The IMA Volumes in Mathematics and its Applications

Eric Carlen
Mokshay Madiman
Elisabeth M. Werner *Editors*

Convexity and Concentration

Springer

The IMA Volumes in Mathematics and its Applications

Volume 161

Series editor

Fadil Santosa, *University of Minnesota, MN, USA*

Institute for Mathematics and its Applications (IMA)

The Institute for Mathematics and its Applications (IMA) was established in 1982 as a result of a National Science Foundation competition. The mission of the IMA is to connect scientists, engineers, and mathematicians in order to address scientific and technological challenges in a collaborative, engaging environment, developing transformative, new mathematics and exploring its applications, while training the next generation of researchers and educators. To this end the IMA organizes a wide variety of programs, ranging from short intense workshops in areas of exceptional interest and opportunity to extensive thematic programs lasting nine months. The IMA Volumes are used to disseminate results of these programs to the broader scientific community.

The full list of IMA books can be found at the Web site of the Institute for Mathematics and its Applications:

http://www.ima.umn.edu/springer/volumes.html.

Presentation materials from the IMA talks are available at

http://www.ima.umn.edu/talks/.

Video library is at

http://www.ima.umn.edu/videos/.

Fadil Santosa, Director of the IMA

More information about this series at http://www.springer.com/series/811

Eric Carlen • Mokshay Madiman
Elisabeth M. Werner
Editors

Convexity and Concentration

Springer

Editors
Eric Carlen
Department of Mathematics
Rutgers University
Piscataway, NJ, USA

Mokshay Madiman
Department of Mathematical Sciences
University of Delaware
Newark, DE, USA

Elisabeth M. Werner
Department of Mathematics
Case Western Reserve University
Cleveland, OH, USA

ISSN 0940-6573 ISSN 2198-3224 (electronic)
The IMA Volumes in Mathematics and its Applications
ISBN 978-1-4939-7004-9 ISBN 978-1-4939-7005-6 (eBook)
DOI 10.1007/978-1-4939-7005-6

Library of Congress Control Number: 2016961208

Mathematics Subject Classification (2010): 46, 52, 58, 60

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Foreword

This volume is based on the research focus at the IMA during the Spring semester of 2015. The Annual Thematic Program covering this period was “Discrete Structures: Analysis and Applications.” The program was organized by Sergey Bobkov, Jerrold Griggs, Penny Haxell, Michel Ledoux, Benny Sudakov, and Prasad Tetali. Many of the topics presented in this volume were discussed in the last five workshops that took place during the year. We thank the organizers of the workshops, the speakers, workshop participants, and visitors to the IMA who contributed to the scientific life at the institute and to the successful program. In particular, we thank volume editors Eric Carlen, Mokshay Madiman, and especially Elisabeth Werner, who also served as associate director of the IMA during the year. Finally, we are grateful to the National Science Foundation for its support of the IMA.

Minneapolis, MN, USA | Fadil Santosa

Preface

The 2014–2015 Annual Thematic Program at the Institute for Mathematics and its Applications (IMA) was *Discrete Structures: Analysis and Applications*. The program was organized by Sergey Bobkov (University of Minnesota), Jerrold Griggs (University of South Carolina), Penny Haxell (University of Waterloo), Michel Ledoux (Paul Sabatier University of Toulouse), Benny Sudakov (University of California, Los Angeles), and Prasad Tetali (Georgia Institute of Technology).

Convexity and concentration phenomena were the focus during the spring semester of 2015, and this volume presents some of the research topics discussed during this period. We have particularly encouraged authors to write surveys of research problems, thus making state-of-the-art results more conveniently and widely available. The volume addresses the themes of five workshops held during the spring semester of 2015:

- *Convexity and Optimization: Theory and Applications*, held February 23–27, 2015, at IMA and organized by Nina Balcan (Carnegie-Mellon University), Henrik Christensen (Georgia Institute of Technology), William Cook (University of Waterloo), Satoru Iwata (University of Tokyo), and Prasad Tetali (Georgia Institute of Technology)
- *The Power of Randomness in Computation*, held March 16–20, 2015, at Georgia Institute of Technology and organized by Dana Randall (Georgia Institute of Technology), Prasad Tetali (Georgia Institute of Technology), Santosh Vempala (Georgia Institute of Technology), and Eric Vigoda (Korea Advanced Institute of Science and Technology (KAIST))
- *Information Theory and Concentration Phenomena*, held April 13–17, 2015, at IMA and organized by Sergey Bobkov (University of Minnesota, Twin Cities), Michel Ledoux (Université de Toulouse III (Paul Sabatier)), and Joel Tropp (California Institute of Technology)
- *Analytic Tools in Probability and Applications*, held April 27–May 01, 2015, at IMA and organized by Sergey Bobkov (University of Minnesota, Twin Cities), Sergei Kislyakov (Russian Academy of Sciences), Michel Ledoux (Université de Toulouse III (Paul Sabatier)), and Andrei Zaitsev (Russian Academy of Sciences)

- *Graphical Models, Statistical Inference, and Algorithms (GRAMSIA)*, held May 18–22, 2015, at IMA and organized by David Gamarnik (Massachusetts Institute of Technology), Andrea Montanari (Stanford University), Devavrat Shah (Massachusetts Institute of Technology), Prasad Tetali (Georgia Institute of Technology), Rüdiger Urbanke (École Polytechnique Fédérale de Lausanne (EPFL)), and Martin Wainwright (University of California, Berkeley)

Discrete Structures: Analysis and Applications attracted intense interest from the mathematical science community at large. Each of the five workshops drew up to or more than 100 visitors. Aside from the workshops, an annual thematic year at the IMA provided an ideal environment for collaborative work. This program drew a mix of experts and junior researchers in various aspects of convex geometry and probability together with numerous people who apply these areas to other fields. This volume reflects many aspects of the semester, with chapters drawn from workshop talks, annual program seminars, and the research interests of many visitors.

The volume is organized into two parts. While the classification is of course arbitrary to some extent given the fluid boundaries between probability and analysis, **Part I: Probability and Concentration** contains those contributions that focus primarily on problems motivated by probability theory, while **Part II: Convexity and Concentration for Sets and Functions** contains those contributions that focus primarily on problems motivated by convex geometry and geometric analysis.

Acknowledgments No single volume could possibly cover all the active and important areas of research in convexity, probability, and related fields that were presented at the IMA, and we make no claim of comprehensiveness. However, we think that this volume presents a reasonable selection of interesting areas, written by leading experts who have surveyed the current state of knowledge and posed conjectures and open questions to stimulate further research. We thank the authors for their generous donation of time and expertise. Needless to say that without them, this volume would not have been possible.

We thank A. Beveridge, J. R. Griggs, L. Hogben, G. Musiker, and P. Tetali, the editors of the 2014–2015 fall semester volume *Recent Trends in Combinatorics*, for their advice and their many helpful comments.

We thank the IMA and their staff for wonderfully stimulating and productive long-term visits. We believe that the IMA is a critical national resource for mathematics. The *Discrete Structures: Analysis and Applications* program will have a lasting impact on research in convexity, probability, and related fields, and we hope this volume will enhance that impact. We are grateful for the opportunity to be part of it.

Piscataway, NJ, USA — Eric Carlen
Newark, DE, USA — Mokshay Madiman
Cleveland, OH, USA — Elisabeth M. Werner

Contents

Part I Probability and Concentration

Interpolation of Probability Measures on Graphs 3
Erwan Hillion

Entropy and Thinning of Discrete Random Variables 33
Oliver Johnson

Concentration of Measure Principle and Entropy-Inequalities 55
Paul-Marie Samson

Structured Random Matrices 107
Ramon van Handel

Rates of Convergence for Empirical Spectral Measures: A Soft Approach 157
Elizabeth S. Meckes and Mark W. Meckes

Concentration of Measure Without Independence: A Unified Approach Via the Martingale Method 183
Aryeh Kontorovich and Maxim Raginsky

Strong Data-Processing Inequalities for Channels and Bayesian Networks 211
Yury Polyanskiy and Yihong Wu

An Application of a Functional Inequality to Quasi-Invariance in Infinite Dimensions 251
Maria Gordina

Borell's Formula on a Riemannian Manifold and Applications 267
Joseph Lehec

Fourth Moments and Products: Unified Estimates 285
Ivan Nourdin and Giovanni Peccati

Asymptotic Expansions for Products of Characteristic Functions Under Moment Assumptions of Non-integer Orders 297
Sergey G. Bobkov

Part II Convexity and Concentration for Sets and Functions

Non-standard Constructions in Convex Geometry: Geometric Means of Convex Bodies 361
Vitali Milman and Liran Rotem

Randomized Isoperimetric Inequalities 391
Grigoris Paouris and Peter Pivovarov

Forward and Reverse Entropy Power Inequalities in Convex Geometry .. 427
Mokshay Madiman, James Melbourne, and Peng Xu

Log-Concave Functions 487
Andrea Colesanti

On Some Problems Concerning Log-Concave Random Vectors 525
Rafał Latała

Stability Results for Some Geometric Inequalities and Their Functional Versions 541
Umut Caglar and Elisabeth M. Werner

Measures of Sections of Convex Bodies 565
Alexander Koldobsky

On Isoperimetric Functions of Probability Measures Having Log-Concave Densities with Respect to the Standard Normal Law 577
Sergey G. Bobkov

Counting Integer Points in Higher-Dimensional Polytopes 585
Alexander Barvinok

The Chain Rule Operator Equation for Polynomials and Entire Functions 613
Hermann König and Vitali Milman

Part I
Probability and Concentration

Interpolation of Probability Measures on Graphs

Erwan Hillion

Abstract These notes are a review of the author's works about interpolation of probability measures on graphs via optimal transportation methods. We give more detailed proofs and constructions in the particular case of an interpolation between two finitely supported probability measures on $\mathbb{Z}$, with a stochastic domination assumption. We also present other types of interpolations, in particular Léonard's entropic interpolations and discuss the relationships between these constructions.

1 Introduction

The main topic of these notes is the theory of optimal transportation on discrete metric spaces, and in particular on graphs. We recall in this introduction some basic facts about this theory. For additional information, the reader is referred to Villani's comprehensive textbooks [Vill03] and [Vill09] or to shorter lectures notes, for instance by [AG13] or [Sant15].

Let (X, d) be a metric space endowed with its Borel σ-algebra. We consider two probability measures μ_0, μ_1 and a parameter $p \geq 1$. The optimal transportation theory is the study of the Monge-Kantorovitch minimization problem

$$\inf_{\pi \in \Pi(\mu_0,\mu_1)} \mathcal{I}_p(\pi) := \inf_{\pi \in \Pi(\mu_0,\mu_1)} \int_{X \times X} d(x,y)^p d\pi(x,y), \tag{1}$$

where $\Pi(\mu_0, \mu_1)$ is the set of couplings between μ_0 and μ_1, i.e. the set of probability measures π on $X \times X$ with marginals μ_0 and μ_1.

Under mild assumptions which are always satisfied in these notes (it suffices, for instance, to assume that (X, d) is Polish, see [Vill09], Theorem 4.1.), the set of optimal couplings, i.e. the set of minimizers for the functional $\mathcal{I}_p(\pi)$, is non-empty. Moreover, the application W_p defined by

E. Hillion (✉)
Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
e-mail: erwan.hillion@univ-amu.fr

E. Carlen et al. (eds.), *Convexity and Concentration*, The IMA Volumes in Mathematics and its Applications 161, DOI 10.1007/978-1-4939-7005-6_1

$$W_p(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0,\mu_1)} \left(\int_{X\times X} d(x,y)^p d\pi(x,y) \right)^{1/p} \tag{2}$$

defines a distance on the set $\mathcal{P}_p(X)$ of probability measures on X having a finite p-th moment. Such distances are called Wasserstein distances.

We are interested in properties of the metric space $(\mathcal{P}_p(X), W_p)$. Recall that the length of a continuous curve $\gamma : [0,1] \to X$ in a metric space (X,d) is given by

$$L(\gamma) := \sup_{0=t_0\leq\cdots\leq t_N=1} \sum_{i=0}^{N-1} d(\gamma(t_i), \gamma(t_{i+1})).$$

This induces a new distance $\tilde{d}$ on X given by

$$\tilde{d}(x,y) := \inf\{L(\gamma) \mid \gamma(0) = x, \gamma(1) = y\}. \tag{3}$$

If the distances d and $\tilde{d}$ coincide, the space (X,d) is said to be a length space. If furthermore the infimum is attained in (3) for some curve γ, then (X,d) is said to be a geodesic space and γ a geodesic curve. Compact length spaces are proven to be geodesic spaces (see [Stu06a], Lemma 2.3). Compact Riemannian manifolds and Euclidean spaces $\mathbb{R}^n$ are other important classes of geodesic spaces.

An important geometric property of Wasserstein spaces is the following :

Proposition 1 *If (X,d) is a geodesic space, then $(\mathcal{P}_p(X), W_p)$ is also a geodesic space.*

The study of Wasserstein W_p-geodesics, in particular for $p = 2$, has gained importance in the last decade, mainly because of the development of Sturm-Lott-Villani theory (see [Stu06a], [Stu06b] and [LV09]), which gives quite unexpected links between some geometric properties of a compact Riemannian manifold (M,g) and convexity properties of some functionals defined on $(\mathcal{P}_2(M), W_2)$. For instance, the Ricci curvature tensor Ric on M satisfies Ric $\geq K$ if and only if every couple of measures $\mu_0, \mu_1 \in \mathcal{P}_2(M)$ can be joined by a W_2-geodesic $(\mu_t)_{t\in[0,1]}$ along which we have

$$H(\mu_t) \geq (1-t)H(\mu_0) + tH(\mu_1) + K\frac{t(1-t)}{2} W_2^2(\mu_0, \mu_1), \tag{4}$$

where the entropy functional $H(\mu)$ is defined by $H(\mu) := -\int_M \rho \log(\rho) dvol$ if $d\mu = \rho dvol$ is absolutely continuous w.r.t. the Riemannian volume measure, and $H(\mu) = -\infty$ elsewhere. (In the original papers by Sturm and Lott-Villani, $H(\mu)$ is defined as $+\int_M \rho \log(\rho) dvol$ and referred to as 'the Boltzmann functional', and equation (4) is thus stated with a different sign.)

Equation (4) is called K-geodesic concavity for the entropy on $\mathcal{P}_2(M)$. The purpose of Sturm-Lott-Villani theory is to use equation (4) to define the notion of a measured length space (i.e. a length space (X,d) with a reference measure ν), satisfying $\mathrm{Ric} \geq K$. This generalized notion of Ricci curvature bounds is consistent with the classical one in the Riemannian setting, and under the assumption $\mathrm{Ric} \geq K > 0$, one can recover geometric properties and functional inequalities holding on (X, d, ν).

The generalization of Sturm-Lott-Villani theory in the case where the underlying space (X, d) is discrete (and thus not a length space) has been the subject of several research works. Among them, papers by Erbar-Maas [EM12] and Léonard [Leo14] will be presented in these notes. These approaches are based on the same idea which can be loosely summed up as follows: given a couple of probability measures f_0, f_1 on a graph G, we first construct an interpolation $(f_t)_{t\in[0,1]}$ which shares some similarities with Wasserstein W_2-geodesics in length spaces. It is then possible to define a notion of 'Ricci curvature bounds' by considering the behaviour of the entropy functional along such interpolations.

A similar approach of discrete Ricci curvature is described in the paper [GRST14] by Gozlan et al. In this article, the authors study the behaviour of the entropy functional along particular interpolations in the space of probability measures on a graph. The notion of discrete Ricci curvature thus obtained is strong enough to imply interesting functional inequalities (for instance, a discrete HWI inequality, see Proposition 5.1. of [GRST14]). The interpolating curves constructed in [GRST14] are seen as mixtures of binomial families, which is also the case of the $W_{1,+}$-geodesics introduced in these notes in Section 4. Whether both interpolating families coincide or not is still an open question, for which there is a positive answer in particular cases (see Remark 2).

Other important works about discrete Ricci curvature which will not be discussed here are Ollivier's Ricci curvature, see [Oll09], Sturm-Bonciocat rough curvature bounds, see [SB09], and the recent Bochner-type approach by Klartag et al, see [KKRT15].

The main purpose of these notes is to present some of the results of the author's papers [Hill14a], [Hill14b], [Hill14c]. These articles are about the construction of an interpolating family $(f_t(x))_{t\in[0,1]}$ between two finitely supported probability measures f_0, f_1 on a graph G, and the study of the concavity properties of the entropy functional $H(t) := H(f_t)$ along this family. In these notes, we mainly focus on the simpler case when the underlying graph is $\mathbb{Z}$ and explain briefly how these constructions can be extended to the general case.

Section 2 is a study of a particular class of interpolations, known as thinning of measures. We define the thinning and give some of its properties, among them a result about the concavity of its entropy. We then give an overview of the paper [Hill14a], about the contraction of probability measures, which is a natural generalization of the thinning in the setting of graphs. In particular, we explain how to adapt the proof of the concavity of the entropy in this new framework.

In Section 3, we recall some more notions on optimal transportation theory in continuous spaces. As in the discrete case, we mainly focus on the one-dimensional setting. We recall the Benamou-Brenier formula and how the description of its solutions by the Hamilton-Jacobi equation can be used to obtain interesting properties about Wasserstein geodesics. In particular, we obtain a concavity of entropy result by applying Proposition 8, which is formally similar to the discrete Benamou-Brenier equation (11) used in Section 2. The methods used in this section, especially Proposition 8, will be extended to the discrete setting in the next section.

In Section 4, we explain how to construct interpolating family between two probability measures on a graph satisfying a generalized version of equation (11). This leads to the definition of $W_{1,+}$-geodesics (see Definition 13). In particular, the similarity between equation (30) and Proposition 8 show that $W_{1,+}$-geodesics share similarities with thinning (or contractions) of measures in the discrete setting and with Wasserstein geodesics in continuous spaces. This section is an overview of the papers [Hill14b] and [Hill14c].

In Section 5, we construct other types of interpolating curves in $\mathcal{P}(\mathbb{Z})$ along which concavity of entropy results hold. These curves, known as entropic interpolations, have been introduced by Léonard in a series of articles among which we can cite [Leo12], [Leo13a], [Leo13b]. We explain with heuristic arguments why $W_{1,+}$-geodesics can be seen as limits of entropic interpolation when a certain parameter is taken to 0.

In Section 6, we explain the proof of the Shepp-Olkin conjecture (see Theorem 7), which is based on the ideas introduced in the theory of $W_{1,+}$-geodesics. The section sums up the results detailed in the papers [HJ14] and [HJ16].

2 A First Example: Entropy and Thinning of Measures

In this section, we study a particular method of interpolation known as the thinning operation, and which is a natural way to interpolate a probability measure finitely supported on $\mathbb{Z}_+$ and the Dirac measure δ_0. Moreover, the entropy along the thinning of a measure is concave. We give a detailed proof of this result, which will serve as a template for other concavity of entropy results.

2.1 The Thinning Operation on $\mathbb{Z}_+$

Let f be a probability measure supported on $\{0, \dots, N\}$ and X be a random variable distributed as f.

Definition 1 *The thinning of f is the family $(T_t f)_{t\in[0,1]}$ of probability measures on $\{0, \dots, N\}$ defined by*

$$\forall k \in \{0, \dots, N\}\ ,\ (T_t f)(k) = \sum_{l\,:\,l\geq k} \mathrm{bin}_{l,t}(k) f(l), \tag{5}$$

where $\mathrm{bin}_{l,t}(k) := \binom{l}{k} t^k (1-t)^{l-k} 1_{k\in\{0,\dots,l\}}$ *is the binomial measure.*

The map $t \mapsto T_t f$ can be seen as a curve in the space $\mathcal{P}(\mathbb{Z})$, interpolating between the Dirac measure $T_0 f = \delta_0$ and $T_1 f = f$. An equivalent point of view on the thinning is the following:

Proposition 2 *Let X be a random variable with probability mass function f. Let $(B_k)_{k\geq 1}$ be an i.i.d. family of Bernoulli variables of parameter t, independent of X. Then the random variable $T_t X := \sum_{i=1}^{X} B_i$ has probability mass function $T_t f$.*

The thinning operation has been introduced by Rényi in [Ren56], and can be seen as a discrete version of the scaling operation, which associates to a random variable X (on $\mathbb{R}$) the random variable tX, or equivalently associates to a density $(f(x))_{x\in\mathbb{R}}$ the density $f_t(x) := 1/tf(x/t)$, see the introduction of [HJK07] for further information. For instance, the thinning operation is used to state the following Poisson limit theorem, known as 'law of thin numbers' (see [HJK07]):

Theorem 1 *Let $f^{\star n}$ denote the n-th convolution of f, or equivalently the probability mass function of the independent sum $X_1 + \cdots + X_n$. Then $T_{1/n}(f^{\star n})$ converges pointwise to the Poisson distribution* $\mathrm{Poi}(\lambda)$*, where $\lambda := \mathbb{E}(f)$.*

2.2 Concavity of the Entropy Along the Thinning

We now state and prove a first concavity of entropy result.

Definition 2 *The entropy $H(f)$ of a finitely supported probabilty measure f on a discrete space E is defined by*

$$H(f) := -\sum_{x\in E} f(x)\log(f(x)),$$

where by convention $0\log(0) = 0$.

Theorem 2 *Let f be a probability measure supported on $\{0,\ldots,N\}$. The function $t \mapsto H(t) := H(T_t f)$ is concave on* $[0,1]$.

Theorem 2 has been first proven by Johnson and Yu in [JY09]. Their proof is based on the decomposition $H(T_t f) = -D(t) - L(t)$, where $D(t)$ is the relative entropy of the measure f_t with respect to the Poisson measure $\mathcal{P}(\lambda_t)$ (with $\lambda_t = \mathbb{E}[T_t f] = t\mathbb{E}[f]$), and $L(t) := \mathbb{E}[\log(\mathcal{P}(X_t,\lambda_t))]$ where X_t as $T_t f$ as p.m.f. The convexity of $D(t)$ follows from the data-processing inequality, and the convexity of $L(t)$ is proven by computing directly $L''(t)$ and by using the formula

$$\frac{\partial}{\partial t} T_t f(k) = -\left(\frac{k+1}{t}(T_t f)(k+1) - \frac{k}{t}(T_t f)(k)\right). \tag{6}$$

We are giving a slightly different proof of Theorem 2, which does not need a decomposition of $H(T_tf)$, but relies on a formula for $\frac{\partial}{\partial t}T_tf(k)$ which is quite similar to (6).

Proof of Theorem 2 Using the equations $k\binom{n}{k} = n\binom{n-1}{k-1}$, $(n-k)\binom{n}{k} = n\binom{n-1}{k}$, we prove the following transport equation for the binomial distributions:

$$\frac{\partial}{\partial t}\mathrm{bin}_{n,t}(k) = -n\left(\mathrm{bin}_{n-1,t}(k) - \mathrm{bin}_{n-1,t}(k-1)\right), \tag{7}$$

where ∇ is the left derivative operator. Setting $f_t(k) := (T_tf)(k)$, it follows from equation (7) that we have the transport equation

$$\frac{\partial}{\partial t}f_t(k) = -\nabla g_t(k)\ , \text{where } g_t(k) := \sum_{l\geq k} l\,\mathrm{bin}_{l-1,t}(k)f(l). \tag{8}$$

There is a similar second-order transport equation:

$$\frac{\partial^2}{\partial t^2}f_t(k) = \nabla_2 h_t(k)\ , \text{where } h_t(k) := \sum_{l\geq k} l(l-1)\mathrm{bin}_{l-2,t}(k)f(l), \tag{9}$$

and where $\nabla_2 := \nabla \circ \nabla$ is the second left derivative operator.

Equations (8) and (9) allow us to express derivatives of $f_t(k)$ with respect to t as 'spatial derivatives' of other families of functions.

$$\begin{aligned}-H''(t) &= \sum_k \left(\frac{\partial^2}{\partial t^2}f_t(k)\right)\log(f_t(k)) + \frac{1}{f_t(k)}\left(\frac{\partial}{\partial t}f_t(k)\right)^2 \\ &= \sum_k \nabla_2 h_t(k)\log(f_t(k)) + \frac{(\nabla_1 g_t(k))^2}{f_t(k)}.\end{aligned}$$

Now, we notice that, for $k \geq 1$ and $l_1, l_2 \geq k+1$, we have

$$\begin{aligned}l_1\binom{l_1-1}{k}l_2\binom{l_2-1}{k-1} &= \frac{l_1!l_2!}{(l_1-1-k)!k!(l_2-k)!(k-1)!} \\ &= l_1(l_1-1)\frac{(l_1-2)!}{((l_1-2)-(k-1))!(k-1)!}\frac{l_2!}{l!(l_2-k)!} \\ &= l_1(l_1-1)\binom{l_1-2}{k-1}\binom{l_2}{k},\end{aligned}$$

which implies

$$l_1(l_1-1)\mathrm{bin}_{l_1-2,t}(k-1)\mathrm{bin}_{l_2,t}(k) = l_1\mathrm{bin}_{l_1-1,t}(k)l_2\mathrm{bin}_{l_2-1,t}(k-1). \tag{10}$$

With the usual convention that $\mathrm{bin}_{l,t}(k) = 0$ if $k \notin \{0, \dots, l\}$, we notice that equation (10) is still true for any $k \geq 0$.

From equation (10) we deduce, by expanding the sums defining $f_t(k)$, $g_t(k)$ and $h_t(k)$, that for every $k \geq 0$,

$$f_t(k)h_t(k-1) = g_t(k)g_t(k-1). \tag{11}$$

Equation (11) allows us to write:

$$\begin{aligned}
\sum_k \nabla_2 h_t(k) \log(f_t(k)) &= \sum_k h_t(k) \left[\log(f_t(k)) - 2\log(f_t(k+1)) + \log(f_t(k+2))\right] \\
&= \sum_k h_t(k) \left[\log(f_t(k)) - \log\left(\frac{g_t(k+1)^2 g_t(k)^2}{h_t(k)^2}\right) + \log(f_t(k+2))\right] \\
&= \sum_k h_t(k) \left[\log\left(\frac{f_t(k)h_t(k)}{g_t(k)^2}\right) + \log\left(\frac{f_t(k+2)h_t(k)}{g_t(k+1)^2}\right)\right] \\
&\geq \sum_k h_t(k) \left[1 - \frac{g_t(k)^2}{f_t(k)h_t(k)} + 1 - \frac{g_t(k+1)^2}{f_t(k+2)h_t(k)}\right] \\
&= \sum_k 2h_t(k) - \frac{g_t(k)^2}{f_t(k)} - \frac{g_t(k+1)^2}{f_t(k+2)}.
\end{aligned}$$

The only inequality we have used is an elementary one: $\log(x) \geq 1 - 1/x$. On the other hand, we have:

$$\begin{aligned}
\sum_k \frac{(\nabla_1 g_t(k))^2}{f_t(k)} &= \sum_k \frac{g_t(k)^2}{f_t(k)} - 2\frac{g_t(k)g_t(k-1)}{f_t(k)} + \frac{g_t(k-1)^2}{f_t(k)} \\
&= \sum_k \frac{g_t(k)^2}{f_t(k)} - 2\frac{g_t(k)g_t(k+1)}{f_t(k+1)} + \frac{g_t(k+1)^2}{f_t(k+2)} \\
&= \sum_k -2h_t(k) + \frac{g_t(k)^2}{f_t(k)} + \frac{g_t(k+1)^2}{f_t(k+2)},
\end{aligned}$$

which finally proves that $-H''(t) \geq 0$. □

Remark 1 *The thinning operation can also be used to define an interpolation* $(f_t)_{t\in[0,1]}$ *between two probability measures* f_0 *and* f_1 *supported on* $\mathbb{Z}_+$*, by defining* f_t *as the convolution* $f_t := (T_{1-t}f_0) \star (T_t f_1)$*. Theorem 2, about the concavity of the entropy, is generalized to these interpolations under a technical assumption: if* f_0 *and* f_1 *are ultra-log-concave, which means that* $(k+1)f_i(k+1)^2 \geq (k+2)f_i(k)f_i(k+2)$ *for* $i = 0, 1$ *and* $k \in \mathbb{Z}_+$*, then the entropy* $H(f_t)$ *is a concave function of* t*. This is the main result of [JY09].*

2.3 Contraction of Measures on Graphs

There is a quite natural way to generalize the notion of thinning to the more general setting of a general connected, locally finite graph G. This has been done by the author in [Hill14a], and we recall here some definitions and theorems from this paper.

A curve on G of length n is any application $\gamma : \{0, \dots, n\} \to G$. We will denote $L(\gamma) = n$. A geodesic between two vertices $x, y \in G$ is a curve which minimizes the length $L(\gamma)$ among the set of curves $\gamma : \{0, \dots, n\} \to G$ with $\gamma(0) = x$ and $\gamma(n) = y$. The length of a geodesic path joining x to y id denoted $d_G(x, y)$, or $d(x, y)$ is there is no ambiguity, and this quantity defines a distance on the vertices of G, called the graph distance.

We will denote by $\Gamma(G)$ the set of geodesic paths on G, and $\Gamma_{x,y}$ the set of geodesic paths on G joining x to y. The cardinality of $\Gamma_{x,y}$ will be denoted by $|\Gamma_{x,y}|$.

Let $o \in G$ be a particular vertex, which will act as the vertex 0 in the thinning case. Let f be a finitely supported distribution on G.

Definition 3 *Let $\gamma \in \Gamma(G)$ be a geodesic path on G, of length n. The binomial family along γ is the family* $(\mathrm{bin}_{\gamma,t})_{t\in[0,1]}$ *of probability distributions supported on* $\{\gamma(0), \dots, \gamma(n)\}$*, defined by* $\mathrm{bin}_{\gamma,t}(\gamma(k)) := \mathrm{bin}_{n,t}(k)$.

Definition 4 *The contraction of f on o is the family $(f_t)_{t\in[0,1]}$ of probability measures defined by*

$$f_t(x) := \sum_{z\in G} \left(\frac{1}{|\Gamma_{o,z}|} \sum_{\gamma\in\Gamma_{o,z}} \mathrm{bin}_{\gamma,t}(x) \right) f(z), \tag{12}$$

where, given a geodesic $\gamma \in \Gamma(G)$, the measure $\mathrm{bin}_{\gamma,t}$ *is defined by*

$$\forall j \in \{0, \dots, l\}\ ,\ \mathrm{bin}_{\gamma,t}(\gamma(j)) := \mathrm{bin}_{l,t}(j),$$

where $l = L(\gamma)$, and by $\mathrm{bin}_{\gamma,t}(z) = 0$ *if $z \neq \gamma(j)$.*

Definition 5 *We orient the graph G as follows: given an edge $(xy) \in E(G)$, we set $x \to y$ if there exists a geodesic $\gamma \in \Gamma_{o,y}$ of length l with $\gamma(l-1) = x$. We then write that $x \in \mathcal{E}(y)$ and $y \in \mathcal{F}(x)$.*

Since G is connected, we have $\mathcal{E}(y) \neq \emptyset$ for every $y \neq o$. However, one may have $\mathcal{F}(x) = \emptyset$. Also notice that some edges may not be oriented with this definition, but they do not play any role in the construction of the contraction of f or in the study of the entropy of f_t.

The oriented graph $G, \to$ is itself naturally oriented as follows: we orient $(x_0y_0) \to (x_1y_1)$ if we have $x_0 \to y_0 = x_1 \to y_1$. The triple (x_0, x_1, y_1) is called an oriented triple. The set $T(G)$ of oriented triples on G can itself be oriented, the graph $(T(G), \to)$ being equal to $(E(E(G), \to), \to)$.

Definition 6 *The divergence of a function $g : (E(G), \to) \to \mathbb{R}$ is the function $\nabla \cdot g : G \to \mathbb{R}$ defined by*

$$\nabla \cdot g(x_1) := -\sum_{x_0 \in \mathcal{E}(x_1)} g(x_0x_1) + \sum_{x_2 \in \mathcal{F}(x_1)} g(x_1x_2).$$

The iterated divergence of $h : (T(G), \to) \to \mathbb{R}$ is the function $\nabla_2 \cdot h : G \to \mathbb{R}$ defined by

$$\begin{aligned}\nabla_2 \cdot h(x_2) :=& \sum_{x_1 \in \mathcal{E}(x_2)} \sum_{x_0 \in \mathcal{E}(x_1)} h(x_0x_1x_2) - 2 \sum_{x_1 \in \mathcal{E}(x_2)} \sum_{x_3 \in \mathcal{F}(x_2)} h(x_1x_2x_3) \\ &+ \sum_{x_3 \in \mathcal{F}(x_2)} \sum_{x_4 \in \mathcal{E}(x_3)} h(x_2x_3x_4).\end{aligned}$$

If we see $(T(G), \to)$ as $(E(E(G), \to), \to)$, then $\nabla_2 \cdot$ is simply $(\nabla \cdot) \circ (\nabla \cdot)$.

Definition 7 *Given a geodesic $\gamma : \{0, \ldots, l\} \to G$, we define the families of functions $g_{t,\gamma} : (E(G), \to) \to \mathbb{R}$ and $h_{t,\gamma} : (T(G), \to) \to \mathbb{R}$ by*

$$\forall j \in \{0, \ldots, l-1\}\ ,\ g_{t,\gamma}((\gamma(j)\gamma(j+1))) := l\mathrm{bin}_{l-1,t}(j)$$

and

$$\forall j \in \{0, \ldots, l-2\}, h_{t,\gamma}((\gamma(j)\gamma(j+1)\gamma(j+2))) := l(l-1)\mathrm{bin}_{l-2,t}(j),$$

the functions $g_{t,\gamma}$ and $h_{t,\gamma}$ taking the value 0 elsewhere.

The families $(g_{\gamma,t})$ and $(h_{\gamma,t})$ have been defined in order to have $\frac{\partial}{\partial t}\mathrm{bin}_{\gamma,t}(x) = -\nabla \cdot g_{\gamma,t}(x)$ and $\frac{\partial^2}{\partial t^2}\mathrm{bin}_{\gamma,t}(x) = \nabla_2 \cdot h_{\gamma,t}(x)$. From this fact we deduce easily:

Proposition 3 *Let (f_t) be a contraction family defined as in equation (12). We define the families of functions $(g_t)_{t\in[0,1]}$ and $(h_t)_{t\in[0,1]}$, respectively, on $(E(G), \to)$ and $(T(G), \to)$ by*

$$\forall (x_0x_1) \in (E(G), \to)\ ,\ g_t(x_0x_1) := \sum_{z\in G} \frac{1}{|\Gamma_{o,z}|} \sum_{\gamma \in \Gamma_{o,z}} g_{\gamma,t}(x_0x_1)f(z), \tag{13}$$

$$\forall (x_0x_1x_2) \in (E(G), \to)\ ,\ h_t(x_0x_1x_2) := \sum_{z\in G} \frac{1}{|\Gamma_{o,z}|} \sum_{\gamma \in \Gamma_{o,z}} h_{\gamma,t}(x_0x_1x_2)f(z). \tag{14}$$

We then have the differential equations:

$$\frac{\partial}{\partial t} f_t(x) = -\nabla \cdot g_t(x)\ ,\ \frac{\partial^2}{\partial t^2} f_t(x) = \nabla_2 \cdot h_t(x). \tag{15}$$

The triple of functions (f_t, g_t, h_t) satisfies a generalized version of equation (11):

Proposition 4 *Let (f_t) be the contraction family of a measure f_1 to a point $o \in G$, and (g_t), (h_t) associated to f by equations (13) and (14). We then have:*

$$\forall (x_0 x_1 x_2) \in T(G)\ ,\ h_t(x_0 x_1 x_2) f_t(x_1) = g_t(x_0 x_1) g_t(x_1 x_2). \tag{16}$$

Equation (16) is then used to prove concavity of entropy results along contraction families on a graph. Explicit bounds on the second derivative $H''(t)$ can be found for particular graphs: the complete graph, the grid $\mathbb{Z}^n$, the cube $\{0,1\}^n$ or trees. The reader is referred to [Hill14a] for detailed proofs and additional information.

Remark 2 *As there is only one coupling between a Dirac measure and another given probability measure on G, it is clear that the interpolation between δ_0 and f, as constructed in the paper [GRST14] (see in particular the beginning of Section 2), is identical to the contraction of f on o, as defined in Definition 4. In more general cases, the links between the interpolating families of [GRST14] and $W_{1,+}$-geodesics (as defined in Section 4) remain unclear.*

3 Optimal Transportation Theory

In this section, we recall some results about optimal transportation theory. We focus on equation (21), or equivalently on equation (22), which is a continuous, generalized version of equation (11) or equation (16) encountered in the previous section. This continuous equation, will be seen as a consequence of a Hamilton-Jacobi type equation (see equation (25)) which is satisfied by the velocity field associated to a Wasserstein geodesic.

The first paragraph is about the one-dimensional case and the second paragraph is about the general Riemannian setting. In both cases, the most important tool is the Benamou-Brenier formula, see equation (18) and equation (23), stated and proven in [BB99].

3.1 Optimal Transportation on the Real Line

We recall here some results from the continuous theory of optimal transportation, in the special case where the underlying metric space is the real line $\mathbb{R}$ with the usual Euclidian distance $d(x, y) = |x - y|$. In order to avoid technical difficulties that will not appear in the discrete setting, we will make the following additional assumption: the densities f_0 and f_1 with respect to the Lebesgue measure on $\mathbb{R}$ are supported on a compact interval K and are such that their respective cumulated distribution functions F_0 and F_1 are smooth bijections between K and $[0, 1]$.

We consider the Monge problem: we want to find

$$\inf_T \int_{\mathbb{R}} |x - T(x)|^p f_0(x)dx, \tag{17}$$

where the infimum is taken over the set of measurable maps $T : \mathbb{R} \to \mathbb{R}$ satisfying $T_*\mu_0 = \mu_1$. If there exists a solution T to the Monge problem, then the coupling $\pi = (Id \times T)_*\mu_0$ is solution to Monge-Kantorovitch problem (1).

It is possible to give an explicit expression for the optimal transport map (see [Vill03] for a proof):

Proposition 5 *The infimum in the Monge problem* (17) *is attained when* $T(x) = F_1^{-1} \circ F_0(x)$. *If* $p > 1$*, then this optimal* T *is unique.*

It is also possible to describe the Wasserstein geodesics:

Proposition 6 *Let* $T := F_1^{-1} \circ F_0$ *be the solution to Monge problem* (17)*. We set, for* $t \in [0, 1]$*,* $T_t(x) := (1 - t)x + tT(x)$ *and* $\mu_t := (T_t)_*\mu_0$*. Then for any* $p \geq 1$*, the family* $(f_t)_{t\in[0,1]}$ *is a* W_p*-Wasserstein geodesic between* f_0 *and* f_1*.*

Proof We have:

$$\begin{aligned}
W_p^p(\mu_s, \mu_t) &\leq \int_{\mathbb{R}} |x - T_t \circ T_s^{-1}(x)|^p d\mu_s \\
&= \int_{\mathbb{R}} |T_s(x) - T_t(x)|^p d\mu_0 \\
&= \int_{\mathbb{R}} |s - t|^p |T_0(x) - T_1(x)|^p d\mu_0 \\
&= |s - t|^p W_p^p(\mu_0, \mu_1).
\end{aligned}$$

In particular, we have, for any $0 \leq s \leq t \leq 1$,

$$\begin{aligned}
W_p(\mu_0, \mu_1) &\leq W_p(\mu_0, \mu_s) + W_p(\mu_s, \mu_t) + W_p(\mu_t, \mu_1) \\
&\leq ((s - 0) + (t - s) + (1 - t))W_p(\mu_0, \mu_1) = W_p(\mu_0, \mu_1),
\end{aligned}$$

so the previous inequalities are actually equalities, and in particular we have $W_p(\mu_s, \mu_t) = |t - s|W_p(\mu_0, \mu_1)$, which proves that $(\mu_t)_{t\in[0,1]}$ is a Wasserstein W_p-geodesic. □

In this one-dimensional framework, the Benamou-Brenier formula is written as follows (see [Sant15], Remark 9):

Theorem 3 *For* $p \geq 1$*, the Wasserstein distance* $W_p(f_0, f_1)$ *satisfies*

$$W_p^p(f_0, f_1) = \inf \int_{\mathbb{R}} \int_0^1 |v_t(x)|^p f_t(x) dx dt, \tag{18}$$

where the infimum is taken over the families $(f_t(x))_{x\in\mathbb{R},t\in[0,1]}$ *joining* f_0 *to* f_1 *and the velocity fields* $(v_t(x))_{x\in\mathbb{R},t\in[0,1]}$ *satisfying the continuity equation*

$$\frac{\partial}{\partial t}f_t(x) + \frac{\partial}{\partial x}(v_t(x)f_t(x)) = 0. \tag{19}$$

The velocity field $(v_t(x))$ associated to a Wasserstein geodesic by equation (19) satisfies interesting properties:

Proposition 7 *The Wasserstein geodesic* $(\mu_t)_{t\in[0,1]}$ *with* $d\mu_t(x) = f_t(x)dx$ *satisfies the continuity equation* (19) *with the velocity field* $(v_t(x))_{t\in[0,1],x\in\mathbb{R}}$ *defined by the equation*

$$\forall t \in [0,1]\ ,\ \forall x \in \mathbb{R}\ ,\ v_t(T_t(x)) = T(x) - x. \tag{20}$$

Moreover, the velocity field satisfies the differential equation

$$\frac{\partial}{\partial t}v_t(x) = -v_t(x)\frac{\partial}{\partial x}v_t(x). \tag{21}$$

Proof We first prove that equation (20) is unambiguous, i.e. that the mapping $x \mapsto T_t(x)$ is injective when t is fixed. The equation $T_t(x) = T_t(y)$ can be rewritten $(1-t)(x-y) + t(T(x) - T(y)) = 0$. This shows that $(x-y)(T(x) - T(y)) \leq 0$, with a strict inequality when $x \neq y$. But this is a contradiction with the fact that $T = F_0 \circ F_1^{-1}$ is increasing, so T_t is injective.

In order to prove equation (21), we differentiate both sides of equation (20) with respect to t. The differential of the right-hand side is clearly zero. To avoid ambiguities, we will use the notations $\partial_1 v(t,x) := \frac{\partial}{\partial t}v_t(x)$ and $\partial_2 v(t,x) := \frac{\partial}{\partial x}v_t(x)$. We then have:

$$\begin{aligned} 0 = \frac{\partial}{\partial t}v_t(T_t(x)) &= \left(\frac{\partial}{\partial t}T_t(x)\right)\partial_2 v(t, T_t(x)) + \partial_1 v(t, T_t(x)) \\ &= v(t, T_t(x))\partial_2 v(t, T_t(x)) + \partial_1 v(t, T_t(x)), \end{aligned}$$

which is equation (21) evaluated in t and $T_t(x)$. □

An equivalent form of Proposition 7 is stated as follows:

Proposition 8 *Let* $(f_t)_{t\in[0,1]}$ *be a Wasserstein geodesic on* $\mathbb{R}$. *We then have*

$$\frac{\partial^2}{\partial t^2}f_t(x) = \frac{\partial^2}{\partial x^2}\left(v_t(x)^2 f_t(x)\right). \tag{22}$$

Proof Equation (22) is simply obtained by differentiating the continuity equation (19) with respect to t and then by using equation (21). □

This description of Wasserstein geodesics can be used to prove a concavity of entropy property. Simple integration by parts allows us to prove the following (see [HJ16], Theorem 2.1. for a detailed proof):

Proposition 9 *Let* $(f_t)_{t\in[0,1]}$ *be a family of probability densities on* $\mathbb{R}$. *We suppose that there exist two families* $(g_t(x))$ *and* $(h_t(x))$ *such that*

$$\frac{\partial}{\partial t}f_t(x) = -\frac{\partial}{\partial x}g_t(x)\ ,\ \frac{\partial^2}{\partial t^2}f_t(x) = \frac{\partial^2}{\partial x^2}h_t(x).$$

Then the entropy $H(t)$ *of* f_t *satisfies*

$$-H''(t) = \int_{\mathbb{R}} \left(h_t(x) - \frac{g_t(x)^2}{f_t(x)}\right)\frac{\partial^2}{\partial x^2}(\log(f_t(x))) + \frac{\left(\frac{\partial g_t(x)}{\partial x}f_t(x) - g_t(x)\frac{\partial f_t(x)}{\partial x}\right)^2}{f_t(x)^3}dx.$$

In the case where $(f_t)_{t\in[0,1]}$ is a Wasserstein geodesic, then $g_t(x) = v_t(x)f_t(x)$ and $h_t(x) = v_t(x)^2f_t(x)$, and we find

$$-H''(t) = \int_{\mathbb{R}}\left(\frac{\partial v_t(x)}{\partial x}\right)^2 f_t(x)dx \geq 0.$$

3.2 Benamou-Brenier Theory in Higher Dimensions

The Benamou-Brenier formula stated in (18) in the one-dimensional setting can be stated in the Riemannian setting, at least for $p = 2$: more precisely, if μ_0, μ_1 are probability measures on a compact Riemannian manifold (M, g), with finite second moment, and with densities f_0,f_1 with respect to the Riemannian volume measure $dvol$, then

$$W_2(\mu, \mu_1)^2 = \inf \int_M \int_0^1 |v_t(x)|^2 f_t(x)dtdvol(x), \tag{23}$$

where the infimum is taken over the set of smooth families $(f_t(x))$ of probability densities joining f_0 to f_1, and where the velocity field $v_t : M \to TM$ satisfies the continuity equation

$$\frac{\partial}{\partial t}f_t(x) = -\nabla \cdot (v_t(x)f_t(x)), \tag{24}$$

where $\nabla\cdot$ is the divergence operator on the tangent bundle TM.

It can then be shown that this infimum is attained when (μ_t) is the Wasserstein W_2-geodesic between μ_0 and μ_1. Moreover, the associated velocity field $v_t(x)$ can be written under the form $v_t(x) = \nabla\Psi_t(x)$, where ∇ is the gradient operator and Ψ_t

is a convex function on M. Moreover, this function $\Psi_t(x)$ is solution to the Hamilton-Jacobi equation

$$\frac{\partial}{\partial t}\Psi_t(x) + \frac{1}{2}|\nabla\Psi_t(x)|^2 = 0. \tag{25}$$

The one-dimensional equation (21) can be seen as a consequence of this general Hamilton-Jacobi equation.

Hamilton-Jacobi equation is a powerful tool to prove concavity of the entropy results or functional inequalities holding on a Riemannian manifolds with Ricci curvature uniformly bounded from below, as done, for instance, in the 'heuristics section' of Otto-Villani's paper [OV00].

Among the several propositions for a definition of discrete curvature bounds, one of the most promising has been made by Erbar-Maas [EM12] and independently by Mielke [Miel13] and is based on a generalization of the Benamou-Brenier formula to the setting of discrete Markov chains:

Let $K : \mathcal{X} \times \mathcal{X}$ be an irreducible Markov Kernel on a finite space $\mathcal{X}$, admitting a reversible measure ν on $\mathcal{X}$. A density on $\mathcal{X}$ is a function $\rho : \mathcal{X} \to \mathbb{R}_+$ with $\sum_{x\in\mathcal{X}} \rho(x)\nu(x) = 1$.

The following definition by Erbar-Maas (see [EM12]) is directly inspired by Benamou-Brenier theory:

Definition 8 *We define a distance $\mathcal{W}$ on the set $\mathcal{P}(\mathcal{X})$ by*

$$\mathcal{W}(\rho_0, \rho_1)^2 := \inf_{\rho,\psi} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y\in\mathcal{X}} (\psi_t(x) - \psi_t(y))^2 \, \hat{\rho}_t(x,y)K(x,y)\nu(x)dt \right\}, \tag{26}$$

where the infimum is taken on the set of regular families of densities $(\rho_t(x))_{t\in[0,1],x\in X}$ joining ρ_0 to ρ_1 and satisfying the transport equation

$$\frac{\partial}{\partial t}\rho_t(x) = -\sum_{y\in\mathcal{X}} (\psi_t(y) - \psi_t(x)) \, \hat{\rho}_t(x,y)K(x,y),$$

where we define $\hat{\rho}_t(x,y) := \int_0^1 \rho(x)^{1-p}\rho(y)^p dp$.

As in classical Sturm-Lott-Villani theory, Erbar and Maas define a notion of Ricci curvature for the Markov kernel K by considering concavity properties of the entropy along generalized Wasserstein geodesics:

Definition 9 *The Markov kernel K satisfies* Ric $\geq \kappa$ *if for any geodesic $(\rho_t)_{t\in[0,1]}$ for the distance $\mathcal{W}$, we have*

$$H(t) \geq (1-t)H(0) + tH(1) - \kappa\frac{t(1-t)}{2}\mathcal{W}(\rho_0,\rho_1)^2,$$

where the entropy function H is defined by $H(t) = -\sum_{x\in\mathcal{X}} \rho_t(x)\log(\rho_t(x))\nu(x)$.

This definition of discrete Ricci curvature bounds is strong enough to recover several results which hold in the continuous setting, for instance a tensorization result ([EM12], Theorem 1.3) and functional inequalities satisfied on spaces with $Ric \geq \kappa > 0$, such as the modified logarithmic Sobolev inequality ([EM12], Theorem 1.5). Moreover, explicit bounds on the Ricci curvature can be computed in explicit fundamental examples, such as the complete graphs, the discrete hypercubes or circle graphs.

There are a lot of similarities and a lot of differences between the geodesics coming from the Erbar-Maas $\mathcal{W}$-distance and the $W_{1,+}$ interpolations presented below. In particular, both approaches are inspired by the continuous Benamou-Brenier theory. An important difference is that Erbar-Maas interpolations are based on a discrete version of the problem (23), whereas $W_{1,+}$-interpolations are based on the generalization of equation (21), which characterizes the solutions.

4 $W_{1,+}$-Geodesics on Graphs

In this section, we construct and study curves (f_t) in the space of finitely supported probability measures on a graph G along which equations similar to (11) and (16) hold. These curves will be called $W_{1,+}$-geodesics on G. We first focus on the case where $G = \mathbb{Z}$ and f_0 is stochastically dominated by f_1 (see Definition 10), before turning to the general case.

4.1 W_1-Geodesics on $\mathbb{Z}$

Before defining and constructing $W_{1,+}$-geodesics on $\mathbb{Z}$, we first recall in this paragraph some results about the geometry of the space $\mathcal{P}(\mathbb{Z})$ with the distance W_1. We consider the Monge-Kantorovitch problem

$$\inf_{\pi\in\Pi(f_0,f_1)} I_1(\pi) := \inf_{\pi\in\Pi(f_0,f_1)} \sum_{i,j} |i-j|\pi(i,j), \tag{27}$$

where f_0, f_1 are two finitely supported probability measures on $\mathbb{Z}$.

The minimization problem (27) has been extensively studied :

Proposition 10 *The set $\Pi_1(f_0,f_1)$ of minimizers for the problem (27) satisfy the following properties:*

- $\Pi_1(f_0,f_1)$ *is non-empty.*
- $\Pi_1(f_0,f_1)$ *is a convex subset of* $\mathcal{P}(\mathbb{Z}\times\mathbb{Z})$.
- *If* $\Pi_1(f_0,f_1)$ *has a unique element* π, *then either* f_0 *or* f_1 *is a Dirac measure.*
- *If neither* f_0 *nor* f_1 *is a Dirac measure, then* $\Pi_1(f_0,f_1)$ *has a non-empty interior.*

The proof of these facts is easy. The second and last points are proven by noticing that, if $\pi(i_1,j_1) \geq a > 0$ and $\pi(i_2,j_2) \geq a > 0$ and $(j_1-i_1)(j_2-i_2) \geq 2$, then the coupling $\tilde{\pi}$ defined by $\tilde{\pi}(i,j) = \pi(i,j) - a$ if $(i,j) = (i_1,j_1)$ or (i_2,j_2), $\tilde{\pi}(i,j) = \pi(i,j) + a$ if $(i,j) = (i_1,j_2)$ or (i_2,j_1) and $\tilde{\pi} = \pi$ elsewhere, satisfies $I_1(\tilde{\pi}) = I_1(\pi)$.

Definition 10 *Let f_0, f_1 be two probability measures on $\mathbb{Z}$. We say that f_0 is stochastically dominated by f_1, and we write $f_0 << f_1$, if we have $\forall k \in \mathbb{Z}$, $F_0(k) \geq F_1(k)$, where $F_i(k) := \sum_{l\leq k} f_i(l)$, $i = 0, 1$, is the cumulative distribution associated to f_i.*

Proposition 11 *We suppose that $f_0 << f_1$. Let π be in $\Pi_1(f_0,f_1)$. Then $\pi(i,j) > 0 \Rightarrow i \leq j$.*

Proof Let π be a coupling between f_0 and f_1. The stochastic domination assumption is equivalent to the following:

$$\forall k \in \mathbb{Z}\ , \sum_{i\leq k, j>k} \pi_{i,j} \geq \sum_{i>k, j\leq k} \pi_{i,j}. \tag{28}$$

Indeed, adding $\sum_{i,j\leq k} \pi_{i,j}$ to both sides of equation (28) gives $F_0(k) \geq F_1(k)$. Suppose that we have $\pi_{i_0 j_0} > 0$ for a couple $i_0 > j_0$. This means that the right-hand side of equation (28), with $k = j_0$, is non-zero, which implies that there exists a couple $i_1 \leq j_0 < j_1$ with $\pi_{i_1,j_1} > 0$. We now define $c(i_0,j_1) = c(i_1,j_0) := 1$, $c(i_0,j_0) = c(i_1,j_1) := -1$ and $c(i,j) := 0$ for other couples (i,j). If $0 < \varepsilon < \min(\pi_{i_0,j_0}, \pi_{i_1,j_1})$, then $\tilde{\pi} := \pi + \varepsilon c$ is a coupling between f_0 and f_1 and we have

$$\sum |j-i|\tilde{\pi}_{i,j} - \sum |j-i|\pi_{i,j} = 2\varepsilon(j_0 - \min(i_0,j_1)) < 0,$$

which means that π is not a W_1-optimal coupling. □

There is a converse to Proposition 11:

Proposition 12 *We suppose that there exists a coupling $\pi \in \Pi_1(f_0,f_1)$ such that for all $i,j \in \mathbb{Z}$, $\pi(i,j) > 0 \Rightarrow i \leq j$. Then $f_0 << f_1$.*

Proof It suffices to notice that the right-hand side of equation (28) is zero. □

We now prove that a W_1-geodesic between $f_0 << f_1$ is monotonic for the stochastic domination order:

Proposition 13 *Let $(f_t)_{t\in[0,1]}$ be a W_1-geodesic with $f_0 << f_1$. Then we have $f_s << f_t$ for any $0 \leq s \leq t \leq 1$. Moreover, we have $\sum_k kf_t(k) = tW_1(f_0,f_1) + \sum_k kf_0(k)$.*

Proof We fix $t \in [0,1]$. Let $\pi_{0,t} \in \Pi_1(f_0,f_1)$, $\pi_{t,1} \in \Pi_1(f_t,f_1)$ be two optimal couplings. In particular, we have $\sum_k \pi_{0,t}(i,k) = f_0(i)$, $\sum_i \pi_{0,t}(i,k) = f_t(k)$ and $\sum_{i,k} |k-i|\pi_{0,t}(i,k) = W_1(f_0,f_t) = tW_1(f_0,f_1)$. We construct a coupling $\pi \in \Pi(f_0,f_1)$ by setting

$$\pi(i,j) := \sum_k \pi_{0,t}(i,k)\pi_{t,1}(k,j).$$

The triangle inequality $|i-j| \leq |i-k| + |k-j|$ easily implies

$$\sum_{i,j} |j-i|\pi(i,j) \leq W_1(f_0,f_t) + W_1(f_t,f_1) = W_1(f_0,f_1),$$

so we have $\pi \in \Pi_1(f_0,f_1)$. In particular, we have by Proposition 11 $\pi(i,j) > 0 \Rightarrow i \leq j$, and moreover there is equality in the triangle inequality: $j - i = |j-i| = |k-i| + |j-k|$ (thus $i \leq k \leq j$), whenever we have $\pi_{0,t}(i,k)\pi_{t,1}(k,j) > 0$, which implies by Proposition 12 that $f_0 << f_t << f_1$. The inequality $f_s << f_t$ for any $s \leq t$ comes from the fact that $(f_{\alpha t})_{\alpha\in[0,1]}$ is a W_1-geodesic between f_0 and f_t, and choosing $\alpha := s/t$ leads to $f_0 << f_s << f_t$.

In order to prove the second point, we write

$$\begin{aligned}\sum_k kf_t(k) &= \sum_k k \sum_i \pi_{0,t}(i,k)\\ &= \sum_i \sum_k (k-i+i)\pi_{0,t}(i,k)\\ &= \sum_{i,k} |k-i|\pi_{0,t}(i,k) + \sum_i i \sum_k \pi_{0,t}(i,k)\\ &= tW_1(f_0,f_1) + \sum_i if_0(i).\end{aligned}$$

□

The following result will be seen as a particular case of Proposition 15:

Proposition 14 *Let $(f_t)_{t\in[0,1]}$ be a W_1-geodesic on $\mathbb{Z}$ with $f_0 << f_1$. Then there exist two families of finitely supported functions $(g_t)_{t\in[0,1]}$, $(h_t)_{t\in[0,1]}$, such that:*

- $\frac{\partial}{\partial t}f_t(k) = -\nabla_1 g_t(k)$.
- $\frac{\partial}{\partial t}g_t(k) = -\nabla_1 h_t(k)$.
- $g_t(k) \geq 0$.

In this setting, the families (g_t) and (h_t) are defined by

$$g_t(k) := -\sum_{l\leq k} \frac{\partial}{\partial t}f_t(l)\ ,\ h_t(k) := -\sum_{l\leq k} \frac{\partial}{\partial t}g_t(l).$$

The fact that (f_t) is a W_1-geodesic between $f_0 << f_1$ is used as follows: we have $\sum_k kf_t(k) = W_1(f_0,f_1).t + \sum_k kf_0(k)$, so $\sum_k g_t(k)$ is constant, so $h_t(k)$ is finitely supported. The stochastic domination is also used to prove that $g_t(k) \geq 0$.

4.2 $W_{1,+}$-Geodesics on $\mathbb{Z}$

The proof of the concavity of the entropy along thinning of measures (Theorem 2) only uses the fact that

$$f_t(k)h_t(k-1) = g_t(k)g_t(k-1), \tag{29}$$

therefore a natural question is to find other families $(f_t(k))_{t\in[0,1],k\in\mathbb{Z}}$ of probability measures on $\mathbb{Z}$ satisfying the same equation (29), because along such families the concavity of the entropy would be already proven.

A very interesting fact is that the most natural continuous analogue of equation (29) is simply the equation $fh = g^2$, which is satisfied by Wasserstein geodesics on the real line. It thus seems natural to use equation (29) as a possible definition of Wasserstein interpolation on $\mathbb{Z}$.

We construct, with Theorem 5, an interpolation $(f_t)_{t\in[0,1]}$ satisfying equation (29), where the families $(g_t(k))$ and $(h_t(k))$ are defined from $(f_t(k))$ by Proposition 14. This interpolation is called $W_{1,+}$-geodesic between f_0 and f_1. We will later explain how to modify equation (29) in the general case to define $W_{1,+}$-geodesics between each couple f_0, f_1 of finitely supported probability measures on a graph.

Definition 11 *Let $f_0 << f_1$ be two finitely supported probability measures on $\mathbb{Z}$. A $W_{1,+}$-geodesic between f_0 and f_1 is a family $(f_t)_{t\in[0,1]}$ of probability measures on $\mathbb{Z}$ such that*

- *(f_t) is a W_1-geodesic.*
- *$f_t(k)h_t(k-1) = g_t(k)g_t(k-1)$.*
- *$g_t(k) > 0$ whenever we have $f_t(k) > 0$,*

where $(g_t)_{t\in[0,1]}$ and $(h_t)_{t\in[0,1]}$ are defined from $(f_t)_{t\in[0,1]}$ by Proposition 14.

The proof of Theorem 2 is still valid in this more general framework, which leads to the following:

Theorem 4 *Let $(f_t)_{t\in[0,1]}$ be a $W_{1,+}$-geodesic on $\mathbb{Z}$. The entropy $H(t)$ of f_t is then a concave function of t.*

In order to make Theorem 4 relevant, an important question is to show the existence of a $W_{1,+}$-geodesic with prescribed f_0 and f_1. This question is answered by the following:

Theorem 5 *Let $f_0 << f_1$ be two probability measures on $\mathbb{Z}$. There exists a unique $W_{1,+}$-geodesic $(f_t)_{t\in[0,1]}$ joining f_0 to f_1. Moreover, (f_t) can be written as a mixture of binomial families:*

$$f_t(k) = \sum_{i\le k\le j} \mathrm{bin}_{i,j,t}(k)\pi(i,j),$$

where the coupling π *is solution to the minimization problem*

$$\inf_{\pi \in \Pi_1(f_0,f_1)} \sum_{i \le j} \pi(i,j) \log \left(\pi(i,j)(j-i)!\right) - \pi(i,j),$$

and where $\mathrm{bin}_{i,j,t}(k) = \mathrm{bin}_{j-i,t}(k-i)$.

The proof of Theorem 5 is quite similar to the proof of Theorem 6, of which a detailed proof is given. It is interesting to notice that the optimal coupling π can be written under the form $\pi(i,j) = \frac{a(i)b(j)}{(j-i)!}$ for a couple of functions $a, b : G \to \mathbb{R}$. A detailed proof of Theorem 5, stated in a more general form, can be found in [Hill14b]: see in particular Theorem 4.5 for the existence and Theorem 3.19 for the binomial mixture.

4.3 $W_{1,+}$-Geodesics on Graphs

In this paragraph we explain briefly how to define $W_{1,+}$-geodesics between a couple of finitely supported probability measures f_0, f_1 on a graph G. If we want an equation similar to (16) to make sense, we first need to define an orientation on G.

Definition 12 *We define the* W_1*-orientation on G, with respect to the couple* f_0, f_1*, in the following way: let* $(xy) \in E(G)$ *be any edge of G. If there exists an optimal coupling* $\pi \in \Pi_1(f_0,f_1)$*, a couple of vertices* $a, b \in G$ *with* $\pi(a,b) > 0$*, a geodesic path* $\gamma \in \Gamma(a,b)$ *and an integer* $l \in \{0, \ldots, d(a,b)-1\}$ *such that* $x = \gamma(l)$ *and* $y = \gamma(l+1)$*, then we orient the edge* (xy) *by* $x \to y$*.*

It is proven in [Hill14b], Theorem 2.18, that the orientation $x \to y$ does not depend on π, a, b, γ, and so this orientation is well defined. As in the case of contraction of measures, some edges may not be oriented by this process, but they do not play any role in the construction of $W_{1,+}$-geodesics.

Having an orientation on G allows us to define the divergence and iterated divergence operators, respectively, on the oriented graphs $(E(G), \to)$ and $(T(G), \to)$. We now associate a family $(g_t)_{t\in[0,1]}$ to each W_1-geodesic as follows (see [Hill14b], Theorem 2.23 and Proposition 2.25):

Proposition 15 *Let* $(f_t)_{t\in[0,1]}$ *be a* W_1*-geodesic on G. We orient G with the* W_1*-orientation with respect to* f_0, f_1*. There exists a family* $(g_t)_{t\in[0,1]}$ *of functions defined on* $(E(G), \to)$ *such that* $\forall x, y \in (E(G), \to)$, $g_t(xy) > 0$ *and* $\frac{\partial}{\partial t} f_t(x) = -\nabla \cdot g_t(x)$.

Definition 13 *Let G be a graph,* W_1*-oriented with respect to* f_0, f_1*. A family* $(f_t)_{t\in[0,1]}$ *of probability measures on G is said to be a* $W_{1,+}$*-geodesic if:*

- $(f_t)_{t\in[0,1]}$ *is a* W_1*-geodesic.*
- *There exist two families* $(g_t)_{t\in[0,1]}$ *and* $(h_t)_{t\in[0,1]}$ *defined, respectively, on* $(E(G), \to)$ *and* $(T(G), \to)$ *such that*