African Journal of Pure & Applied Sciences

ISSN: 2710-0758
https://doi.org/10.33886/ajpas.v7i1.821
Journal homepage: http://journal.ku.ac.ke/index.php/AJPAS

Research Article Open Access

Optimal Transport and Convexity in Orlicz Spaces

Evans Mogoi and Robert Obogi

Department of Mathematics and Actuarial Science
Kisii University, Kenya robogi@kisiiuniversity.ac.ke

Corresponding author: Evans Mogoi: Email: mogoievans4020@gmail.com

Received 17th February, 2026; Accepted 28th May, 2026; Published online 30th June, 2026

How to Cite:
Obogi, R., & Mogoi, E. N. (2026). Optimal Transport and Convexity in Orlicz Spaces. *African Journal of Pure and Applied Sciences, 7*(1), 70-77. https://doi.org/10.33886/ajpas.v7i1.821

ABSTRACT

Optimal transport theory has traditionally been developed in the context of classical L^p spaces, where Wasserstein metrics provide a natural geometry on the space of probability measures with finite p-moments. However, in many applications-such as in finance, statistical mechanics, and the modelling of heavy-tailed phenomena-classical L^p spaces prove insufficient. This paper proposes a novel generalization of optimal transport to Orlicz spaces, which naturally extend the L^p framework by accommodating more general growth conditions via Young functions. We introduce and rigorously define the *Wasserstein-Orlicz metric* and explore its topological and geometric properties on spaces of probability measures with finite Orlicz norm. Key results include the extension of Kantorovich duality to the Orlicz setting, as well as the derivation of convexity and lower semicontinuity properties of entropy-type functionals under the Wasserstein-Orlicz geometry. Applications are presented in the context of stochastic processes with heavy-tailed distributions, where the classical assumptions of finite p-moments are violated. The proposed framework opens a new avenue for analysing optimal transport problems in nonstandard settings and provides tools for applications in areas where data exhibits non-Gaussian behaviour.

Keywords: Optimal Transport, Orlicz Spaces, Wasserstein-Orlicz Metric, Convexity, Kantorovich Duality, and Heavy-Tailed Distributions.

INTRODUCTION

The study of optimal transport has undergone remarkable developments in recent decades, with profound implications across mathematics, statistics, and data science [12, 9]. This paper investigates the interplay between Orlicz spaces and Wasserstein distances, introducing a novel framework for probability measures with heavy-tailed distributions where classical L^p moments may not exist [7, 10].

Theoretical Foundations

Our work builds on three fundamental pillars:

1. **Geometric Analysis in Measure Spaces**: Following the seminal gradient flow perspective developed in [1], we extend the theory of geodesics to probability measures equipped with Orlicz norms. The recent advances in weak transport costs [4] and displacement convexity [11] provide crucial technical tools.
2. **Orlicz Space Theory**: The structural properties of Orlicz spaces, as comprehensively treated in [10], allow us to handle growth conditions more general than polynomial tails. This is particularly relevant for heavy-tailed distributions in stochastic modelling [6].
3. **Computational Optimal Transport**: While our focus is theoretical, we maintain connections to computational methods through the Sinkhorn divergence framework [5] and the numerical analysis perspective in [3].

Key Contributions

Our main results advance the field in several directions:

- We establish existence and uniqueness of geodesics in Wasserstein-Orlicz space (Theorem 3), generalizing the L^p theory from [1] to Orlicz norms. The proof leverages measurable selection theorems and the Δ_2-condition as in [8].
- The dual formulation (Theorem 4) extends Kantorovich duality via Fenchel Rockafellar theory, with particular attention to the Orlicz-Legendre transform as developed in [10].
- For heavy-tailed models (Theorem 10), we demonstrate how Wasserstein-Orlicz metrics remain finite where classical W_p diverges, complementing recent work on nonparametric testing [6].
- The topological properties (Theorem 5), lower semicontinuity of Orliczentropy functionals (Theorem 6), stability under perturbations (Theorems 7-8), compactness criteria (Theorem 9), and convexity of optimal transport paths (Theorem 10) further establish the robustness of the Wasserstein-Orlicz framework.

Applications and Implications

The theoretical framework developed here has immediate consequences for:

- **Stochastic Processes**: Our results enable rigorous analysis of transport distances between heavy-tailed distributions, addressing limitations noted in [7].
- **Functional Inequalities**: The displacement convexity results (Theorem 9) extend the McCann theory [11] to Orlicz-entropy functionals.
- **Statistical Learning**: The stability theorems (Theorems 7-8) provide foundations for robust inference in high-dimensional settings, connecting to recent advances in [9].

Preliminaries

Orlicz Spaces and Young Functions

Definition 1 (Young Function). *A function* $\Phi : [0,\infty) \rightarrow [0,\infty]$ *is called a* Young function *if it is convex, lower semicontinuous, and satisfies:*

- $\Phi(0) = 0$
- $\lim\limits_{t\to\infty} \Phi(t) = \infty$
- $\Phi \not\equiv 0$ *and* $\Phi \not\equiv \infty$ *on* $(0,\infty)$

We say Φ *satisfies the* Δ_2*-condition if* $\exists\, C > 0$ *and* $t_0 \geq 0$ *such that* $\Phi(2t) \leq C\Phi(t)$ *for all* $t \geq t_0$ *[10].*

Definition 2 (Orlicz Space). *For a Young function* Φ *and measure space* (X, B, λ), *the* Orlicz space $L^\Phi(X)$ *is:*

$$L^\Phi(X) = \left\{ f : X \rightarrow \mathbb{R}\ measurable : \int_X \Phi\left(\frac{|f(x)|}{\lambda}\right) d\lambda < \infty \right\}$$

equipped with the Luxemburg norm:

$$\| f \|_\Phi = \inf\left\{ \lambda > 0 : \int_X \Phi\left(\frac{|f(x)|}{\lambda}\right) d\lambda < \infty \right\}$$

Wasserstein Metrics

Definition 3 (p-Wasserstein Distance). *For* $p \geq 1$ *and* $\mu, \nu \in P_p(X)\ \mu, \nu \in P_p(X)$ *(probability measures with finite* $p-th\ p-th$ *moments):*

$$W_p(\mu,\nu) = \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} d(x,y)^p \, d\gamma(x,y) \right)^{1/p}$$

where $\Gamma(\mu,\nu)$ *denotes the set of couplings [12].*

Definition 4 (Wasserstein-Orlicz Distance). *Given a Young function* Φ, *the* Wasserstein-Orlicz distance *is:*

$$W_\Phi(\mu,\nu) = \inf\left\{ \lambda > 0 : \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} \left(\frac{d(x,y)}{\lambda}\right) d\gamma(x,y) \leq 1 \right\}$$

This generalizes both W_p *(when* $\Phi(t) = t^p$ *) and classical Orlicz norms [1].*

Key Functional Inequalities

Theorem 1 (Kantorovich Duality). *For lower semicontinuous cost* $c : X \times X \rightarrow [0,\infty]$:

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int c\, d\gamma = \sup_{\substack{\varphi,\psi\in L^1(\mu)\times L^1(\nu) \\ \varphi(x)+\psi(y)<c(x,y)}} \left(\int \varphi d\mu + \int \psi d\nu \right)$$

The Orlicz-space version holds when $c(x,y) = \Phi(d(x,y))$ *with* $\varphi, \psi \in L^{\Phi^*}$ *[11].*

Theorem 2 (Displacement Convexity). *For* Φ *strictly convex and* μ_t *a* W_Φ *geodesic:*

$$H_\Phi(\mu_t) \leq (1-t)H_\Phi(\mu_0) + t\, H_\Phi(\mu_1)$$

where $H_\Phi(\mu) := \int \Phi(d\mu/d\lambda)d\lambda$ *is the Orlicz entropy [1].*

Technical Tools

Lemma 1 (Δ_2-Condition Implications). *For Φ satisfying Δ_2 :*

- *L^{Φ} is reflexive and separable*
- *Strong and weak topologies coincide on bounded sets*
- *Uniform integrability: $\sup_n \int \Phi(f_n)d\lambda \leq C \Rightarrow \{f_n\}$ uniformly integrable [10]*

Lemma 2 (Geodesic Spaces). *A metric space (X, d) is* geodesic *if $\forall x, y \in X, \exists$ a curve $\gamma : [0,1] \to X$ with:*

$d\left(\gamma(s), \gamma(t)\right) = |s - t| d(x, y), \gamma(0) = x, \gamma(1) = y$

For such spaces, $P_{\Phi}(X)$ inherits geodesic structure [8].

Lemma 3 (Prokhorov's Theorem). *A family $K \subset P(X)$ is tight iff it is relatively compact in the weak topology. For $\Phi - \Phi -$bounded sets:*

$\sup_{\mu \in K} \int \Phi\left(d(x, x0)\right) d\mu \leq C \Rightarrow K$ *tight* [2].

Notation Table

$P(X)$	Probability measures on X
$P\Phi(X)$	$\{\mu \in P(X) : \int \Phi(d(x, x_0))d\mu < \infty\}$
$\Gamma(\mu, \nu)$	Couplings with marginals μ, ν
Φ^*	Legendre transform of Φ
e_t	Evaluation map $\gamma \mapsto \gamma(t)$
Δ_2	Doubling condition for Φ
$H_{\Phi}(\mu)$	Orlicz entropy functional $\int \Phi(d\mu/d\lambda)d\lambda$
$\Gamma_{opt}^{\Phi}(\mu, \nu)$	Set of optimal transport plans for W_{Φ}
ρ_t	Density of μ_t with respect to reference measure

RESULTS AND DISCUSSIONS

Theorem 3 (Existence and Uniqueness of Wasserstein-Orlicz Geodesics). *Let $P_{\Phi}(X)$ be the space of probability measures with finite Orlicz norm on a Polish space X. Assume Φ is a Young function satisfying the Δ_2-condition. Then:*

(i) For any two measures $\mu, \nu \in P_{\Phi}(X)$, there exists a geodesic curve $(\mu_t)_{t\in[0,1]}$ in $P_{\Phi}(X)$ connecting μ and ν with respect to the Wasserstein-Orlicz metric W_{Φ}.

(ii) If Φ is strictly convex, the geodesic is unique.

Proof. Let (X, d) be a complete, separable, geodesic metric space and let $\Phi :$ $[0, \infty) \to [0, \infty)$ be a Young function satisfying the Δ_2-condition. The Wasserstein-Orlicz distance between $\mu, \nu \in P_{\Phi}(X)$ is defined by

$$W_{\Phi}(\mu, \nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \inf\left\{\lambda > 0 : \int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\gamma(x, y) \leq 1\right\},$$

where $\Gamma(\mu, \nu)$ is the set of transport plans with marginals μ and ν.

Step 1: Existence of an optimal plan. Since Φ satisfies the Δ_2-condition, the topology induced by the Orlicz norm is compatible with weak convergence and moments of order Φ. Consider a minimizing sequence (γ_n) in $\Gamma(\mu, \nu)$ such that

$$\int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda_n}\right) d\gamma_n(x, y) \leq 1$$

$$\lambda_n \to W_{\Phi}(\mu, \nu)\lambda_n \to W_{\Phi}(\mu, \nu)$$

By Prokhorov's theorem, (γ_n) is tight in $P(X \times X)$, hence has a narrowly convergent subsequence $(\gamma_{nk}) \to \gamma$. Lower semicontinuity of $\gamma \to \int \Phi(d(x, y)/\lambda)d\gamma$ yields

$$\int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\gamma(x, y) \leq 1$$

for $\lambda = W_{\Phi}(\mu, \nu). (\mu, \nu)$. Thus γ is optimal.

Step 2: Displacement interpolation. Let γ be an optimal plan. Since X is geodesic, for each (x, y) there exists a constant-speed geodesic $\sigma_{x,y} : [0,1] \to X$ with $\sigma_{x,y}(0) = x$ *and* $\sigma_{x,y}(1) = y$. By the measurable selection theorem, we can choose $(x, y) 7 \to \sigma_{x,y}$ measurably. Define

$$\mu_t = (e_t)_{\#}\Pi, \quad \Pi = \int_{X\times X} \delta_{\delta\sigma_{x,y}} d\gamma(x, y),$$

where $e_t(\sigma) = \sigma(t)$. This $(\mu_t)_{t\in[0,1]}$ is called the *displacement interpolation* between μ and ν.

Step 3: Geodesic property. By constant-speed parametrization of $\sigma_{x,y}$, we have

$$d(\sigma_{x,y}(s), \sigma_{x,y}(t)) = |s - t| d(x, y).$$

Using the definition of W_{Φ} and Jensen's inequality applied to Φ, one obtains

$$W_{\Phi}(\mu_s, \mu_t) = |s - t| W_{\Phi}(\mu, \nu).$$

showing (μ_t) is a geodesic in $(P_\Phi(X), W_\Phi)$.

Step 4: Uniqueness under strict convexity. If Φ is strictly convex, then the cost functional

$$\gamma \to \int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\gamma(x,y) \le 1$$

is strictly convex in γ. Thus, the optimal transport plan γ is unique. In a nonbranching space, the displacement interpolation determined by γ is also unique. Hence the Wasserstein-Orlicz geodesic between μ and ν is unique.

Example 1 (Power-type case $\Phi(t) = t^p$, $p > 1$)**.** *Here* $P_\Phi(X)$ *coincides with* $P_p(X)$*, the space of measures with finite* $p-th$ *moment, and*

$$W_\Phi(\mu,\nu) = \left(\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} d(x,y)^p \, d\gamma(x,y)\right)^{1/p}$$

If $X = R^n$ *and* $\mu = \delta x_0, \nu = \delta x_1$ *are Dirac masses, the optimal plan is* $\gamma = \delta(x_0, x_1)$ *, and the geodesic is* $\mu_t = \delta_{(1-t)x_0+tx_1}$ *. Strict convexity of* $\Phi(t) = t^p$ *ensures uniqueness.*

Theorem 4 (Dual Formulation of Wasserstein-Orlicz Distance)**.** *Let Φ be a lower semicontinuous, strictly convex Young function and let* W_Φ *be the Wasserstein Orlicz metric on* $P_\Phi(X)$*. Then:*

(i) *The optimal transport cost admits a dual representation involving Orlicz Legendre transforms.*

(ii) *The duality holds with equality under mild regularity conditions on the cost function.*

Proof. Let Φ be a strictly convex, lower semicontinuous Young function and let Φ^* denote its convex conjugate:

$$\Phi^*(s) = \sup_{t\ge0}\{st - \Phi(t)\} \ .$$

Let $c(x,y) := \Phi(d(x,y))$ be the transport cost.

Step 1: Primal formulation. The Wasserstein-Orlicz distance is given by

$$W_\Phi(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \inf\left\{\lambda > 0 : \int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\gamma(x,y) \le 1\right\}.$$

Step 2: Kantorovich duality with Orlicz costs. By Fenchel-Rockafellar duality applied to the convex functional

$$F(\gamma)\int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\gamma(x,y)$$

subject to marginal constraints, we obtain

$$\inf_{\gamma\in\Gamma(\mu,\nu)} c(x,y)\, d\gamma(x,y) = \sup_{\substack{\varphi,\psi\in L^1(\mu)\times L^1(\nu)\\ \varphi(x)+\psi(y)<c(x,y)}} \left(\int_X \varphi d\mu + \int_X \psi d\nu\right)$$

Here $L^{\Phi*}$ is the Orlicz space associated to Φ^*.

Step 3: Equality and regularity. Under the assumptions that Φ is lower semicontinuous and strictly convex and that c is continuous and finite, the supremum is attained and the duality holds with equality. The maximizing (φ, ψ) are Φ^*-integrable and satisfy

$\varphi(x) + \psi(y) = c(x,y)\gamma - a.e.$ for any optimal γ.

This is the desired dual representation of W_Φ.

Example 2 (Dual potentials for $\Phi(t) = t^p$, $p > 1$)**.** *Here* $\Phi^*(s) = \frac{s^q}{q}$ *where q is the conjugate exponent* ($1/p + 1/q = 1$)*. The Kantorovich dual formulation reduces to*

$$W_p^p(\mu,\nu) = \sup_{\varphi,\psi}\left\{\int \varphi d\mu + \int \psi d\nu : \varphi(x) + \psi(y) \le d(x,y)^p\right\} \cdot$$

If $\mu = \delta_0$ *and* $\nu = \delta_a$ *on* $\mathbb{R}$*, an optimal choice is* $\varphi(x) = -|x-a|^p$ *and* $\psi(y) =$ $|a|^p$*, giving* $W_p(\delta_0, \delta_a) = |a|$.

Theorem 5 (Topological Properties of $P_\Phi(X)$)**.** *Let X be a separable metric space and Φ a convex, increasing Young function satisfying the* Δ_2*-condition. Then:*

(i) $P_\Phi(X)$ *is a complete and separable metric space under* W_Φ.

(ii) *Convergence in* W_Φ *implies convergence in distribution.*

Proof. We begin by recalling the *Orlicz-Wasserstein distance* associated with a Young function Φ:

$$W_\Phi(\mu,\nu) = \inf\left\{\lambda > 0 : \exists\pi \in \Pi(\mu,\nu) \ s.t \int_{X\times X} \Phi\left(\tfrac{d(x,y)}{\lambda}\right) d\pi(x,y) \le 1\right\},$$

where $\prod(\mu,\nu)$ is the set of couplings of μ and ν, and d is the metric on X. Under the assumptions:

1. X is separable and metric.
2. Φ is convex, increasing, and satisfies the Δ_2-condition (ensuring the associated Orlicz space L^Φ is reflexive and separable).

(i) Completeness and separability. Separability of $(P_\Phi(X), W_\Phi)$ follows from:

The separability of X implies separability of $P_p(X)$ for every $p \ge 1$ under the p-Wasserstein metric.

The Δ_2-condition ensures that W_Φ is topologically equivalent to some W_p (see [1, Theorem. 3.1] adapted to Orlicz growth), hence separability carries over.

For completeness, let (μ_n) be a W_Φ-Cauchy sequence. By the definition of W_Φ, uniform integrability of $\Phi(d(x,x_0))$ for some $x_0 \in X$ follows:

$$\sup_n \int_X \Phi\left(d\left(x, x_0\right)\right) d\mu_n(x) < \infty \ .$$

This guarantees tightness (by the de la Vallee Poussin criterion). By Prokhorov's theorem, (μ_n) is relatively compact in the topology of weak convergence, so there exists μ such that $\mu_{n_k} \rightharpoonup \mu$. The uniform Φ-moment bound plus weak convergence imply $W_\Phi(\mu_{n_k}, \mu) \to 0$ (cf. [8]), hence (μ_n) converges in W_Φ. Thus $P_\Phi(X), W_\Phi$ is complete.

(ii) W_Φ-convergence ⇒ convergence in distribution. Let $\mu_n \to \mu$ in W_Φ. Then:
$W_\Phi(\mu_n, \mu) \to 0 \Rightarrow \{\mu_n\}$ is tight and has uniformly bounded Φ-moments.

By the tightness criterion, there exists a subsequence $\mu_{n_k} \rightharpoonup \nu$. Lower semicontinuity of W_Φ implies $W_\Phi(\nu, \mu) \leq \liminf W_\Phi(\mu_n, \mu) = 0$, hence $\nu = \mu$. Since every subsequence has a further subsequence converging weakly to μ, we have $\mu_n \rightharpoonup \mu$.

Theorem 6 (Lower Semicontinuity of Orlicz-Entropy Functionals)**.** *Let $\mu_n \to \mu$ in W_Φ and let $H_\Phi(\mu) = \int \Phi\left(\frac{d\mu}{d\lambda}\right) d\lambda$ denote the Orlicz entropy relative to a reference measure λ. Then:*

(i) H_Φ is lower semicontinuous with respect to W_Φ-convergence.

(ii) If Φ is strictly convex, H_Φ is also strictly convex.

Proof. Let λ be a σ-finite reference measure on X and define the *Orlicz entropy functional*:

$$H_\Phi(\mu) = \begin{cases} \int_X \Phi\left(\frac{d\mu}{d\lambda}\right) d\lambda, \mu \leq \lambda, \\ +\infty, \quad otherwise \end{cases}$$

We assume Φ is convex, increasing, and satisfies the Δ_2-condition.

(i) Lower semicontinuity. Let $\mu_n \to \mu$ in W_Φ. Then $\mu_n \rightharpoonup \mu$ and the family $\{d\mu_n/d\lambda\}$ is uniformly integrable in $L^\Phi(\lambda)$ (by the Φ-moment bound).

The map $u \to \int \Phi(u)d\lambda$ is convex and lower semicontinuous on $L^\Phi(\lambda)$ under the weak topology of L^Φ (cf. [10, Prop. 2.2.6]). Since W_Φ-convergence implies weak convergence in $L^\Phi(\lambda)$ for densities, we have: $H_\Phi(\mu) \leq \liminf_{n\to\infty} H_\Phi \ H\Phi(\mu_n)$.

(ii) Strict convexity. If Φ is strictly convex, then the map $u \to \int \Phi(u)d\lambda$ is strictly convex on $L^\Phi(\lambda)$. If $\mu_0 \neq \mu_1$ are absolutely continuous $w.r.t$. λ with densities f_0, f_1 then for $t \in (0,1)$:
$H_\Phi(t\mu_0) + \int \Phi(tf_0 + tf_1)d\lambda < H_\Phi(t\mu_0) + (1-t)H_\Phi(t\mu_1)$,
where the inequality is strict because $f_0 \neq f_1$ on a set of positive λ-measure. Thus, H_Φ is strictly convex.

Theorem 7 (Wasserstein-Orlicz Stability Under Perturbations)**.** *Let $\{\mu_n\} \subset P_\Phi(X)$ and let Φ satisfy the Δ_2-condition. Then:*

(i) If $\mu_n \to \mu$ in W_Φ, then any Φ-bounded functional $F : P_\Phi(X) \to R$ is continuous at μ. (We note that this notion of Φ-boundedness is sufficient for continuity, but weaker conditions may suffice in specific applications.)

(ii) Small perturbations of Φ induce uniformly small changes in the Wasserstein Orlicz metric.

Proof. We prove parts (i) and (ii) separately.

(i) Continuity of Φ-bounded functionals. Let $F : P_\Phi(X) \to R$ be Φ bounded, meaning that there exists $M > 0$ such that

$$|F(\nu)| \leq M \text{ whenever } \int_X \Phi(d(x,x_0))d\nu(x) \leq C$$

for some $x_0 \in X$ and fixed $C > 0$. Assume $\mu_n \to \mu$ in W_Φ. By definition of the W_Φ metric, this convergence implies:

$$\lim_{n\to\infty} W_\Phi(\mu_n, \mu) = 0$$

which, in turn, ensures (see e.g. [3]) that:

1. $\mu_n \to \mu$ weakly;
2. $\{\mu_n\}$ is uniformly Φ-integrable, i.e.

$$\sup_n \int_X \Phi\left(d\left(x, x_0\right)\right) d\mu_n(x) < \infty$$

The Δ_2-condition on Φ guarantees that Φ-integrability is stable under uniform bounds and weak convergence (see [10]). Since F is Φ-bounded, the uniform Φ-integrability bound ensures $\{F(\mu_n)\}$ is uniformly bounded. Furthermore, the weak convergence $\mu_n \rightharpoonup \mu$ and boundedness imply continuity of F at μ by the dominated convergence principle adapted to Orlicz spaces (cf. [3, Prop. 3.2]). Therefore:

$$\lim_{n\to\infty} F(\mu_n) = F(\mu) \cdot$$

(ii) Stability under perturbations of Φ. Let $\Phi, \tilde{\Phi}$ be two Young functions satisfying the Δ_2-condition and such that

$$\sup_{t\geq 0} \frac{|\Phi(t) - \tilde{\Phi}(t)|}{1 - \Phi(t)} \leq \varepsilon$$

for some small $\varepsilon > 0$. Let W_Φ and $W_{\tilde{\Phi}}$ be the corresponding Wasserstein-Orlicz metrics. For any coupling π between $\mu, \nu \in P_\Phi(X)$, we have:

$$\int_{X\times X}\Phi(d(x,y))d\pi(x,y) \leq (1+\varepsilon)\int_{X\times X}\tilde{\Phi}(d(x,y))d\pi(x,y) + \varepsilon.$$

Taking infima over all couplings and using the definition of W_Φ, we deduce:

$$|W_\Phi(\mu,\nu) - W_{\tilde{\Phi}}(\mu,\nu)| \leq C\varepsilon$$

for some constant C depending on the uniform Φ-moment bound of μ and ν. Thus, small perturbations in Φ yield uniformly small perturbations in the W_Φ metric.

Theorem 8 (Compactness in Wasserstein-Orlicz Space)**.** *Let Φ be a super linear Young function and $K \subset P_\Phi(X)$. Then the set K is relatively compact in the W_Φ-topology if:*

(i) $\{\mu \in K : \int \Phi(d(x,x_0))d\mu(x) \leq C\}$ *for some $x_0 \in X$ and constant $C > 0$;*

(ii) K is tight in the weak topology.

Proof of Theorem 8. Let Φ be a super linear Young function, i.e.

$$\lim_{t\to\infty}\frac{\Phi(t)}{t} = \infty.$$

Let $K \subset P_\Phi(X)$ satisfy (i) and (ii) in the theorem.

Step 1: Uniform Φ-moment bound. From condition (i) there exists $C > 0$ and $x_0 \in X$ such that

$$\sup_{\mu\in K}\int_X \Phi(d(x,x_0))d\mu(x) \leq C.$$

This ensures that K is uniformly Φ-integrable. By the Δ_2-condition (which follows for super linear Φ), this also implies uniform integrability in any smaller Orlicz space corresponding to a Young function $\Psi \ll \Phi$.

Step 2: Tightness in the weak topology. By assumption (ii), K is tight. By Prokhorov's theorem, tightness implies that K is relatively compact in the weak topology of $P(X)$.

Step 3: Upgrade weak convergence to W_Φ convergence. Let $(\mu_n) \subset K$ be a sequence converging weakly to some $\mu \in P(X)$. We claim that $\mu \in P_\Phi(X)$ and $\mu_n \to \mu$ in W_Φ.

Indeed, uniform Φ-moment bounds and weak convergence imply (by [3, Theorem. 3.1]) that

$$\lim_{n\to\infty}\int_X \Phi\left(d\left(x,x_0\right)\right)d\mu_n\left(x\right) = \int_X\Phi(d(x,x_0))d\mu(x)$$

hence $\mu \in P_\Phi(X)$.

The equivalence between weak convergence plus convergence of Φ-moments and W_Φ-convergence holds under the Δ_2-condition on Φ (see [3, Prop. 3.2]). This condition ensures that the Orlicz norm topology is compatible with weak convergence on uniformly Φ-integrable families.

Step 4: Conclusion. Since every sequence in K has a W_Φ-convergent subsequence, K is relatively compact in the W_Φ topology.

Theorem 9 (Convexity of Optimal Transport Paths)**.** *Let Φ be strictly convex and differentiable, and let $(\mu_t)_{t\in[0,1]}$ be the W_Φ-geodesic between $\mu_0, \mu_1 \in P_\Phi(X)$. Then:*

(i) The map $t \to \int_X f(x)d\mu_t(x)$ is convex for every convex $f \in L^\Phi(X)$.

(ii) The entropy functional $\mu \to H_\Phi(\mu)$ is displacement convex along W_Φ geodesics.

Proof. Let $(\mu_t)_{t\in[0,1]}$ be a W_Φ-geodesic between μ_0 and μ_1. By definition of a Wasserstein-Orlicz geodesic, there exists an optimal coupling $\pi \in \Gamma^{\Phi}_{opt}(\mu_0,\mu_1)$ and a measurable map $\gamma : X \times X \times [0,1] \to X$ such that:

$$\mu_t = (\gamma_t)_\#\pi,$$

where $\gamma_t(x,y)$ denotes the constant-speed W_Φ-geodesic between x and y.

(i) Convexity for convex integrands. Let $f \in L^\Phi(X)$ be convex. For πalmost every $(x,y) \in X \times X$, the curve $t \to \gamma_t(x,y)$ is a geodesic in X. By convexity of f in X:

$$f(\gamma_t(x,y)) \leq (1-t)f(x) + tf(y),$$

$t \in [0,1]$.

Integrating with respect to π and using the pushforward relation$\mu_t = (\gamma_t)_\#\pi$ we obtain:

$$\int_X f(z)d\mu_t(z) = \int_{X\times X} f\left(\gamma_t(x,y)\right)d\pi(x,y)$$
$$\leq (1-t)\int_X f(z)d\mu_0 + t\int_X f(z)d\mu_1$$

.

This proves convexity of the map $t \to \int_X f\, d\mu_t$.

(ii) Displacement convexity of entropy. The entropy functional in the Orlicz setting is defined by:

$$H_\Phi(\mu) = \begin{cases}\int_X \Phi\left(\frac{d\mu}{d\lambda}\right)d\lambda, \mu \ll dx,\ \rho = \frac{d\mu}{dx}, \\ +\infty, \quad otherwise\end{cases},$$

Let ρ_t denote the density of μ_t with respect to the reference measure dx. By the continuity equation and the Benamou-Brenier formulation of W_Φ, the evolution (ρ_t) satisfies:

$$\partial_t\rho_t + \nabla\cdot(\rho_t v_t) = 0,$$

where v_t is the optimal velocity field. Unlike the classical L^p case, the Benamou Brenier formulation for W_Φ requires careful justification. However, for strictly convex Φ satisfying the Δ_2-condition, the functional $\rho \to \int_X \Phi(\rho)dx$ is convex. Following the standard argument in [11] (Chapter 17), displacement convexity follows from the convexity of Φ and the fact that W_Φ-geodesics are obtained as pushforwards of optimal couplings. More concretely, for any W_Φ-geodesic μ_t with optimal coupling π, we have $\mu_t = (T_t)_{\#}\mu_0$ where $T_t(x) = \gamma_t(x, y)$ is a displacement interpolation. Then by convexity of Φ and Jensen's inequality:

$$\int \Phi(\rho_t)\, dx = \int \Phi\left(\frac{d(T_t)_{\#}\mu_0}{dx}\right) dx \le (1-t)\int \Phi(\rho_0)\, dx + \int \Phi(\rho_1)\, dx \ ,$$

provided the appropriate change-of-variables formula holds. This yields the desired displacement convexity:

$$H_\Phi(\mu t) \le (1 - t)H_\Phi(\mu 0) + tH_\Phi(\mu 1).$$

Theorem 10 (Application to Heavy-Tailed Stochastic Models)**.** *Let Φ be a regularly varying function at infinity and X a probability space supporting a heavy-tailed stochastic process $\{X_t\}$. Then:*

(i) *The law of X_t belongs to $P_\Phi(X)$ even if $X_t \notin L^p$ for any $p > 0$.*

(ii) *The Wasserstein-Orlicz metric provides finite transport distances between distributions of heavy-tailed processes where classical W_p fails.*

Proof. Let Φ be regularly varying at infinity with index $\alpha > 0$, i.e.,

$$\lim_{x\to\infty} \frac{\Phi(\lambda x)}{\Phi(x)} = \lambda^\alpha, \forall \alpha > 0$$

.

(i) Membership in $P_\Phi(X)$. Suppose $\{X_t\}$ is a heavy-tailed process such that $\mathbb{E}[\Phi(|X_t|)] = \infty$ for all $p > 0$. By regular variation, there exists $\beta > 0$ such that $\Phi(x) = o(x^p)$ as $x \to \infty$ for all $p > \alpha$. If the tail of X_t decays like $x^{-\gamma}$ with $\gamma > \alpha$, then:

$$\mathbb{E}\left[\Phi\left(\left|X_t\right|\right)\right] = \int_0^\infty \Phi(r)\, dF_{|X_t|}(R) < \infty \ ,$$

even though $\mathbb{E}\left[\Phi\left(\left|X_t\right|\right)\right] = \infty$ for all $p > 0$. Thus Law (X_t) $\in P_\Phi(X)$.

(ii) Finite W_Φ where W_p fails. Recall that:

$$W_\Phi(\mu, \nu) = \inf\left\{\lambda > 0 : \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times X} \Phi\left(\frac{d(x,y)}{\lambda}\right) d\pi(x,y) \le 1\right\}.$$

If μ, ν are heavy-tailed with finite Φ-moment but infinite p-moment for every $p > 0$, then the W_p distance is $+\infty$ (since it requires finite p-moment), whereas $W_\Phi(\mu,\nu) < \infty$ because the cost functional only demands integrability with respect to Φ. Hence W_Φ is well-suited for comparing heavy-tailed laws.

Concrete Example: Consider Pareto distributions with CDF $F(x) = 1 - (x_m/x)^\gamma$ for $x \ge x_m > 0$, where $\gamma > 0$ is the tail index. For μ = Pareto(x_m,γ) and $\Phi(t) = \log(1+t)$ (which grows slower than any power), we have $E[\Phi(|X|)] < \infty$ when $\gamma > 1$. However, $E[|X|p] = \infty$ for all $p \ge \gamma$. In this case, $W_\Phi(\mu, \delta_0)$ is finite, while $W_p(\mu, \delta_0) = \infty$ for $p \ge \gamma$. This illustrates the practical utility of the Wasserstein-Orlicz framework.

CONCLUSION

Summary of Contributions

This work has established a comprehensive framework for Wasserstein-Orlicz geometry, with three fundamental advances:

- **Geodesic Structure**: Theorem 3 resolves the existence and uniqueness problem for geodesics in $(P_\Phi(X), W_\Phi)$ extending the classical L^p theory of [1] to general Young functions. Our results show that strict convexity of Φ plays an analogous role to $p > 1$ in the power case.
- **Dual Theory**: The duality framework in Theorem 4 provides the first systematic treatment of Kantorovich duality for Orlicz costs, complementing recent advances in weak transport [4]. The explicit connection with Orlicz-Legendre transforms (Example 2) bridges convex analysis and optimal transport.
- **Heavy-Tailed Applications**: Theorem 10 demonstrates that Wasserstein Orlicz metrics remain effective for distributions beyond $P_p(X)$, addressing limitations noted in [6] for statistical applications.

Theoretical Implications

Our results reveal several fundamental insights:

- The Δ_2-condition (Definition 1) emerges as the critical threshold for good topological properties, mirroring its role in Orlicz space theory [10] but with new probabilistic interpretations.
- Displacement convexity (Theorem 9) holds for general Orlicz entropies, suggesting unified approaches to functional inequalities beyond the L^p setting [11].
- The stability results (Theorems 7-8) show that

Wasserstein-Orlicz metrics are robust to both measure perturbations (i) and cost function deformations (ii), crucial for data science applications [9].

Open Problems and Future Work

- **Numerical Computation**: Can Sinkhorn-type algorithms converge for W_Φ when Φ grows exponentially (e.g., $\Phi(t) = e^{t\alpha} - 1$)? Preliminary work suggests regularization via [5] may fail.
- **Infinite Dimensions**: Does W_Φ metrize weak convergence on $P(C[0,1])$ for path-valued processes? This would extend [8].
- **Statistical Limits**: What are the minimax estimation rates for W_Φ between empirical measures of heavy-tailed data? Known results for W_p [6] likely break down.
- **Breaking Δ_2**: Our theory requires $\Phi(2t) \leq C\Phi(t)$. Can this be relaxed for slowly varying Φ (e.g., $\Phi(t) = e^{\log 1+\epsilon(1+t)}$)?

Limitations

While W_Φ resolves moment restrictions, challenges remain:

- **Topology**: W_Φ may not metrize weak convergence if Φ grows too slowly (e.g., $\Phi(t) = \log(1 + t)$).
- **Computation**: No known polynomial-time algorithms exist for W_Φ with general Φ, unlike W_2's linear programming.
- **Curse of Dimensionality**: Rates for empirical W_Φ estimation likely worsen in high dimensions, analogous to W_p [9].

Remark 1. *The Wasserstein-Orlicz framework developed here provides a versatile toolkit for analysing probability measures across the full spectrum of tail behaviours. By unifying optimal transport with Orlicz space theory, we open new possibilities for both theoretical analysis and practical applications in data science and stochastic modelling.*

REFERENCES

L. Ambrosio, N. Gigli, and G. Savar´e. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkh¨auser, Basel, 2nd edition, 2008.

L. Ambrosio, E. Brue, and D. Semola. *Lectures on Optimal Transport*. Springer, 2020.

L. Ambrosio. *Lecture Notes on Optimal Transport Problems*. Springer, 2021.

J. Backhoff-Veraguas, M. Beiglb¨ock, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 59(6):1-32, 2020.

L. Chizat, G. Peyr´e, B. Schmitzer, and F.-X. Vialard. Faster Wasserstein distance estimation with the Sinkhorn divergence. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 2020.

N. Deb, P. Ghosh, and B. Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association*, 116(536):1923-1937, 2021.

M. N. Evans and R. Obogi. Geometry of Norm Attainability in Orlicz Spaces. *Mathematical Analysis and its Contemporary Applications*, 7(1):7178, 2025.

M. Liero, A. Mielke, and G. Savar´e. Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869-2911, 2018.

G. Peyr´e and M. Cuturi. *Computational Optimal Transport: With Applications to Data Science*. Foundations and Trends in Machine Learning, 11(5-6):355-607, 2019.

M. M. Rao and Z. D. Ren. *Theory of Orlicz Spaces*. Marcel Dekker, New York, 1991.

F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modelling*. Birkh¨auser, 2017.

C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2nd edition, 2021.