mathbb{Q}_{M,N}[w(\mathfrak{M}_\nu)]\leq\mu_0$ for all $N\in\mathbb{N}$,

(b) $\lim_{\delta \searrow 0} \limsup_{N\to\infty} \mathbb{Q}_{M,N}[\mathfrak{M}_\nu(\{x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta\})] = 0$ for every $\varepsilon > 0$.

(a) Define the finite measure

$$\mu_0 := \delta_0 + \mathrm{Exp}(\gamma_s) + \mathrm{Exp}(\gamma_g) * \mathrm{Exp}(\gamma_s), \tag{E5.8}$$

where $\mathrm{Exp}(\gamma)$ denotes the exponential distribution with parameter $\gamma > 0$ and $\mu * \eta$ denotes the convolution of two distributions μ and η. The law of $r(\boldsymbol{x},\boldsymbol{y})$ is bounded from above by the finite measure μ_0 for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{N}^2$ (cf. (E5.1)). Since $w(\mathfrak{M}_{\nu_{M,N}}) = r_* \mathfrak{M}_{\nu_{M,N}}$, we get the desired inequality

$$\mathbb{Q}_{M,N}[w(\mathfrak{M}_\nu)] = \int w(\mathfrak{M}_{\nu_{M,N}})\,\mathrm{d}\mathbb{P} \le \mu_0.$$

(b) Let $\varepsilon > 0$. Then we have

$$\begin{aligned} &\mathbb{Q}_{M,N}[\mathfrak{M}_\nu(\{x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta\})] \\ &= \int \mathfrak{M}_{\nu_{M,N}}(\{x \in Z \mid \mathfrak{M}_{\nu_{M,N}}(\overline{B}(x,\varepsilon)) < \delta\})\,\mathrm{d}\mathbb{P} \\ &= \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{j=1}^{N}\int \mathbb{1}_{[0,\delta)}\left(\mathfrak{M}_{\nu_{M,N}}(\overline{B}((i,j),\varepsilon))\right)\mathrm{d}\mathbb{P} \\ &= \mathbb{P}(\mathfrak{M}_{\nu_{M,N}}(\overline{B}((1,1),\varepsilon)) < \delta) \\ &\le \mathbb{P}(\mathfrak{M}_{\nu_{M,N}}(\overline{B}((1,1),\varepsilon)) \le \delta), \end{aligned} \tag{E5.9}$$

where we used the fact that $\mathbb{P}(\mathfrak{M}_{\nu_{M,N}}(\overline{B}((i,j),\varepsilon)) < \delta)$ is the same for all $i,j \in \mathbb{N}$. By the definition of the relative frequencies $\mathfrak{f}_{1,i}(\varepsilon)$ we have

$$\mathfrak{M}_{\nu_{M,N}}(\overline{B}((1,1)),\varepsilon)) \to \frac{1}{M}\sum_{l=1}^{M}\mathfrak{f}_{1,l}(\varepsilon)$$

almost surely for $N \to \infty$ and Fatou's lemma yields

$$\limsup_{N\to\infty} \mathbb{P}(\mathfrak{M}_{\nu_{M,N}}(\overline{B}((1,1),\varepsilon)) \le \delta) \le \mathbb{P}\left(\frac{1}{M}\sum_{l=1}^{M}\mathfrak{f}_{1,l}(\varepsilon) \le \delta\right). \tag{E5.10}$$

Combining (E5.9), (E5.10) and (E5.3) we obtain

$$\begin{aligned} &\lim_{\delta \searrow 0}\limsup_{N\to\infty} \mathbb{Q}_{M,N}[\mathfrak{M}_\nu(\{x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta\})] \\ &\le \lim_{\delta\searrow 0}\mathbb{P}\left(\frac{1}{M}\sum_{l=1}^{M}\mathfrak{f}_{1,l}(\varepsilon) \le \delta\right) \\ &= \mathbb{P}\left(\frac{1}{M}\sum_{l=1}^{M}\mathfrak{f}_{1,l}(\varepsilon) = 0\right) \\ &\le \mathbb{P}(\mathfrak{f}_{1,1}(\varepsilon) = 0) \\ &= 0. \end{aligned}$$

5.3.2 Weak convergence of $(\mathbb{Q}_M)_M$

We have no information about $\mathbb{Q}_M$ other than that it is the weak limit of $(\mathbb{Q}_{M,N})_N$. Thus we must derive its properties from the approximating measures $\mathbb{Q}_{M,N}$.

Uniqueness: We show that the sequence $(\mathbb{Q}_M)_M$ has at most one limit point by proving existence of the limit $\lim_{M\to\infty}\mathbb{Q}_M[\Phi]$ for each m2pm-polynomial $\Phi \in \mathcal{T}_1^{(2)}$. Because the m2pm-polynomials are convergence determining, it follows that the sequence $(\mathbb{Q}_M)_M$ has at most one limit point.

Fix an m2pm-polynomial $\Phi \in \mathcal{T}_1^{(2)}$ of the form (E5.6). Since Φ is continuous and bounded, we have $\lim_{M\to\infty}\mathbb{Q}_M[\Phi] = \lim_{M\to\infty}\lim_{N\to\infty}\mathbb{Q}_{M,N}[\Phi]$. Using (E5.7) and the partial exchangeability of the distances (Remark 5.4) one can show that

$$\begin{aligned}\lim_{M\to\infty}\mathbb{Q}_M[\Phi] &= \lim_{M\to\infty}\lim_{N\to\infty}\mathbb{Q}_{M,N}[\Phi] \\ &= \int \varphi\big(R\big((1,1),\dots,(1,n_1),(2,1),\dots,(m,n_m)\big)\big)\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Again, we omit the cumbersome proof. Heuristically, if $M\to\infty$, then the probability to sample a species more than once goes to 0.

Relative compactness: Again, we use Corollary 4.46. We will show the following:

(a) $\mathbb{Q}_M[w(\mathfrak{M}_\nu)] \le \mu_0$ for every $M\in\mathbb{N}$, where μ_0 is defined in (E5.8),

(b) $\lim_{\delta\searrow 0}\limsup_{M\to\infty}\mathbb{Q}_M[\mathfrak{M}_\nu(\{x\in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta\})] = 0$ for every $\varepsilon>0$.

(a) Fix $M\in\mathbb{N}$. Observe that for finite measures $\eta,\mu\in\mathcal{M}_f(\mathbb{R}_+)$ we have $\eta\le\mu$ if and only if $\int f\,\mathrm{d}\eta \le \int f\,\mathrm{d}\mu$ for every non-negative function $f\in\mathcal{C}_b(\mathbb{R}_+)$. For such f the function

$$\begin{aligned}\mathbb{M}_1^{(2)} &\to \mathbb{R}_+ \\ [X,r,\nu]^{(2)} &\mapsto \int f\,\mathrm{d}w(\mathfrak{M}_\nu)\end{aligned}$$

is bounded and continuous as a concatenation of bounded and continuous functions (cf. Lemma 4.39 and Corollary 3.36). Because $\mathbb{Q}_M$ is the weak limit of $(\mathbb{Q}_{M,N})_N$ and because $\int f\,\mathrm{d}\mathbb{Q}_{M,N}[w(\mathfrak{M}_\nu)] \le \int f\,\mathrm{d}\mu_0$, we obtain

$$\begin{aligned}\int f\,\mathrm{d}\mathbb{Q}_M[w(\mathfrak{M}_\nu)] &= \int\int f\,\mathrm{d}w(\mathfrak{M}_\nu)\,\mathrm{d}\mathbb{Q}_M \\ &= \lim_{N\to\infty}\int\int f\,\mathrm{d}w(\mathfrak{M}_\nu)\,\mathrm{d}\mathbb{Q}_{M,N} \\ &= \lim_{N\to\infty}\int f\,\mathrm{d}\mathbb{Q}_{M,N}[w(\mathfrak{M}_\nu)] \\ &\le \int f\,\mathrm{d}\mu_0.\end{aligned}$$

Because this holds for every non-negative function $f \in \mathcal{C}_b(\mathbb{R}_+)$, we obtain

$$\mathbb{Q}_M[w(\mathfrak{M}_\nu)] \le \mu_0$$

for every $M \in \mathbb{N}$.

(b) Fix $\varepsilon > 0$. Because $\mathbb{Q}_{M,N}$ converges weakly to $\mathbb{Q}_M$, we have

$$\liminf_{N\to\infty} \mathbb{Q}_{M,N}[f] \ge \mathbb{Q}_M[f]$$

for every bounded lower semi-continuous function $f\colon \mathbb{M}_1^{(2)} \to \mathbb{R}$ (cf. [Bog07, Corollary 8.2.5]). By Lemma 4.39 and Lemma 3.38 the function

$$\begin{aligned} \mathbb{M}_1^{(2)} &\to \mathbb{R}_+ \\ [X,r,\nu]^{(2)} &\mapsto \mathfrak{M}_\nu(\{\, x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon) < \delta) \,\}) \end{aligned}$$

is lower semi-continuous (and obviously bounded by 1). Therefore, we have

$$\begin{aligned} &\mathbb{Q}_M[\mathfrak{M}_\nu(\{\, x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta \,\})] \\ &\le \liminf_{N\to\infty} \mathbb{Q}_{M,N}[\mathfrak{M}_\nu(\{\, x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta \,\})]. \end{aligned}$$

Using inequalities (E5.9) and (E5.10) we obtain

$$\mathbb{Q}_M[\mathfrak{M}_\nu(\{\, x \in X \mid \mathfrak{M}_\nu(\overline{B}(x,\varepsilon)) < \delta \,\})] \le \mathbb{P}\left(\frac{1}{M} \sum_{l=1}^{M} \mathfrak{f}_{1,l}(\varepsilon) \le \delta \right). \tag{E5.11}$$

$(\mathfrak{f}_{1,l}(\varepsilon))_{l\ge 2}$ is a sequence of exchangeable random variables. By de Finetti's Theorem (cf. [Ald85, Theorem 3.1]) there exists a random probability measure Ξ with values in $\mathcal{M}_1([0,1])$ such that $\Xi^{\otimes\infty}$ is a regular conditional distribution of $(\mathfrak{f}_{1,l}(\varepsilon))_{l\ge 2}$ given $\sigma(\Xi)$. In other words, $(\mathfrak{f}_{1,l}(\varepsilon))_{l\ge 2}$ is conditionally i. i. d. given $\sigma(\Xi)$. It follows that

$$\frac{1}{M} \sum_{l=2}^{M} \mathfrak{f}_{1,l}(\varepsilon) \xrightarrow{M\to\infty} \mathbb{E}(\mathfrak{f}_{1,2}(\varepsilon) \mid \Xi) = \int x \,\mathrm{d}\Xi(x)$$

almost surely (cf. [Ald85, Equation 2.24]). Fatou's lemma yields

$$\begin{aligned} \lim_{\delta\searrow 0} \limsup_{M\to\infty} \mathbb{P}\left(\frac{1}{M} \sum_{l=1}^{M} \mathfrak{f}_{1,l}(\varepsilon) \le \delta \right) &\le \lim_{\delta\searrow 0} \mathbb{P}\left(\int x \,\mathrm{d}\Xi(x) \le \delta \right) \\ &= \mathbb{P}\left(\int x \,\mathrm{d}\Xi(x) = 0 \right) \\ &= \mathbb{P}(\Xi = \delta_0). \end{aligned} \tag{E5.12}$$

Since $\Xi^{\otimes\infty}$ is the conditional distribution of $(\mathfrak{f}_{1,l}(\varepsilon))_{l\geq 2}$ given $\sigma(\Xi)$, we have

$$\mathbb{P}(\mathfrak{f}_{1,l}(\varepsilon) = 0 \text{ for all } l \geq 2 \mid \Xi = \delta_0) = 1.$$

It follows that

$$\begin{aligned} \mathbb{P}(\Xi = \delta_0) &= \mathbb{P}(\Xi = \delta_0) \cdot \mathbb{P}(\mathfrak{f}_{1,l}(\varepsilon) = 0 \text{ for all } l \geq 2 \mid \Xi = \delta_0) \\ &= \mathbb{P}\left(\mathfrak{f}_{1,l}(\varepsilon) = 0 \text{ for all } l \geq 2\right). \end{aligned} \tag{E5.13}$$

The latter probability is 0 by Lemma 5.8. Combining the inequalities (E5.11), (E5.12) and (E5.13) yields our claim.

Appendix A

Supplementary facts

A.1 Gromov-Hausdorff metric and topology

This section is a short overview over the Gromov-Hausdorff metric and topology. For a comprehensive treatment of this topic we refer the reader to [BBI01, chapter 7]. Most of the results of this section are taken from this book.

First we define the Hausdorff distance between subsets of a fixed metric space.

Definition A.1 (Hausdorff metric)
Let E, F be subsets of a metric space (X, r). We define the Hausdorff distance between *E and F by*

$$d_{\mathrm{H}}^X(E,F) := \inf\{\,\varepsilon > 0 \mid E \subset B(F,\varepsilon) \text{ and } F \subset B(E,\varepsilon)\,\}.$$

Sometimes we suppress the space X in our notation and write d_{H} instead of d_{H}^X.

It is easy to show that d_{H}^X is a metric on the *compact* subsets of X. Thus, we have a notion of distance and convergence for compact subsets of the same metric space. Mikhail Gromov extended this notion to compact metric spaces which are defined on totally different sets. His idea was to embed the spaces into a common metric space and then measure the Hausdorff distance between the embeddings. Thereby, we get a distance which measures how far compact metric spaces are from being isometric.

Definition A.2 (Gromov-Hausdorff metric)
Let $\mathbb{X}_c$ denote the isometry classes of the set

$$\{\,(X,r) \mid (X,r) \text{ compact metric space with } X \subset \mathbb{R}^{\mathbb{N}}\,\}$$

and let $[X, r_X], [Y, r_Y] \in \mathbb{X}_c$. We define the Gromov-Hausdorff distance *between $[X, r_X]$ and $[Y, d_Y]$ by*

$$d_{\mathrm{GH}}([X,r_X],[Y,r_Y]) := \inf_{Z,\varphi_X,\varphi_Y} d_{\mathrm{H}}^Z(\iota_X(X),\iota_Y(Y)),$$

where the infimum is taken over all isometric embeddings $\iota_X : X \to Z$, $\iota_Y : Y \to Z$ into a common metric space (Z, r_Z).

Recall that every compact metric space is complete and separable and thus Polish. In light of Proposition 2.13 we see that the property $X \subset \mathbb{R}^{\mathbb{N}}$ is not a restriction, since any compact metric space is homeomorphic to a subset of $\mathbb{R}^{\mathbb{N}}$.

It is shown in [BBI01, Theorem 7.3.30] that d_{GH} is a metric on $\mathbb{X}_c$. The topology induced by d_{GH} is called the *Gromov-Hausdorff topology*. By a slight abuse of terminology we sometimes say that a set (or a sequence) of compact metric spaces is relatively compact (or convergent) in the Gromov-Hausdorff topology, when the corresponding set of isometry classes is relatively compact (or convergent).

It is easy to see that Gromov-Hausdorff convergence generalizes Hausdorff convergence: If $(C_n)_n$ are compact subsets of a metric space (X,r) which converge to a compact set $C \subset X$ with respect to the Hausdorff metric d_{H}^X, then $[C_n,r]$ converges to $[C,r]$ in the Gromov-Hausdorff topology. The following lemma from [GPW09, Lemma A.1] shows that this statement can be reversed: A convergent sequence in $(\mathbb{X}_c, d_{\mathrm{GH}})$ can always be isometrically embedded in a common metric space such that we have d_{H} convergence of the embedded sets.

Lemma A.3 (Embedding of d_{GH}-convergent sequences)
Let $[X,r],[X_1,r_1],[X_2,r_2],\dots$ be elements of $\mathbb{X}_c$. The metric spaces $[X_n,r_n]$ converge to $[X,r]$ in the Gromov-Hausdorff topology if and only if there is a compact metric space (Z,r_Z) and isometries $\iota,\iota_1,\iota_2,\dots$ from $X,X_1,X_2,\dots$ into Z, respectively, such that

$$d_{\mathrm{H}}^Z(\iota_n(X_n),\iota(X)) \to 0.$$

In the remainder of this section we characterize relatively compact subsets of $\mathbb{X}_c$. Recall that a metric space (X,r) is called *totally bounded* if for every $\varepsilon > 0$ there is a number $N_\varepsilon \in \mathbb{N}$ such that X can be covered by N_ε balls of radius ε. This notion can be generalized to *uniform* totally boundedness of *a set* Γ of metric spaces, if the number N_ε is the same for all the metric spaces of Γ.

Definition A.4 (Uniformly totally bounded)
A set Γ of metric spaces is called uniformly totally bounded *if*

(a) there is a constant $D > 0$ such that $\operatorname{diam} X \leq D$ *for all $(X,r) \in \Gamma$,*

(b) for every $\varepsilon > 0$ there is a positive integer N_ε such that each $(X,r) \in \Gamma$ can be covered by at most N_ε balls of radius ε.

Since a metric space (X,r) is compact if and only if X is totally bounded and complete, it is not surprising that uniform totally boundedness is some sort of "uniform compactness". The following proposition is based on [BBI01, section 7.4.2].

Proposition A.5 (Characterization of relatively compact subsets of $\mathbb{X}_c$)
A set $\Gamma \subset \mathbb{X}_c$ is relatively compact in the Gromov-Hausdorff topology if and only if Γ is uniformly totally bounded.

A.2 Semi-continuous functions

In this section we define and characterize semi-continuity. The definitions and statements are based on [Bou89, section IV.6.2].

Definition A.6
Let X be a topological space and let f be a real-valued function on X. We say that f is upper semi-continuous *if for every point $x_0 \in X$ and every $\varepsilon > 0$ there is a neighborhood N of x_0 such that*

$$f(x) \leq f(x_0) + \varepsilon$$

for every $x \in N$. We say that f is lower semi-continuous *if the function $-f$ is upper semi-continuous.*

The semi-continuity of functions can be characterized in terms of nets and sequences, as the following lemma shows.

Lemma A.7 (Characterization of semi-continuous functions)
Let X be a topological space and let f be a real-valued function on X. The following are equivalent:

(a) The function f is upper semi-continuous.

(b) For every $a \in \mathbb{R}$ the set $\{ x \in X \mid f(x) < a \}$ is open in X.

(c) For every $a \in \mathbb{R}$ the set $\{ x \in X \mid f(x) \geq a \}$ is closed in X.

(d) For every $x_0 \in X$ and every net $(x_\alpha)_\alpha$ in X which converges to x_0 we have

$$\limsup_\alpha f(x_\alpha) \leq f(x_0).$$

If X is first-countable, then in addition the following is equivalent to the previous statements:

(e) For every $x_0 \in X$ and every sequence $(x_n)_n$ in X which converges to x_0 we have

$$\limsup_n f(x_n) \leq f(x_0).$$

It follows that a function f is lower semi-continuous if and only if

$$\liminf_\alpha f(x_\alpha) \geq f(x_0)$$

for every $x_0 \in X$ and every net $(x_\alpha)_\alpha$ in X which converges to x_0. Moreover, if X is metrizable, we can replace nets by sequences.

Proof (of Lemma A.7): The equivalence of (a), (b) and (c) can be found in [Bou89, section IV.6.2].

(a) $\Rightarrow$ (d): Let $(x_\alpha)_\alpha$ be a net in X which converges to $x_0 \in X$. It follows from Definition A.6 that for every $\varepsilon > 0$ we eventually have $f(x_\alpha) \leq f(x_0) + \varepsilon$. This implies

$$\limsup_\alpha f(x_\alpha) \leq f(x_0) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this establishes (d).

(d) $\Rightarrow$ (c): For $a \in \mathbb{R}$ define $A_a := \{x \in X \mid f(x) \geq a\}$. The set A_a is closed if and only if the limit of every convergent net in A_a is also in A_a. Let $(x_\alpha)_\alpha \subset A_a$ be a convergent net with limit $x_0 \in X$. By the definition of A_a and (d) we obtain

$$a \leq \limsup_\alpha f(x_\alpha) \leq f(x_0)$$

and this shows that $x_0 \in A_a$.

(d) $\Rightarrow$ (e): This is obvious since every sequence is also a net.

(e) $\Rightarrow$ (c): If X is first-countable, then the set A_a is closed if and only if the limit of every *sequence* in A_a is also in A_a. We can now proceed in the same way as we did in the proof of "(d) $\Rightarrow$ (c)" to obtain the result. □

Bibliography

[Ald85] David J. Aldous. Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer-Verlag, Berlin, 1985.

[ALW17] Siva Athreya, Wolfgang Löhr, and Anita Winter. Invariance principle for variable speed random walks on trees. *Ann. Probab.*, 45(2):625–667, 2017.

[Ban08] Vincent Bansaye. Proliferating parasites in dividing cells: Kimmel's branching model revisited. *Ann. Appl. Probab.*, 18(3):967–996, 2008.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BDLSJ18] Airam Blancas, Jean-Jil Duchamps, Amaury Lambert, and Arno Siri-Jégousse. Trees within trees: simple nested coalescents. *Electron. J. Probab.*, 23:27 pp., 2018.

[Bil68] Patrick Billingsley. *Convergence of probability measures*. Wiley, New York, 1968.

[BK10] Douglas Blount and Michael A. Kouritzin. On convergence determining and separating classes of functions. *Stochastic Process. Appl.*, 120(10):1898–1907, 2010.

[Bla16] Airam Blancas. *Two contributions to the theory of stochastic population dynamics*. PhD thesis, Centro de Investigación en Matemáticas A.C. (CIMAT), Guanajuato, Mexico, 2016.

[Bog07] Vladimir I. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.

[Bou89] Nicolas Bourbaki. *General topology. Chapters 1–4*. Elements of Mathematics. Springer-Verlag, Berlin, 1989. Translated from the French, Reprint of the 1966 edition.

[BRSSJ18] Airam Blancas, Tom Rogers, Jason Schweinsberg, and Arno Siri-Jégousse. The Nested Kingman Coalescent: Speed of Coming Down from Infinity. 2018. preprint available at `https://arxiv.org/abs/1803.08973`.

[BT11] Vincent Bansaye and Viet Chi Tran. Branching feller diffusion for cell division with parasite infection. *ALEA Lat. Am. J. Probab. Math. Stat.*, 8:95–127, 2011.

[Cro16] David A. Croydon. Scaling limits of stochastic processes associated with resistance forms. 2016. preprint available at `https://arxiv.org/abs/1609.05666`.

[Daw93] Donald A. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour, XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer-Verlag, Berlin, 1993.

[Daw18] Donald A. Dawson. Multilevel mutation-selection systems and set-valued duals. *Journal of Mathematical Biology*, 76(1):295–378, 2018.

[DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber. Marked metric measure spaces. *Electron. Commun. Probab.*, 16:174–188, 2011.

[DGP12] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber. Tree-valued Fleming-Viot dynamics with mutation and selection. *Ann. Appl. Probab.*, 22(6):2560–2615, 2012.

[DHW90] Donald A. Dawson, Kenneth J. Hochberg, and Yadong Wu. Multilevel branching systems. In *White noise analysis*, pages 93–107. World Sci. Publ., River Edge, NJ, 1990.

[Dud02] Richard M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley, New York, 1986.

[Eng89] Ryszard Engelking. *General topology.* Heldermann Verlag, Berlin, second edition, 1989.

[Glö12] Patric K. Glöde. *Dynamics of Genealogical Trees for Autocatalytic Branching Processes.* PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2012.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields*, 145(1-2):285–322, 2009.

[GPW13] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Tree-valued resampling dynamics martingale problems and applications. *Probab. Theory Related Fields*, 155(3-4):789–838, 2013.

[Gro99] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1999.

[Guf17] Stephan Gufler. Pathwise construction of tree-valued Fleming-Viot processes. 2017. preprint available at `https://arxiv.org/abs/1404.3682`.

[HJ77] Jorgen Hoffmann-Jørgensen. Probability in Banach spaces. In *École d'Été de Probabilités de Saint-Flour, VI-1976*, volume 598 of *Lecture Notes in Math.*, pages 1–186. Springer-Verlag, Berlin, 1977.

[Kel55] John L. Kelley. *General topology.* D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.

[Kim97] Marek Kimmel. Quasistationarity in a branching model of division-within-division. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 157–164. Springer-Verlag, New York, 1997.

[Kle08] Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin, 2008.

[KW17] Sandra Kliem and Anita Winter. Evolving phylogenies of trait-dependent branching with mutation and competition. Part I: Existence. 2017. preprint available at `https://arxiv.org/abs/1705.03277`.

[LV09] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[Mei18] Roland Meizis. Convergence of metric two-level measures spaces. 2018. preprint available at `https://arxiv.org/abs/1805.00853`.

[MR13] Sylvie Méléard and Sylvie Rœlly. Evolutive two-level population process and large population approximations. *Ann. Univ. Buchar. Math. Ser.*, 4(LXII)(1):37–70, 2013.

[Pit99] Jim Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999.

[Pro56] Yuri V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability and Its Applications*, 1(2):157–214, 1956.

[ST43] James A. Shohat and Jacob D. Tamarkin. *The Problem of Moments.* American Mathematical Society Mathematical surveys, vol. I. American Mathematical Society, New York, 1943.

[Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[Top70] Flemming Topsøe. *Topology and measure.* Lecture Notes in Mathematics, Vol. 133. Springer-Verlag, Berlin-New York, 1970.

[Top74] Flemming Topsøe. Compactness and tightness in a space of measures with the topology of weak convergence. *Math. Scand.*, 34:187–210, 1974.

[Vil09] Cédric Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 2009.

[vQ79] Boto von Querenburg. *Mengentheoretische Topologie.* Springer-Verlag, Berlin-New York, second edition, 1979.

[Wil70] Stephen Willard. *General topology.* Addison-Wesley Publishing Co., Reading, Massachusetts, 1970.

[Wu91] Yadong Wu. *Dynamic Particle Systems and Multilevel Measure Branching Processes.* PhD thesis, Carleton University, Ottawa, Canada, 1991.

[Zes83] Hans Zessin. The method of moments for random measures. *Z. Wahrsch. Verw. Gebiete*, 62(3):395–409, 1983.

Index

bicontinuous function, 31
blocks of an equivalence relation, 101

compact net, 12
completely regular, 18
convergence determining, 17

degree of an mm-monomial, 29
directed set, 10
distance array, 62
distance array distribution, 65
    n-point, 64
distance distribution, 37
distance map, 27, 62
distance matrix, 26
distance matrix distribution, 28
    n-point, 28

empirical distribution, 25
    n-th, 25
equidistributed sequence, 25
eventually, 10

first-countable space, 11
frequently, 11

gene coalescent, 102
Gromov-Hausdorff metric, 113
Gromov-Hausdorff topology, 114
Gromov-Prokhorov metric
    one-level, 34
    two-level, 68, 86
Gromov-Prokhorov topology
    one-level, 34
    two-level, 68, 87
Gromov-weak topology
    one-level, 30
    two-level, 66, 84

Hausdorff metric, 113

initial topology, 14
intensity measure, 52

Le Cam's theorem, 18
lower semi-continuous, 115

m2m-isomorphic, 48
m2m-monomial, 81
m2m space, 49
m2m triple, 48
m2pm-polynomial, 65
m2pm space, 49
m2pm triple, 48
mass distribution, 55
mass of a measure, 16
metric measure space, 24
metric measure triple, 24
metric probability measure space, 24
metric probability measure triple, 24
metric two-level measure space, 49
metric two-level measure triple, 48
metric two-level probability measure space, 49
metric two-level probability measure triple, 48
mm-isomorphic, 24
mm-monomial, 29
mm space, 24
mm triple, 24
modulus of mass distribution, 38
moment measure

- $\boldsymbol{n}$-mixed, 52
- m-th, 52
- first, 52
- infinite, 52
- infinite mixed, 52

mpm space, 24
mpm triple, 24

nested equivalence relations, 101
nested Kingman coalescent
- finite, 102
- infinite, 102

nested Kingman coalescent measure tree, 104
net, 10
- compact, 12
- tight, 13

normalization of a measure, 16
null measure, 16

order of a two-level measure, 52

partial exchangeability, 104
point separating, 14
Polish metric space, 14
Polish space, 14
Prokhorov's theorem, 17
Prokhorov metric, 15
push-forward
- one-level, 20
- two-level, 20

random distance distribution, 39
reconstruction theorem
- for m2pm spaces, 63
- for mm spaces, 27

semi-continuous, 115
separating, 17
separating points, 14
sequentially compact, 10
sequentially continuous, 10
species coalescent, 102
strongly separating points, 18
subnet, 12
supporting measure, 53

tail of a net, 10
tightness
- of a net, 13
- of a set, 17

totally bounded, 114
two-level Gromov-Prokhorov metric, 68, 86
two-level Gromov-Prokhorov topology
- on $\mathbb{M}^{(2)}$, 87
- on $\mathbb{M}_1^{(2)}$, 68

two-level Gromov-weak topology
- on $\mathbb{M}^{(2)}$, 84
- on $\mathbb{M}_1^{(2)}$, 66

two-level measure, 48
two-level probability measure, 48
two-level push-forward, 20

uniformly totally bounded, 114
upper semi-continuous, 115

weak convergence, 15
weak topology, 15

List of symbols

Throughout the whole thesis we use bold letters to denote tuples, vectors or matrices. For instance, $\boldsymbol{x}$ denotes a tuple (or matrix) of points of a set X and $\boldsymbol{\mu}$ denotes a tuple (or matrix) of measures.

In this list of symbols we use the following dummy notation: (X,r) a Polish metric space, Y a topological space, $x \in X$, $\boldsymbol{x} \in X^{\mathbb{N}}$, $A \subset X$, $\varepsilon, \delta > 0$, $K \geq 0$, $a,b \in \mathbb{R}$, $m \in \mathbb{N}$, $\boldsymbol{n} = (n_1,\dots,n_m) \in \mathbb{N}^m$, $f\colon X \to Y$ a measurable function, $\mu \in \mathcal{M}_f(X)$, $\boldsymbol{\mu} = (\mu_1,\mu_2,\dots)$ a finite or infinite tuple of measures in $\mathcal{M}_f(X)$, $\nu \in \mathcal{M}_f(\mathcal{M}_f(X))$.

Notation	**Description**	**Page List**
$\preceq$, $\preceq_{\mathcal{A}}$	Partial order on a directed set $\mathcal{A}$	10
$\xrightarrow{w}$	Weak convergence	15
$\complement A$	Complement of the set A	
$\mathbb{1}_A$	Characteristic function of the set A	
$\lvert A \rvert$	Cardinality of the set A	
$\lvert \boldsymbol{n} \rvert$	$:= \sum_{i_1}^{m} n_i$	
$[m]$	$:= \{1,\dots,m\}$	
$[\boldsymbol{n}]$	$:= \{(i,j) \mid i \in [m], j \in [n_i]\}$	
$\lVert \varphi \rVert_\infty$	Supremum of a real-valued function φ	
$[X,r,\mu]$	Metric measure space (mm space)	24
$[X,r,\nu]^{(2)}$	Metric two-level measure space (m2m space)	49
$(X,r,\nu)^{(2)}$	Metric two-level measure triple (m2m triple)	48
$f \circ g$	Concatenation of two functions f and g	
$f_*\mu$	Push-forward of μ with respect to the function f	20
$f_{**}\nu$	Two-level push-forward of ν with respect to the function f	20
$\overline{\mu}$	$:= \begin{cases} \frac{\mu}{\mathfrak{m}(\mu)}, & \mu \neq o \\ o, & \mu = o \end{cases}$, normalization of μ	16
$\overline{\boldsymbol{\mu}}$	$:= (\overline{\mu_1}, \overline{\mu_2}, \dots)$ for a (finite or infinite) tuple of measures $\boldsymbol{\mu}$	16
$\boldsymbol{\mu}^{\otimes \boldsymbol{n}}$	$:= \bigotimes_{i=1}^{m} \mu_i^{\otimes n_i} \in \mathcal{M}(X^{\lvert \boldsymbol{n} \rvert})$ for a finite tuple of measures $\boldsymbol{\mu}$	52
$\boldsymbol{\mu}^{\otimes \infty}$	$:= \bigotimes_{i=1}^{\infty} \mu_i^{\otimes \infty} \in \mathcal{M}_1(X^{\mathbb{N} \times \mathbb{N}})$ for an infinite tuple of measures $\boldsymbol{\mu}$	52
$\mu * \eta$	Convolution of two measures $\mu, \eta \in \mathcal{M}_1(\mathbb{R})$	

Notation	Description	Page List
$(x_n)_n$	Short form of a sequence $(x_n)_{n\in\mathbb{N}}$	
$(x_{ij})_{ij}$	Short form of an infinite matrix $(x_{ij})_{i,j\in\mathbb{N}}$	
$(x_\alpha)_\alpha$	Short form of a net $(x_\alpha)_{\alpha\in\mathcal{A}}$	
$X\sqcup Y$	Disjoint union of X and Y	
$\mathbb{A}_N$	A certain compact subset of $\mathbb{M}^{(2)}$	91
$B(X)$	Set of bounded Borel-measurable functions from X to $\mathbb{R}$	
$\mathcal{B}(X)$	Set of Borel sets of X	
$B(x,\varepsilon)$	Open ε-ball around x	
$\overline{B}(x,\varepsilon)$	Closed ε-ball around x	
$B(A,\varepsilon)$	$:=\bigcup_{x\in A}B(x,\varepsilon)=\{y\in X\mid r(A,y)<\varepsilon\}$, ε-environment of A	
$\mathcal{C}_b(X)$	Set of bounded, continuous functions from X to $\mathbb{R}$	
$\mathbb{D}_m$	Set of (m-point) distance matrices	26
$\mathbb{D}_\mathbb{N}$	Set of infinite distance matrices	26
$\mathbb{D}_{\boldsymbol{n}}$	Set of ($\boldsymbol{n}$-point) distance arrays	62
$\mathbb{D}_{\mathbb{N}\times\mathbb{N}}$	Set of inifinite distance arrays	62
$d_{2\mathrm{GP}}$	Two-level Gromov-Prokhorov metric	68, 86
d_{GH}	Gromov-Hausdorff metric	113
d_{GP}	Gromov-Prokhorov metric	34
d_{H}	Hausdorff metric	113
d_{P}	Prokhorov metric	15
d_{P}^X, $d_{\mathrm{P}}^{(X,r)}$	Prokhorov metric on $\mathcal{M}_f(X)$ (with respect to the metric r)	15
δ_x	Dirac measure at the point x	
$\mathrm{Exp}(\gamma)$	Exponential distribution with parameter $\gamma>0$	
$\mathcal{E}(I)$	Set of equivalence relations on I	101
f_K	Density function used in the approximations $f_K\cdot\nu$	56
$f_K\cdot\nu$	Approximation of the two-level measure ν via density	56
$\mathfrak{f}_{i,l}(t)$	Relative frequency of a block in the nested Kingman coalescent	105
g_K	A function used in the approximations $f_K\cdot\nu$	56

Notation	Description	Page List
$\mathfrak{L}(\xi)$	Law of a random variable ξ	
$M(X)$	Set of Borel-measurable functions from X to $\mathbb{R}$	
$\mathcal{M}(X)$	Set of Borel measures on X	
$\mathcal{M}_f(X)$	Set of finite Borel measures on X equipped with the Prokhorov metric d_{P}	15
$\mathcal{M}_f(X,r)$	Set of finite Borel measures on (X,r) equipped with the Prokhorov metric $d_{\mathrm{P}}^{(X,r)}$	
$\mathcal{M}_K(X)$	$:=\{\mu\in\mathcal{M}_f(X)\mid\mu(X)=K\}$, Borel measures on X with mass K	16
$\mathcal{M}_{\leq K}(X)$	$:=\{\mu\in\mathcal{M}_f(X)\mid\mu(X)\leq K\}$, Borel measures on X with mass less or equal than K	16
$\mathcal{M}_f(\mathcal{M}_f(X))$	Set of finite Borel measures on $\mathcal{M}_f(X)$ (i. e. two-level measures on X)	
$\mathbb{M}$	Set of metric measure spaces	24
$\mathbb{M}_1$	Set of metric probability measure spaces	24
$\mathbb{M}^{(2)}$	Set of metric two-level measure spaces	49
$\mathbb{M}_1^{(2)}$	Set of metric two-level probability measure spaces	49
$\mathfrak{M}_\nu$	First moment measure of ν	52
$\mathfrak{M}_\nu^m$	m-th moment measure of ν	52
$\mathfrak{M}_\nu^\infty$	Infinite moment measure of ν	52
$\mathfrak{M}_\nu^{\boldsymbol{n}}$	$\boldsymbol{n}$-mixed moment measure of ν	52
$\mathfrak{M}_\nu^{\infty,\infty}$	Infinite mixed moment measure of ν	52
$\mathfrak{m}(\mu)$	$:=\mu(X)$, mass of the measure μ	16
$\mathfrak{m}(\boldsymbol{\mu})$	$:=(\mathfrak{m}(\mu_1),\mathfrak{m}(\mu_2),\dots)$ for the (finite or infinite) tuple of measures $\boldsymbol{\mu}$	16
$\mathfrak{m}_*\nu$	Mass distribution of ν	55
$\mathcal{N}(I)$	Set of nested equivalence relations on I	101
$\mathbb{N}$	$:=\{1,2,\dots\}$	
$\mathbb{N}_0$	$:=\{0,1,2,\dots\}$	
$\mathbb{N}_{\geq 2}$	$:=\{2,3\dots\}$	
o	Null measure, which is 0 on all sets	
$P[f]$	$:=\int f\,\mathrm{d}P$, expectation of f with respect to a distribution P	
R, R^X	Distance map (on X)	27, 62
$\mathbb{R}_+$	$:=[0,\infty)$	

Notation	**Description**	**Page List**
$\operatorname{supp}\mu$	Support of the measure μ	
$\sigma(\xi)$	Sigma algebra of a random variable ξ	
$\mathcal{T}^{(1)}$	Set of mm-monomials (test functions on $\mathbb{M}$)	29
$\mathcal{T}^{(2)}$	Set of m2m-monomials (test functions on $\mathbb{M}^{(2)}$)	81
$\mathcal{T}_1^{(2)}$	Set of m2pm-monomials (test functions on $\mathbb{M}_1^{(2)}$)	65
τ_{2GP}	Two-level Gromov-Prokhorov topology on $\mathbb{M}^{(2)}$	87
τ'_{2GP}	Two-level Gromov-Prokhorov topology on $\mathbb{M}_1^{(2)}$	68
τ_{2Gw}	Two-level Gromov-weak topology on $\mathbb{M}^{(2)}$	84
τ'_{2Gw}	Two-level Gromov-weak topology on $\mathbb{M}_1^{(2)}$	66
τ_{GP}	Gromov-Prokhorov topology on $\mathbb{M}$	34
τ_{Gw}	Gromov-weak topology on $\mathbb{M}$	30
$V_\delta(\mu)$	Modulus of mass distribution of the measure μ	38
$W(\mu)$	Random distance distribution of the measure μ	39
$w(\mu)$	Distance distribution of the measure μ	37
$\text{w-lim}_n \mu_n$	Weak limit of a sequence $(\mu_n)_n$ of measures	
$\Xi_m(\boldsymbol{x})$	$:= \frac{1}{m}\sum_{i=1}^m \delta_{x_i}$, m-th empirical distribution of the sequence $\boldsymbol{x}$	25
$\Xi_\infty(\boldsymbol{x})$	$:= \begin{cases} \text{w-lim}_n \Xi_n(\boldsymbol{x}), & \text{if the limit exists} \\ o, & \text{else} \end{cases}$, infinite empirical distribution of the sequence $\boldsymbol{x}$	25