Internship report
MAP594

Around Caffarelli's contraction theorem

Anna Bahrii

supervised by
Professor Nathael Gozlan

August 25, 2023

Déclaration d'intégrité relative au plagiat

I, the undersigned, Anna Bahrii, certify that

– I am the author of the report;

– The results described in the report are the culmination of my work;

– I have not used third-party sources or results without clearly citing and referencing them according to the recommended bibliographic rules.

Done in Paris, 25/08/2023

Signature:

Abstract

We study the entropic optimal transport formulation of the Caffarelli's contraction theorem which states that the optimal map transporting the standard Gaussian measure on $\mathbb{R}^d$ on a strongly log-concave probability measure is a contraction. We study the same formulation of the theorem by replacing the Gaussian measure by the discrete Poisson measure. We prove convergence of regularized entropic cost to W_1 distance, monotonicity of W_1 distance between ultra log-concave and ultra log-convex measures in the discrete space and restriction to convex functions in the dual form of the W_1 distance between ultra log-concave and ultra log-convex measures.

Contents

1 Introduction

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$. Luis Caffarelli proved in [3] that the map that transports γ_d to a probability measure having a log-concave density with respect to γ_d is a 1-Lipschitz map for the Euclidian norm (i.e. a *contraction*). Let us state the generalized version of Caffarelli's theorem ([5]).

Theorem 1.1. *(Caffarelli, 2000, [3]) Let μ be a probability measure on $\mathbb{R}^d$ with a finite second moment and ν be a probability measure of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$, $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ with V and W convex functions. Then there exists a continuously differentiable and convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\varphi$ is 1-Lipschitz and $\nu = \nabla\varphi_{\#}\mu$.*

Here, $\nu = \nabla\varphi_{\#}\mu$ means that ν is a *push-forward measure* of μ under the *Brenier's map* $\nabla\varphi$ (explained in details in Section 2). We recall that a function $f : \mathbb{R}^d \to \mathbb{R}_+$ is called *log-concave* if $\log f$ is concave. In other words, if for any $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta y)) \geq f(x)^\theta f(y)^{1-\theta}, \quad x, y \in \mathbb{R}^d.$$

One way of proving Theorem 1.1, as it was described in [8], consists in proving the following statement: the transport map between measures μ and ν given as in Theorem 1.1 is a contraction if and only if for any η probability measure on $\mathbb{R}^d$ with finite second moment such that $\int f d\eta \leq \int f d\nu$ for $f : \mathbb{R}^d \to \mathbb{R}$ convex,

$$W_2(\mu, \nu) \leq W_2(\mu, \eta). \tag{1}$$

Caffarelli's contraction theorem has found its applications in functional inequalities theory. It is known that Gaussian measure satisfies many functional inequalities such as, for example, logarithmic Sobolev inequality, Poincaré inequality, isoperimetric inequalities, etc. Having Theorem 1.1 in hand, we can almost readily deduce that any measure that has a log-concave density with respect to the Gaussian measure also satisfies these inequalities (see Section 3).

The question we have asked ourselves is if the Gaussian measure is the only measure for which this contraction result is true. Though the clear answer is not given yet, in this report we are trying to study the behaviour of the optimal transport map by replacing the Gaussian measure with the *Poisson measure* and by adopting (1). As follows, we switch to measures defined on the discrete space. As it will be shown in Section 6, this requires more careful analysis of convexity of discrete functions.

The main and original contribution of this report is proving monotonicity of Wasserstein 1-distance for discrete measures where one measure is *ultra log-concave* and the other one is *ultra log-convex* (see Section 6). This in particular includes proving many preliminary results such as log-concavity and log-convexity preservation properties of $M/M/\infty$ process semigroup, convergence of regularized entropic transport cost, monotonicity properties of entropic transport cost. Unlike for the Gaussian measure, where monotonicity of Wasserstein 2-distance is equivalent to the contraction of the transport map, for the Poisson measure we cannot deduce the same. We give an application in terms of monotonic W_1 convergence in the *law of thin numbers*. We share our hypothesis and reasoning on this matter at the end of the report. In the report we mostly omit proving known or classical results and focus on proving original contributions and the statements that will provide a better understanding of these results.

Organization This report is organized as follows. In Section 2 we list fundamental definitions and lemmas of the optimal transport. In Section 3 we explain how exactly Caffarelli's theorem can be used for transferring functional inequalities between measures. In Section 4 we briefly explain the main ideas of entropical optimal transport (EOT). Section 5 explains how Caffarelli's contraction theorem can be proved with means of EOT. Finally, in Section 6 we develop EOT approach towards Caffarelli's theorem for the Poisson measure.

2 Optimal transport background

Optimal transport studies optimal transportation and allocation of resources. Gaspard Monge was the first to describe mathematically a very practical question: how to transport given resources from one point to another in the most optimal way. Optimality here is described by a *cost function* that can mean price, energy, etc, and transportation is described by a *transport map*, or in other words *deterministic coupling*. This optimization problem is known as *Monge problem*. Years later, Leonid Kantorovich was studying a modified problem: what is the most optimal way to transport given resources from n initial locations to m destinations. *Kantorovich problem* is not deterministic anymore since every resource is transported to many destinations by parts that give the optimality of the transportation.

Let X, Y be some locally compact, separable, and complete metric spaces. We denote by $\mathcal{P}(X)$ the set of Borel probability measures on X, and by $\mathcal{P}_p(X) \subset \mathcal{P}(X)$, $p \geq 1$, the set of Borel probability measures with finite p-th moment, i.e.

$$\mu \in \mathcal{P}_p(X) \iff \int_X d(x_0, x)^p d\mu(x) < +\infty, \quad \text{where } x_0 \in \mathbb{R}^d \text{ is a given point.}$$

We start with basic definitions of optimal transport theory.

Definition 2.1. *(Transport map) Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. A map $T : X \to Y$ is called a transport map from μ to ν if ν is a push-forward measure of μ under the map T, i.e. if for any Borel set $A \in \mathcal{B}(Y)$,*

$$(T_{\#}\mu)(A) := \mu(T^{-1}(A)) = \nu(A).$$

Lemma 2.1. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. A map $T : X \to Y$ is a transport map if and only if for any Borel bounded function $\varphi : Y \to \mathbb{R}$ it holds,*

$$\int_Y \varphi(y) d\nu(y) = \int_X \varphi(T(x)) d\mu(x).$$

Definition 2.2. *(Coupling) Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. A measure $\pi \in \mathcal{P}(X \times Y)$ is called a coupling if*

$$\pi(A \times Y) = \mu(A), \quad \forall A \in \mathcal{B}(X),$$

and

$$\pi(X \times B) = \nu(B), \quad \forall B \in \mathcal{B}(Y).$$

We denote by $\Pi(\mu,\nu)$ the set of all couplings between μ and ν. Note that this set is always non-empty. Indeed, we always have that $\mu \otimes \nu \in \Pi(\mu,\nu)$. At the same time, it is not always true that there exists a transport map between all μ and ν. Consider, for example, the case when μ is a Dirac measure and ν is not.

Remark 2.1. *Let $T : X \to Y$ be a transport map between $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then*

$$\pi_T := (Id \times T)_{\#}\mu \in \Pi(\mu,\nu).$$

Indeed, if $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$, then

$$\pi_T(A \times Y) = ((Id \times T)_{\#}\mu)(A \times Y) = \mu(A),$$

and

$$\pi_T(X \times B) = ((Id \times T)_{\#}\mu)(X \times B) = \nu(B).$$

In this manner, a transport map always induces a coupling.

Remark 2.2. *Let $\pi \in \Pi(\mu,\nu)$ be of the form $\pi = (Id \times S)_{\#}\mu$ where $S : X \to Y$ is some map. Then S is a transport map between μ and ν. Indeed, if $B \in \mathcal{B}(Y)$, then*

$$\nu(B) = \pi(X \times B) = ((Id \times S)_{\#}\mu)(X \times B) = S_{\#}\mu(B).$$

Now, having all necessary definitions, we can formulate Monge and Kantorovich problems rigorously. Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, +\infty]$ be a lower semicontinuous cost function. Then,

$$\begin{aligned} \textit{Monge problem: } & \inf\Big\{ \int_X c(x,T(x))d\mu(x) : \ T : X \to Y,\ T_{\#}\mu = \nu \Big\}, \\ \textit{Kantorovich problem: } & \inf\Big\{ \int_{X\times Y} c(x,y)d\pi(x,y) : \ \pi \in \Pi(\mu,\nu) \Big\}. \end{aligned} \tag{2}$$

Theorem 2.1. *(Existence of optimal coupling) Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, +\infty]$ be a lower semicontinuous function called cost function. Then there exists an optimal coupling for Kantorovich problem* (2)*.*

In Section 5 we will be working with quadratic cost $c(x,y) = \frac{|x-y|^2}{2}$ in spaces $X = Y = \mathbb{R}^d$, in this manner we state a couple results we will need. Quadratic cost is a non-negative and continuous function, and thus satisfies the conditions of Theorem 2.1. First, we show that by assuming that measures $\mu \in \mathcal{P}(\mathbb{R}^d)$, $\nu \in \mathcal{P}(\mathbb{R}^d)$ have finite second moments, studying the quadratic cost is equivalent to studying the cost $c(x,y) = -x \cdot y$. Let $\pi \in \Pi(\mu,\nu)$. Then

$$\begin{aligned} \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|x-y|^2}{2} d\pi(x,y) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \Big(\frac{|x|^2}{2} + \frac{|y|^2}{2} - x\cdot y\Big) d\pi(x,y) \\ &= \int_{\mathbb{R}^d} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^d} \frac{|y|^2}{2} d\nu(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y d\pi(x,y). \end{aligned}$$

Since the last two terms do not depend on the coupling π, we obtain the claimed equivalence of optimization problems.

Now, we are interested in the conditions that guarantee the uniqueness of the optimal coupling in (2) and the cases when optimal coupling is induced by an optimal transport map. Answers to these questions is known as *Brenier's theorem* and is formulated specifically for the quadratic transport cost, but can be generalized to more general costs with some additional work.

Theorem 2.2. *(Brenier's theorem) Let $X = Y = \mathbb{R}^d$, $c(x,y) = \frac{|x-y|^2}{2}$. Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and assume that $\mu \ll dx$. Then there exists a unique optimal transport map $T : \mathbb{R}^d \to \mathbb{R}^d$, called Brenier's map, such that $\nu = T_{\#}\mu$. Moreover, $T = \nabla\varphi$ with $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ a convex function called Brenier's potential.*

As a problem of convex optimization, (2) admits a dual problem. For this we introduce the notion of the *convex conjugate* of a function.

Definition 2.3. *(Convex conjugate) The convex conjugate of the function* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *is defined as*

$$\varphi^*(y) := \sup_{x \in \mathbb{R}^d} \{x \cdot y - \varphi(x)\}, \quad y \in \mathbb{R}^d.$$

If φ *is convex,* φ^* *is called Legendre transform of* φ*.*

Definition 2.4. *(Proper function) A convex function* $\varphi : \mathbb{R}^d \to \mathbb{R}$ *is called proper if*

$$\varphi(x) > -\infty, \quad \forall x \in \mathbb{R}^d,$$

and there exists $x_0 \in \mathbb{R}^d$ *such that*

$$\varphi(x_0) < +\infty.$$

Theorem 2.3. *(Fenchel–Moreau theorem, [19]) The biconjugate* φ^{**} *of* φ *is the greatest lower semi-continuous convex function such that*

$$\varphi^{**} \leq \varphi.$$

Moreover,

$$\varphi^{**} = \varphi \iff \varphi \textit{ is a proper, convex, lower semi-continuous function.}$$

We now state the dual representation of Kantorovich problem where the quadratic cost is considered again.

Theorem 2.4. *(Kantorovich duality) Let* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$*,* $\nu \in \mathcal{P}_2(\mathbb{R}^d)$*. Then for any* $\varphi, \psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *it holds*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|x-y|^2}{2} d\pi(x,y) = \max_{\varphi \ convex} \Big\{ \int_{\mathbb{R}^d} \Big(\frac{|x|^2}{2} - \varphi(x)\Big) d\mu(x) + \int_{\mathbb{R}^d} \Big(\frac{|y|^2}{2} - \varphi^*(y)\Big) d\nu(y) \Big\} \tag{3}$$

Last but not least, we recall the definition of *Wasserstein distance.*

Definition 2.5. *(Wasserstein p-distance) Let* $\mu \in \mathcal{P}_p(X)$*,* $\nu \in \mathcal{P}_p(Y)$*,* $p \geq 1$*. Wasserstein p-distance between* μ *and* ν *is defined as*

$$W_p(\mu,\nu) := \Big(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} |x-y|^p d\pi(x,y) \Big)^{\frac{1}{p}}.$$

Note that a transport map $T : X \to Y$ is optimal if

$$W_p(\mu,\nu) = \Big(\int_X |x - T(x)|^p d\mu(x) \Big)^{\frac{1}{p}}.$$

In this work we are only interested in cases $p = 1, 2$. In particular, in section 6 we will be working with W_1 distance, therefore we state the following useful result.

Proposition 2.1. *(Dual representation of Wasserstein 1-distance, [20]) Let* $\mu \in \mathcal{P}_1(X)$*,* $\nu \in \mathcal{P}_1(Y)$*. The dual form of Wassertein 1-distance has the following form:*

$$W_1(\mu,\nu) = \sup \Big\{ \int_X f(x) d(\mu - \nu)(x) : f : X \to \mathbb{R} \textit{ 1-Lipschitz} \Big\}.$$

3 Applications of Caffarelli's contraction theorem

For the moment, three approaches for proving Caffarelli's theorem are known. Originally, Caffarelli proved it exploiting the fact that Brenier's maps are solutions to a Monge-Ampère equation and by using maximum principle-type estimates. Another approach was described by Chewi and Pooladian in [5], and it consists in applying inequalities for covariance matrices: the Brascamp-Lieb inequality and the Cramér-Rao inequality. We are, however, interested in the third approach as it was briefly described in the introduction: formulating the problem as the entropy-regularized optimal transport problem and exploiting directly ideas from optimal transport theory. But first, let us see the applications of Caffarelli's contraction theorem.

3.1 Functional inequalities

First, we provide a few basic definitions from information theory.

Definition 3.1. *(Relative entropy) The relative entropy of $\mu \in \mathcal{P}(X)$ with respect to a measure ν defined on Y, $\mu \ll \nu$, is defined as*

$$H(\mu|\nu) := \int \log\Big(\frac{d\mu}{d\nu}\Big) d\mu.$$

Definition 3.2. *(Entropy) Entropy of $\mu \in \mathcal{P}(X)$ is a relative entropy of μ with respect to Lebesgue measure if $X = \mathbb{R}^d$ and with respect to the counting measure if $X = \mathbb{Z}^d$, $d \geq 1$, and is denoted by $H(\mu)$.*

Definition 3.3. *(Fisher information) Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ be such that $\mu \ll \nu$. Then Fisher information of μ with respect to ν is defined as*

$$I(\mu|\nu) := \int \Big|\nabla \log \frac{d\mu}{d\nu}\Big|^2 d\mu = \int \frac{|\nabla \frac{d\mu}{d\nu}|^2}{\frac{d\mu}{d\nu}} d\nu.$$

Logarithmic Sobolev inequality For all $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu \ll \gamma_d$ and $\frac{d\mu}{d\gamma_d}$ is smooth on $\mathbb{R}^d$, γ_d satisfies the *log-Sobolev inequality*:

$$H(\mu|\gamma_d) \leq \frac{1}{2} I(\mu|\gamma_d). \tag{4}$$

Denoting $f := \frac{d\mu}{d\gamma_d}$, we can rewrite (4) as

$$\int f \log f d\gamma_d \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\gamma_d. \tag{5}$$

One way to prove log Sobolev inequality involves exploiting optimal transport; it was proved by Dario in [6].

In the setup of Theorem 1.1, let $\mu(dx) = \gamma_d(dx)$, i.e. $V \equiv 0$, and $\nu \in \mathcal{P}(\mathbb{R}^d)$ be of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ where W is convex. Let $T = \nabla\varphi$ be an optimal map between γ_d and ν, i.e. $\nu = \nabla\varphi_{\#}\gamma_d$. Note that by Caffarelli's theorem we have the following bound on the Hessian of the Brenier potential:

$$0 \leq \mathrm{Hess}\varphi \leq 1. \tag{6}$$

Let us show how log-Sobolev inequality can be transported from γ_d to ν, i.e. is also true for measure ν. Let $\xi \in \mathcal{P}(\mathbb{R}^d)$ be such that $\xi \ll \nu$, and denote $g := \frac{d\xi}{d\nu}$. We apply log-Sobolev inequality (5) for the function $f(x) = g(\nabla\varphi(x))$ and use bound (6).

$$\begin{aligned}
H(\xi|\nu) &= \int \log\Big(\frac{d\xi}{d\nu}\Big) d\xi \\
&= \int g(x) \log g(x) d\nu(x) \\
&= \int g(\nabla\varphi(x)) \log g(\nabla\varphi(x)) d\gamma_d(x) \\
&\leq \int \frac{|\nabla g(\nabla\varphi(x))\mathrm{Hess}\varphi|^2}{g(\nabla\varphi(x))} d\gamma_d(x) \\
&\leq \int \frac{|\nabla g(\nabla\varphi(x))|^2}{g(\nabla\varphi(x))} d\gamma_d(x) \\
&= \int \frac{|\nabla g(x)|^2}{g(x)} d\nu(x) \\
&= I(\xi|\nu).
\end{aligned}$$

Poincaré Inequality ([2]) For all locally Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ the standard Gaussian measure γ_d satisfies the *Poincaré inequality*:

$$\mathrm{Var}_{\gamma_d}(f) = \int f^2 d\gamma_d - \Big(\int f d\gamma_d\Big)^2 \leq \int |\nabla f|^2 d\gamma_d. \tag{7}$$

Let us show that Poincaré inequality is also true for the measure ν defined as before. To show that measure ν satisfies Poincaré inequality for any locally Lipschitz function g, we apply (7) to the function $f(x) = g(\nabla\varphi(x))$ and use bound (6).

$$\begin{aligned} Var_\nu(g) = \int g^2 d\nu - \Big(\int g d\nu\Big)^2 &= \int g(\nabla\varphi(x))^2 d\gamma_d(x) - \Big(\int g(\nabla\varphi(x)) d\gamma_d(x)\Big)^2 \\ &\leq \int |\nabla g(\nabla\varphi(x))\mathrm{Hess}\varphi|^2 d\gamma_d(x) \\ &\leq \int |\nabla g(\nabla\varphi(x))|^2 d\gamma_d(x) \\ &= \int |\nabla g(x)|^2 d\nu(x). \end{aligned}$$

Caffarelli's theorem can be used for transporting many other functional inequalities, for example, Gaussian isoperimetric inequalities, Talagrand transport inequality, HWI inequality. For more details see, for example, [6].

4 Entropic optimal transport background

Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. We introduce a *reference measure* $R^\epsilon \in \mathcal{P}(X \times Y)$ such that $dR^\epsilon(x,y) \propto e^{\frac{-c(x,y)}{\epsilon}} d(\mu \otimes \nu)(x,y)$. The optimization problem of the type

$$\inf_{\pi\in\Pi(\mu,\nu)} \Big\{ \int_{X\times Y} c(x,y) d\pi(x,y) + \epsilon H(\pi|\mu\otimes\nu) \Big\} \tag{8}$$

is called an *entropically regularized optimal transport problem* with $\epsilon > 0$ a regularization parameter. Problem (8) is equivalent to the following optimization problem:

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R^\epsilon), \tag{9}$$

which is known as *static Schrödinger bridge problem*, where optimal π is *Schrödinger bridge* between μ and ν. For any $\pi \in \Pi(\mu,\nu)$, equivalence of (8) and (9) follows from the following identities:

$$\begin{aligned} H(\pi|R^\epsilon) = \int \log\Big(\frac{d\pi}{dR^\epsilon}\Big) d\pi &= \int \log\Big(\frac{d\pi}{e^{\frac{-c}{\epsilon}} d(\mu\otimes\nu)}\Big) d\pi + \mathrm{C} \\ &= H(\pi|\mu\otimes\nu) + \frac{1}{\epsilon}\int c(x,y) d\pi(x,y) + \mathrm{C}, \quad \mathrm{C} \geq 0. \end{aligned}$$

And so optimizing $H(\pi|R^\epsilon)$ over π is the same as optimizing $H(\pi|\mu\otimes\nu) + \frac{1}{\epsilon}\int c(x,y)d\pi(x,y)$ over π, or equivalently (8). The value of $\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R^\epsilon)$ we denote as $\mathcal{T}_H^\epsilon(\mu,\nu)$ and call *entropic transport cost*.

Adding a regularizing term in (8) has many practical advantages, such as, for example, getting a better convergence rate of computational algorithms for finding an optimal coupling as $\epsilon \to 0$. However, we are only interested in the properties of the optimal coupling that have been discovered by studying Schrödinger problem (9) ([14]).

As before, let R^ϵ be a reference measure. Now we impose the following conditions on R^ϵ:

$$\mu\otimes\nu \ll R^\epsilon \quad \text{and} \quad \log\frac{d(\mu\otimes\nu)}{dR^\epsilon} \in L^1(\mu\otimes\nu).$$

Equivalently, R^ϵ should be such that $H(\mu\otimes\nu|R^\epsilon) < +\infty$. Under these conditions the following theorem is true.

Theorem 4.1. *1. There exists a unique coupling $\pi^\epsilon \in \Pi(\mu,\nu)$ such that*

$$\mathcal{T}_H^\epsilon(\mu,\nu) = H(\pi^\epsilon|R^\epsilon).$$

2. *There exist two functions $f^\epsilon : X \to \mathbb{R}_+$, $g^\epsilon : Y \to \mathbb{R}_+$, unique up to a multiplicative constant, such that*

$$\pi^\epsilon(dxdy) = f^\epsilon(x)g^\epsilon(y)R^\epsilon(dxdy). \tag{10}$$

In particular, functions f^ϵ, g^ϵ are of the form $f^\epsilon = e^\varphi$, $g^\epsilon = e^\psi$, where $\varphi : X \to \mathbb{R}$, $\psi : Y \to \mathbb{R}$ are Schrödinger potentials.

We also state the following useful lemma about duality form of the entropic transport cost.

Lemma 4.1. *Let β be a σ-finite positive measure on a measurable space $(X, \mathcal{A})$. For all measurable bounded from above functions $h : X \to \mathbb{R}$ it holds*

$$\sup_{\substack{\alpha\in\mathcal{P}(X),\\ H(\alpha|\beta)<+\infty}} \left\{ \int h d\alpha - H(\alpha|\beta) \right\} = \log \int e^h d\beta.$$

5 Proof of Caffarelli's theorem using EOT

We start with the Prékopa-Leindler theorem and its corollary that will be an important tool for proving results in this section.

5.1 Prékopa-Leindler theorem

Theorem 5.1. *(Prékopa-Leindler inequality) Let $\omega, u, v : \mathbb{R}^d \to \mathbb{R}_+$ be measurable functions such that for any $\theta \in [0,1]$ and $x, y \in \mathbb{R}^d$ it holds*

$$\omega(\theta x + (1-\theta)y) \geq u(x)^\theta v(y)^{1-\theta}. \tag{11}$$

Then

$$\int_{\mathbb{R}^d} \omega dx \geq \Big(\int_{\mathbb{R}^d} u dx \Big)^\theta \Big(\int_{\mathbb{R}^d} v dx \Big)^{1-\theta}. \tag{12}$$

Corollary 5.1. *Let $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ be a log-concave function. Then function*

$$F(x) := \int_{\mathbb{R}^d} f(x,y)dy, \quad x \in \mathbb{R}^d,$$

is also log-concave.

Proof. In Theorem 5.1, take

$$\begin{aligned} \omega(y) &:= f(\theta x_1 + (1-\theta)x_2, y), \\ u(y) &:= f(x_1, y), \\ v(y) &:= f(x_2, y), \end{aligned} \tag{13}$$

for $x_1, x_2 \in \mathbb{R}^d$, $\theta \in [0,1]$. By assumption, f is a log-concave function, i.e.

$$f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \geq f(x_1,y_1)^\theta f(x_2,y_2)^{1-\theta}, \quad x_1, x_2, y_1, y_2 \in \mathbb{R}^d.$$

In notation of the previous theorem it reads as

$$\omega(\theta y_1 + (1-\theta)y_2) \geq u(y_1)^\theta v(y_2)^{1-\theta},$$

which is condition (11). Thus, by (12) and (13),

$$\int_{\mathbb{R}^d} f(\theta x_1 + (1-\theta)x_2, y)dy \geq \Big(\int_{\mathbb{R}^d} f(x_1,y)dy \Big)^\theta \Big(\int_{\mathbb{R}^d} f(x_2,y)dy \Big)^{1-\theta},$$

and so

$$F(\theta x_1 + (1-\theta)x_2) \geq F(x_1)^\theta F(x_2)^{1-\theta}, \quad x_1, x_2 \in \mathbb{R}^d.$$

This completes the proof. □

5.2 Main theorems and ideas of EOT approach

We start with the following useful definition.

Definition 5.1. *(Domination of a measure in convex order) Let $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$. We say that η is dominated by ν in convex order if for any convex function f we have $\int f d\eta \le \int f d\nu$, and we write $\eta \le_c \nu$.*

Next theorem is one of the main results in order to prove Caffarelli's theorem via entropic optimal transport.

Theorem 5.2. *(Gozlan, Juillet, [10]) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Then the following assertions are equivalent:*

1. *There exists a continuously differentiable and convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\varphi$ is 1-Lipschitz and $\nu = \nabla\varphi_{\#}\mu$;*
2. *For all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\eta \le_c \nu$,*
$$W_2(\mu, \nu) \le W_2(\mu, \eta).$$

First, we will prove statement 2 of Theorem 5.2. Here, the main idea is to use entropic optimal transport theory. For this, we should choose a "good" reference measure that will allow us to get more information about Schrödinger potentials in (10) and about entropic transport cost. We will use the convergence of scaled by ϵ entropic transport cost to $\frac{1}{2}W_2^2$ as $\epsilon \to 0$ (proved in [4]) to obtain the monotonicity property of W_2 distance. Then, we will explain why one can expect the equivalence of monotonicity of W_2 and Caffarelli's theorem. Due to technicality of the original proof, here we present only key propositions and theorems. We also write in more detail those proofs that will guarantee a better understanding of Section 6. For full proof see [8].

Consider the Ornstein-Uhlenbeck process $(Z_t)_{t\ge 0}$ defined by

$$dZ_t = -\frac{1}{2}Z_t dt + dB_t, \quad t \ge 0,$$

where $(B_t)_{t\ge 0}$ is a standard d-dimensional Brownian motion and $Z_0 \sim \gamma_d$. By Itô's formula it can be shown that

$$Z_t = Z_0 e^{-\frac{t}{2}} + e^{-\frac{t}{2}} \int_0^t e^{\frac{s}{2}} dW_s, \quad t \ge 0.$$

Then as a reference measure we choose the law of (Z_0, Z_ϵ):

$$R^\epsilon = Law(Z_0, Z_\epsilon) = Law(X, Xe^{-\frac{\epsilon}{2}} + \sqrt{1 - e^{-\epsilon}}Y), \quad X, Y \sim \gamma_d,\ \epsilon > 0.$$

In other words,

$$R^\epsilon(dxdy) = \gamma_d(dx) r_x^\epsilon(dy),$$

where $r_x^\epsilon = \mathcal{N}(xe^{-\frac{\epsilon}{2}}, (1 - e^{-\epsilon})I_d)$.

The semigroup of the Orstein-Uhlenbeck process is given by

$$\begin{aligned}
\mathcal{P}_\epsilon\varphi(x) &= \mathbb{E}[\varphi(Z_\epsilon)|Z_0 = x] \\
&= \int_{\mathbb{R}^d} \varphi(y) r_x^\epsilon(dy) \\
&= \int_{\mathbb{R}^d} \varphi(y + e^{-\frac{\epsilon}{2}}x) r_0^\epsilon(dy) \\
&= \frac{1}{(2\pi(1 - e^{-\epsilon}))^{\frac{d}{2}}} \int_{\mathbb{R}^d} \varphi(y + e^{-\frac{\epsilon}{2}}x) e^{-\frac{|y|^2}{2(1-e^{-\epsilon})}} dy, \quad x \in \mathbb{R}^d,
\end{aligned} \tag{14}$$

for any non-negative measurable function φ. In the next two propositions we formulate important properties of the Orstein-Uhlenbeck semigroup.

Proposition 5.1. *1. If $f : \mathbb{R}^d \to \mathbb{R}$ is log-convex, then $\mathcal{P}_\epsilon f$ is also log-convex.*

2. If $g : \mathbb{R}^d \to \mathbb{R}$ is log-concave, then $\mathcal{P}_\epsilon g$ is also log-concave.

Proof. 1. Log-convexity follows from Hölder inequality and log-convexity of f:

$$\begin{aligned}
\mathcal{P}_\epsilon f(\theta x_1 + (1-\theta)x_2) &= \int_{\mathbb{R}^d} f(\theta(y + e^{-\frac{\epsilon}{2}}x_1) + (1-\theta)(y + e^{-\frac{\epsilon}{2}}x_2)) r_0^\epsilon(dy) \\
&\le \int_{\mathbb{R}^d} f(y + e^{-\frac{\epsilon}{2}}x_1)^\theta f(y + e^{-\frac{\epsilon}{2}}x_2)^{1-\theta} r_0^\epsilon(dy) \\
&\le \Big(\int_{\mathbb{R}^d} f(y + e^{-\frac{\epsilon}{2}}x_1) r_0^\epsilon(dy)\Big)^\theta \Big(\int_{\mathbb{R}^d} f(y + e^{-\frac{\epsilon}{2}}x_2) r_0^\epsilon(dy)\Big)^{1-\theta} \\
&= \mathcal{P}_\epsilon f(x_1)^\theta \mathcal{P}_\epsilon f(x_2)^{1-\theta}, \quad \forall x_1, x_2 \in \mathbb{R}^d.
\end{aligned}$$

2. Log-concavity follows from the corollary of Prékopa-Leindler theorem 5.1. Since g is log-concave by assumption, we have that

$$h(x,y) := g(y + e^{-\frac{\epsilon}{2}}x) e^{-\frac{|y|^2}{2(1-e^{-\epsilon})}}$$

is also log-concave for any $x, y \in \mathbb{R}^d$, $\epsilon > 0$. Then, by Theorem 5.1, $\int_{\mathbb{R}^d} h(x,y)dy$ is log-concave, which is equivalent to the log-concavity of

$$\mathcal{P}_\epsilon g(x) = \frac{1}{(2\pi(1-e^{-\epsilon}))^{\frac{d}{2}}} \int_{\mathbb{R}^d} h(x,y)dy.$$

□

Proposition 5.2. *If $\mu(dx) = e^{V(x)}\gamma_d(dx)$, $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ and f^ϵ, g^ϵ are such that (10) is satisfied, then*

$$f^\epsilon(x)\mathcal{P}_\epsilon g^\epsilon(x) = e^{V(x)}, \quad g^\epsilon(y)\mathcal{P}_\epsilon f^\epsilon(y) = e^{-W(y)}, \quad \forall x, y \in \mathbb{R}^d.$$

Proof. Let $h : \mathbb{R}^d \to \mathbb{R}$ be a test function. By definition of coupling,

$$\iint h(x)\pi^\epsilon(dxdy) = \int h(x)\mu(dx), \tag{15}$$

and

$$\iint h(y)\pi^\epsilon(dxdy) = \int h(y)\nu(dy). \tag{16}$$

By (10) and (14),

$$\begin{aligned}
\iint h(x)\pi^\epsilon(dxdy) &= \iint h(x)f^\epsilon(x)g^\epsilon(y)R^\epsilon(dxdy) \\
&= \iint h(x)f^\epsilon(x)g^\epsilon(y)\gamma_d(dx)r_x^\epsilon(dy) \\
&= \int \Big(\int g^\epsilon(y)r_x^\epsilon(dy)\Big) h(x)f^\epsilon(x)\gamma_d(dx) \\
&= \int h(x)f^\epsilon(x)\mathcal{P}_\epsilon g^\epsilon(x)\gamma_d(dx).
\end{aligned}$$

Comparing last integral with (15) and using the representation of μ, we get $\mu(dx) = f^\epsilon(x)\mathcal{P}_\epsilon g^\epsilon(x)\gamma_d(dx)$, thus

$$f^\epsilon(x)\mathcal{P}_\epsilon g^\epsilon(x) = e^{V(x)}.$$

Similarly for the measure ν:

$$\begin{aligned}
\iint h(y)\pi^\epsilon(dxdy) &= \iint h(y)f^\epsilon(x)g^\epsilon(y)R^\epsilon(dydx) \\
&= \iint h(y)f^\epsilon(x)g^\epsilon(y)\gamma_d(dy)r_y^\epsilon(dx) \\
&= \int \Big(\int f^\epsilon(x)r_y^\epsilon(dx)\Big) h(y)g^\epsilon(y)\gamma_d(dy) \\
&= \int h(y)g^\epsilon(y)\mathcal{P}_\epsilon f^\epsilon(y)\gamma_d(dy).
\end{aligned}$$

Comparing last integral to (16) we get that $g^\epsilon(y)\mathcal{P}_\epsilon f^\epsilon(y)\gamma_d(dy) = \nu(dy) = e^{-W(y)}\gamma_d(dy)$, thus

$$g^\epsilon(y)\mathcal{P}_\epsilon f^\epsilon(y) = e^{-W(y)}.$$

$\square$

From Proposition 6.3 it follows that

$$g^\epsilon(y) = \frac{e^{-W(y)}}{\mathcal{P}_\epsilon f^\epsilon(y)}, \quad \forall y \in \mathbb{R}^d,$$

and thus

$$f^\epsilon(x) = \frac{e^{V(x)}}{\mathcal{P}_\epsilon\left(\frac{e^{-W(x)}}{\mathcal{P}_\epsilon f^\epsilon(x)}\right)}, \quad \forall x \in \mathbb{R}^d.$$

We define a functional

$$\Phi(\varphi)(x) := \frac{e^{V(x)}}{\mathcal{P}_\epsilon\left(\frac{e^{-W(x)}}{\mathcal{P}_\epsilon \varphi(x)}\right)}, \quad x \in \mathbb{R}^d.$$

We note that f^ϵ is a fixed point of Φ, and Φ preserves log-convexity (follows from the Proposition 5.1 and definition of log-convexity).

Now, we prove one of the main results about Schrödinger potentials that is the key for proving Theorem 5.2.

Theorem 5.3. *Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ be of the form $\mu(dx) = e^{V(x)}\gamma_d(dx)$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$ be of the form $\nu(dx) = e^{-W(x)}\gamma_d(dx)$ with $V, W : \mathbb{R}^d \to \mathbb{R}$ convex functions. Then there exist a log-convex function $f^\epsilon : \mathbb{R}^d \to [1, +\infty)$ and a log-concave function $g^\epsilon : \mathbb{R}^d \to [0, +\infty)$ such that the unique optimal coupling $\pi^\epsilon \in \Pi(\mu, \nu)$ is of the form $\pi^\epsilon(dxdy) = f^\epsilon(x)g^\epsilon(y)R^\epsilon(dxdy)$. Moreover, $\log f^\epsilon \in L^1(\mu)$, $\log g^\epsilon \in L^1(\nu)$ and*

$$\mathcal{T}_H^\epsilon(\mu, \nu) = H(\pi^\epsilon | R^\epsilon) = \int \log f^\epsilon d\mu + \int \log g^\epsilon d\nu.$$

Proof. From Proposition 4.1, it is known that functions f^ϵ, g^ϵ that satisfy (10) exist. We have to show that they are log-convex and log-concave respectively. Since f^ϵ and g^ϵ are related to each other through (27) and $\mathcal{P}_\epsilon$ preserves log-convexity, it is enough to show that f^ϵ is log-convex. Since it is equivalent to showing that $\log f^\epsilon$ is convex, we will show that $\log f^\epsilon = (\log f^\epsilon)^{**}$ (Theorem 2.3).

Define $h_0 = (\log f^\epsilon)^{**}$ and $f_0 = e^{h_0}$. We define an iterative process for f_0 as

$$f_1 = \max(\Phi(f_0), f_0).$$

By construction, f_0 is the greatest log-convex function below f^ϵ, i.e.

$$f_0 \leq f^\epsilon. \tag{17}$$

Since $\max(\Phi(f_0), f_0) \geq f_0$,

$$f_0 \leq f_1. \tag{18}$$

Since Φ preserves log-convexity, we get that f_1 is also log-convex. By monotonicity of Φ, the fact that f^ϵ is a fixed point of Φ and (17), we get $\max(\Phi(f_0), f_0) \leq \max(\Phi(f^\epsilon), f^\epsilon) = f^\epsilon$, thus

$$f_1 \leq f^\epsilon,$$

and since f_0 is the greatest log-convex function below f^ϵ, we have:

$$f_1 \leq f_0. \tag{19}$$

From (18) and (19) we get that $f_0 = f_1$ and thus,

$$f_0 = \max(\Phi(f_0), f_0) \geq \Phi(f_0).$$

Exploiting the definition of the functional Φ and symmetry of the semigroup operator in $L^2(\gamma_d)$ we write

$$1 \leq \int \frac{f_0}{\Phi(f_0)} d\mu = \int f_0 \mathcal{P}_\epsilon\Big(\frac{e^{-W}}{\mathcal{P}_\epsilon f_0}\Big) d\gamma_d = \int \mathcal{P}_\epsilon f_0 \frac{e^{-W}}{\mathcal{P}_\epsilon f_0} d\gamma_d = \int e^{-W} d\gamma_d = 1,$$

since $e^{-W} d\gamma_d$ is a probability measure on $\mathbb{R}^d$. And so $\Phi(f_0) = f_0$ almost everywhere but since Φ is continuous, $\Phi(f_0) = f_0$ everywhere. Since $f_0 \leq f^\epsilon$ is the convex envelope of f^ϵ, there exists $x^* \in \mathbb{R}^d$ such that

$$f_0(x^*) = f^\epsilon(x^*). \tag{20}$$

Recall that by Proposition 4.1 a pair of functions (f, g) that represent a unique optimal coupling $\pi^\epsilon = fgdR^\epsilon$ is always of the form $f = \lambda f^\epsilon$, $g = \frac{g^\epsilon}{\lambda}$, $\lambda \in \mathbb{R}$. From (20) we find that $\lambda = 1$, therefore, $f_0(x) = f^\epsilon(x)$, $\forall x \in \mathbb{R}^d$, and $\log f^\epsilon = (\log f^\epsilon)^{**}$. This concludes the proof. □

Now, knowing the representation of the entropic transport cost, we can prove its monotonicity property that formulates as follows.

Proposition 5.3. *Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\nu \in \mathcal{P}(\mathbb{R}^d)$. Then for any $\eta \in \mathcal{P}(\mathbb{R}^d)$ such that $\eta \leq_c \nu$ it holds*

$$\mathcal{T}_H^\epsilon(\mu, \nu) \leq \mathcal{T}_H^\epsilon(\mu, \eta), \quad \forall \epsilon > 0.$$

Proof. First, we note that from Lemma 4.1 under the given assumptions on h and β it follows that for any $\alpha \in \mathcal{P}(\mathbb{R}^d)$

$$H(\alpha|\beta) \geq \int h d\alpha - \log \int e^h d\beta.$$

Let $\pi_2 \in \Pi(\mu, \eta)$ be some coupling between μ and η, $\eta \leq_c \nu$. Take $\alpha = \pi_2$, $\beta = R^\epsilon$, $h = \log(f^\epsilon g^\epsilon)$, and let $\pi_1^* \in \Pi(\mu, \nu)$ be an optimal coupling between μ and ν. Noting that $\int f^\epsilon(x) g^\epsilon(y) dR^\epsilon(x, y) = \int d\pi_2^*(x, y) = 1$, we write

$$\begin{aligned}
H(\pi_2|R^\epsilon) &\geq \int \log(f^\epsilon(x) g^\epsilon(y)) d\pi_2(x, y) - \log \int f^\epsilon(x) g^\epsilon(y) dR^\epsilon(x, y) \\
&= \int \log f^\epsilon(x) d\pi_2(x, y) + \int \log g^\epsilon(y) d\pi_2(x, y) \\
&= \int \log f^\epsilon(x) d\mu(x) + \int \log g^\epsilon(y) d\eta(y) \\
&\geq \int \log f^\epsilon(x) d\mu(x) + \int \log g^\epsilon(y) d\nu(y) \\
&= H(\pi_1^*|R^\epsilon) \\
&= \mathcal{T}_H^\epsilon(\mu, \nu).
\end{aligned}$$

For the second inequality we used the fact that $\log g^\epsilon$ is concave and $\eta \leq_c \nu$. Since $\mathcal{T}_H^\epsilon(\mu, \nu)$ is independent of π_2, we take infimum over all couplings $\pi_2 \in \Pi(\mu, \eta)$ in $H(\pi_2|R^\epsilon)$, and obtain

$$\mathcal{T}_H^\epsilon(\mu, \nu) \leq \mathcal{T}_H^\epsilon(\mu, \eta), \quad \forall \epsilon > 0.$$

□

Last but not least, we have the following convergence result.

Theorem 5.4. *(Carlier et al., [4]) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $H(\mu|\gamma_d) < +\infty$, $H(\nu|\gamma_d) < +\infty$. Then,*

$$\epsilon \mathcal{T}_H^\epsilon(\mu, \nu) \xrightarrow[\epsilon \to 0]{} \frac{1}{2} W_2^2(\mu, \nu).$$

Idea of proof. For any $\pi \in \Pi(\mu, \nu)$ we have,

$$\begin{aligned}\epsilon H(\pi|R^\epsilon) &= \epsilon \int \log\Big(\frac{d\pi}{dx}\Big) d\pi - \epsilon \int \log\Big(\frac{dR^\epsilon}{dx}\Big) d\pi \\ &= \epsilon \int \log\Big(\frac{d\pi}{dx}\Big) d\pi + \frac{\epsilon}{2(1-e^{-\epsilon})} \int |y - e^{-\frac{\epsilon}{2}} x|^2 d\pi(x,y) + \frac{\epsilon}{2} \int |x|^2 d\mu(x) - \epsilon \log((2\pi)^d \det((1-e^{-\epsilon})^{\frac{d}{2}} I_d)) \\ &\xrightarrow[\epsilon \to 0]{} \frac{1}{2} \int |y - x|^2 d\pi(x,y).\end{aligned}$$

So, minimizing $H(\pi|R^\epsilon)$ over all π is the same that minimizing $\frac{1}{2}\int |x-y|^2 d\pi(x,y)$. Rigorously, one has to prove Γ-convergence which is done in [4]. □

From Theorem 5.4 and Proposition 5.3 it follows that for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and for all $\eta \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\eta \leq_c \nu$,

$$W_2(\mu, \nu) \leq W_2(\mu, \eta).$$

Thus, we have proved statement 2 of Theorem 5.2.

Sketch of an alternative proof Recall that by Kantorovich duality (3),

$$\frac{1}{2} W_2^2(\mu, \nu) = \int_{\mathbb{R}^d} \Big(\frac{|x|^2}{2} - \varphi(x)\Big) d\mu(x) + \int_{\mathbb{R}^d} \Big(\frac{|y|^2}{2} - \varphi^*(y)\Big) d\nu(y),$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ is a convex function such that the optimal transport map is given by $T = \nabla\varphi$. Then since by Theorem 5.3 the optimal entropic cost is given by

$$\mathcal{T}_H^\epsilon(\mu, \nu) = \int \log f^\epsilon d\mu + \int \log g^\epsilon d\nu,$$

from Theorem 5.4 we can expect that

$$\epsilon \log f^\epsilon \xrightarrow[\epsilon \to 0]{} \frac{|x|^2}{2} - \varphi(x), \quad \forall x \in \mathbb{R}^d, \tag{21}$$

up to some additive constant. But by Theorem 5.3 $\log f^\epsilon$ is convex and convexity is preserved by pointwise convergence, and so we expect $\frac{|x|^2}{2} - \varphi(x)$ to be convex. Convexity of $\frac{|x|^2}{2} - \varphi(x)$ means

$$\text{Hess}\varphi \leq 1.$$

Thus, $\nabla\varphi$ is 1-Lipschitz and we get Caffarelli's contraction theorem.

6 Towards Caffarelli's theorem for Poisson distribution

The goal of this section is to develop the same method as in Section 5 for Poisson measure in dimension 1. For this, we need to introduce a stronger notion of log-concavity for discrete functions – *ultra log-concavity.* We will prove interesting properties of the $M/M/\infty$ process that later on will be used for applying EOT theory. In particular, we will prove convergence of regularized entropic cost and monotonicity properties of W_1 distance under specific assumptions in the discrete set-up.

Let ξ be the counting measure on $\mathbb{Z}_+$. We define the Poisson measure $\rho \in \mathcal{P}(\mathbb{Z}_+)$ with parameter $\lambda > 0$ as

$$\rho^\lambda(dn) = \frac{\lambda^n e^{-\lambda}}{n!} \xi(dn).$$

In other words, ρ^λ has a Radon-Nikodym derivative with respect to ξ that is equal to the probability mass function of the Poisson distribution. For $\lambda = 1$ we set $\rho := \rho^1$.

6.1 Convexity of discrete functions

We first recall the definition of convex (concave) and log-convex (log-concave) functions in $\mathbb{Z}_+$ as well as introduce a definition of an ultra log-convex (ultra log-concave) function.

Definition 6.1. *A function $f : \mathbb{Z}_+ \to \mathbb{R}$ is called convex if*

$$f(n) \leq \frac{1}{2}(f(n+1) + f(n-1)), \quad n \geq 1.$$

A function $f : \mathbb{Z}_+ \to \mathbb{R}$ is concave if $-f$ is convex.

Definition 6.2. *A sequence of non-negative numbers $(u(n))_{n\in\mathbb{Z}_+}$ is called*

- *log-convex if*

$$u(n) \leq \big(u(n+1)u(n-1)\big)^{\frac{1}{2}}, \quad \forall n \geq 1,$$

- *log-concave if*

$$u(n) \geq \big(u(n+1)u(n-1)\big)^{\frac{1}{2}}, \quad \forall n \geq 1,$$

- *ultra log-concave (ULC) if $(n!u(n))_{n\in\mathbb{Z}_+}$ is log-concave, i.e.*

$$u(n) \geq \Big(\Big(1 + \frac{1}{n}\Big)u(n+1)u(n-1)\Big)^{\frac{1}{2}}, \quad \forall n \geq 1,$$

  or equivalently if $\big(u(n)/\frac{d\rho^\lambda}{d\xi}(n)\big)_{n\in\mathbb{Z}_+}$ is a log-concave sequence $\forall\lambda \geq 1$, where $\frac{d\rho^\lambda}{d\xi}$ is density of the Poisson measure,

- *ultra log-convex (ULCX) if $(n!u(n))_{n\in\mathbb{Z}_+}$ is log-convex.*

Remark 6.1. *We say that a random variable is ultra log-concave (convex) if its probability mass function defines an ultra log-concave (convex) sequence.*

For proving the results of this section, we have to introduce a conjugate function of a discrete function. We use results introduced and proved by Murota in [16]. For arbitrary functions $f : \mathbb{Z}_+ \to \mathbb{R}\cup\{+\infty\}$, $g : \mathbb{Z}_+ \to \mathbb{R}\cup\{-\infty\}$ such that $\mathrm{dom} f, \mathrm{dom} g \neq \emptyset$ we define

$$\begin{aligned} f^{\triangle}(m) &= \sup_{n\in\mathbb{Z}_+}\{f(n) + nm\}, \quad m \in \mathbb{R}, \\ g^{\triangledown}(m) &= \inf_{n\in\mathbb{Z}_+}\{g(n) - nm\}, \quad m \in \mathbb{R}, \end{aligned} \tag{22}$$

where $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$.

Theorem 6.1. *(Murota, [16]) For a function $f : \mathbb{Z}_+ \to \mathbb{R}$ that is discretely convex, the conjugate function $f^{\triangle} : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ is discretely concave, and*

$$(f^{\triangle})^{\triangledown}(n) = f(n), \quad n \in \mathbb{Z}_+.$$

Remark 6.2. *Note that $(f^{\triangle})^{\triangledown}$ is the greatest convex function that bounds f from below. Indeed, from (22),*

$$f^{\triangle}(m) \leq f(n) + nm, \quad \forall n \in \mathbb{Z}_+, m \in \mathbb{R},$$

or equivalently,

$$f(n) \geq f^{\triangle}(m) - nm, \quad \forall n \in \mathbb{Z}_+, m \in \mathbb{R}.$$

Since left-hand side does not depend on m, we can optimize over $m \in \mathbb{R}$. Thus, we obtain

$$f(n) \geq \sup_{m\in\mathbb{R}}\{f^{\triangle}(m) - nm\} = (f^{\triangle})^{\triangledown}(n), \quad \forall n \in \mathbb{Z}_+.$$

6.2 $M/M/\infty$ process and its properties

We consider the $M/M/\infty$ process on $\mathbb{Z}_+$ that describes a queue where customers arrive according to a Poisson point process with intensity $\lambda = 1$. The process is characterized in the following way:

- Arrival time of i-th client T_i, $i \geq 1$, is a Poisson point process with parameter $\lambda = 1$;
- The service time S_i of i-th client follows exponential distribution with parameter $\mu > 0$, $S_i \perp\!\!\!\perp S_j$ and $S_i \perp\!\!\!\perp T_j$, $i, j \geq 1$;
- There is an infinite amount of servers, thus there is no waiting time in the queue and clients are served immediately;
- At time $t = 0$ there are $N_0 \sim \mu$ clients being served, where $\mu \in \mathcal{P}(\mathbb{Z}_+)$;
- The service time $(R_i)_{i\geq 1}$, of the clients at time $t = 0$ is exponentially distributed, and $R_i \perp\!\!\!\perp S_j$, $R_i \perp\!\!\!\perp T_j$, $i, j \geq 1$.

By N_t we denote the number of clients being served at time $t \geq 0$, and by $A_t = \sum_{i=1}^{+\infty} 1_{t\leq T_i}$ we denote the number of arrivals at time $t \geq 0$. Then the $M/M/\infty$ process is defined as

$$N_t = \sum_{i=0}^{N_0} 1_{t\leq R_i} + \sum_{i=1}^{A_t} 1_{t\leq T_i+S_i}.$$

It can be shown that the law of N_t can be represented as a convolution of Bernoulli distribution with parameter e^{-t} and Poisson distribution with parameter $1 - e^{-t}$. In other words,

$$N_t \overset{law}{=} \sum_{i=1}^{N_0} X_i + Y, \quad t \geq 0, \tag{23}$$

with initial law $N_0 \sim \mu$, where $X_i \sim \mathit{Bernoulli}(e^{-t})$, $Y \sim \mathit{Poisson}(1 - e^{-t})$, and $X_i \perp\!\!\!\perp Y$, $X_i \perp\!\!\!\perp X_j$, $i, j \geq 1$. In the scope of the report, we will be working with the representation of N_t given by (23).

Reference measure for EOT We make the following remark that will be useful later. Let $N_0 \sim \rho$. Then as a reference measure R^ϵ in (9) we choose the joint law of (N_0, N_ϵ), and it is given by

$$R^\epsilon(dndm) = \rho(dn)r_n^\epsilon(dm),$$

where $r_n^\epsilon = \mathit{Binomial}(n, e^{-\epsilon}) * \mathit{Poisson}(1 - e^{-\epsilon})$, $\epsilon > 0$.

Definition 6.3. *(α-thinning) Let $Y \sim \mu$ be a $\mathbb{Z}_+$-valued random variable, and $X_i \sim \mathit{Bernoulli}(\alpha)$ for $\alpha \in (0,1)$, $1 \leq i \leq n$, $n \geq 1$. The law of the sum $\sum_{i=1}^{Y} X_i$ is called an α-thinning of the random variable Y. We denote it as $T_\alpha\mu$ and it has the following explicit form,*

$$T_\alpha\mu(k) = \sum_{j=k}^{\infty} \mathbb{P}(Y = j)\binom{j}{k}\alpha^k(1-\alpha)^{j-k}. \tag{24}$$

Remark 6.3. *Define operator $E_p d(k) = d(k)p^k$ for any non-negative sequence $(d(k))_{k\geq 1}$. Then (24) can be rewritten as*

$$T_\alpha\mu = E_{\frac{\alpha}{1-\alpha}} \circ T \circ E_\alpha\mu,$$

where $T\mu(k) = \sum_{j=k}^{\infty} \binom{j}{k} \mathbb{P}(Y=j)$. *Indeed,*

$$\begin{aligned} E_{\frac{\alpha}{1-\alpha}} \circ T \circ E_{1-\alpha}\mu(k) &= E_{\frac{\alpha}{1-\alpha}} \circ T\mu(k)(1-\alpha)^k \\ &= E_{\frac{\alpha}{1-\alpha}} \Big(\sum_{j=k}^{\infty} \binom{j}{k} \mathbb{P}(Y=j)(1-\alpha)^j \Big) \\ &= \sum_{j=k}^{\infty} \mathbb{P}(Y=j) \binom{j}{k} \frac{\alpha^k}{(1-\alpha)^k} (1-\alpha)^j \\ &= \sum_{j=k}^{\infty} \mathbb{P}(Y=j) \binom{j}{k} \alpha^k (1-\alpha)^{(j-k)} \\ &= T_\alpha \mu(k). \end{aligned}$$

Remark 6.4. *We note that the law of the M/M/∞ process can be then written as*

$$\begin{aligned} Law(N_t) &= T_{e^{-t}}\rho * \rho^{1-e^{-t}} \\ &= \Big(E_{\frac{e^{-t}}{1-e^{-t}}} \circ T \circ E_{e^{-t}}\rho \Big) * \rho^{1-e^{-t}}. \end{aligned}$$

For proving our next results we will be using the facts that log-concavity and ultra log-concavity are preserved by convolution. While first result is rather simple, the demonstration of the second result is very technical, and thus is omitted in this report. It was first stated and proved by Walkup, see [17] or [15] for details. We formulate both theorems.

Theorem 6.2. *Let $(a_i)_{i\geq 0}$, $(b_j)_{j\geq 0}$ be two log-concave sequences. Then their convolution $(c_k)_{k\geq 0}$, defined by $c_k = (a*b)(k) = \sum_{k=i+j} a_i b_j$, $k \geq 0$, is also log-concave.*

Proof. For fixed $1 \leq i \leq j-1$ log-concavity of $(a_i)_{i\geq 0}$, $(b_j)_{j\geq 0}$, we have

$$a_i a_{j-1} \geq a_{i-1} a_j, \tag{25}$$

and

$$b_j b_{i+1} \geq b_{j+1} b_i, \quad 1 \leq i \leq j-1. \tag{26}$$

Indeed, multiply inequalities $a_k^2 \geq a_{k-1}a_{k+1}$ for $k = i, ..., j-1$:

$$a_i^2 a_{i+1}^2 ... a_{j-1}^2 a_{j-1}^2 \geq a_{i-1} a_i a_{i+1} ... a_{j-2} a_{i+1} ... a_{j-1} a_j.$$

Now, divide both sides of the inequality by $a_i a_{i-1}^2 ... a_{j-2}^2 a_{j-1}$, and we get exactly (25). Same for (26). Therefore, from (25) and (26) we get,

$$\sum_{j>i} (a_i a_{j-1} - a_{i-1} a_j)(b_{k-i} b_{k-j+1} - b_{k-i+1} b_{k-j}) \geq 0, \quad i \geq 1.$$

Multiplying out this expression, and interchanging the roles of i and j in two of the four resulting terms, we obtain

$$\begin{aligned} &\sum_{i\neq j} a_i a_{j-1} b_{k-i} b_{k-j+1} - \sum_{i\neq j} a_i a_{j-1} b_{k-i+1} b_{k-j} \\ =&\sum_{i\neq j} (a_i a_{j-1} b_{k-i} b_{k-j+1} + a_i a_{i-1} b_{k-i} b_{k-i+1}) - \sum_{i\neq j} (a_i a_{j-1} b_{k-i+1} b_{k-j} + a_i a_{i-1} b_{k-i+1} b_{k-i}) \\ =&\sum_{i=0}^{j-1} a_i a_{j-1} b_{k-i} b_{k-j+1} - \sum_{i=1}^{j-1} a_i a_{j-1} b_{k-i+1} b_{k-j}, \quad j \leq k+1. \end{aligned}$$

Summing up for all $j \geq 1$, we obtain

$$\begin{aligned}&\sum_{j=1}^{k+1}\sum_{i=0}^{j-1} a_i a_{j-1} b_{k-i} b_{k-j+1} + \sum_{j=1}^{k+1}\sum_{i=0}^{j-1} a_i a_{j-1} b_{k-i+1} b_{k-j}\\ =&\Big(\sum_{j=0}^{k} a_j^2 b_{k-j} + \sum_{i\neq j} a_i a_j b_{k-i} b_{k-j}\Big) + \sum_{j=0}^{k}\sum_{i=0}^{j} a_i a_j b_{k-i+1} b_{k-j-1}\\ =&c_k^2 - c_{k-1}c_{k+1} \geq 0, \quad k \geq 1.\end{aligned}$$

Thus, $(c_k)_{k\geq 0}$ is a log-concave sequence. □

Theorem 6.3. *(Walkup theorem) Let $(a_i)_{i\geq 0}$, $(b_i)_{i\geq 0}$ be two ULC sequences. Then their convolution $(c_i)_{i\geq 0}$ is also ULC.*

In the next proposition we prove that if initially the law of the number of clients in the $M/M/\infty$ queue being served is ULC, then at any time $t \geq 0$ it stays ULC. This result is a key point that will be used for proving some properties of the semigroup of $M/M/\infty$ process.

Proposition 6.1. *If N_0 is an ultra log-concave random variable, then N_t defined by* (23) *is also an ultra log-concave random variable.*

Proof. Thanks to Remark 6.4 and Theorem 6.3 we just have to show that $E_{\frac{e^{-t}}{1-e^{-t}}} \circ T \circ E_{e^{-t}}\rho$ forms a ULC sequence. First note that for any $k \geq 1$, $p \in (0,1)$, if $d(k)$ is a ULC sequence, then $E_p d(k)$ is also a ULC sequence. Indeed,

$$\begin{aligned}E_p d(k) = d(k)p^k &\geq p^k\sqrt{\Big(1+\frac{1}{k}\Big)d(k+1)d(k-1)}\\ &= \sqrt{\Big(1+\frac{1}{k}\Big)d(k+1)p^{k+1}d(k-1)p^{k-1}}\\ &= \sqrt{\Big(1+\frac{1}{k}\Big)E_p d(k+1)E_p d(k-1)}.\end{aligned}$$

Thus, $E_{e^{-t}}\rho$ is ULC. Let us show that T preserves ultra log-concavity. Now, let $Y \sim \mu$ be an ULC random variable.

$$\begin{aligned}T\mu(k) = \sum_{j=k}^{\infty}\binom{j}{k}\mathbb{P}(Y=j) &= \sum_{j=0}^{\infty}\binom{j+k}{k}\mathbb{P}(Y=j+k)\\ &= \sum_{j=0}^{\infty}\frac{(j+k)!}{j!k!}\mathbb{P}(Y=j+k).\end{aligned}$$

Define a log-concave measure $\tilde{\mu}$ on $\mathbb{Z}_+$ as $\tilde{\mu}(k) = k!\mathbb{P}(Y=k)$, $\forall k \geq 0$. Then

$$\begin{aligned}k!T\mu(k) = \sum_{j=0}^{+\infty}\frac{(j+k)!}{j!}\mathbb{P}(Y=j+k) &= \sum_{j=0}^{+\infty}\frac{\tilde{\mu}(j+k)}{j!}\\ &= \sum_{j=0}^{+\infty}\tilde{\mu}(j+k)r(-j)\\ &= \sum_{j=-\infty}^{+\infty}\tilde{\mu}(k-j)r(j)\\ &= \tilde{\mu} * r(k),\end{aligned}$$

where

$$r(k) = \begin{cases}\frac{1}{(-k)!}, & k \leq 0,\\ 0, & k > 0.\end{cases}$$

$(r(k))_{k\leq 0}$ forms a log-concave sequence. Indeed, from the inequality $-k+1 \geq -k$ for $k \leq -1$ we construct

$$(-k+1)!(-k-1)! \geq (-k)!(-k)!$$

And thus it follows that

$$\frac{1}{(-k)!} \geq \frac{1}{\sqrt{(-k+1)!(-k-1)!}}, \quad k \leq -1.$$

Finally, since convolution of log-concave sequences is log-concave by Theorem 6.2, we obtain that $(k!T\mu(k))_{k\geq 0}$ is log-concave, and therefore, $(T\mu(k))_{k\geq 0}$ is ultra log-concave. Finally, $\left(E_{\frac{e^{-t}}{1-e^{-t}}} \circ T \circ E_{e^{-t}}\rho(k)\right)_{k\geq 0}$ is ULC, and by Theorem 6.3 N_t is a ULC random variable. $\square$

6.3 Properties of $M/M/\infty$ semigroup

The semigroup of the $M/M/\infty$ process given by

$$\mathcal{P}_t\varphi(n) = \mathbb{E}[\varphi(N_t)|N_0 = n] = \int \varphi(m) r_n^t(dm),$$

has the following important properties.

Proposition 6.2. *1. If $f : \mathbb{Z}_+ \to \mathbb{R}$ is log-convex, then $\mathcal{P}_t f$ is also log-convex.*

2. If $g : \mathbb{Z}_+ \to \mathbb{R}$ is log-concave, then $\mathcal{P}_t g$ is also log-concave.

Proof. 1. Using log-convexity of f and Cauchy-Schwarz inequality we get

$$\begin{aligned}
\mathcal{P}_t f(n) = \mathbb{E}[f(N_t)|N_0 = n] &= \int f(m+n) r_0^t(dm) \\
&\leq \int \big(f(m+n+1)f(m+n-1)\big)^{\frac{1}{2}} r_0^t(dm) \\
&\leq \Big(\int f(m+n+1) r_0^t(dm)\Big)^{\frac{1}{2}} \Big(\int f(m+n-1) r_0^t(dm)\Big)^{\frac{1}{2}} \\
&= \Big(\mathbb{E}[f(N_t)|N_0 = n+1]\Big)^{\frac{1}{2}} \Big(\mathbb{E}[f(N_t)|N_0 = n-1]\Big)^{\frac{1}{2}} \\
&= \big(\mathcal{P}_t f(n+1)\mathcal{P}_t f(n-1)\big)^{\frac{1}{2}}.
\end{aligned}$$

2. Assume that $N_0 \overset{law}{=} g d\rho$, i.e. N_0 is ULC. Then $\forall t \geq 0$, N_t is also ULC thanks to Theorem 6.1. Let $\varphi : \mathbb{Z}_+ \to \mathbb{R}$ be a test function. We write,

$$\begin{aligned}
\mathbb{E}[\varphi(N_t)] &= \mathbb{E}[\mathbb{E}[\varphi(N_t)|N_0]] = \mathbb{E}[\mathcal{P}_t\varphi(N_0)] \\
&= \int \mathcal{P}_t\varphi(n) g(n)\rho(dn) = \int \varphi(n)\mathcal{P}_t g(n)\rho(dn),
\end{aligned}$$

where last equality is obtained due to the symmetry of $\mathcal{P}_t$ in $L^2(\rho)$. It follows that N_t has density $\mathcal{P}_t g$ with respect to the measure ρ, and by definition of ultra log-concavity it means that $\mathcal{P}_t g$ is log-concave. $\square$

Next proposition is a counterpart of Proposition 5.2 for the Poisson measure. Thus, we will present a shortened proof, since its idea is identical to the one in the proof of Proposition 5.2. Note that Theorem 4.1 is still valid in our set-up, thus the optimal coupling $\pi^\epsilon \in \Pi(\mu, \nu)$ between any measures $\mu \in \mathcal{P}(\mathbb{Z}_+)$ and $\nu \in \mathcal{P}(\mathbb{Z}_+)$ can be represented as

$$\pi^\epsilon(dndm) = f^\epsilon(n) g^\epsilon(m) R^\epsilon(dndm),$$

for functions $f^\epsilon : \mathbb{Z}_+ \to \mathbb{R}_+$, $g^\epsilon : \mathbb{Z}_+ \to \mathbb{R}_+$ that are unique up to some multiplicative constant.

Proposition 6.3. *If $\mu(dn) = e^{V(n)}\rho(dn)$, $\nu(dn) = e^{-W(n)}\rho(dn)$ and f^ϵ, g^ϵ are such that (10) is satisfied, then*

$$f^\epsilon(n)\mathcal{P}_\epsilon g^\epsilon(n) = e^{V(n)}, \quad g^\epsilon(m)\mathcal{P}_\epsilon f^\epsilon(m) = e^{-W(m)}, \quad \forall n, m \in \mathbb{Z}_+.$$

Proof. Let $h : \mathbb{Z}_+ \to \mathbb{R}$ be a test function. Using definition of coupling, representation of the optimal coupling and definition of the $M/M/\infty$ semigroup we write,

$$\begin{aligned}\int h(n)\mu(dn) &= \iint h(n)\pi^\epsilon(dndm)\\ &= \iint h(n)f^\epsilon(n)g^\epsilon(m)\rho(dn)r_n^\epsilon(dm)\\ &= \int h(n)f^\epsilon(n)\mathcal{P}_\epsilon g^\epsilon(n)\rho(dn).\end{aligned}$$

Comparing first and last integral, we get

$$f^\epsilon(n)\mathcal{P}_\epsilon g^\epsilon(n) = e^{V(n)}.$$

Similarly for the measure ν:

$$\begin{aligned}\int h(m)\nu(dm) &= \iint h(m)\pi^\epsilon(dndm)\\ &= \iint h(m)f^\epsilon(n)g^\epsilon(m)\rho(dm)r_m^\epsilon(dn)\\ &= \int h(m)g^\epsilon(m)\mathcal{P}_\epsilon f^\epsilon(m)\rho(dm).\end{aligned}$$

Comparing first and last integral, we get

$$g^\epsilon(m)\mathcal{P}_\epsilon f^\epsilon(m) = e^{-W(m)}.$$

□

From Proposition 6.3 it follows that

$$g^\epsilon(m) = \frac{e^{-W(m)}}{\mathcal{P}_\epsilon f^\epsilon(m)}, \quad \forall m \in \mathbb{Z}_+, \tag{27}$$

and thus

$$f^\epsilon(n) = \frac{e^{V(n)}}{\mathcal{P}_\epsilon\Big(\frac{e^{-W(n)}}{\mathcal{P}_\epsilon f^\epsilon(n)}\Big)}, \quad \forall n \in \mathbb{Z}_+.$$

As before, we define a functional

$$\Phi(\varphi)(n) := \frac{e^{V(n)}}{\mathcal{P}_\epsilon\Big(\frac{e^{-W(n)}}{\mathcal{P}_\epsilon \varphi(n)}\Big)},$$

and f^ϵ is a fixed point of the functional Φ.

Proposition 6.4. *If $f : \mathbb{Z}_+ \to \mathbb{R}$ is log-convex, then $\Phi(f)$ is also log-convex.*

Proof. Since $\mathcal{P}_\epsilon$ preserves log-convexity, $\mathcal{P}_\epsilon f$ is log-convex. Then $\frac{1}{\mathcal{P}_\epsilon f}$ is log-concave. Indeed,

$$\begin{aligned}&\log \mathcal{P}_\epsilon f(n) \le \frac{1}{2}\Big(\log \mathcal{P}_\epsilon f(n+1) + \log \mathcal{P}_\epsilon f(n-1)\Big)\\ \iff &-\log \mathcal{P}_\epsilon f(n) \ge \frac{1}{2}\Big(-\log \mathcal{P}_\epsilon f(n+1) - \log \mathcal{P}_\epsilon f(n-1)\Big)\\ \iff &\log\Big(\frac{1}{\mathcal{P}_\epsilon f(n)}\Big) \ge \frac{1}{2}\Big(\log\Big(\frac{1}{\mathcal{P}_\epsilon f(n+1)}\Big) + \log\Big(\frac{1}{\mathcal{P}_\epsilon f(n-1)}\Big)\Big).\end{aligned}$$

Since e^{-W} is a log-concave function, $\frac{e^{-W}}{\mathcal{P}_\epsilon f}$ is still log-concave as a product of log-concave functions:

$$\begin{aligned}\log\Big(\frac{e^{-W(n)}}{\mathcal{P}_\epsilon f(n)}\Big) &= -W(n) + \log\Big(\frac{1}{\mathcal{P}_\epsilon f(n)}\Big)\\ &\geq \frac{1}{2}(-W(n+1) - W(n-1)) + \frac{1}{2}\Big(\log\Big(\frac{1}{\mathcal{P}_\epsilon f(n+1)}\Big) + \log\Big(\frac{1}{\mathcal{P}_\epsilon f(n-1)}\Big)\Big)\\ &= \frac{1}{2}\Big(\log\Big(\frac{e^{-W(n+1)}}{\mathcal{P}_\epsilon f(n+1)}\Big) + \log\Big(\frac{e^{-W(n-1)}}{\mathcal{P}_\epsilon f(n-1)}\Big)\Big).\end{aligned}$$

Then $\mathcal{P}_\epsilon\Big(\frac{e^{-W}}{\mathcal{P}_\epsilon f}\Big)$ is log-concave, and $\frac{e^V}{\mathcal{P}_\epsilon\Big(\frac{e^{-W}}{\mathcal{P}_\epsilon f^\epsilon}\Big)}$ is log-convex reasoning as before.

□

As in the case with Gaussian measure in Theorem 5.3, under similar conditions for Poisson measure we prove that the functions that give the representation of the optimal coupling, is a pair of log-concave and log-convex functions.

Theorem 6.4. *Let $\mu(dn) = e^{V(n)}\rho(dn)$ be an ultra log-convex measure and $\nu(dn) = e^{-W(n)}\rho(dn)$ be an ultra log-concave probability measure with $V, W : \mathbb{Z}_+ \to \mathbb{R}$ convex functions. Then there exist a log-convex function $f^\epsilon : \mathbb{Z}_+ \to [1, +\infty)$ and a log-concave function $g^\epsilon : \mathbb{Z}_+ \to (0, +\infty)$ such that the unique optimal coupling $\pi^\epsilon \in \Pi(\mu, \nu)$ is of the form $\pi^\epsilon(dndm) = f^\epsilon(n)g^\epsilon(m)R^\epsilon(dndm)$.*

Proof. We embrace the same algoritm of proof as in Theorem 5.3. Since the chosen reference measure R^ϵ satisfies the conditions of Proposition 4.1, functions f^ϵ, g^ϵ that satisfy (10) exist. We just have to show that they are log-convex and log-concave respectively in discrete sense. As before, it is enough to show that $\log f^\epsilon$ is convex, or equivalently by Theorem 6.1, $\log f^\epsilon = (\log f^\epsilon)^{\triangle\triangledown}$.

Define $h_0 = (\log f^\epsilon)^{\triangle\triangledown}$ and $f_0 = e^{h_0}$. We define an iterative process for f_0 as

$$f_1 = \max(\Phi(f_0), f_0).$$

Function f_0 satisfies the following properties:

1. $f_0 \leq f^\epsilon$ as greatest log-convex function by Remark 6.2;
2. $f_0 \leq f_1$ by $\max(\Phi(f_0), f_0) \geq f_0$;
3. f_1 is log-convex since Φ preserves log-convexity.

By monotonicity of Φ, the fact that f^ϵ is a fixed point of Φ and first property of f_0, we get $\max(\Phi(f_0), f_0) \leq \max(\Phi(f^\epsilon), f^\epsilon) = f^\epsilon$, thus

$$f_1 \leq f^\epsilon,$$

and since f_0 is the greatest log-convex function below f^ϵ, we have:

$$f_1 \leq f_0. \tag{28}$$

From second property of f_0 and (28) we get that $f_0 = f_1$ and thus,

$$f_0 = \max(\Phi(f_0), f_0) \geq \Phi(f_0).$$

By definition of the functional Φ and symmetry of the semigroup operator in $L^2(\rho)$ we write

$$1 \leq \int \frac{f_0}{\Phi(f_0)} d\mu = \int f_0 \mathcal{P}_\epsilon\Big(\frac{e^{-W}}{\mathcal{P}_\epsilon f_0}\Big) d\rho = \int \mathcal{P}_\epsilon f_0 \frac{e^{-W}}{\mathcal{P}_\epsilon f_0} d\rho = \int e^{-W} d\rho = 1,$$

since $e^{-W}d\rho$ is a probability measure on $\mathbb{Z}_+$. And so $\Phi(f_0) = f_0$ on $\mathbb{Z}_+$. Since $f_0 \leq f^\epsilon$ is the convex envelope of f^ϵ, there exists $n^* \in \mathbb{Z}_+$ such that

$$f_0(n^*) = f^\epsilon(n^*). \tag{29}$$

By Proposition 4.1 a pair of functions (f, g), $f, g : \mathbb{Z}_+ \to \mathbb{R}_+$, that represents a unique optimal coupling $\pi^\epsilon(dndm) = f(n)g(m)R^\epsilon(dndm)$ is of the form $f = \lambda f^\epsilon$, $g = \frac{g^\epsilon}{\lambda}$, $\lambda \in \mathbb{R}$. From (29) we find that $\lambda = 1$, therefore, $f_0(n) = f^\epsilon(n)$, $\forall n \in \mathbb{Z}_+$, and $\log f^\epsilon = (\log f^\epsilon)^{\triangle\triangledown}$. This concludes the proof. □

6.4 Convergence and other properties of entropic transport cost

As in the case with Gaussian measure, we would like to have convergence of $h(\epsilon)\mathcal{T}_H^\epsilon(\mu,\nu)$ with $h(\epsilon) \xrightarrow{\epsilon\to 0} 0$ to $\frac{1}{2}W_2^2(\mu,\nu)$ in order to get a version of Caffarelli's theorem for Poisson measure. However, in this case the limit is different as it will be shown in the next theorem.

Theorem 6.5. *Assume that $\mu,\nu \in \mathcal{P}(\mathbb{Z}_+)$ are such that*

$$\int n\log n\mu(dn) < +\infty, \qquad \int m\log m\nu(dm) < +\infty,$$

and

$$\int \log\Big(\frac{d\xi}{d\mu}\Big)d\mu < +\infty, \qquad \int \log\Big(\frac{d\xi}{d\nu}\Big)d\nu < +\infty.$$

Then for $0<\epsilon<1$

$$-\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) \xrightarrow[\epsilon\to 0^+]{} W_1(\mu,\nu).$$

Proof. For proving this limit we aim at bounding $-\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu)$ from above and below by a function that tends to $\int|n-m|\pi(dndm)$ as $\epsilon\to 0^+$ plus some function of ϵ that goes to zero as $\epsilon\to 0^+$. We first rewrite relative entropy in the following way.

$$\begin{aligned}
-\frac{1}{\log\epsilon}H(\pi|R^\epsilon) &= -\frac{1}{\log\epsilon}\int \log\Big(\frac{d\pi}{dR^\epsilon}\Big)d\pi \\
&= -\frac{1}{\log\epsilon}\int \log\Big(\frac{d\pi}{d\xi d\xi}\Big)d\pi + \frac{1}{\log\epsilon}\int \log\Big(\frac{dR^\epsilon}{d\xi d\xi}\Big)d\pi \\
&= -\frac{1}{\log\epsilon}\int \log\Big(\frac{d\pi}{d\xi d\xi}\Big)d\pi + \frac{1}{\log\epsilon}\int \log\Big(\frac{d\rho}{d\xi}\Big)d\mu + \frac{1}{\log\epsilon}\int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)d\pi.
\end{aligned} \tag{30}$$

We analyze every of the three terms separately. For the first terms we have

$$\int \log\Big(\frac{d\pi}{d\xi d\xi}\Big)d\pi \le \int \log(1)d\pi = 0,$$

and

$$\begin{aligned}
\int \log\Big(\frac{d\pi}{d\xi d\xi}\Big)d\pi &= \int \log\Big(\frac{d\pi}{d\mu d\nu}\Big)d\pi + \int \log\Big(\frac{d\mu d\nu}{d\xi d\xi}\Big)d\pi \\
&= \int \log\Big(\frac{d\pi}{d\mu d\nu}\Big)d\pi + \int \log\Big(\frac{d\mu}{d\xi}\Big)d\mu + \int \log\Big(\frac{d\nu}{d\xi}\Big)d\nu \\
&= H(\pi|\mu\times\nu) + H(\mu|\xi) + H(\nu|\xi) \\
&\ge H(\mu|\xi) + H(\nu|\xi).
\end{aligned}$$

We get the upper and lower bound of the first term in (30):

$$H(\mu|\xi) + H(\nu|\xi) \le \int \log\Big(\frac{d\pi}{d\xi d\xi}\Big)d\pi \le 0. \tag{31}$$

Moreover, since by assumption $H(\mu|\xi) < +\infty$, $H(\nu|\xi) < +\infty$, we get

$$-\frac{1}{\log\epsilon}(H(\mu|\xi) + H(\nu|\xi)) \xrightarrow[\epsilon\to 0^+]{} 0.$$

For the second term we use inequality $n! \le n^n$ and write

$$\begin{aligned}
\int \log\Big(\frac{d\rho}{d\xi}\Big)d\mu &= \int \log\Big(\frac{e^{-1}}{n!}\Big)\mu(dn) = -\int \mu(dn) - \int \log(n!)\mu(dn) \\
&\ge -1 - \int n\log n\mu(dn).
\end{aligned}$$

Since $n! \geq 1$, $\forall n \in \mathbb{Z}_+$,

$$\int \log\Big(\frac{d\rho}{d\xi}\Big) d\mu = -1 - \int \log(n!)\mu(dn) \leq -1 - \int \log(1)\mu(dn) = -1.$$

We get the upper and lower bound of the second term in (30):

$$-1 - \int n \log n \mu(dn) \leq \int \log\Big(\frac{d\rho}{d\xi}\Big) d\mu \leq -1. \tag{32}$$

Also, by assumption $\int n \log n \mu(dn)$ is finite, therefore,

$$-\frac{1}{\log \epsilon}\Big(1 + \int n \log n \mu(dn)\Big) d\mu \xrightarrow[\epsilon \to 0^+]{} 0.$$

We introduce the following functions,

$$\begin{aligned}
a(\epsilon) &:= -\frac{1}{\log \epsilon}, \\
b(\epsilon) &:= -\frac{1}{\log \epsilon}(H(\mu|\xi) + H(\nu|\xi)) - \frac{1}{\log \epsilon}(1 + \int n \log n \mu(dn)).
\end{aligned}$$

Note that $a(\epsilon), b(\epsilon) \xrightarrow[\epsilon \to 0^+]{} 0$, and by (31), (32),

$$\frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi + b(\epsilon) \leq -\frac{1}{\log \epsilon} H(\pi|R^\epsilon) \leq \frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi + a(\epsilon). \tag{33}$$

We are left to show that $\frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi \xrightarrow[\epsilon \to 0^+]{} \int |n - m| \pi(dndm)$, $\forall n, m \geq 1$.

By definition of r_n^ϵ we have

$$\begin{aligned}
\frac{dr_n^\epsilon}{d\xi}(m) &= \sum_{k=0}^{m} \mathbb{P}(Y = k)\mathbb{P}(X = m - k) \\
&= \sum_{k=0}^{m} \frac{(1-\epsilon^k)^k e^{-(1-e^{-\epsilon})}}{k!} \binom{n}{m-k} e^{-(m-k)\epsilon}(1 - e^{-\epsilon})^{n-(m-k)} \\
&= \frac{(1 - e^{-\epsilon})^{n-m}}{e^{m\epsilon} e^{-(1-e^{-\epsilon})}} \sum_{k=0}^{m} \binom{n}{m-k} (1 - e^{-\epsilon})^{2k} e^{k\epsilon}.
\end{aligned}$$

Then, we can write

$$\begin{aligned}
\frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi &= \frac{1}{\log \epsilon} \int (1 - e^{-\epsilon})\pi(dndm) - \frac{1}{\log \epsilon} \int m\epsilon\pi(dndm) \\
&\quad + \frac{1}{\log \epsilon} \int \log(1 - e^\epsilon)^{(n-m)} \pi(dndm) \\
&\quad + \frac{1}{\log \epsilon} \int \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1 - e^{-\epsilon})^{2k} e^{k\epsilon}}{k!}\Big) \pi(dndm) \\
&= \frac{1 - e^{-\epsilon}}{\log \epsilon} - \frac{\epsilon}{\log \epsilon} \int m\nu(dm) + \frac{\log(1 - e^\epsilon)}{\log \epsilon} \int (n - m)\pi(dndm) \\
&\quad + \int \frac{1}{\log \epsilon} \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1 - e^{-\epsilon})^{2k} e^{k\epsilon}}{k!}\Big) \pi(dndm).
\end{aligned} \tag{34}$$

We note that

$$\lim_{\epsilon \to 0^+} \frac{1 - e^\epsilon}{\log \epsilon} = 0, \tag{35}$$

$$\lim_{\epsilon \to 0^+} \frac{\epsilon}{\log \epsilon} = 0, \tag{36}$$

$$\lim_{\epsilon \to 0^+} \frac{\log(1 - e^\epsilon)}{\log \epsilon} = 1.$$

Since by assumptions $\int m\nu(dm)$ and $\int(n-m)\pi(dndm)$ are finite, we conclude that

$$-\frac{\epsilon}{\log\epsilon}\int m\nu(dm)\to 0,$$
$$\frac{\log(1-e^{\epsilon})}{\log\epsilon}\int(n-m)\pi(dndm)\to\int(n-m)\pi(dndm),\quad \text{as } \epsilon\to 0^+.$$

For finding the limit of the last term in (34), we bound the integrand from above and below. For this we have to distinguish two cases of relations between n and m. Assume that $n\geq m$. Then

$$\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\leq\sum_{k=0}^{m}\binom{n}{m-k}e^{m\epsilon}\leq\sum_{i=0}^{n}\binom{n}{i}e^{m\epsilon}=2^n e^{m\epsilon},$$
$$\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\geq\binom{n}{m}.$$

Therefore,

$$\frac{1}{\log\epsilon}(n\log 2+m\epsilon)\leq\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\Big)\leq\frac{1}{\log\epsilon}\log\binom{n}{m}.$$

Letting ϵ to 0^+, we get that lower and upper bound tend to zero, and thus,

$$\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\Big)\xrightarrow[\epsilon\to 0^+]{}0,$$

and

$$\begin{aligned}&\lim_{\epsilon\to 0^+}\int\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\Big)1_{n\geq m}\pi(dndm)\\
=&\lim_{\epsilon\to 0^+}\sum_{(n,m)\in\mathbb{Z}_+\times\mathbb{Z}_+}\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\Big)1_{n\geq m}\pi(n,m)\\
=&\sum_{(n,m)\in\mathbb{Z}_+\times\mathbb{Z}_+}\lim_{\epsilon\to 0^+}\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{2k}e^{k\epsilon}}{k!}\Big)1_{n\geq m}\pi(n,m)\\
=&0,\end{aligned}$$

for $n\geq m$.

Finally, putting all the results together,

$$\frac{1}{\log\epsilon}\int\log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)1_{\{n\geq m\}}d\pi\xrightarrow[\epsilon\to 0^+]{}\int(n-m)1_{\{n\geq m\}}d\pi. \tag{37}$$

Suppose now that $n\leq m$. We rewrite (34) as

$$\begin{aligned}\frac{1}{\log\epsilon}\int\log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)1_{m\geq n}d\pi=&\frac{1-e^{-\epsilon}}{\log\epsilon}-\frac{\epsilon}{\log\epsilon}\int m\nu(dm)\\
&+\int\frac{1}{\log\epsilon}\log\Big(\sum_{k=0}^{m}\binom{n}{m-k}\frac{(1-e^{-\epsilon})^{n-m+2k}e^{k\epsilon}}{k!}\Big)\pi(dndm).\end{aligned}$$

We only have to estimate the terms that contain both n and m, i.e. the last term. First, we estimate the sum in the logarithm from above and below. Let $j:=m-k, k=m-j$. Since for $k>n$ the binomial coefficient $\binom{n}{m-k}$ is zero (because $n\leq m$), we can consider sum only till index n. Thus, we write,

$$\sum_{j=0}^{n}\binom{n}{j}\frac{(1-e^{-\epsilon})^{n+m-2j}e^{(m-j)\epsilon}}{(m-j)!}=(1-e^{-\epsilon})^{m-n}\sum_{j=0}^{n}\binom{n}{j}\frac{(1-e^{-\epsilon})^{2(n-j)}e^{(m-j)\epsilon}}{(m-j)!}.$$

Similarly to the case when $n \geq m$, we can upper bound this expression with $(1-e^{-\epsilon})^{m-n}2^n e^{\epsilon m}$. Also, since

$$\sum_{j=0}^{n} \binom{n}{j} \frac{(1-e^{-\epsilon})^{2(n-j)} e^{(m-j)\epsilon}}{(m-j)!} \geq \frac{e^{(m-n)\epsilon}}{(m-n)!} \geq \frac{1}{(m-n)!},$$

we get

$$\frac{(1-e^{-\epsilon})^{m-n}}{(m-n)!} \leq (1-e^{-\epsilon})^{m-n} \sum_{j=0}^{n} \binom{n}{j} \frac{(1-e^{-\epsilon})^{2(n-j)} e^{(m-j)\epsilon}}{(m-j)!} \leq (1-e^{-\epsilon})^{m-n} 2^n e^{\epsilon m},$$

and

$$\begin{aligned} \frac{1}{\log \epsilon} \log((1-e^{-\epsilon})^{m-n} 2^n e^{\epsilon m}) &\leq \frac{1}{\log \epsilon} \log\Big((1-e^{-\epsilon})^{m-n} \sum_{j=0}^{n} \binom{n}{j} \frac{(1-e^{-\epsilon})^{2(n-j)} e^{(m-j)\epsilon}}{(m-j)!}\Big) \\ &\leq \frac{1}{\log \epsilon} \log \frac{(1-e^{-\epsilon})^{m-n}}{(m-n)!}. \end{aligned}$$

Since

$$\frac{1}{\log \epsilon} \log \frac{(1-e^{-\epsilon})^{m-n}}{(m-n)!} \xrightarrow[\epsilon \to 0^+]{} m-n,$$

and

$$\frac{1}{\log \epsilon} \log((1-e^{-\epsilon})^{m-n} 2^n e^{\epsilon m}) \xrightarrow[\epsilon \to 0^+]{} m-n,$$

we get

$$\frac{1}{\log \epsilon} \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1-e^{-\epsilon})^{n-m+2k} e^{k\epsilon}}{k!}\Big) \xrightarrow[\epsilon \to 0^+]{} m-n, \quad m \geq n.$$

We are left to show that

$$\begin{aligned} &\lim_{\epsilon \to 0^+} \int \frac{1}{\log \epsilon} \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1-e^{-\epsilon})^{n-m+2k} e^{k\epsilon}}{k!}\Big) 1_{\{m \geq n\}} \pi(dndm) \\ &= \lim_{\epsilon \to 0^+} \sum_{(n,m) \in \mathbb{Z}_+ \times \mathbb{Z}_+} \frac{1}{\log \epsilon} \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1-e^{-\epsilon})^{n-m+2k} e^{k\epsilon}}{k!}\Big) 1_{\{m \geq n\}} \pi(n,m) \\ &= \sum_{(n,m) \in \mathbb{Z}_+ \times \mathbb{Z}_+} \lim_{\epsilon \to 0^+} \frac{1}{\log \epsilon} \log\Big(\sum_{k=0}^{m} \binom{n}{m-k} \frac{(1-e^{-\epsilon})^{n-m+2k} e^{k\epsilon}}{k!}\Big) 1_{\{m \geq n\}} \pi(n,m) \\ &= \int (m-n) 1_{\{m \geq n\}} \pi(dndm). \end{aligned} \tag{38}$$

Combining (35), (36) and (38), we get

$$\frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) 1_{\{m \geq n\}} d\pi \xrightarrow[\epsilon \to 0^+]{} \int (m-n) 1_{\{m \geq n\}} d\pi. \tag{39}$$

Finally, from (37) and (39) we obtain

$$\frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi \xrightarrow[\epsilon \to 0^+]{} \int |m-n| d\pi, \quad \forall n, m \geq 1.$$

Let $\pi_1^\epsilon, \pi_1 \in \Pi(\mu, \nu)$ be such that π_1^ϵ converges weakly to π_1 as $\epsilon \to 0^+$ and $H(\pi_1^\epsilon | R^\epsilon) = \mathcal{T}_H^\epsilon(\mu, \nu)$. Then from (33),

$$-\frac{1}{\log \epsilon} \mathcal{T}_H^\epsilon(\mu, \nu) \geq \frac{1}{\log \epsilon} \int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big) d\pi_1^\epsilon + b(\epsilon),$$

and

$$\begin{aligned}\liminf_{\epsilon\to 0^+} -\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) &\geq \liminf_{\epsilon\to 0^+}\Big(\frac{1}{\log\epsilon}\int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)d\pi_1^\epsilon + b(\epsilon)\Big)\\ &= \int |n-m| d\pi_1\\ &\geq W_1(\mu,\nu).\end{aligned}$$

Now let $\pi_2 \in \Pi(\mu,\nu)$ be such that $\int |n-m|d\pi_2 = W_1(\mu,\nu)$. Then,

$$-\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) \leq \frac{1}{\log\epsilon}\int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)d\pi_2 + a(\epsilon),$$

and

$$\begin{aligned}\limsup_{\epsilon\to 0^+} -\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) &\leq \limsup_{\epsilon\to 0^+}\Big(\frac{1}{\log\epsilon}\int \log\Big(\frac{dr_n^\epsilon}{d\xi}\Big)d\pi_2 + a(\epsilon)\Big)\\ &= \int |n-m| d\pi_2\\ &= W_1(\mu,\nu).\end{aligned}$$

So we get

$$W_1(\mu,\nu) \leq \liminf_{\epsilon\to 0^+} -\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) \leq \limsup_{\epsilon\to 0^+} -\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) \leq W_1(\mu,\nu),$$

and it follows that

$$\lim_{\epsilon\to 0^+} -\frac{1}{\log\epsilon}\mathcal{T}_H^\epsilon(\mu,\nu) = W_1(\mu,\nu).$$

□

We have the following monotonicity property of the entropic transport cost.

Proposition 6.5. *Let $\mu,\nu \in \mathcal{P}(\mathbb{Z}_+)$. Then for any $\tau,\eta \in \mathcal{P}(\mathbb{Z}_+)$ such that $\mu \leq_c \tau$, $\eta \leq_c \nu$ and $H(\eta|\xi) < +\infty$, $H(\tau|\xi) < +\infty$, it holds*

$$\mathcal{T}_H^\epsilon(\mu,\nu) \leq \mathcal{T}_H^\epsilon(\tau,\eta), \quad \forall\epsilon > 0.$$

Proof. The proof is of the same idea as in Proposition 5.3. Let $\pi_2 \in \Pi(\tau,\eta)$ be some coupling between τ and η, $\eta \leq_c \nu$. To apply Lemma 4.1, we take $\alpha = \pi_2$, $\beta = R^\epsilon$, $h = \log(f^\epsilon g^\epsilon)$, and let $\pi_1^* \in \Pi(\mu,\nu)$ be an optimal coupling between μ and ν, $\mu \leq_c \tau$. As before, we write

$$\begin{aligned}H(\pi_2|R^\epsilon) &\geq \int \log(f^\epsilon(n)g^\epsilon(m))d\pi_2(n,m) - \log\int f^\epsilon(n)g^\epsilon(m)dR^\epsilon(n,m)\\ &= \int \log f^\epsilon(n)d\pi_2(n,m) + \int \log g^\epsilon(m)d\pi_2(n,m)\\ &= \int \log f^\epsilon(n)d\tau(n) + \int \log g^\epsilon(m)d\eta(m)\\ &\geq \int \log f^\epsilon(n)d\mu(n) + \int \log g^\epsilon(m)d\nu(m)\\ &= H(\pi_1^*|R^\epsilon)\\ &= \mathcal{T}_H^\epsilon(\mu,\nu).\end{aligned}$$

For the second inequality we used the fact that $\log f^\epsilon$ is convex and $\mu \leq_c \tau$, $\log g^\epsilon$ is concave and $\eta \leq_c \nu$. Optimize $H(\pi_2|R^\epsilon)$ over all coupling $\pi_2 \in \Pi(\tau,\eta)$ in $H(\pi_2|R^\epsilon)$, and we obtain

$$\mathcal{T}_H^\epsilon(\mu,\nu) \leq \mathcal{T}_H^\epsilon(\tau,\eta), \quad \forall\epsilon > 0.$$

□

Corollary 6.1. *Taking $\mu = \tau = \rho^\lambda$ for $\lambda \in \mathbb{R}_+$ in Proposition 6.5, we get*

$$\mathcal{T}_H^\epsilon(\rho^\lambda,\nu) \leq \mathcal{T}_H^\epsilon(\rho^\lambda,\eta), \quad \forall\eta \leq_c \nu.$$

6.5 Monotonicity properties of W_1 distance

Proposition 6.6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{Z}_+)$ satisfy assumptions of Theorem 6.5. Then*

1. *For all $\eta \in \mathcal{P}(\mathbb{Z}_+)$ such that $\forall \eta \leq_c \nu$,*

$$W_1(\rho^\lambda, \nu) \leq W_1(\rho^\lambda, \eta)$$

2. *For all $\tau, \eta \in \mathcal{P}(\mathbb{Z}_+)$ that satisfy assumptions of Theorem 6.5 and such that $\mu \leq_c \tau$, $\eta \leq_c \nu$,*

$$W_1(\mu, \nu) \leq W_1(\tau, \eta).$$

Proof. The result follows from the convergence of $-\frac{1}{\log \epsilon}\mathcal{T}_H^\epsilon(\mu, \nu)$ to $W_1(\mu, \nu)$ as $\epsilon \to 0^+$ under given assumptions, and Propositions 6.5 and 6.1. □

In what follows, we equip the space of measures $\mathcal{P}_1(\mathbb{Z}_+)$ with the topology generated by metric W_1. In this topology if $(\eta_k)_{k\geq 1} \subset \mathcal{P}_1(\mathbb{Z}_+)$ and $\eta \in \mathcal{P}_1(\mathbb{Z}_+)$, then $W_1(\eta_k, \eta) \to 0$ if and only if $\int f d\eta_k \to \int f d\eta$ as $k \to +\infty$ for any 1-Lipschitz function $f : \mathbb{Z}_+ \to \mathbb{R}$.

Lemma 6.1. *The set $C_\nu = \{\eta \in \mathcal{P}(\mathbb{Z}_+) : \ \eta \leq_c \nu\}$ is compact in $\mathcal{P}_1(\mathbb{Z}_+)$ for the topology induced by W_1.*

Proof. Let us show that C_ν is closed. Rewrite the set C_ν as follows:

$$C_\nu = \Big\{\eta \in \mathcal{P}(Z_+) : \ \int f d\eta \leq \int f d\nu, \ \forall f : \mathbb{Z}_+ \to \mathbb{R} \text{ convex and 1-Lipschitz}\Big\}.$$

Since functionals $\eta \mapsto \int f d\eta$ with f convex and 1-Lipschitz are continuous for the W_1 topology, C_ν is closed.

Now let us show that C_ν is precompact, i.e. that the closure of C_ν is compact. We can use a variant of Prokhorov theorem (see, for example, Theorem 9.9 in [11]) that can be deduced from the classical Prokhorov Theorem 2.1.11 in [9]. According to this theorem, a set of measures C_ν is precompact in the topology generated by W_1 if and only if

$$\sup_{\eta \in C_\nu} \int |x| 1_{|x|>R} d\eta(x) \to 0, \quad R \to +\infty.$$

Add to the inequality $|x| > R$ term $-2|x|$, and we obtain that

$$|x|1_{|x|>R} \leq \max\{0, 2|x| - R\}, \quad x \in \mathbb{R}^d, \ R \in \mathbb{R}_+.$$

Then

$$\sup_{\eta \in C_\nu} \int |x|1_{|x|>R} d\eta(x) \leq \sup_{\eta \in C_\nu} \int \max\{0, 2|x| - R\} d\eta(x) := f(R).$$

Note that $\max\{0, 2|x| - R\} \leq 2|x|$ and $\max\{0, 2|x| - R\} \xrightarrow[R \to +\infty]{} 0$, $x \in \mathbb{R}^d$, thus by monotone convergence theorem we conclude that $f(R) \xrightarrow[R \to +\infty]{} 0$. Therefore, C_ν is precompact. Since from closedness $\bar{C}_\nu = C_\nu$, we obtain that C_ν is compact. □

We state the following result about duality of W_1 distance between ULC and ULCX measures.

Proposition 6.7. *For all ULCX measures $\mu \in \mathcal{P}(\mathbb{Z}_+)$ and ULC measures $\in \mathcal{P}(\mathbb{Z}_+)$ that satisfy assumptions of Theorem 6.5 it holds*

$$W_1(\mu, \nu) = \sup\Big\{\int f d(\mu - \nu) \text{ for } f \text{ 1-Lipschitz and convex on } \mathbb{Z}_+\Big\}.$$

Proof. By definition of $W_1(\mu, \nu)$ and Proposition 6.6 we write:

$$\begin{aligned} W_1(\mu, \nu) &= \inf_{\eta \leq_c \nu} W_1(\mu, \eta) \\ &= \inf_{\eta \leq_c \nu} \sup_{f 1-Lipschitz} \int f d(\mu - \eta). \end{aligned} \tag{40}$$

Our next step is to show that in (40) infimum and supremum can be exchanged. For this, we introduce the function of two variables $g(\eta, f) = \int f d(\mu - \eta)$ and apply Sion's theorem ([18]) for which we have to show that:

- The set $\{\eta \in \mathcal{P}(\mathbb{Z}_+) : \eta \leq_c \nu\}$ is compact (proved in Lemma 6.1);

- $g(\cdot, f)$ is lower semicontinuous and quasi-convex for any f 1-Lipschitz;

- $g(\eta, \cdot)$ is upper semicontinuous and quasi-concave for any $\eta \in \mathcal{P}(\mathbb{Z}_+)$ such that $\eta \leq_c \nu$.

We choose the following topology: we say that f_n converges to f as $n \to +\infty$ if and only if $f_n(i)$ converges to $f(i)$ as $n \to +\infty$ for all $i \in \mathbb{Z}_+$. Assuming without a loss of generality that $f(0) = 0$, by 1-Lipschitzianity of f we get that $|f(i)| \leq i$. Then by applying dominated convergence theorem we get,

$$\int f_n d\eta = \sum_{i=0}^{+\infty} f_n(i)\eta(i) \xrightarrow[n\to+\infty]{} \sum_{i=0}^{+\infty} f(i)\eta(i) = \int f d\eta.$$

Thus, $g(\eta, \cdot)$ is continuous in chosen topology. Also, if $(\eta_k)_{k\geq 1}$ is a sequence of measures that converges in W_1 distance to η as $k \to \infty$ and $\eta_k \leq_c \nu$, $\forall k \geq 1$, then

$$\lim_{k\to\infty} \int f d\eta_k = \int f d\eta,$$

i.e. for any fixed 1-Lipschitz $f : \mathbb{Z}_+ \to \mathbb{R}$, $g(\cdot, f)$ is continuous. Next, we note that $g(\cdot, f)$ is convex on the set of all measures $\eta \in \mathcal{P}(\mathbb{Z}_+)$ such that $\eta \leq_c \nu$, $g(\eta, \cdot)$ concave on the set of all 1-Lipschitz functions $f : \mathbb{Z}_+ \to \mathbb{R}$. Therefore, we are in the framework of Sion's minimax theorem ([19], Theorem 36.3), and we have that

$$\inf_{\eta\leq_c\nu} \sup_{f 1-Lipschitz} \int f d(\mu - \eta) = \sup_{f 1-Lipschitz} \inf_{\eta\leq_c\nu} \int f d(\mu - \eta).$$

Then we can rewrite (40) as

$$W_1(\mu, \nu) = \sup_{f 1-Lipschitz} \Big\{ \int f d\mu - \sup_{\eta\leq_c\nu} \int f d\eta \Big\}.$$

Note that since $f \mapsto \int f d\mu$ is continuous and infimum of a continuous function is upper semicontinuous, we get that $f \mapsto \int f d\mu - \sup_{\eta\leq_c\nu} \int f d\eta$ is upper semicontinuous. From this and from the fact that the set of functions $\{f : \mathbb{Z}_+ \to \mathbb{R} : \ f \text{ 1-Lipschitz and } f(0) = 0\}$ is compact, there exists $\bar{f} : \mathbb{Z}_+ \to \mathbb{R}$ such that

$$W_1(\mu, \nu) = \int \bar{f} d\mu - \sup_{\eta\leq_c\nu} \int \bar{f} d\eta, \tag{41}$$

Now, by contradiction we will prove that $\bar{f}$ is convex. Assume that $\bar{f}$ is not convex. Then there exists $n \in \mathbb{N}$ such that

$$2\bar{f}(n) > \bar{f}(n+1) + \bar{f}(n-1). \tag{42}$$

We construct a new probability measure η_ϵ in the following way.

$$\eta_\epsilon(k) = \begin{cases} \nu(k), & k \notin \{n-1, n, n+1\}, \\ \nu(k) + 2\epsilon, & k = n, \\ \nu(k) - \epsilon, & k \in \{n-1, n+1\}, \end{cases}$$

where $\epsilon = \min\{\nu(n+1), \nu(n-1)\}$. Then using (42),

$$\begin{aligned} \int \bar{f} d\eta_\epsilon &= \int \bar{f} d\nu + \epsilon(2\bar{f}(n) - (\bar{f}(n+1) + \bar{f}(n-1))) \\ &> \int \bar{f} d\nu. \end{aligned}$$

In particular, $\sup_{\eta\leq_c\nu} \int \bar{f} d\eta \geq \int \bar{f} d\eta_\epsilon$. Substituting this result into (41), we obtain

$$\begin{aligned} W_1(\rho, \nu) &= \int \bar{f} d\rho - \sup_{\eta\leq_c\nu} \int \bar{f} d\eta \\ &< \int \bar{f} d\rho - \int \bar{f} d\nu \\ &\leq W_1(\rho, \nu), \end{aligned}$$

and we get a contradiction. Thus, $\bar{f}$ is convex, and this completes the proof. □

Let $Y_i \sim \mu$, $i \geq 1$, be a sequence of i.i.d. random variables with values in $\mathbb{Z}_+$, $\mathbb{E}[Y_i] = \lambda$. Let $X_i \sim Bernoulli(1/n)$ be independent random variables for $1 \leq i \leq n$, $n \geq 1$, and $X_i \perp\!\!\!\perp Y_j$, $i = 1, ..., n$, $j \geq 0$. We consider the $1/n$-thinning of μ^{*n} defined as before in Definition 6.3. As μ^{*n} we denote the n-th convolution of μ.

Harremoes, Johnson and in Kontoyiannis proved in [12] the so called *law of thin numbers* which can be understood as a discrete version of the central limit theorem for Poisson distribution. It says that the thinning of the sum of i.i.d. random variables tends pointwise to the Poisson measure with parameter being equal to the mean of the given random variable. More precisely:

Theorem 6.6. *(Law of thin numbers, [12]) Let $Y_1, ..., Y_n$ be i.i.d. random variables distributed according to $\mu \in \mathcal{P}(\mathbb{Z}_+)$ with $\mathbb{E}[Y_i] = \lambda < +\infty$. Then*

$$T_{\frac{1}{n}}\mu^{*n}(k) \xrightarrow[n \to +\infty]{} \rho^{\lambda}(k), \quad \forall k \in \mathbb{Z}_+.$$

We state monotonicity properties of entropy that were proven by Yu in [22], using in particular the law of thin numbers, which we will be using for elaborating monotonicity properties of W_1 distance.

Theorem 6.7. *(Yu, [22]) Let $Y_1, ..., Y_n$ be i.i.d. $\mathbb{Z}_+$-valued random variables with $\mathbb{E}[Y_i] = \lambda < +\infty$ distributed according to μ.*

1. *If $Y_1, ..., Y_n$ are ultra log-concave, then*

$$T_{\frac{1}{n}}\mu^{*n} \leq_c T_{\frac{1}{n+1}}\mu^{*(n+1)}$$

and

$$H(T_{\frac{1}{n}}\mu^{*n}) \downarrow H(\rho^{\lambda}), \quad n \to \infty.$$

2. *If $Y_1, ..., Y_n$ are ultra log-convex and log-concave, then*

$$T_{\frac{1}{n+1}}\mu^{*(n+1)} \leq_c T_{\frac{1}{n}}\mu^{*n}$$

and

$$H(T_{\frac{1}{n}}\mu^{*n}) \uparrow H(\rho^{\lambda}), \quad n \to \infty.$$

Remark 6.5. *(Information-theoretic CLT) Theorem 6.7 is a discrete counterpart of a celebrated result by Artstein, Ball, Barthe and Naor ([1]) showing monotonicity in entropy in the usual central limit theorem. More precisely, given i.i.d. random variables $X_1, ..., X_n$ such that $\mathbb{E}[|X_i|^2] < +\infty$, it holds*

$$H\Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\Big) \downarrow H(\gamma), \quad n \to \infty.$$

Theorem 6.8. *Let $Y_1, ..., Y_n$ be i.i.d. $\mathbb{Z}_+$-valued random variables with $\mathbb{E}[Y_i] = \lambda < +\infty$ distributed according to μ. If $Y_1, ..., Y_n$ are ultra log-concave, then*

$$W_1(\rho^{\lambda}, T_{\frac{1}{n}}\mu^{*n}) \downarrow 0, \quad n \to \infty.$$

Proof. Note that by Theorem 6.7 $T_{\frac{1}{n+1}}\mu^{*(n+1)} \leq_c T_{\frac{1}{n}}\mu^{*n}$, thus we can use (6.6). So we get

$$W_1(\rho^{\lambda}, T_{\frac{1}{n+1}}\mu^{*(n+1)}) \leq W_1(\rho^{\lambda}, T_{\frac{1}{n}}\mu^{*n}), \quad \forall n \geq 1.$$

Convergence follows from the pointwise convergence of $T_{\frac{1}{n}}\mu^{*n}$ to ρ^{λ} as $n \to \infty$ ([22]). □

Theorem 6.9. *Let $Y_1, ..., Y_n$ and $Z_1, ..., Z_n$ be i.i.d. $\mathbb{Z}_+$-valued random variables with $\mathbb{E}[Y_i] = \lambda_1 < +\infty$, $\mathbb{E}[Z_i] = \lambda_2 < +\infty$ such that $Y_i \perp\!\!\!\perp Z_j$, $i, j = 1, ..., n$, and let $Y_i \sim \mu$, $Z_j \sim \nu$ have log-concave densities. If $Y_1, ..., Y_n$ are ultra log-concave and $Z_1, ..., Z_n$ are ultra log-convex random variables, then*

$$W_1(T_{\frac{1}{n}}\mu^{*n}, T_{\frac{1}{n}}\nu^{*n}) \downarrow W_1(\rho^{\lambda_1}, \rho^{\lambda_2}), \quad n \to \infty.$$

Proof. By Theorem 6.7, $T_{\frac{1}{n}}\mu^{*n} \leq_c T_{\frac{1}{n+1}}\mu^{*(n+1)}$ and $T_{\frac{1}{n+1}}\nu^{*(n+1)} \leq_c T_{\frac{1}{n}}\nu^{*n}$. Writing $\mu = T_{\frac{1}{n+1}}\nu^{*(n+1)}$, $\tau = T_{\frac{1}{n}}\nu^{*n}$, $\nu = T_{\frac{1}{n+1}}\mu^{*(n+1)}$, $\eta = T_{\frac{1}{n}}\mu^{*n}$, we are in the setup of Proposition 6.6, and therefore

$$W_1(T_{\frac{1}{n+1}}\mu^{*(n+1)}, T_{\frac{1}{n+1}}\nu^{*(n+1)}) \leq W_1(T_{\frac{1}{n}}\mu^{*n}, T_{\frac{1}{n}}\nu^{*n}), \quad n \geq 1.$$

As in Theorem 6.8, convergence follows from the pointwise convergence of $T_{\frac{1}{n}}\mu^{*n}$ to ρ^{λ_1} and $T_{\frac{1}{n}}\nu^{*n}$ to ρ^{λ_2} as $n \to \infty$.

□

Remark 6.6. *(Monotonicity of W_2 for Gaussian measure) Monotonicity for W_2 metric along the usual central limit theorem is still an open question. The question of showing that $W_2\Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}, \frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\Big)$, where $X_1, ..., X_n$ and $Y_1, ..., Y_n$ mutually independent random variables, is monotonic in n was first raised by Villani in [21], and that it is known to be true along the sequence 2^n. In [7], page 131, Schachermayer, Schmock and Teichmann prove that there is no such monotonicity without additional assumptions by constructing a counterexample. Though, this monotonicity could be true with some additional assumptions, for example log-concavity. At the same time, it is still an open question if the following monotonocity is not always true:*

$$W_2\Big(Z, \frac{\sum_{i=1}^n X_i}{\sqrt{n}}\Big) \downarrow 0, \quad n \to \infty, \text{ with } Z \sim \gamma.$$

Corollary 6.2. *Under assumptions of Theorem 6.9, if $\lambda_1 = \lambda_2 =: \lambda$, then*

$$W_1(T_{\frac{1}{n}}\mu^{*n}, T_{\frac{1}{n}}\nu^{*n}) \downarrow 0, \quad n \to \infty.$$

Corollary 6.3. *As in Theorem 6.9, let $Y_1, ..., Y_n$ be i.i.d. $\mathbb{Z}_+$-valued random variables with $\mathbb{E}[Y_i] = \lambda_1 < +\infty$, $Y_i \sim \mu$. If $Y_1, ..., Y_n$ are ULC and $Z \sim \nu$ is an ULCX random variable, then*

$$W_1(T_{\frac{1}{n}}\mu^{*n}, \nu) \downarrow W_1(\rho^{\lambda}, \nu), \quad n \to \infty.$$

Remark 6.7. *If in Corollary 6.3 we assume that ν is not ULCX but ULC, then the monotonicity of W_1 distance is no longer true. Indeed, assume that it still holds. Then by taking $\nu = T_1\mu$, we get,*

$$W_1(T_{\frac{1}{2}}\mu^{*2}, T_1\mu) \leq W_1(T_1\mu, T_1\mu) = 0.$$

Contradiction.

Remark 6.8. *It can be shown that in the set-up of Theorem 6.9, W_1 distance is monotonic along the sequence $\frac{1}{2^k}$, $k \geq 1$. In other words,*

$$W_1(T_{\frac{1}{2^{n+1}}}\mu^{*2^{n+1}}, T_{\frac{1}{2^{n+1}}}\nu^{*2^{n+1}}) \leq W_1(T_{\frac{1}{2^n}}\mu^{*2^n}, T_{\frac{1}{2^n}}\nu^{*2^n}), \quad n \geq 1.$$

It is natural to ask if all the assumptions stated in Theorem 6.8 and 6.9 are mandatory to still have monotonicity properties of Wasserstein distance. To understand this, we are demonstrating some simulations that have been conducted using Python. As examples, we will be considering random variables with randomly generated probability mass functions. This guarantees that PMF is indeed arbitrary and it is easier to understand the behaviour of the Wasserstein distance outside of ultra log-convexity and ultra log-concavity conditions.

First, we are trying to understand the behaviour of W_1 distance between Poisson measure and thinning of $Y_1 + ... + Y_n$, $n \geq 1$, when $Y_i \sim \mu$ are not ULC. We consider random variables Y_i, $i = 1, ..., n$, $n = 1, ..., 6$, supported on $0, ..., 6$ with the following PMF:

$$p = [0.17504764, 0.17053155, 0.01344013, 0.223762220.28507222, 0.06412496, 0.06802126]$$

Note that $\lambda := \mathbb{E}[Y_i] = 2.737739794318436$, and p is a sequence that is neither ULC nor ULCX. Then we observe the following convergence which is always true:

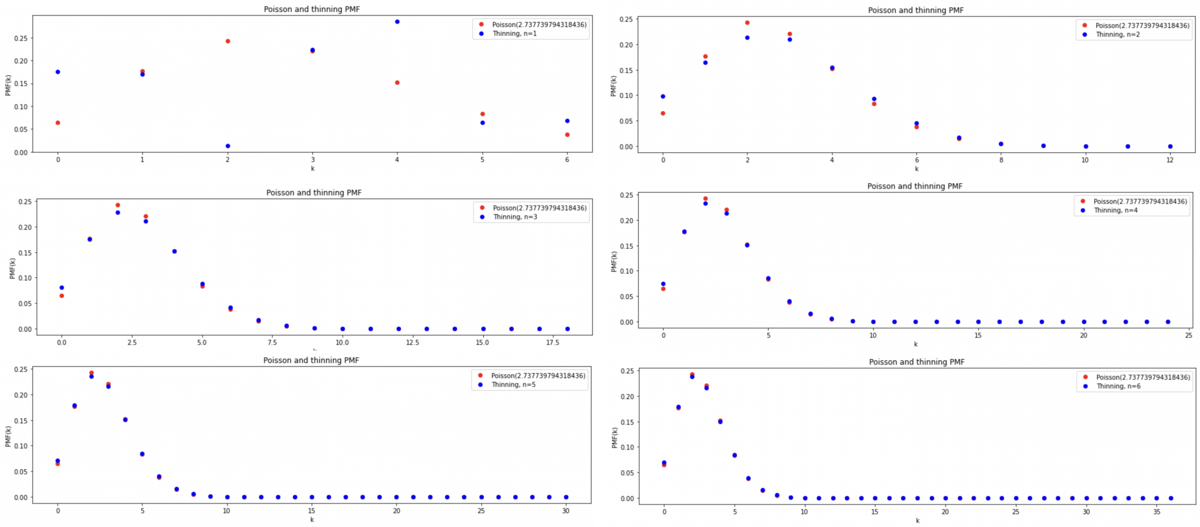

Figure 1: Convergence of $T_{\frac{1}{n}}p$ to ρ^λ as n grows

We also observe that $W_1(\rho^\lambda, T_{\frac{1}{n}}p) \downarrow 0$ as $n \to +\infty$:

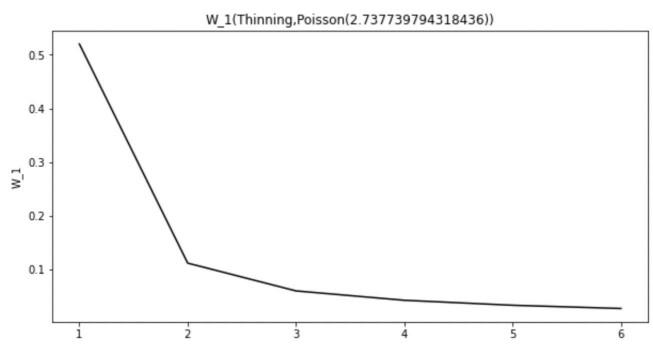

Figure 2: $W_1(\rho^\lambda, T_{\frac{1}{n}}p)$, $n = 1, ..., 6$

Now, we will be analyzing the behaviour of W_1 distance between two arbitrary thinning. We consider two sequences of i.i.d. random variables Y_i and Z_i, $i = 1, ..., n$, $n = 1, ..., 6$, $supp\{Y_i\} = supp\{Z_i\} = \{0, ..., 4\}$, $Y_i \perp\!\!\!\perp Z_i$, such that Y_i has PMF p_1 and Z_i has PMF p_2, where

$$p_1 = [0.09403830825484012, 0.08390511435409455, 0.47178599996427194, 0.1337160743029783, 0.21655450344536767],$$
$$p_2 = [0.1615592, 0.29216927, 0.02648327, 0.39563212, 0.12415614].$$

Both p_1 and p_2 are not ULC or ULCX, and $\mathbb{E}[Y_i] = 2.294843350329939$, $\mathbb{E}[Z_i] = 2.1124084377431873$. As n grows, we have the following evolution of the thinning:

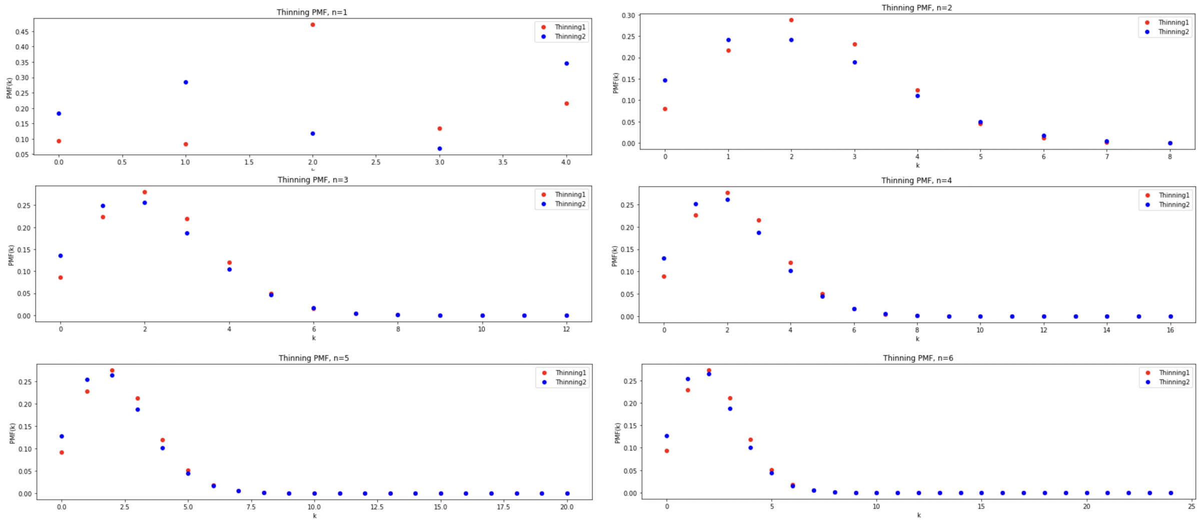

Figure 3: Convergence of $T_{\frac{1}{n}}p_1$ and $T_{\frac{1}{n}}p_2$ as n grows

Visually, we observe that $T_{\frac{1}{n}}p_1$ converges to $\rho^{2.29}$ and $T_{\frac{1}{n}}p_1$ converges to $\rho^{2.11}$ as n grows. For the Wasserstein distance between these two thinning we have the following observation:

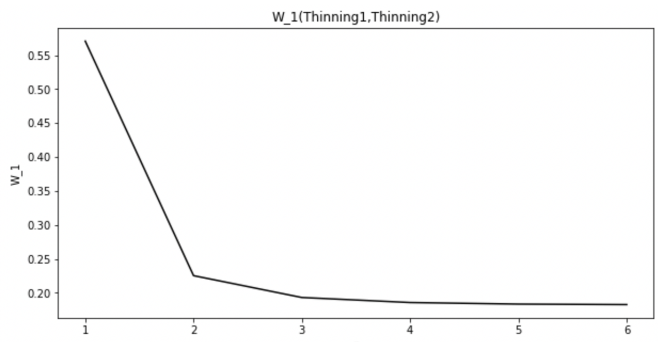

Figure 4: $W_1(T_{\frac{1}{n}}p_1, T_{\frac{1}{n}}p_2)$, $n = 1, ..., 6$

So, the distance is still monotically decreasing to some constant even though Y_i are not ULC and Z_i are not ULCX, $i = 1, ..., n$.

6.5.1 Interpretation in the context of Caffarelli's theorem

Thanks to Proposition 6.7, we know that the function $\bar{f}$ on which W_1 distance between ULCX and ULC measures attains its maximum is not only 1-Lipschitz but convex. It means that without a loss of generality a typical optimal function can be represented as follows: there exists $a \in \mathbb{Z}_+$ and $b \in \mathbb{Z}_+$, $a \leq b$, such that

$$\bar{f}(x) = \begin{cases} a - x, & x \leq a, \\ 0, & a \leq x \leq b, \\ x - b, & x \geq b. \end{cases}$$

Recall, that a transport map $T : \mathbb{Z}_+ \to \mathbb{Z}_+$ between $\mu \in \mathcal{P}(\mathbb{Z}_+)$ ULC and $\nu \in \mathcal{P}(\mathbb{Z}_+)$ ULCX is optimal if

$$W_1(\mu, \nu) = \int |n - T(n)| d\mu(n).$$

Using 1-Lipschitzianity of $\bar{f}$ and definition of a transport map, we write:

$$\begin{aligned} W_1(\mu,\nu) &= \int \bar{f}d(\mu-\nu) = \int (\bar{f}(n) - \bar{f}(T(n)))d\mu(n) \\ &\le \int |n - T(n)|d\mu(n) = W_1(\mu,\nu). \end{aligned}$$

Therefore, we have

$$\bar{f}(n) - \bar{f}(T(n)) = |n - T(n)|, \quad n \in \mathbb{Z}_+.$$

Now, there are three different cases based on the value of the function $\bar{f}$:

- If $n \le a$, then

$$\begin{aligned} a - n - a + T(n) = |n - T(n)| &\implies -n + T(n) = |n - T(n)| \\ &\implies T(n) - n \ge 0; \end{aligned}$$

- If $a \le n \le b$, then

$$0 = |n - T(n)| \implies T(n) - n = 0;$$

- If $n \ge b$, then

$$\begin{aligned} n - b - T(n) + b = |n - T(n)| &\implies n - T(n) = |n - T(n)| \\ &\implies T(n) - n \le 0. \end{aligned}$$

Thus, the sign of the difference $T(n) - n$ is the only information we have gotten about the transport map, and it does not tell us a lot about its more sophisticated properties.

7 Conclusions and open questions

In this report we have studied Caffarelli's contraction theorem and have developed some theory towards adapting this theorem to the Poisson measure. We made a short introduction to optimal transport and entropic optimal transport. We have seen how Caffarelli's theorem can be applied to functional inequalities. We have seen how one can prove Caffarelli's theorem via entropic regularization. For this we introduced the Orstein-Uhlenbeck process $(Z_t)_{t\ge 0}$ and took the law of (Z_0, Z_ϵ) as a reference measure for entropic transport cost. We showed that Orstein-Uhlenbeck semigroup preserves log-convexity and log-concavity. We proved that the Schrödinger potentials in the representation of the optimal coupling between a log-convex and a log-concave measures are convex and concave respectively. We proved monotonicity property of the entropic cost. Using known result of convergence of the regularized by ϵ entropic cost to W_2 distance as $\epsilon \to 0$, we showed the same monotonicity property for W_2 distance for measures where one is dominated by the other in convex order.

We have tried to adapt Caffarelli's result for the Poisson measure. For this we introduced the $M/M/\infty$ process $(N_t)_{t\ge 0}$ and took the law of (N_0, N_ϵ), $\epsilon > 0$, as a reference measure for the entropic transport cost. We proved that N_t stays log-concave if at time $t = 0$ the process was log-concave, using the facts that both log-concavity and ultra log-concavity are preserved by convolution. We studied the $M/M/\infty$ semigroup, in particular, we proved that it preserves log-convexity and log-concavity. We proved that the functions in the representation of the optimal coupling between ultra log-concave and ultra log-convex measures are log-convex and log-concave respectively. Then, we showed that the entropic transport cost regularized by $-1/\log\epsilon$ converges to W_1. We showed that similarly to the Gaussian measure case entropic transport cost is monotone in measure that admits a convex domination in order by other measure. Finally, we proved, thanks to the convergence of the regularized entropic cost, the same monotonicity property for W_1 distance. We also studied monotonicity of W_1 when measures are given by a thinning of discrete random variables. We tried to understand if the conditions for monotonicity can be relaxed by performing Python simulations.

This report does not give an adapted Caffarelli's theorem for the Poisson measure but it gives a better understanding of what next steps can be. For example, as we have seen in Section 6.5.1, knowing the dual form of W_1 for ULC measures and the monotonicity between ULC and ULCX measures does not give us a lot of information about transport. One way to analyze this problem is to look at the entropic tranport cost instead of Wasserstein distance. So, the question about a version of Caffarelli's theorem for Poisson measure still remains open.

Another unanswered question is validity of Theorem 6.8 and 6.9 under relaxed conditions. As we have seen with our simulations, it seems that ULC and ULCX conditions are not required but the possibility of the mistake in the code is not excluded, and so the question of finding a counterexample or proving that it does not exist remains open.

We note that originally Caffarelli's theorem was formulated for d-dimensional measures while we were studying only 1-diminsional Poisson measure. A question of possible interest is to adapt the developed method for d-dimensional Poisson measure. This problem is much more delicate as we have to understand the convexity of discrete functions in higher dimensions. Discrete convex analysis is well developed in [16] by Kazuo Murota; it involves introducing other notions of convexity such as $M^\natural$-convexity, M-convexity, $L^\natural$-convexity, L-convexity, separable convexity. One would have to understand what are the analogous of Walkup theorem and other theorems that are among the main ingredients for proving the log-concavity results.

Acknowledgements

I want to thank my supervisor Nathael Gozlan for attentive advising, productive discussions and for guidance into various mathematical topics. I also want to thank the lab MAP5 for a warm welcome.

References

[1] Shiri Artstein et al. "Solution of Shannon's problem on the monotonicity of entropy". In: *Artstein, S. and Ball, K.M. and Barthe, F. and Naor, A. (2004) Solution of Shannon's problem on the monotonicity of entropy. Journal of the American Mathematical Society, 17 (4). pp. 975-982. ISSN 08940347* 17 (Oct. 2004). DOI: 10.1090/S0894-0347-04-00459-X.

[2] Sergey G. Bobkov and Michel Ledoux. "From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities". In: *Geometric and Functional Analysis* 10.5 (2000), pp. 1028–1052. DOI: 10.1007/PL00001645. URL: https://doi.org/10.1007/PL00001645.

[3] Luis Caffarelli. "Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities". In: *Communications in Mathematical Physics* 214 (Jan. 2000), pp. 547–563. DOI: 10.1007/s002200000257.

[4] Guillaume Carlier et al. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418. DOI: 10.1137/15M1050264. eprint: https://doi.org/10.1137/15M1050264. URL: https://doi.org/10.1137/15M1050264.

[5] Sinho Chewi and Aram-Alexandre Pooladian. "An entropic generalization of Caffarelli's contraction theorem via covariance inequalities". In: (2022). arXiv: 2203.04954 [math.PR].

[6] Dario Cordero-Erausquin. "Some Applications of Mass Transport to Gaussian-Type Inequalities". In: *Archive for Rational Mechanics and Analysis* 161.3 (2002), pp. 257–269. DOI: 10.1007/s002050100185. URL: https://doi.org/10.1007/s002050100185.

[7] Claude Dellacherie, Paul-André Meyer, and Miguel Weil, eds. *Séminaire de Probabilités XIII*. Nov. 2006, p. 647. DOI: 10.1007/BFb0070846.

[8] Max Fathi, Nathael Gozlan, and Maxime Prod'homme. "A proof of the Caffarelli contraction theorem via entropic regularization". In: *Calculus of Variations and Partial Differential Equations* 59.3 (2020), p. 96. DOI: 10.1007/s00526-020-01754-0. URL: https://doi.org/10.1007/s00526-020-01754-0.

[9] Alessio Figalli, Federico. Glaudo, and European Mathematical Society Publishing House ETH-Zentrum SEW A27. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. EMS textbooks in mathematics. EMS Press, 2021. ISBN: 9783985470105. URL: https://books.google.fr/books?id=WhiUzgEACAAJ.

[10] Nathael Gozlan and Nicolas Juillet. "On a mixture of Brenier and Strassen Theorems". In: *Proceedings of the London Mathematical Society* 120 (Aug. 2018). DOI: 10.1112/plms.12302.

[11] Nathael Gozlan et al. "Kantorovich duality for general transport costs and applications". In: *Journal of Functional Analysis* 273.11 (2017), pp. 3327–3405. ISSN: 0022-1236. DOI: https://doi.org/10.1016/j.jfa.2017.08.015. URL: https://www.sciencedirect.com/science/article/pii/S0022123617303294.

[12] Peter Harremoës, Oliver Johnson, and Ioannis Kontoyiannis. “Thinning and the Law of Small Numbers”. In: *Journal of Knot Theory and Its Ramifications - JKTR* (July 2007), pp. 1491–1495. DOI: 10.1109/ISIT.2007.4557433.

[13] Oliver Johnson. “Log-concavity and the maximum entropy property of the Poisson distribution”. In: *Stochastic Processes and their Applications* 117.6 (June 2007), pp. 791–802. DOI: 10.1016/j.spa.2006.10.006. URL: https://doi.org/10.1016%2Fj.spa.2006.10.006.

[14] Christian Léonard. “A survey of the Schrödinger problem and some of its connections with optimal transport”. In: (2013). arXiv: 1308.0215 [math.PR].

[15] Thomas M. Liggett. “Ultra Logconcave Sequences and Negative Dependence”. In: *Journal of Combinatorial Theory, Series A* 79.2 (1997), pp. 315–325. ISSN: 0097-3165. DOI: https://doi.org/10.1006/jcta.1997.2790. URL: https://www.sciencedirect.com/science/article/pii/S0097316597927904.

[16] Kazuo Murota. “Discrete Convex Analysis: A Tool for Economics and Game Theory”. In: (2022). arXiv: 2212.03598 [math.CO].

[17] Piotr Nayar and Krzysztof Oleszkiewicz. “Khinchine type inequalities with optimal constants via ultra log-concavity”. In: *Positivity* 16.2 (2012), pp. 359–371. DOI: 10.1007/s11117-011-0130-z. URL: https://doi.org/10.1007/s11117-011-0130-z.

[18] Sehie Park. “A simple proof of the Sion minimax theorem”. In: *Bulletin of the Korean Mathematical Society* 47 (Sept. 2010). DOI: 10.4134/BKMS.2010.47.5.1037.

[19] Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton: Princeton University Press, 1970. ISBN: 9781400873173. DOI: doi:10.1515/9781400873173. URL: https://doi.org/10.1515/9781400873173.

[20] Cédric Villani. “Optimal Transport: Old and New”. In: 2008. URL: https://api.semanticscholar.org/CorpusID:118347220.

[21] Cédric Villani. “Topics in Optimal Transportation Theory”. In: 58 (Jan. 2003). DOI: 10.1090/gsm/058.

[22] Yaming Yu. “Monotonic Convergence in an Information-Theoretic Law of Small Numbers”. In: *IEEE Transactions on Information Theory* 55.12 (Dec. 2009), pp. 5412–5422. DOI: 10.1109/tit.2009.2032727. URL: https://doi.org/10.1109%2Ftit.2009.2032727.