# An Entropic Optimal Transport Numerical Approach to the Reflector Problem

Jean-David Benamou, Wilbert Ijzerman, Giorgi Rukhaia

# AN ENTROPIC OPTIMAL TRANSPORT NUMERICAL APPROACH TO THE REFLECTOR PROBLEM

JEAN-DAVID BENAMOU, WILBERT L. IJZERMAN, AND GIORGI RUKHAIA

ABSTRACT. The point source far field reflector design problem is one of the main classic optimal transport problems with a non-euclidean displacement cost [Wang, 2004] [Glimm and Oliker, 2003]. This work describes the use of Entropic Optimal Transport and the associated Sinkhorn algorithm [Cuturi, 2013] to solve it numerically. As the reflector modelling is based on the Kantorovich potentials, several questions arise. First, on the convergence of the discrete entropic approximation and here we follow the recent work of [Berman, 2017] and in particular the imposed discretization requirements therein. Secondly, the correction of the Entropic bias induced by the Entropic OT, as discussed in particular in [Ramdas et al., 2017] [Genevay et al., 2018] [Feydy et al., 2018], is another important tool to achieve reasonable results. The paper reviews the necessary mathematical and numerical tools needed to produce and discuss the obtained numerical results. We find that Sinkhorn algorithm may be adapted, at least in simple academic cases, to the resolution of the far field reflector problem. Sinkhorn canonical extension to continuous potentials is needed to generate continuous reflector approximations. The use of Sinkhorn divergences [Feydy et al., 2018] is useful to mitigate the entropic bias.

## CONTENTS

## 1. INTRODUCTION

1.1. **The optical reflector design problem and its optimal transportation formulation.** A light source, also called "illuminance", is sufficiently small compared to the reflecting surface so that
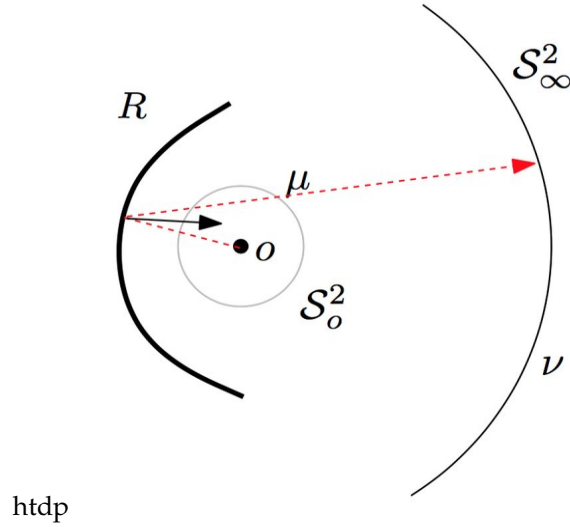
htdp

FIGURE 1.1. Reflector problem from Point source O to Far Field.

it can be regarded as a point in space. It can therefore be modelled as a probability distribution on the sphere, it will be denoted $\mu$ in this paper. The light hits a perfect mirror and we are also given a desired target light distribution, the "illumination" in the far field. From the far field the reflecting surface can be regarded as a point and the illumination again modelled as a probability distribution, denoted $\nu$, on the sphere. Total light conservation is assumed. The design problem is to determine the shape of the mirror which produces the specular reflection from the source to the target distribution. This can be interpreted as the inverse problem of generating some illumination given an illuminance and a reflector (see figure 1.1).

This problem has an elegant mathematical modelization and solution based on the optimal transportation (OT) theory due to [Glimm and Oliker, 2003] and [Wang, 2004]. We briefly recall the main result as presented in [Wang, 2004]. In its Kantorovich primal and dual form (see [Villani, 2008]) :

**Theorem 1** (Kantorovich duality). *Given two compact manifold $X$ and $Y$ endowed with a continuous, bounded from below cost function $c : X \times X \to \mathbf{R}$ and two borel probability measures $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Then, Kantorovich problem in primal and dual forms (1.1) has solutions.*

(1.1)
$$OT(\mu, \nu) := \min_{\gamma \in \Pi(\mu, \nu)} \langle c, \gamma \rangle_{X \times Y} = \max_{f,g \in C} \langle f, \mu \rangle_X + \langle g, \nu \rangle_Y$$

*with respectively primal :*

$$\Pi(\mu, \nu) := \{ \gamma \in \mathcal{P}(X \times Y), \langle 1_X, \gamma \rangle_Y = \nu \ \langle 1_Y, \gamma \rangle_X = \mu \},$$

*and dual :*

$$C = \{ (f, g) \in \mathcal{C}(X) \times \mathcal{C}(Y), f \oplus g \leq c \},$$

*constraints sets*

The notation $\langle f, \alpha \rangle_\Omega$ stands for the duality product $\int_\Omega f \, d\alpha$ between bounded continuous functions $f \in \mathcal{C}(\Omega)$ and probability measures $\alpha \in \mathcal{P}(\Omega)$, $\{ f \oplus g \}(x, y) = f(x) + g(y)$ is the direct sum and $\mu \otimes \nu \in \mathcal{P}(X \times Y)$ the tensor product. Finally $1_\Omega$ is the characteristic function, i.e. a constant 1 on $\Omega$.

Under suitable hypothesis $c$ (as they are technical and satisfied for the costs in this paper, we skip this part), the OT problem is well posed and the the optimal transference plan $\gamma$ is concentrated on

a graph of the OT map $y = T(x)$ implicitly defined by the saturation of the dual constraint :

$$(1.2) \qquad\qquad f(x) + g(T(x)) = c(x, T(x)), \ \mu \ a.e.$$

The pair $(f, g)$ are called the Kantorovich potentials and is unique up to a constant .

By construction $T$ is a measure preserving map characterizing the transport. The measure preserving property is usually denoted $\nu = T\#\mu$ ($T$ pushes forward $\mu$ to $\nu$). The pushforward of $\mu$ is the measure defined as

$$(1.3) \qquad\qquad \nu(A) = T_\#\mu(A) = \mu(T^{-1}(A)) \text{ for all } \nu \text{ measurable subset } A$$

**Remark 1** ($L^p$ Wasserstein metric)**.** *For complete separable metric space X and $L^p$ costs $c := 1/pd^p(x, y)$, this OT problems defines a separable metric on the set of probability measures with finite second moments: the "Wasserstein" distance, which is given by $W_p^p(\mu, \nu) := OT(\mu, \nu)$.*

*This metric metrizes weak convergence of measures is a fundamental tool in image processing (see [Peyré and Cuturi, 2018]).*

In [Wang, 2004], Wang showed that the point source reflector model can be translated to an OT problem. More precisely, he proved the following theorem :

**Theorem 2.** *Let $S_0 \in \mathbb{S}^{d-1}$ and $S_\infty \in \mathbb{S}^{d-1}$ be connected domains in northern and southern hemispheres respectively, $\mu$ and $\nu$ which represent the given illuminance and illumination probability distributions. Then theorem 1 applies to the cost function*

$$(1.4) \qquad\qquad c(x, y) = -\log(1 - x \cdot y).$$

*A transport map T satisfying (1.2) exists and the solution of the corresponding OT problem can be used to build the desired reflector.*

The construction of the reflector can be summarized as follows : Taking the exponential of the dual constraints and the saturation property (1.2) we get

$$(1.5) \qquad\qquad \frac{e^{-g(T(x))}}{1 - x \cdot T(x)} = e^{f(x)} \leq \frac{e^{-g(y)}}{1 - x \cdot y}, \ \mu \otimes \nu \ a.e.$$

We now define in $\mathbb{R}^d$ a family of parabolic reflectors with axis $y \in S_\infty : x \in S_0 \rightarrow P_y(x) := \frac{e^{-g(y)}}{1 - x \cdot y}$.

And directly infer that the reflector shape parameterized over the directions in $S_0$ and given as :

$$(1.6) \qquad\qquad R = \{x\, e^{f(x)} \,|\, x \in S_0\}.$$

Under this choice the map $x \rightarrow T(x)$ can be interpreted as the specular reflection of an optical ray at $R(x)$ onto a parabola of axis $T(x)$ while the illumination and illuminance constraints are enforced by (1.3).

1.2. **Numerical resolution of (1).** A good overview of the available numerical methods applying to the general problem (1) are described in [Peyré and Cuturi, 2018]. We summarize the general classes of methods which can be used, in particular, to tackle the reflector cost (1.4).

*Linear Programming.* The first and obvious one, as already suggested in [Wang, 2004] is the linear programming approach. This assumes that the data in discrete form is given by

$$(1.7) \qquad\qquad \mu_N = \sum_{i=1}^{N} p_i \delta_{x_i}, \quad \nu_N = \sum_{j=1}^{N} q_j \delta_{y_j}, \quad \text{where} \quad \sum_{i=1}^{N} p_i = \sum_{j=1}^{N} q_j = 1.$$

Of course the number of discrete points for $\mu$ and $\nu$ may differ, we keep $N$ for both to simplify the presentation.

This discretisation provides a natural discretization of of the OT problem (1). Settting $X_N = \{x_i\}_{i=1..N}$ and $Y_N = \{y_j\}_{j=1..N}$ we can again use the $\langle .,. \rangle$ notation :

$$(1.8) \qquad OT_N(p,q) := \min_{\gamma_N \in \Pi(p,q)} \langle c_N, \gamma_N \rangle_{X_N \otimes Y_N}$$

where

$$(1.9) \qquad \Pi(p,q) := \left\{ \gamma_N \in \mathbb{R}_+^{N \times N} | \langle 1_{X_N}, \gamma_N \rangle_{Y_N} = p, \ \langle 1_{Y_N}, \gamma_N \rangle_{X_N} = q \right\}$$

Here $c_N = \{c(x_i, y_j)\}_{i,j=1..N}$, $q = \{q_j\}_{j=1..N}$ and $p = \{p_i\}_{i=1..N}$ .

(1.8) is a discrete linear programming problem which can be solved numerically using standard linear programming solvers. The main drawback of this method is it's high dimensionality. It is a linear problem with $N \times N$ unknowns and $2N$ constraints. Numerical resolution with linear solvers which have cubic complexity in the number of unknowns is therefore out of reach for reasonable discretizations (typically $N > 100$). For more details see [Peyré and Cuturi, 2018].

*A Partial Differential Approach.* At least formally, taking the gradient of (1.2) gives the OT map as a function of $f$ :

$$(1.10) \qquad T(x) = \{y \mapsto \nabla_x c(x,y)\}^{-1}(\nabla f(x))$$

Assuming sufficient smoothness to interpret (1.3) in a pointwise sense and plugging this expression gives a Monge-Ampère PDE for $L^2$ cost. This formulation has been studied in detail by Urbas [Urbas, 1997]. For the reflector cost (1.4) the PDE (see equation (1.5) in [Wang, 2004]) is even more non-linear and there is no available numerical approximation theory. A B-spline collocation method was nevertheless proposed and implemented in [Brix et al., 2014], [Brix et al., 2015] with convincing numerical results. An alternative solution method of the Monge-Ampere problem based on a least square approach is presented [Romijn et al., 2020] [Romijn et al., 2019]. Wu [Wu et al., 2013] derives the Monge-Ampere equation for a lens surface and solve the equations using standard finite differences and Newton iteration.

*Semi-discrete OT.* Semi-discrete OT deals with the special case where $\mu$ has continuous density and the other $\nu_N$ is discrete as in (1.7). Then the dual formulation in 1) can be reduced to discrete optimisation problem through the elimination of the constraints replacing $f$ by its $c$-transform :

$$(1.11) \qquad f(x) = g^c(x) = \min_y \left( c(x,y) - g(y) \right) = \min_j \left( c(x,y_j) - g(y_j) \right)$$

and optimizing over the vector $\{g(y_j)\}_{j=1..N}$.

For the $L^2$ cost, It is now well understood that a Newton method can be used [Mérigot, 2011]. The implementation relies on fast (linear time) computations of a tesselation of the target domain called Laguerre cells :

$$Lag_j = \{x, s.t.\, c(x,y_j) - g(y_j) < c(x,y_i) - g(y_i), \forall i = 1..N\} \ \forall j = 1..N$$

. The method was applied to the reflector cost (1.4) in [Machado Manhães De Castro et al., 2016] but the Laguerre cells can be only computed in $\mathbb{R}^d$ and not on $\mathbb{S}^{d-1}$ thus lifting by one the dimension of the problem. Solvers in dimension 2 and 3 are available [Leclerc, 2019] [Lévy, 2015]. See also the pioneering work of [Kochengin and Oliker, 2003] using this methods on the reflector problem.

1.3. **Our contribution.** Our contribution is based on a strict convexification of the OT problem called "entropic regularization". In a nutshell :

It has been introduced for OT computations in [Cuturi, 2013] (see [Peyré and Cuturi, 2018] for a comprehensive review). This topic has been very active since for two reasons : first, the method comes with a simple computational method call Sinkhorn algorithm for which there are now countless variants and acceleration techniques. It can be applied to discretization on the order of $N = 10^6$. Secondly, it can be applied to a wide range of costs but does not seem to have been investigated yet

for the reflector cost. The "entropic regularization" of the Kantorovich problem (1) is based on the following KullBack-Leibler divergence or "relative entropy" (KL) penalization :

(1.12)
$$OT_\varepsilon(\mu, \nu) := \min_{\gamma_\varepsilon \in \Pi(\mu,\nu)} \langle c, \gamma_\varepsilon \rangle_{X \times Y} + \varepsilon \, \mathrm{KL}(\gamma_\varepsilon \,|\, \mu \otimes \nu) =$$

$$\max_{f_\varepsilon, g_\varepsilon} \langle f, \mu \rangle_X + \langle g_\varepsilon, \nu \rangle_Y - \varepsilon \, \langle \exp(\tfrac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c)) - 1, \mu \otimes \nu \rangle_{X \times Y}$$

where $\varepsilon > 0$ is a small "temperature" parameter (see [Léonard, 2013] for a Statistical Physics interpretation of this problem due to Schroedinger) and

$$\mathrm{KL}(\gamma \,|\, \mu \otimes \nu) := \int_{X \times Y} \log(\frac{d\gamma}{d\mu \otimes d\nu}) \, d\gamma \text{ if } \gamma \text{ is absolutely continuous w.r.t to } \mu \otimes \nu \text{ and } +\infty \text{ else.}$$

The primal-dual optimality condition is given by

(1.13)
$$\gamma_\varepsilon = \exp(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c)) \, \mu \otimes \nu.$$

The optimal entropic plan is therefore the scaling by the Kantorovich potentials of a fixed Kernel $\exp(-\frac{1}{\varepsilon}c)$. It is diffuse, i.e. not concentrated on a map, $\varepsilon$ can be interpreted as a bandwidth under which transport is blurred.

Numerical solutions are produced under the discretization (1.7) of this problem, i.e. replacing $(X, Y, c, \mu, \nu)$ by $(X_N, Y_N, c_N, \mu_N \nu_N)$ in the above formula. The resulting discrete non-linear system of optimality is solvable in particular with Sinkhorn algorithm.

In section 2 we recall basic convergence properties of this approximation when $\varepsilon \to 0$ and how $N$ and $\varepsilon$ are connected. We also discuss the resulting bias when $\varepsilon > 0$ and how the notion of Sinkhorn Divergence introduced in [Ramdas et al., 2017] [Genevay et al., 2018] [Feydy et al., 2018] can produce a correction.

We then adapt this framework and Sinkhorn algorithm in section 3 to the reflector cost OT problem and discuss specifics that need to be addressed for this transition, such as de-biasing for reflector cost, choice of discretization and interpolation. We present several example of re-simulations using ray tracing.

Wasserstein distances is a natural distance to measure discrepancies between measures. In section 4 we therefore propose to evaluate it between the illumination $\nu$ and a sampling of the computed approximation. The sampling is obtained via ray tracing reflections on a continuous interpolation of the reflector as a residual error. We use this setting to study the convergence with $N$ and $\varepsilon$. We also show how the continuous approximation of the reflector behaves when the number of rays used for the resimulation increase.

## 2. ENTROPIC OT

2.1. **Entropic OT and Sinkhorn algorithm.** We will restrict from now on to the discrete case (see (1.7)) and the dual optimal transport problem :

(2.1)
$$OT_{\varepsilon,N} := \max_{f_\varepsilon, g_\varepsilon} \langle f_\varepsilon, \mu_N \rangle_{X_N} + \langle g_\varepsilon, \nu_N \rangle_{Y_N} - \varepsilon \, \langle \exp(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c_N)) - 1, \mu_N \otimes \nu_N \rangle_{X_N \times Y_N}.$$

where we use the same notation $(f_\varepsilon, g_\varepsilon)$ for discrete vectors in $\mathbb{R}^N$.

We solve (2.1) with Sinkhorn algorithm. It corresponds to a block coordinate ($f_\varepsilon$ and $g_\varepsilon$) ascent : Initialize with $g_\varepsilon^0 = 0_Y$ and then iterate (in $k$) :

(2.2)
$$f_\varepsilon^{k+1} = -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(g_\varepsilon^k - c_N)), \nu_N \rangle_{Y_N})$$

$$g_\varepsilon^{k+1} = -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_\varepsilon^{k+1} - c_N)), \mu_N \rangle_{X_N})$$

Sinkhorn linear convergence for fixed $\varepsilon$ has been largely studied (see [Peyré and Cuturi, 2018]) again). Its numerical stability depends on typical transport scale of the data $\tau = \sup_{x \in \mathcal{X}} c(x, T(x))$

and $\varepsilon$. The computer hits overflows or underflows when the ratio $\tau/\varepsilon$ is too large. A good review of existing hacks and methods to mitigate this problem can be found in [Schmitzer, 2016]. The convergence as $\varepsilon \to 0$ of $OT_\varepsilon(\mu,\nu)$ toward $OT(\mu,\nu)$ is also well understood in the continuous [Léonard, 2013] and discrete setting [Cominetti and Martin, 1994].

The potentials computed using the iterations (2.2) are defined on the discrete sets $X_N$ and $Y_N$ but they admit a canonical extention by replacing $c_N(x_i, y_j)$ respectively by $c(x, y_j), x \in X$ and $c(x_i, y), y \in Y$. Omitting the iteration index :

$$(2.3) \qquad f[g_\varepsilon](x) = -\varepsilon \log(\sum_{j=1..N} \exp(\frac{1}{\varepsilon}(g_\varepsilon(y_j) - c(x, y_j)))\nu_N(y_j)), \ \ \forall x \in X$$

As our goal is to form the reflector (1.6) from the discrete numerical potential $f_\varepsilon$, we are interested in the convergence of $f_\varepsilon$ as $\varepsilon \to 0$ and $N \to \infty$. To the best of our knowledge the joint convergence in $N$ and $\varepsilon$ has only be studied in [Berman, 2017], we reproduce partially his results :

**Theorem 3** (Berman joint convergence - corollary 1.3 [Berman, 2017] ). *We assume $\mu$ and $\nu$ are in $C^{2,\alpha}$ and positive, and that $N$ and $\varepsilon$ are dependent parameters : $N = (1/\varepsilon)^d$ where $d$ is the dimension of the problem. A technical condition on the sequence of discretization $(X_N, Y_N, c_N, \mu_N \nu_N)$ called "density propery" (see remark 3 below) is also necessary. Then there exists a positive constant $A_0$ such that for any $A > A_0$ the folowing holds : setting $m_\varepsilon = [-A\log(\varepsilon)/\varepsilon]$ the continuous interpolation provided by $f[g_\varepsilon^{m_\varepsilon}]$, built using the cannonical extension (2.3) from the discrete Sinkhorn iterate at $k = m_\varepsilon$, satisfies the estimate*

$$(2.4) \qquad \sup_X |f[g_\varepsilon^{m_\varepsilon}] - f| \le -C\varepsilon \log(\varepsilon)$$

*for some constant $C$ (depending on $A$) and $f$ is an optimal potential for (1).*

The following remarks can also be extracted from [Berman, 2017] :

**Remark 2** (Domains and Cost ). *Theorem 3 holds for compact manifold $X$ and $Y$ and in particular the Euclidean torus endowed with the periodic $L^2$ Euclidean costs. But it also on the sphere for the reflector cost (see section 6.3.3 [Berman, 2017] ).*

**Remark 3** ( Density property Lemma 3.1 [Berman, 2017] ). *For any given open set $U$ intersecting the support $X$ of $\mu$ (same for $Y$ and $\nu$)*

$$\liminf_{\varepsilon \to 0} \varepsilon \log(\mu_N(U)) = 0$$

*For the flat space $X \subset \mathbb{R}^d$, this condition is enough. For curved surfaces, a technical generalization is required. But in both cases, this density property ensures the discretization of $X$ and $\mu$ (1.7) is such that, for $U$ the sequence of approximations $\mu_N(U)$ never converges faster to 0 than $\varepsilon$ (remember that $N = (1/\varepsilon)^d$). For $X = \mathbb{R}^d$, uniform grids are fine. See section 3.2 for the reflector case.*

**Remark 4** (Domains and Cost ). *Theorem 3 holds for compa particular for the torus and the $L^2$ Euclidean costs but also on the sphere for the reflector cost (see section 6.3.3 [Berman, 2017] ).*

**Remark 5** (Transportation plan convergence). *The support of the entropic transportation plan $\gamma_\varepsilon$ (see (1.13)), built trough the interpolation procedure in theorem (3), converges exponentially fast to the graph of the non entropic OT map when $\varepsilon \to 0$.*

**Remark 6** (Gradient convergence). *The convergence rate in (2.4) is sublinear in $\varepsilon$. As $\varepsilon$ is the "edge" length of the discretization, discrete approximations of the derivative of $f$ may be non convergent.*

2.2. **Entropic bias and Sinkhorn Divergence.** From now on, the potentials computed using Sinkhorn iterations (2.2) are denoted by $f_{OT_\varepsilon}$ and $g_{OT_\varepsilon}$.

Entropic OT is popular to approximate transport distances between probability measures. The Entropic cost $OT_\varepsilon(\mu,\nu)$ is know to converge exponentially fast to $OT(\mu,\nu) = OT_0(\mu,\nu)$ with $\varepsilon$

[Cominetti and Martin, 1994]. Regarding the reflector problem, we are not interested in the transport cost but in $f_{OT_\varepsilon}$ itself, the approximation of $f$. As can be seen formally on the dual formulations of (1) and (1.12) the gradient of the transport and entropic transport costs as a function of measures $\mu$ and $\nu$ are given by the Kantorovich potentials : $\nabla_{\mu,\nu} OT(\mu,\nu) = (f,g)$ and $\nabla_{\mu,\nu} OT_\varepsilon(\mu,\nu) = (f_{OT_\varepsilon}, g_{OT_\varepsilon})$ (see also Proposition 2 [Feydy et al., 2018]). The convergence rate in theorem 3 with an infinite slope at $\varepsilon = 0$ and the numerical stability limit imposed when decreasing $\varepsilon$ is a difficulty. The pollution of the potential by the entropy induced bias is well known issue.

It arises for instance when using the $L^2$ Wasserstein distance as a loss over the space of probability measures. Using the easier to compute $f_\varepsilon$ as a gradient proxy for $f$ to minimize $\mu \to OT(\mu,\nu)$ will therefore be biased by the entropy. This problem is discussed in depth in [Feydy et al., 2018] where it is proposed, in order to correct the bias, to add "diagonal terms" to correct the entropic cost :

$$(2.5) \qquad S_\varepsilon(\mu,\nu) = OT_\varepsilon(\mu,\nu) - \frac{1}{2}(OT_\varepsilon(\mu,\mu) + OT_\varepsilon(\nu,\nu)).$$

Quite remarkably, the authors show that this quantity, called Sinkhorn divergence, remains positive and is convex. It also obviously vanishes for $\mu = \nu$ wich is not the case for $OT_\varepsilon$. Thanks to the symmetry, there is only one dual potentials for each of diagonal problems. We denote them $f_\varepsilon^\mu$ and $f_\varepsilon^\nu$. They can be computed using independent Sinkhorn iterations :

$$(2.6) \qquad \begin{aligned} f_{OT_\varepsilon}^{\mu,k+1} &= -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_{OT_\varepsilon}^{\mu,k} - c_N)), \mu \rangle_X) \\ f_{OT_\varepsilon}^{\nu,k+1} &= -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_{OT_\varepsilon}^{\nu,k} - c_N)), \nu \rangle_Y) \end{aligned}$$

The $\mu$ gradient of $S_\varepsilon$, denoted $f_{S_\varepsilon}$ may be formed by a simple substraction. An open question is whether this correction :

$$(2.7) \qquad f_{S_\varepsilon} = f_{OT_\varepsilon} - f_{OT_\varepsilon}^\mu$$

is a better approximation to $f$ than $f_{OT_\varepsilon}$. Numerical simulations of gradient flows in [Feydy et al., 2018] do indicate this is the case.

**Remark 7** (Asymptotics of $OT_\varepsilon$). *The entropic bias may be related formally to asymptotic results on the difference between $OT$ and $OT_\varepsilon$, see [Conforti and Tamanini, 2019] [Pal, 2019] for recent publications on this subject. For the $L^2$ cost (Theorem 1 [Pal, 2019] treats more costs in the form $c := g(x-y)$, $g$ convex which includes the 1D reflector cost), it gives the following assymptotic behavior for small $\varepsilon$ :*

$$(2.8) \qquad \frac{1}{\varepsilon}\{OT_\varepsilon(\mu,\nu) - OT(\mu,\nu)\} \simeq \frac{1}{2}\left\{ \int_Y \log(\nu)\, d\nu - \int_X \log(\mu)\, d\mu) \right\}$$

*This result is for continuous and compactly supported densities $\mu$ and $\nu$ with respect to $\mathcal{L}(\mathbb{R}^d)$, the Lebesgue measure. Formally taking the gradient in $\mu$ of (2.8) we get for small $\varepsilon$ :*

$$(2.9) \qquad f_{OT_\varepsilon} \simeq f - \frac{\varepsilon}{2}\log(\mu)$$

*In a smooth continuous setting and after convergence of Sinkhorn (2.6) we get*

$$(2.10) \qquad f_{OT_\varepsilon}^\mu(x) = -\varepsilon \log(\int_X exp(\frac{1}{\varepsilon}(f_{OT_\varepsilon}^\mu(x') - c(x,x')))d\mu(x'))$$

*Assuming that $f_{OT_\varepsilon}^\mu(x') - c(x,x')$ has maximum at $x' = x$. This is expected to be asymptotically true as $\varepsilon \to 0$ (remark that $T(x) = x$ in (1.2) for $\mu = \nu$). Then, the Laplace method (also used in [Berman, 2017] and [Pal, 2019]) gives for $\varepsilon \to 0$*

$$f_{OT_\varepsilon}^\mu(x) \simeq -\varepsilon \log\{(C\varepsilon)^{\frac{1}{2}} exp(\frac{1}{\varepsilon}(f_{OT_\varepsilon}^\mu(x)\mu(x)\}$$

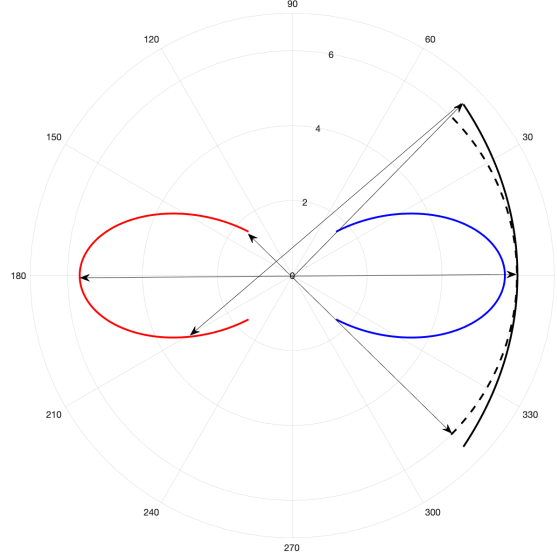FIGURE 3.1. Polar plot of a 1D computation. In blue and red are $\mu$ and $\nu$. Solid and dashed black curves are $R_\varepsilon$ and $R$. Also shown are rays traced. The bottom and middle rays are reflected if the exaxt opposite "identity" directon. The top rays, reflecting on the biased entropic reflector will send light strictly inside the support of $\nu$.

for some constant $C$ depending on $x$ and the hessian of $f_{OT_\varepsilon}^\mu$. It gives :

$$f_{OT_\varepsilon}^\mu(x) \simeq -\frac{\varepsilon}{2}\log(\mu(x)) - \frac{\varepsilon}{4}\log(C\,\varepsilon)$$

which shows that the Sinkhorn divergence de-biasing (2.7) is consistent with (2.9).

This correction can also be observed on the numerical simulation presented in section 3.1.

Remark finally that the density property discussed in remark 3 is requiring for the first order bias terms to converge to 0 when discrete mesures $\mu_N$ converge to original measure $\mu$.

## 3. APPLICATION OF ENTROPIC REGULARIZATION TO THE REFLECTOR PROBLEM

3.1. **Entropic Bias and de-biasing.** The Entropic bias may be observed on numerical solutions. In the simplest identity reflector case for example, where each ray is reflected in its opposite direction, i.e. when $\nu(x) = \mu(x + \pi)$, $x \in \mathsf{S}$, the exact potentials $f$ and $g$ are constant and the reflector $R = \{x\,e^{f(x)}|x \in X\}$ is a portion of circle. It is represented, for the plane problem ($d = 2$), in dashed line in figure 3.1. The solid line is the entropic reflector $R_\varepsilon = \{x\,e^{f_{OT_\varepsilon}(x)}|x \in X\}$ computed with Sinkhorn algorithm. The discretization is $N = 128$ and $\varepsilon = 0.1$ is chosen large enough to see the bias. Remember that the potentials are defined up to a constant, we therefore adjust the plot such that the reflectors superimpose at $x = 0$. The illuminance and illumination $\mu$ and $\nu$ are perfect centered Gaussians and represented respectively in blue and red. This example shows the effect of the entropic bias (2.9). We also trace rays : the bottom and middle rays are reflected if the exact opposite "identity" direction. The top rays, reflecting on the biased entropic reflector will send light strictly inside the support of $\nu$. This shrinking effect of the re-simulation offers a former explanation of the third plot of figure (5.1) where the density is a characteristic function which decreases to 0 at the boundary.

In order to deal with this bias, the notion of Sinkhorn divergences (2.5) is required. When the spaces $X$ and $Y$ are different and a for a general cost $c(x,y)$, we suggest the following extension :

$$(3.1) \qquad S_\varepsilon(\mu, \nu) = OT_\varepsilon(\mu, \nu) - \frac{1}{2}(OT_\varepsilon(\mu, \mu') + OT_\varepsilon(\nu', \nu)).$$

where $\mu' = argminOT(\mu, \cdot)$ and $\nu' = argminOT(\cdot, \nu)$. This is consistent with the distance costs $c(x,y) = d_p^p(x,y)$ as $\mu' = \mu$ and $\nu' = \nu$ in this case. Simillarly, in the reflector cost case , $\mu'$ and $\nu'$ correspond to the reflections of respectively $\mu$ and $\nu$ on the circular reflector mentioned above. Indeed one has $arg \min_{y \in Y} -log(1- <x,y>) = -x$.

The de-biased Kantorovich potentials for the Sinkhorn divergence follow the same formula as for $L_2$-distance cost :

$$(3.2) \qquad \begin{aligned} f_{S_\varepsilon} &= -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(g_{OT_\varepsilon} - c_N)), \nu_N \rangle_{Y_N}) + \varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_{OT_\varepsilon}^\mu - c_N)), \mu \rangle_{X_N}) \\ g_{S_\varepsilon} &= -\varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_{OT_\varepsilon} - c_N)), \mu_N \rangle_{X_N}) + \varepsilon \log(\langle \exp(\tfrac{1}{\varepsilon}(f_{OT_\varepsilon}^\nu - c_N)), \nu \rangle_{Y_N}) \end{aligned}$$

3.2. **Choice of Discretization.** In theorem 3, the requirement that $\epsilon$ is of order $(1/N)^{\frac{1}{d}}$ and that $\mu_N, \nu_N$ satisfy the density condition (remark 3) are closely related. Intuitively this condition means that while choosing the discretizations $\mu_N, \nu_N$ with $N$ points, it is important to make sure that they approximate corresponding distributions $\mu, \nu$ with integration error of order $(1/N)^{\frac{1}{d}}$ for functions which do not oscillate on finer scale then $(1/N)^{\frac{1}{d}}$. This density property is satisfied in particular by Quasi Monte-Carlo point clouds ([Berman, 2017]), defined by bounding the worst case error of integration over a desired function space $W$ :

**Definition 1.** Given a Riemmanian manifold $X$ of dimension $d$, a corresponding normalized volume form $dV$, and a discretization $X_N = x_{i\{i \leq N\}}$, the worst case error of $X_N$ with respect to function space $W$ is:

$$(3.3) \qquad WCE(X_N, W) = \sup_{f \in W} \left\{ \int_X f dV - \sum_{i=0}^{N} f(x_i) \right\}$$

Then, $X_N = x_{i\{i \leq N\}}$ is a Quasi Monte-Carlo discretization of $X$ on lengthscale $(1/N)^{\frac{1}{d}}$ if for any $p \in [1, \infty)$ and $s > \frac{d}{p}$ :

$$(3.4) \qquad WCE(X_N, W_p^s) \leq \frac{C_{s,p}}{\left(N^{\frac{1}{d}}\right)^s}$$

Where $W_p^s$ is a Sobolev space of functions $f$ such that all derivatives of order $s$ are in $L_p(X)$ and $C_{s,p}$ is a uniform constant depending only on $s$ and $p$.

[Berman, 2017] shows that for a density $\mu$ which is absolutely continuous with respect to the volume form $dV$ with density $\rho_\mu$, the discrete approximation $\mu_N$ which satisfies the density requirements of theorem 3 can be constructed using Quasi Monte-Carlo discretization $X_N$ of $X$ as the empirical measure :

$$(3.5) \qquad \mu_N = \frac{1}{\sum_{x_i \in X_N} \rho_\mu(x_i)} \sum_{x_i \in X_N} \rho_\mu(x_i) \delta_{x_i}$$

When the space $X$ is linear, a standard uniform square grids with step-size $h$ is a Quasi Monte-Carlo systems on lengthscale $h$. But for curved spaces, such as a sphere in the case of the reflector problem, constructing a Quasi Monte-carlo systems is not straightforwad and usually they do not have such simple structure.

One way of constructing such a discretization on the sphere, which was used for the simulations in this work, is given in [Womersley, 2017].

3.3. **Ray tracing, Re-Simulation and Interpolation.** The reflector problem is the inverse problem associated to the following forward problem : Given the source illumination $\mu$ as a distribution on the unit sphere $S^d$ and the reflecting surface $R$ as a radial graph over $S^d$, what is the distribution of illuminance after $\mu$ is reflected from $R$? The geometric optics answer to this problem is the simulation of the light propagation known as Ray-Tracing. It can be simply described as follows :

The distribution $\mu$ over the sphere is sampled with an empirical measure $\mu^M := \sum_{i=1..M} p_i \delta_{x_i}$. Then $M$ rays are shot according to their direction $\{x_i\}$s and reflected in the directions $\{y_i = x_i - 2(x_i \cdot n_i)n_i\}$s according to Snells's law ($n_i$ is the normal vector to $R$ at $R(x_i)$). The resulting directions provide a sample $\nu^M := \sum_{i=1..M} p_i \delta_{y_i}$ of the illumination distribution in the far field $\nu = T_\# \mu$. In the optics community, ray tracing is used to evaluate visually the quality of the computed reflector. This procedure is called *Re-Simulation*. The quality of course also depends on the sampling method and the number of samples $M$.

**Remark 8** (Ray tracing and Push-Forward). *Since by construction, the exact continuous reflector reflects the ray $x$ in the direction $T(x)$ (the associated OT map), the above described procedure is the same as computing the empirical push-forward $\nu^M = T_\# \mu^M$ and this is a sampling of $\nu = T_\# \mu$.*

In order to use this procedure with our numerical approximation of the reflector, we need a continuous extension of the reflector from its discrete version computed on the $N$ discretization (3.5). For standard cartesian or structured grids this is not a problem, since there are various well-studied interpolation techniques. Our solutions on the contrary are computed on an unstructured Quasi Monte-Carlo grid. We use the Optimal transport and entropic OT canonical continuous extension procedures based on duality and using the $g$ potentials (respectively (1.11) and (2.3)).

**Definition 2** (c-concave interpolation). For a discrete pair of Kantorovich potentials $(f_N, g_N)$, defined on the discrete grids $X_N$ and $Y_N$ respectively, the c-convex interpolation of the first potential $f_N$ is

$$(3.6) \qquad \hat{f}(x) := \min_{y_j \in Y_N} \{c(x, y_j) - g_N(y_j)\}, \ \forall x \in X.$$

As discussed in section 1.1 (see (1.5)), the continuous reflector $R = \{x e^{\hat{f}(x)} | x \in X\}$. is the envelope of $N$ paraboloids with focal directions $\{y_j\}$. This means that this approach is able to generate a continuous reflector with discontinuous derivatives. This may be an advantage when the target distribution is discontinuous itself. The downside is that in order to achieve continuous effect for human eye, a very fine discretization (a large $N$) is needed.

The second approach uses either the smoothed transform (2.3) for $OT_\epsilon$ or (2.10) for $S_\epsilon$

**Definition 3** (Entropic interpolation). For a discrete pair of Kantorovich potentials $(f_{OT_\epsilon,N}, g_{OT_\epsilon,N})$, defined respectively on the discrete grids $X_N$ and $Y_N$ and solutions of the Sinkhorn algorithm applied to the $OT_\epsilon$ problem, the extension of the first potential $f_{OT_\epsilon,N}$ is

$$(3.7) \qquad \tilde{f}_{OT_\epsilon}(x) := -\epsilon \log(\sum_{j=1..N} \exp(\frac{1}{\epsilon}(g_{OT_\epsilon}(y_j) - c(x, y_j)))\nu_N(y_j)), \ \forall x \in X.$$

Note that (3.7) is the "softmin" version of the mininization (3.6).

Finally using the "'identity" potentials $(f_{OT_\epsilon,N}^\mu, f_{OT_\epsilon,N}^\nu)$ computed when applying Sinkhorn algorithm applied to $S_\epsilon$, the smooth interpolation is :

$$(3.8) \qquad \tilde{f}_{S_\epsilon}(x) := \tilde{f}_{OT_\epsilon} + \epsilon \log(\sum_{j=1..N} \exp(\frac{1}{\epsilon}(f_{OT_\epsilon}^\mu(y_j) - c(x, y_j)))\mu_N'(y_j))$$

This approach would result in a smoother reflector than in the previous case but will also blur any sharp details in the target image.

3.4. **Summary of the numerical procedure.** Our input are analytical descriptions of the illumination/illuminance $\mu$ and $\nu$ described in the test cases section below. All test cases presented in this paper will have the same source and target domains $X$ and $Y$. The source domain $X \subset \mathbb{S}^2$ will be the inverse stereographic projection in the northern hemisphere of the square domain centered at the origin $\{(x_1, x_2) \in \mathbb{R}^2 | -0.6 \leq x_1 \leq 0.6, -0.6 \leq x_2 \leq 0.6\}$. Similarly, $Y \subset \mathbb{S}^2$ will be the inverse stereographic projection in the southern hemisphere of same domain. The outputs are re-simulations/RT computed according to the following procedure :

(1) Computation of $f_N$ : The potentials, optimal maps and reflectors obtained from the minimization of the regularized functional (2.1) will be denoted using the subscript $\cdot_{OT_\varepsilon}$, and those obtained from the minimization of the de-biased functional (3.1) by subscript $\cdot_{S_\varepsilon}$. The discrete Kantorovich potentials are computed for the discretizations $(\mu_N, \nu_N)$ induced by the Quasi Monte-Carlo grids $X_N$ and $Y_N$ respectively according to the formula (3.5).

(2) Interpolation/ray tracing: The choice of interpolation when building the continuous reflector for RT will be denoted with the upperscript $\hat{\cdot}$ when using the $c$-concave interpolation (3.6), and $\tilde{\cdot}$ when using the Entropic interpolation (3.7,3.8). The ray tracing/re-simulation will be shot using a Quasi Monte-Carlo sampling of the source domain of size $M$.

(3) Domain of computation: The distribution of reflected directions (obtained as distribution of directions in $\mathbb{R}^3$) is projected using the stereographic projection from the south pole to the equator plane:

$$(x, y, z) \in \mathbb{R}^3 \to (\frac{x}{1+z}, \frac{y}{1+z}) \in \mathbb{R}^2$$

**Definition 4** ( Parameters and notation for the forward map). In order to discuss the numerics we need to introduce a notation for the forward map induced by the above procedure. This is cumbersome as many parameters are involved : $\varepsilon$ the entropic regularization, $N$ the discretization of the distributions $\mu$ and $\nu$ used in the Sinkhorn algorithm. The choice of using or not the Sinkhorn divergence correction. The choice and notation of the interpolation for the reflector has discussed above and in section 3.3.

We will use the notation $\hat{R}^N_{OT_\varepsilon}$, $\tilde{R}^N_{OT_\varepsilon}$, $\hat{R}^N_{S_\varepsilon}$, $\tilde{R}^N_{S_\varepsilon}$ for the reflectors defined using (1.6) and respectively $\hat{f}^N_{OT_\varepsilon}$, $\tilde{f}^N_{OT_\varepsilon}$, $\hat{f}^N_{S_\varepsilon}$, $\tilde{f}^N_{S_\varepsilon}$.

We will denote $\nu^M = \hat{R}^N_{OT_\varepsilon}[\mu^M]$ the ray traced sampling with $M$ rays (and a similar notation for the other versions of the reflector). Of course $\nu^M$ also depends on the choice of the method and other parameters. Finally the $N$ and $M$ discretizations of the domain are QMC discretizations as explained in section 3.2.

3.5. **Numerical illustration.** We give here several re-simulations as an illustration of the different approximation of the reflectors described in section 3.3. See section 3.4 and 4 for a precise description of the numerical procedure.

The test cases we used are described below and are set on bounded subsets of the sphere. We are not exactly in the setting of theorem 3 and remark (2).

**Test Case 1: Square To Circle.** The source distribution $\mu$ will be the uniform distribution over the set with a square stereographic projection $StP(supp(\mu)) = \{(x_1, x_2) \in \mathbb{R}^2 | -0.5 \leq x_1 \leq 0.5, -0.5 \leq x_2 \leq 0.5\}$. The target distribution $\nu$ in the first example will be an uniform distribution over the set with a disk (circle) stereographic projection $StP(supp(\mu)) = \{(x_1, x_2) \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 0.5^2\}$. Even though the densities are constant, mapping from a non smooth support geometry of

the square to the smooth geometry of a circle is not a trivial task.

We show in figure 5.1, from left to right and from top to bottom the re-simulations using the reflectors $\hat{R}_{OT_\varepsilon}$, $\tilde{R}_{OT_\varepsilon}$, $\hat{R}_{S_\varepsilon}$, $\tilde{R}_{S_\varepsilon}$ and finally also the QMC discretization used for $\nu_N$, $N = 64 * 64$ points. The regularization parameter was taken to be $\varepsilon = \frac{1}{2*64}$.

Figures 5.1 (a) and (c) correspond to the $c$-concave interpolations. This builds the minimal envelope of the family of parabolae with focal axis given by the $\nu$ discretization. All rays hitting one parabola end up at the same position. So up to numerical errors we indeed recover (e), that is $\nu_N$ even though $M >> N$. Also the $S_\varepsilon$ solution performs better at the boundary.

Figures 5.1 (b) and (d) correspond to the entropic interpolations. This is a smoothing of the the $c$-concave interpolation. The $M$ rays are therefore distributed more evenly over the support of the target. We observe, for (b), at the boundary of the support, an important shrinking effect. It can be explained by the $OT_\varepsilon$ entropic bias as in section 3.1, which smoothing the density at the boundary of the support would produce, according to formula (2.9) and the effect observed in figure 3.1. As expected, the $S_\varepsilon$ (d) solution is effective to de-bias the solution. Except, at the corners of the square where the the map is singular, we see extra artefacts probably induced by the smooth interpolation.

**Test Case 2 : Square to Gaussian.** The target distribution $\nu$ is a gaussian distribution with density on the projected domain $\rho(x_1, x_2) = e^{-\frac{x_1^2 + x_2^2}{2}}$ over whole target domain $Y$, and the source distribution will be the same as in Test Case 1.

**Test Case 3 : Square to Square.** The source and target distribution is a uniform distribution over the set with stereographic projection (square) $StP(supp(\nu)) = \{(x_1, x_2) \in \mathbb{R}^2| -0.5 \leq x_1 \leq 0.5, -0.5 \leq x_2 \leq 0.5\}$.

**Test Case 4 : Circle To Steep Gaussian.** The source distribution $\mu$ is uniform distribution over the set with stereographic projection (circle) $StP(supp(\mu)) = \{(x_1, x_2) \in \mathbb{R}^2| x_1^2 + x_2^2 \leq 0.5^2\}$. The target distribution $\nu$ is gaussian distribution with density on the projected domain $\rho(x_1, x_2) = e^{-16*(x_1^2 + x_2^2)}$ over whole target domain $Y$.

**Test Case 5 : Circle To Two Steep Gaussians.** The source distribution $\mu$ is the same as in the fourth. The target distribution $\nu$ is a gaussian distribution with density on the projected domain $\rho(x_1, x_2) = e^{-16*((x_1-0.25)^2 + (x_2-0.25)^2)} + e^{-16*((x_1-0.25)^2 + (x_2+0.25)^2)}$ over whole target domain $Y$.

Figures 5.2- 5.6 show the re-simulations for the 5 test cases.

## 4. ERROR ESTIMATIONS AND CONVERGENCE

4.1. **Wasserstein metrics as error estimator.** From the OT point of view, a straightforward evaluation of our numerical approach would be to build an approximate transport map $T^{app}$ obtained by plugging the different approximations of the potential $\hat{f}^N_{OT_\varepsilon}$, $\tilde{f}^N_{OT_\varepsilon}$, $\hat{f}^N_{S_\varepsilon}$, $\tilde{f}^N_{S_\varepsilon}$ into formula (1.3) and compare it against the exact solution $T$. For the reflector problem, however, only trivial analytical solutions such as the circle (identity) or a parabola (dirac target) are known. On the other hand and from the optics application point of view, the main concern is not the shape of reflector itself (except for designing constraints ignored here), but rather the quality of the re-simulation. As mentioned in remark 8, mathematically this re-simulation is a sampling of the push-forward $T^{app}{}_\#\mu$. Therefore, in order to build an error estimator we will consider the difference between the push-forward $T^{app}{}_\#\mu$ and the desired distribution $\nu$. For this we will use the standard $L^2$ Wasserstein distance $W_2(T^{app}{}_\#\mu, \nu)$, or its entropic counterparts presented in section 2. They are known to provide a smooth estimators between on empirical distributions [Peyré and Cuturi, 2018]. In order to build a numerically computable error estimate we will use the ray traced sample $\mu^M$ of the source measure $\mu$.

First, using the triangle inequality for the Wasserstein metrics one gets :

$$(4.1) \qquad W_2(T^{app}{}_{\#}\mu, \nu) \leq W_2(T^{app}{}_{\#}\mu, T^{app}{}_{\#}\mu^M) + W_2(T^{app}{}_{\#}\mu^M, \nu)$$

The first term on the right hand side $(W_2(T^{app}{}_{\#}\mu, T^{app}{}_{\#}\mu^M))$ can be estimated using Lemma 4 and inequality (5.3) in the appendix 5 giving :

$$(4.2) \qquad W_2(T^{app}{}_{\#}\mu, T^{app}{}_{\#}\mu^M) \leq K W_2(\mu, \mu^M)$$

where $K$ depends on the data of the problem. The convergence in $W^2$ norm of such sampling has been studied in [Fournier and Guillin, 2015] and is known to behave asymptotically as $M^{-\frac{1}{2}}$ in expectation.

The second term on the right hand side of (4.1) can be approximated using $\nu M = \hat{R}^N_{OT_\varepsilon}[\mu^M]$ the sample obtained using the different combinations of ray tracing and interpolation (we only show one of these combinations there see definition 4 and remark 8) :

$$(4.3) \qquad W_2(T^{app}{}_{\#}\mu^M, \nu) \simeq W_2(\nu^M, \nu)$$

The continuous densities $\mu$ and $\nu$ still appear in estimates (4.2-4.3). The $W_2$ distance can be computed either using semi-discrete OT (section 1.2) which relies on a $P1$ discretization of the continuous densities or again using Sinkhorn divergence but this time for the $L^2$ cost. This is the choice we made in this paper and $\mu$ and $\nu$ are discretized on a finer $M_\infty = 512^2$ grid.

The error estimates are therefore computed using the Sinkhorn divergence approximation of the $L^2$ Wasserstein distance on the projection plane with a smaller $\varepsilon : W_2(.,.) \simeq S_{\varepsilon=1e-06}(.,.)$, but we will keep the $W^2$ notation below.

### 4.2. Numerical convergence Study in $N$.
In this section $M = 128^2$ the number of rays for the resimulation is fixed. The densities $\mu$ and $\nu$ are discretized with a finer $M_\infty = 512^2$ points orthogonal grid on the plane.

In (4.2), for $\mu$ as in Test Cases 1,2 and 3 , we obtained with the above parameters :

$$(4.4) \qquad W_2(\mu^M, \mu^{M_\infty}) = 2.355e - 03$$

For Test cases 4 and 5, we obtained

$$(4.5) \qquad W_2(\mu^M, \mu^{M_\infty}) = 2.310e - 03$$

Regarding $W_2(\nu^M, \nu^{M_\infty})$ in (4.2), we plot for different tests cases described below the convergence curves in $N$ for the different reflector approximations methods explained in definition 4 and two values of $\varepsilon : 1/2\sqrt{N}, 1/8\sqrt{N}$. See figures 5.7- 5.11 for the 5 test cases. We observe :
  (1) Convergence to error levels comparable to (4.4-4.5) which makes sense as $\nu_M$ is at best a QMC sampling of $\nu$.
  (2) Decreasing $\varepsilon$ improves the $OT_\varepsilon$ reflectors.
  (3) Sinkhorn divergence $S_\varepsilon$ de-besiasing is effective, both in the sense of obtaining lower errors, and being less dependent on the choice of $\varepsilon$.
  (4) As the target is smooth,the entropic interpolation of $S_\varepsilon$ solutions are less dependent on the choice of the discretization, and moderate values of $N$ are enough to achieve the same error as when using the highest value of $N$ used here.

### 4.3. Numerical convergence Study in $M$.
In this section, we use the Entropic interpolation method and generate the potential with the Sinkhorn divergence method as they seemed to performed the best. We then fix $N = 128^2$ and $\epsilon = 128/8$ and study the dependence of the error term $W_2(\nu^M, \nu)$ with $M$ where $\nu^M = \tilde{R}^N_{S_\varepsilon}[\mu^M]$.

In figures 5.12-5.13 we plot the error curves for Test Cases 1 and 5 in original and log scales. The curves demonstrates that the computed continuous numerical approximation of the reflector preserves after ray tracing the quality of the illumination/source ray sampling. In particular :

(1) The convergence curves closely are similar to the convergence curves in $N$ obtained using the Sinkhorn divergence method with $\epsilon = 1/(8\sqrt{N})$ and the c-concave interpolation. Indeed, with this interpolation method and a good approximation of the potential, the reflector will send all the rays onto $\nu_N$ wich is also discretized using a QMC system (see also section 3.3). So increasing $N$ there and increasing $M$ here results in the same empirical measure $\nu^M = \tilde{R}^N_{S_\epsilon}[\mu^M]$.

(2) In logarithmic scales the curves agree with the $M^{-\frac{1}{2}}$ rate predicted by the theory [Fournier and Guillin, 2015].

## 5. CONCLUSION

We have studied numerically the use of entropic regularization to solve the OT/Reflector problem. Based on theorem 3 [Berman, 2017] we checked the convergence on the method the entropic regularization and space discretization parameter. Error quantities on the reflector shape itself are not available. We therefore defined and used an error criterium based on re-simulations/ray tracing and the Wasserstein 2 distance, a standard tool to compare images and empirical distribution. We also tested a novel approach [Feydy et al., 2018] called Sinkhorn Divergence to correct the bias induced by the entropic regularization.

Our experiments indicate first that the method converges. It also shows that the combination of a QMC discretization satisfying the density property (remark 3) the entropic interpolation (2.1) and correcting the Entropic bias with the Sinkhorn divergence technique produce the best results.

## ACKNOWLEDGEMENTS

FIGURES

(a) $\hat{R}_{OT_\varepsilon}$

(b) $\tilde{R}_{OT_\varepsilon}$

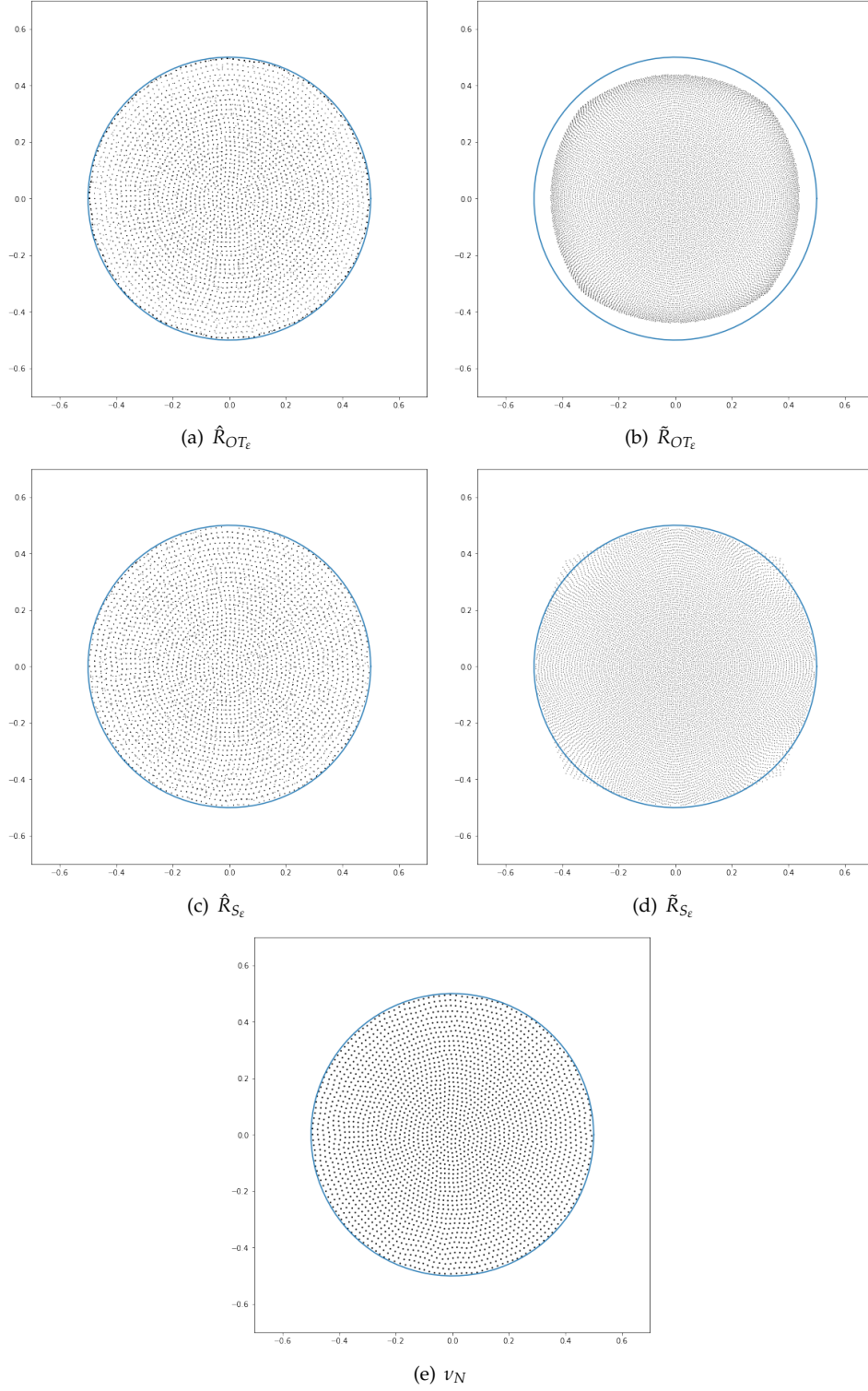(c) $\hat{R}_{S_\varepsilon}$

(d) $\tilde{R}_{S_\varepsilon}$

(e) $\nu_N$

FIGURE 5.1. Test Case 1 : From left to right and from top to bottom the re-simulations of $\mu^M$ (M=128*128) using the reflectors $\hat{R}_{OT_\varepsilon}$, $\tilde{R}_{OT_\varepsilon}$, $\hat{R}_{S_\varepsilon}$, $\tilde{R}_{S_\varepsilon}$ (check definition 4 for explanation of the notations) and finally also the QMC discretization used for $\nu_N$, $N = 64 * 64$ points. The regularization parameter was taken to be $\varepsilon = \frac{1}{2*64}$ for all four solutions.
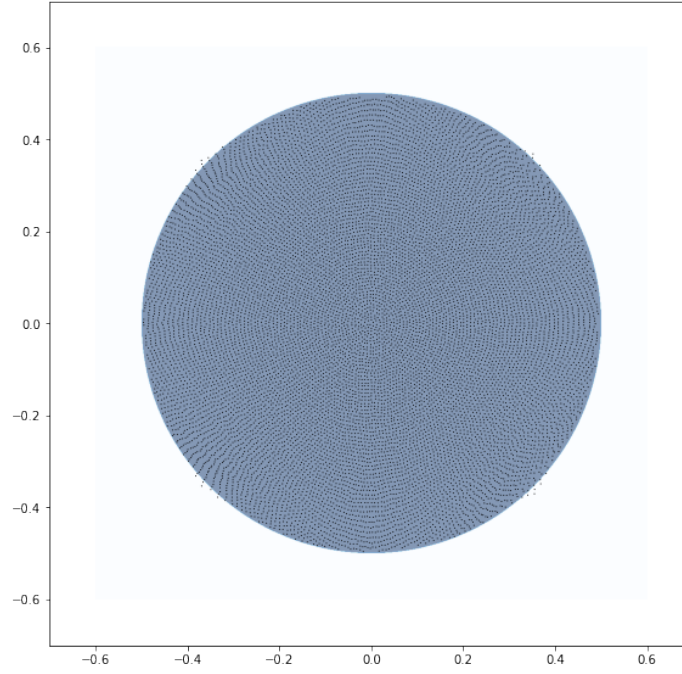
FIGURE 5.2. Test Case 1 : Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)
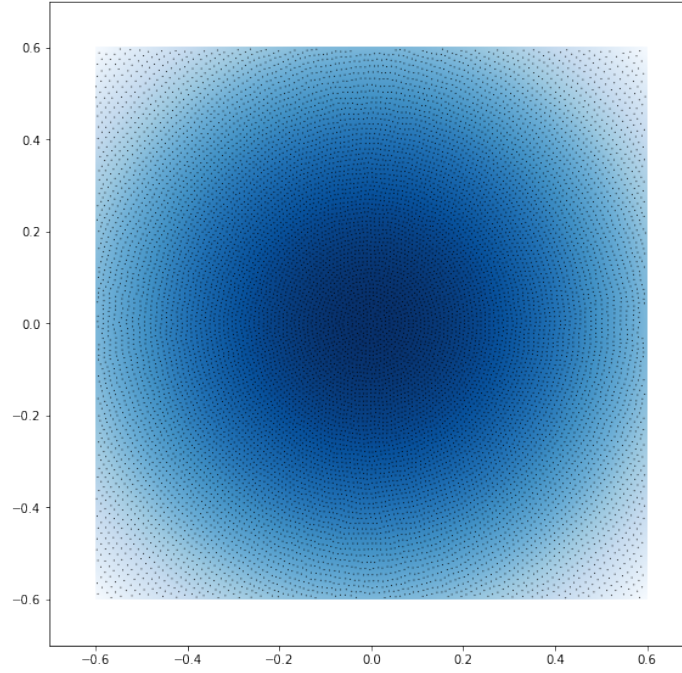


FIGURE 5.3. Test Case 2 : Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)
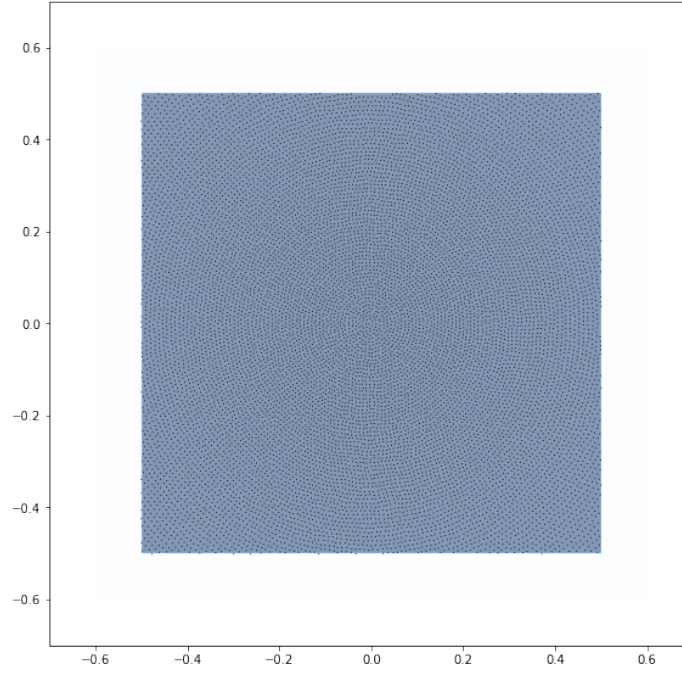
FIGURE 5.4. Test Case 3 : Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$(point cloud) together with the exact desired distribution (color graph)
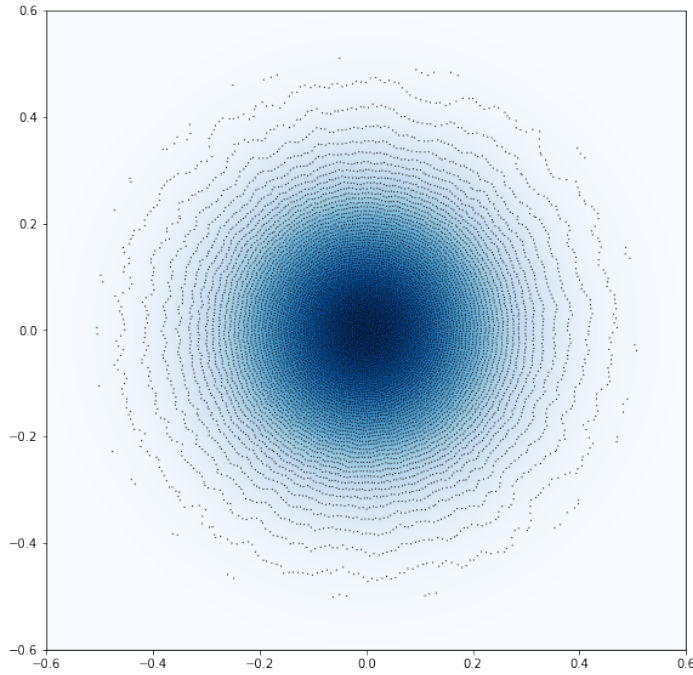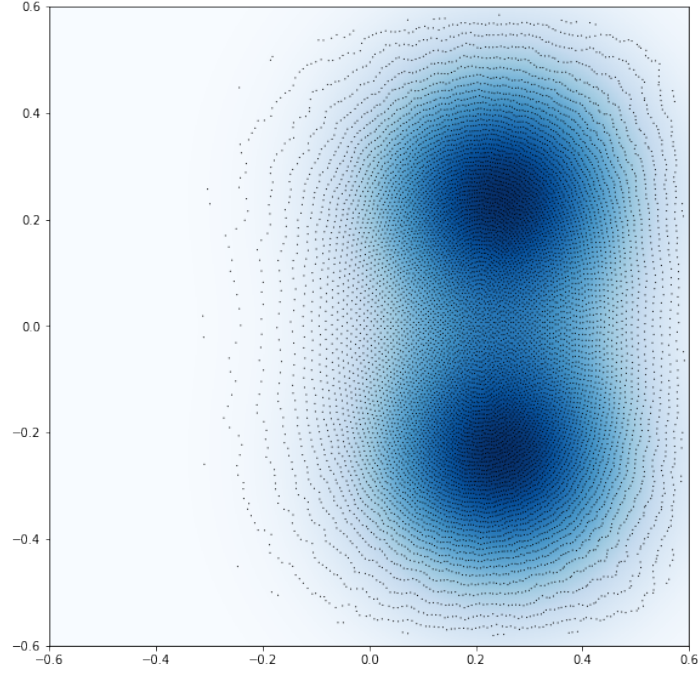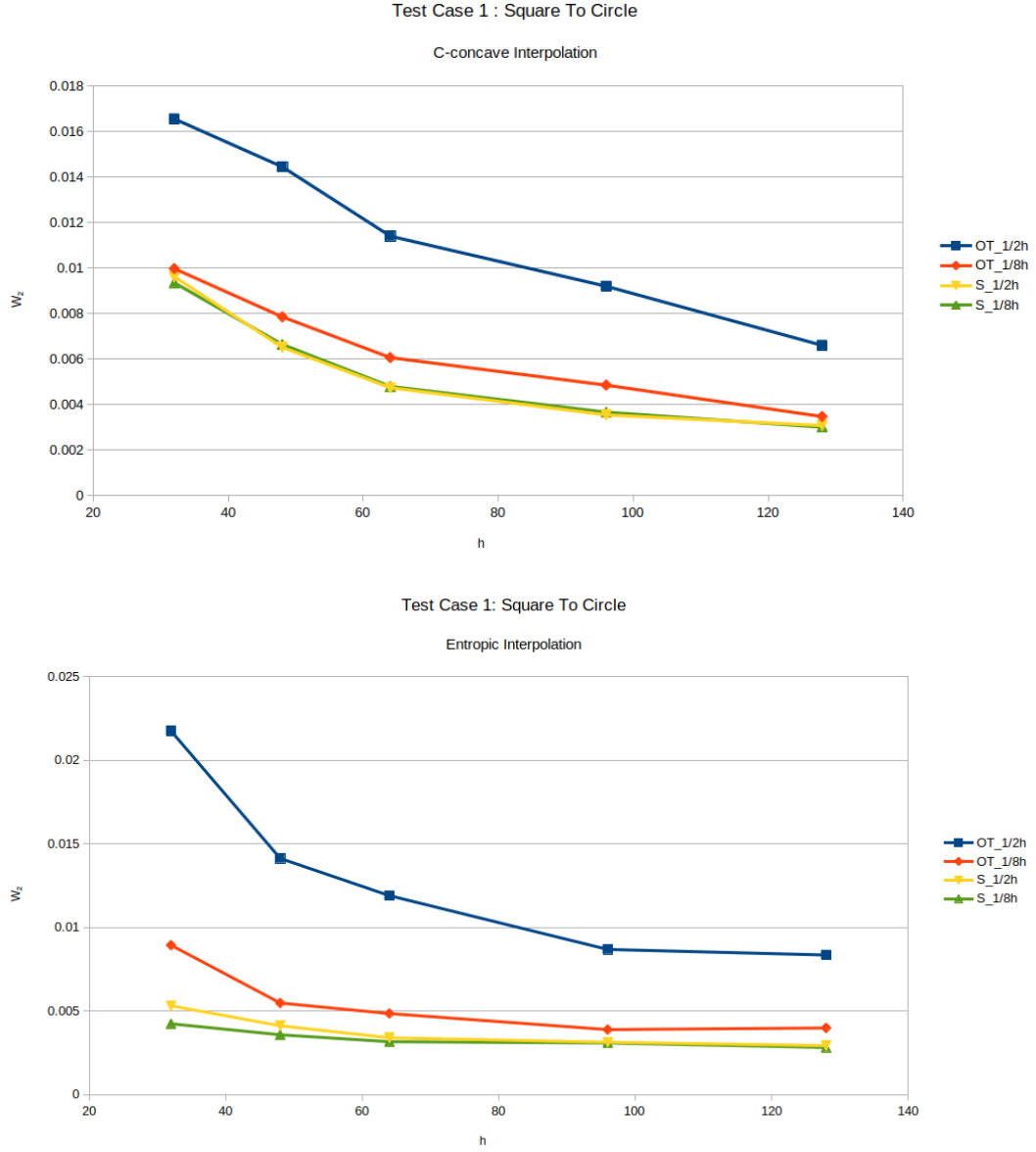


FIGURE 5.5. Test Case 4 : Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$(point cloud) together with the exact desired distribution (color graph)

FIGURE 5.6. Test Case 5 : Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)

FIGURE 5.7. Test Case 1 ($h := \sqrt{N}$) : $W_2$ distance between Re-simulated Push-Forward with $128^2$ points and exact target.
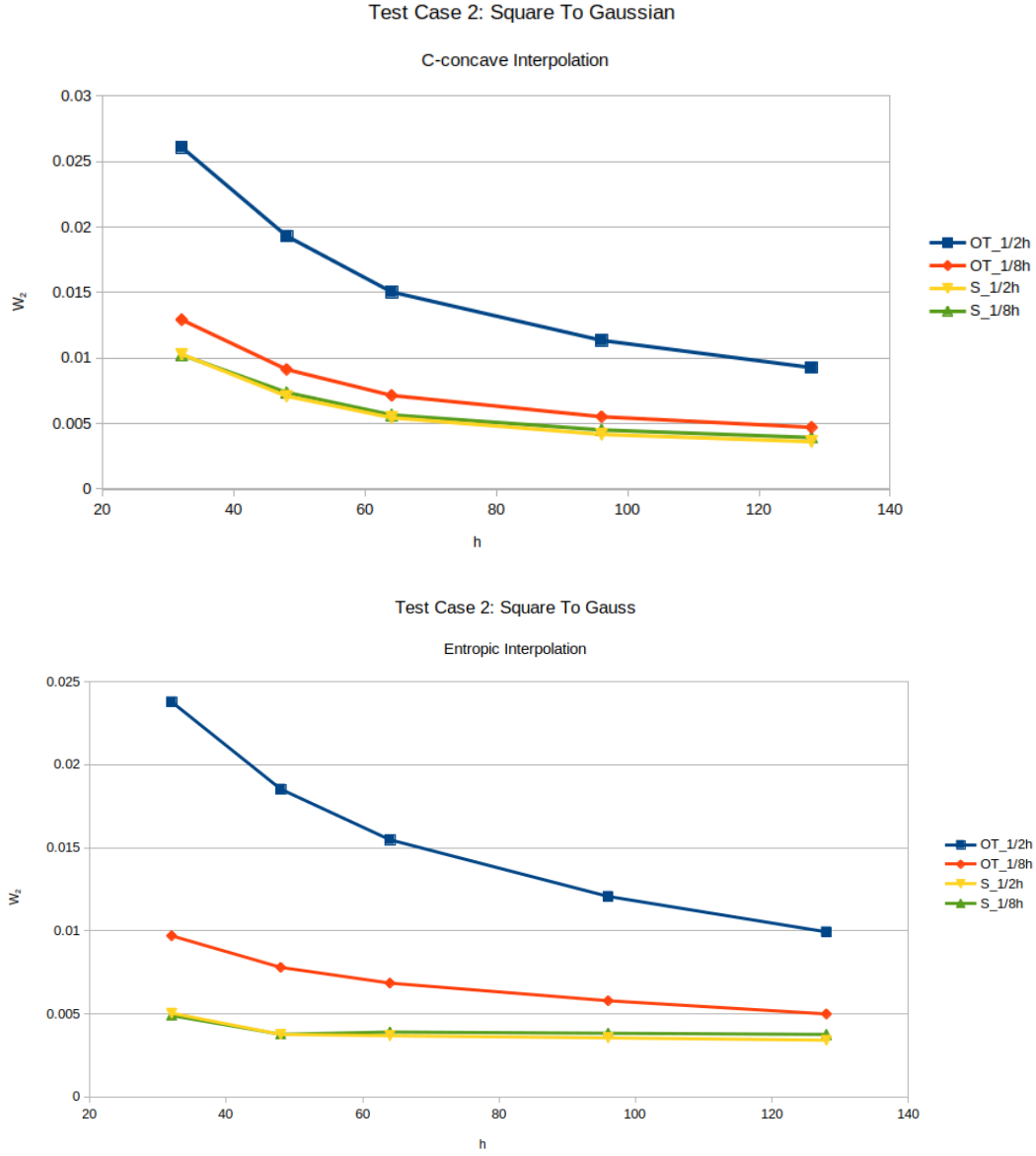
FIGURE 5.8. Test Case 2 ($h = \sqrt{N}$) : $W_2$ distance between Re-simulated Push-Forward with $128^2$ points and exact target.
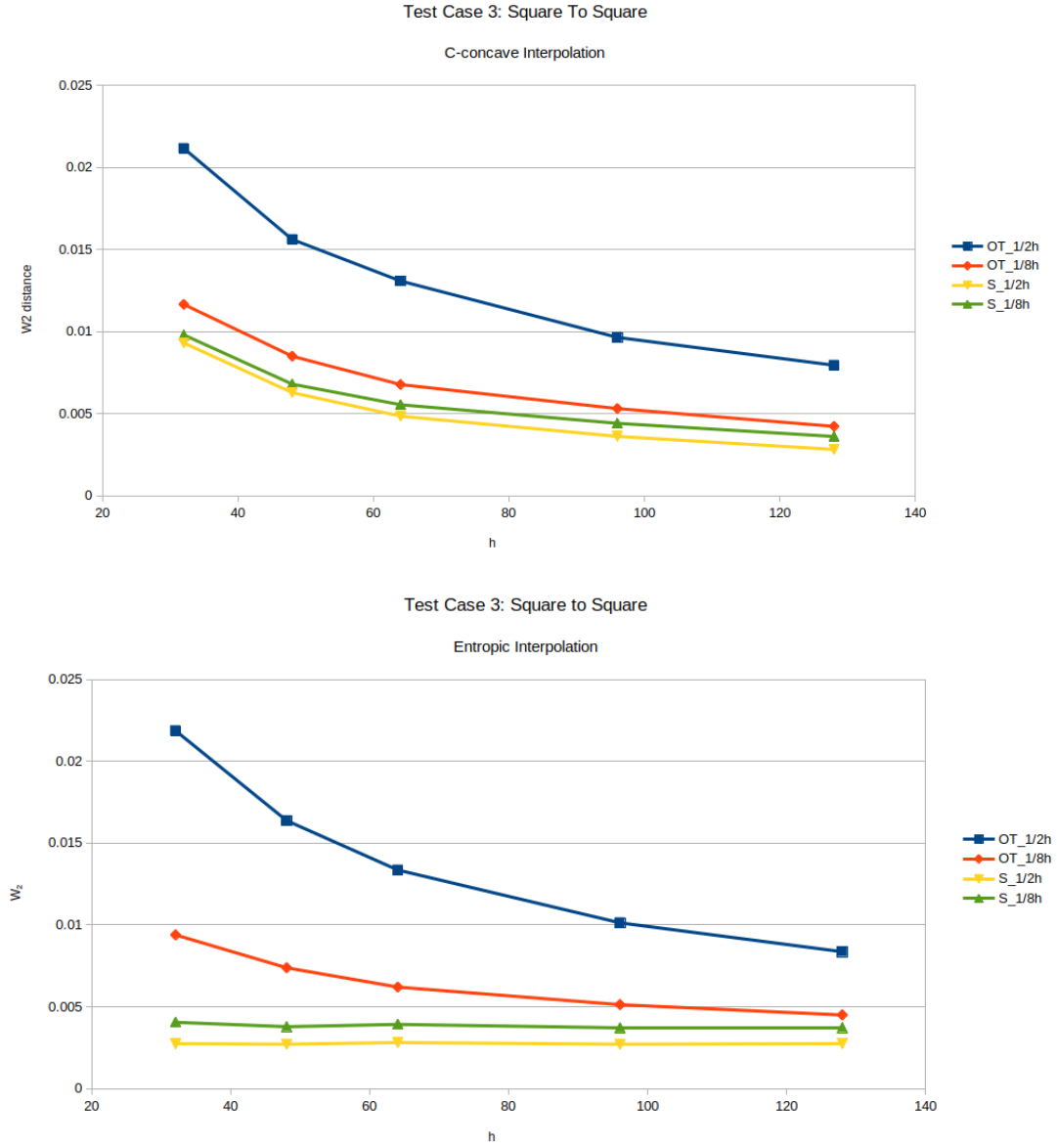
Test Case 3: Square To Square

C-concave Interpolation



Test Case 3: Square to Square

Entropic Interpolation



FIGURE 5.9. Test Case 3 ($h = \sqrt{N}$) : $W_2$ distance between Re-simulated Push-Forward with $128^2$ points and exact target.

FIGURE 5.10. Test Case 4 ($h = \sqrt{N}$) : $W_2$ distance between Re-simulated Push-Forward with $128^2$ points and exact target.
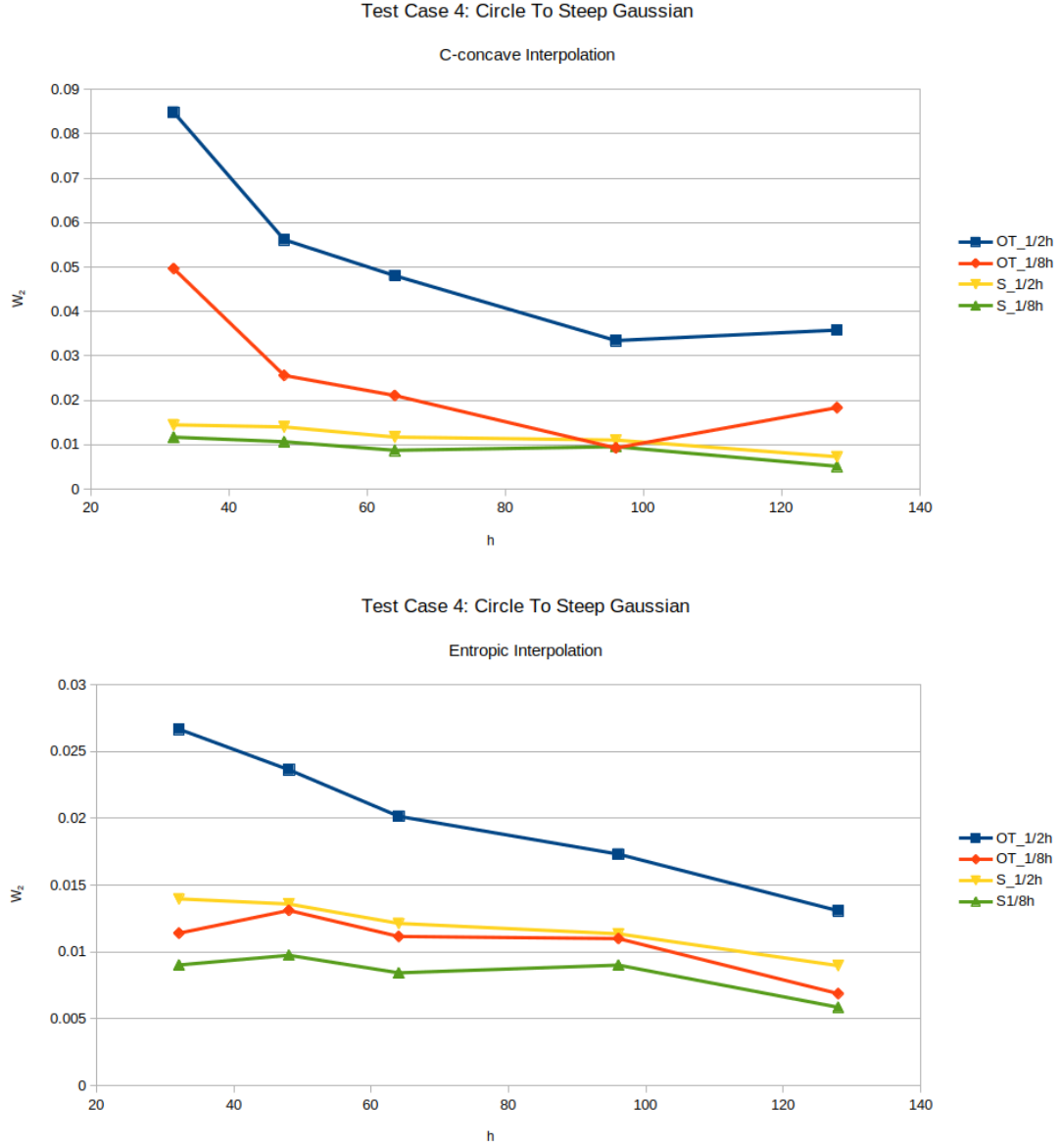
Test Case 5: Circle To Two Gaussians

C-concave Interpolation



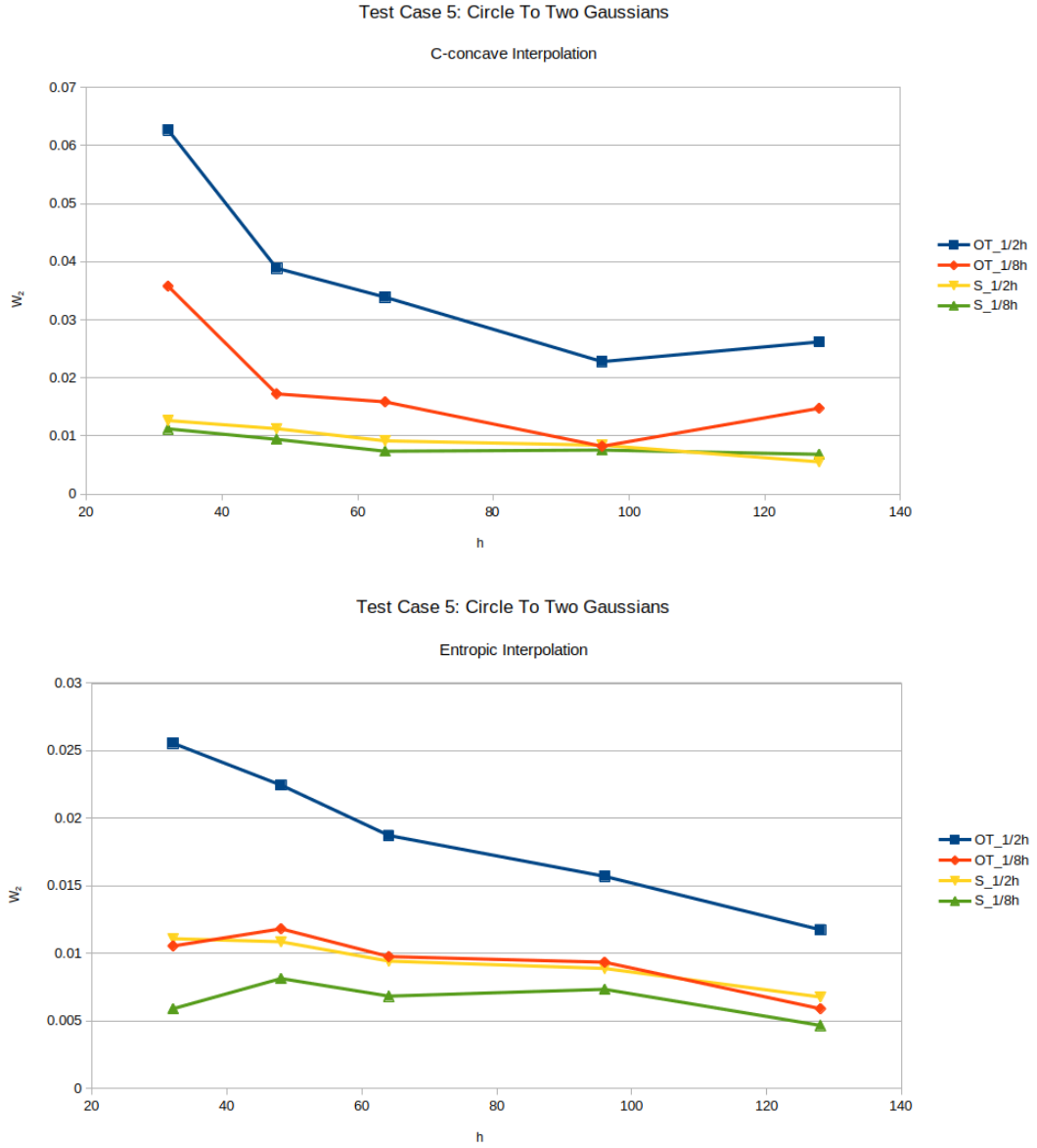Test Case 5: Circle To Two Gaussians

Entropic Interpolation



FIGURE 5.11. Test Case 5 ($h = \sqrt{N}$) : $W_2$ distance between Re-simulated Push-Forward with $128^2$ points and exact target.
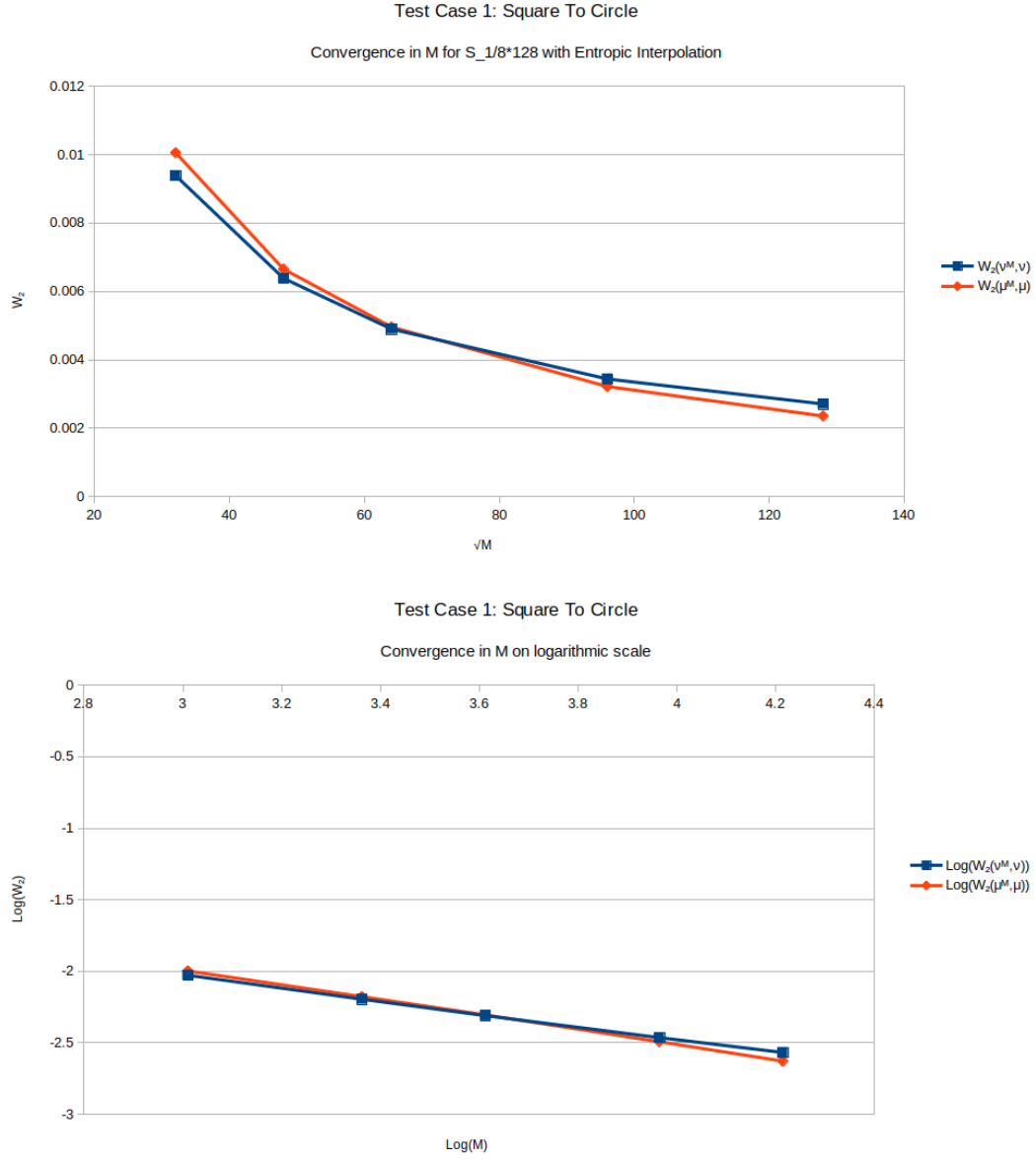
FIGURE 5.12. Test Case 1 : Convergence of error terms $W_2(\mu^M, \mu)$, $W_2(\nu^M, \nu)$, original (top) and logarithmic (bottom) scales.
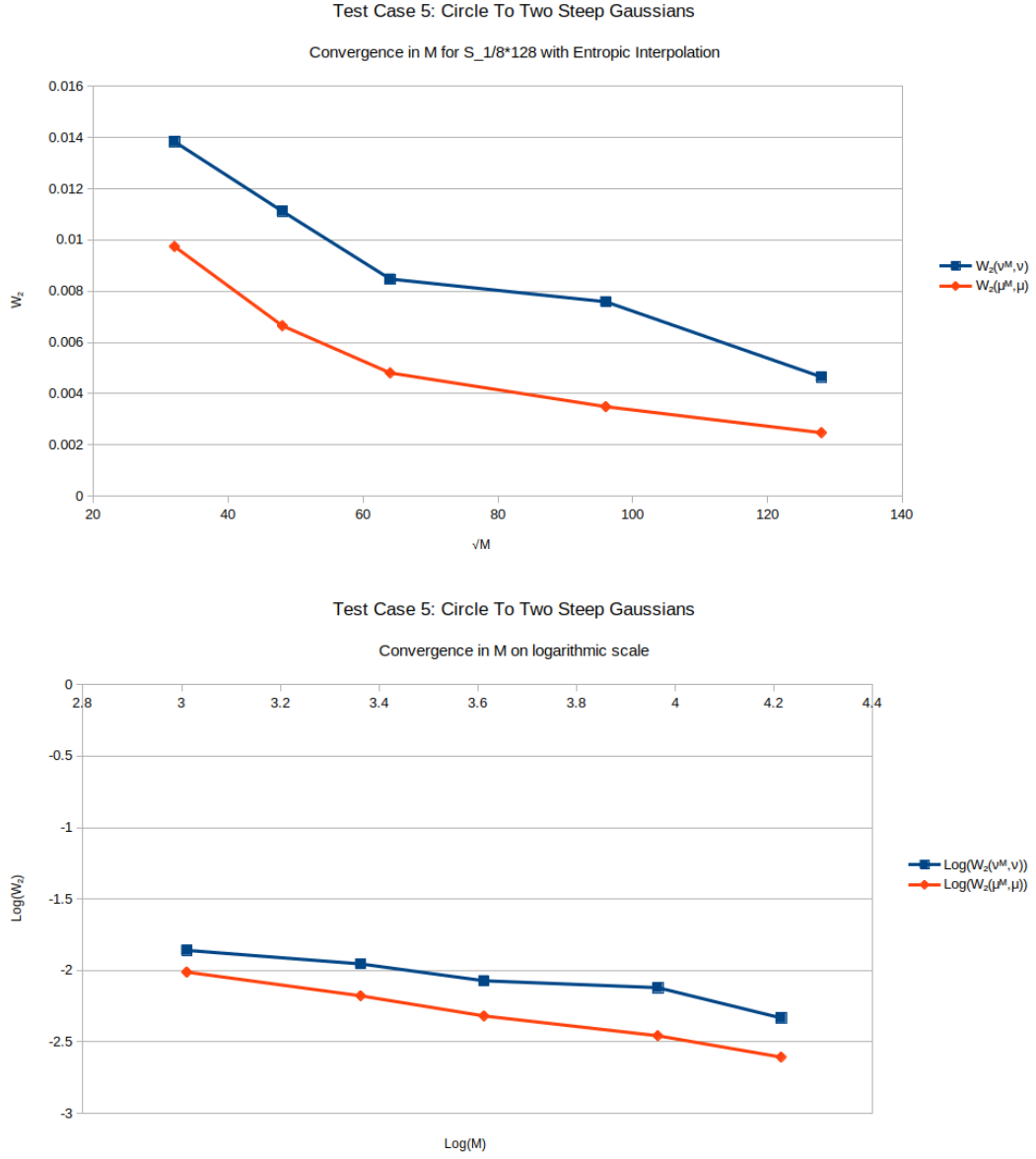
FIGURE 5.13. Test Case 5 : Convergence of error terms $W_2(\mu^M, \mu)$, $W_2(\nu^M, \nu)$, original (top) and logarithmic (bottom) scales.

## APPENDIX A

**Lemma 4.** *Given two Polish spaces $(X, d_Y)$ and $(Y, d_Y)$, with Borel probability measures $\mu_1, \mu_2$ on X and Lipschitz continuous map $T : X \to Y$ with Lipschitz constant Lip(T), Then*

$$(5.1) \qquad W_p(T_\# \mu_1, T_\# \mu_2) \leq Lip(T) W_p(\mu_1, \mu_2)$$

*Proof.* Let $(f, g)$ be optimal pair of Kantorovich potentials for $W_p(T_\# \mu_1, T_\# \mu_2)$. Then, for all $x, x' \in X$

$$(5.2) \qquad f(T(x)) + g(T(x')) \leq d_Y^p(T(x), T(x')) \leq Lip(T)^p d_X^p(x, x')$$

Where first inequality holds due to fact that admissible pairs for maximization in Kantorovich duality approach cost function from below.

Inequality (5.2) implies that functions $\frac{f(T(\cdot))}{Lip(T)^p}$ and $\frac{g(T(\cdot))}{Lip(T)^p}$ are admissible pair in the dual form of $W_p(\mu_1, \mu_2)$. This leads to following:

$$
\begin{aligned}
W_p^p(T_\# \mu_1, T_\# \mu_2) &= \min_\gamma \int_{Y \times Y} d_Y^p((T_\# \mu_1, T_\# \mu_2) d\gamma \\
&= \int_Y f(y) dT_\# \mu_1(y) + \int_Y g(y') T_\# \mu_2(y') \\
&= \int_X f(T(x)) d\mu_1(x) + \int_X g(T(x')) \mu_2(x') \\
&= Lip(T)^p \int_X \frac{f(T(x))}{Lip(T)^p} d\mu_1(x) + \int_X \frac{g(T(x'))}{Lip(T)^p} d\mu_2(x') \\
&\leq Lip(T)^p W_2^p(\mu_1, \mu_2)
\end{aligned}
$$

Taking $p$-th root on both sides leads to the desired inequality.

$\square$

When the target support $Y$ is convex and the densities of $\mu$ and $\nu$ smooth enough ($C^{1,\alpha}$ is enough) and bounded below and above by positive constants, a classic regularity result by Caffarelli [Caffarelli, 1992] gives sufficient regularity to get

$$(5.3) \qquad Lip(T) \leq K$$

where the constant $K$ only depends on the dimension $d$ and the the data $\mu$ and $\nu$. See [Philippis and Figalli, 2014] for further refinements and discussions.

## References

[Berman, 2017] Berman, R. J. (2017). The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampere equations. *ArXiv e-prints*.

[Brix et al., 2014] Brix, K., Hafizogullari, Y., and Platen, A. (2014). Solving the monge-ampère equations for the inverse reflector problem. *Mathematical Models and Methods in Applied Sciences*, 25.

[Brix et al., 2015] Brix, K., Hafizogullari, Y., and Platen, A. (2015). Designing illumination lenses and mirrors by the numerical solution of monge–ampère equations. *Journal of the Optical Society of America A*, 32(11):2227.

[Caffarelli, 1992] Caffarelli, L. (1992). The regularity of mappings with a convex potential. *Journal of The American Mathematical Society - J AMER MATH SOC*, 5.

[Cominetti and Martin, 1994] Cominetti, R. and Martin, J. S. (1994). Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1):169–187.

[Conforti and Tamanini, 2019] Conforti, G. and Tamanini, L. (2019). A formula for the time derivative of the entropic cost and applications.

[Cuturi, 2013] Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc.

[Feydy et al., 2018] Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouvé, A., and Peyré, G. (2018). Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. working paper or preprint.

[Fournier and Guillin, 2015] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707.

[Genevay et al., 2018] Genevay, A., Peyre, G., and Cuturi, M. (2018). Learning generative models with sinkhorn divergences. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1608–1617, Playa Blanca, Lanzarote, Canary Islands. PMLR.

[Glimm and Oliker, 2003] Glimm, T. and Oliker, V. (2003). Optical design of single reflector systems and the monge–kantorovich mass transfer problem. *Journal of Mathematical Sciences*, 117:4096–4108.

[Kochengin and Oliker, 2003] Kochengin, S. A. and Oliker, V. I. (2003). Computational algorithms for constructing reflectors. *Computing and Visualization in Science*, 6(1):15–21.

[Leclerc, 2019] Leclerc, H. (2019). *Pysdot software - https://pypi.org/project/pysdot/*.

[Lévy, 2015] Lévy, B. (2015). A numerical algorithm for L2 semi-discrete optimal transport in 3D. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693 – 1715.

[Léonard, 2013] Léonard, C. (2013). A survey of the schrödinger problem and some of its connections with optimal transport.

[Machado Manhães De Castro et al., 2016] Machado Manhães De Castro, P., Mérigot, Q., and Thibert, B. (2016). Far-field reflector problem and intersection of paraboloids. *Numerische Mathematik*, 134(2):389–411.

[Mérigot, 2011] Mérigot, Q. (2011). A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1584–1592. 18 pages.

[Pal, 2019] Pal, S. (2019). On the difference between entropic cost and the optimal transport cost.

[Peyré and Cuturi, 2018] Peyré, G. and Cuturi, M. (2018). Computational Optimal Transport. *ArXiv e-prints*.

[Philippis and Figalli, 2014] Philippis, G. D. and Figalli, A. (2014). The monge-ampère equation and its link to optimal transportation.

[Ramdas et al., 2017] Ramdas, A., Trillos, N., and Cuturi, M. (2017). On wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47.

[Romijn et al., 2019] Romijn, L., ten Thije Boonkkamp, J., and IJzerman, W. (2019). Freeform lens design for a point source and far-field target. In *Optical Design and Fabrication 2019 (Freeform, OFT)*, page FT1B.2. Optical Society of America.

[Romijn et al., 2020] Romijn, L., ten Thije Boonkkamp, J., and IJzerman, W. (2020). Inverse reflector design for a point source and far-field target. *Journal of Computational Physics*, 408.

[Schmitzer, 2016] Schmitzer, B. (2016). Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems. *arXiv e-prints*, page arXiv:1610.06519.

[Urbas, 1997] Urbas, J. (1997). On the second boundary value problem for equations of monge-ampère type. *Journal für die reine und angewandte Mathematik*, 487:115–124.

[Villani, 2008] Villani, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

[Wang, 2004] Wang, X.-J. (2004). On the design of a reflector antenna ii. *Calculus of Variations and Partial Differential Equations*, 20(3):329–341.

[Womersley, 2017] Womersley, R. S. (2017). Efficient Spherical Designs with Good Geometric Properties. *ArXiv e-prints*.

[Wu et al., 2013] Wu, R., Xu, L., Liu, P., Zhang, Y., Zheng, Z., Li, H., and Liu, X. (2013). Freeform illumination design: a nonlinear boundary problem for the elliptic monge&#x2013;amp&#xe9;re equation. *Opt. Lett.*, 38(2):229–231.

INRIA-Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, FRANCE and CEREMADE, UMR CNRS 7534, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16

*Email address*: `jean-david.benamou@inria.fr`

Signify Research High Tech Campus 7, 5656 AE Eindhoven, The Netherlands

*Email address*: `wilbert.ijzerman@signify.com`

INRIA-Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, FRANCE and CEREMADE, UMR CNRS 7534, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16

*Email address*: `giorgi.rukhaia@inria.fr`