

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 a classic example of an optimal transport problem with a non-euclidean displacement cost [Wang, 2004] [Glimm and Oliker, 2003]. This work discusses 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 bias induced by the Entropic Optimal Transport using the recent notion of Sinkhorn divergences [Ramdas et al., 2017] [Genevay et al., 2018] [Feydy et al., 2018] is shown to be necessary to achieve reasonable results. The paper does not establish any new theoretical result but discusses the necessary mathematical and numerical tools needed to produce and analyse 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.

Key words. Inverse reflector problem, optimal transportation, non-linear optimization.

Mathematics Subject Classification. 78A46, 49Q22, 65K10.

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 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 μ 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 ν , 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 designing a reflector, i.e. a surface which reflects a given illuminance towards a given illumination (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]. The general Kantorovich primal and dual form of the OT problem is (see [Villani, 2008]) :

THEOREM 1 (Kantorovich duality). *Given two compact manifolds X and Y endowed with a continuous, bounded from below “displacement” cost function $c : X \times Y \rightarrow \mathbf{R}$ and two borel probability measures $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Then, Kantorovich*

*Received April 9, 2020; accepted for publication November 6, 2020.

[†]INRIA-Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, France; & CEREMADE, UMR CNRS 7534, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, France (jean-david.benamou@inria.fr).

[‡]Signify Research High Tech Campus 7, 5656 AE Eindhoven, The Netherlands (wilbert.ijzerman@signify.com).

[§]INRIA-Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, France; & CEREMADE, UMR CNRS 7534, Université Paris Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, France (giorgi.rukhaia@inria.fr).

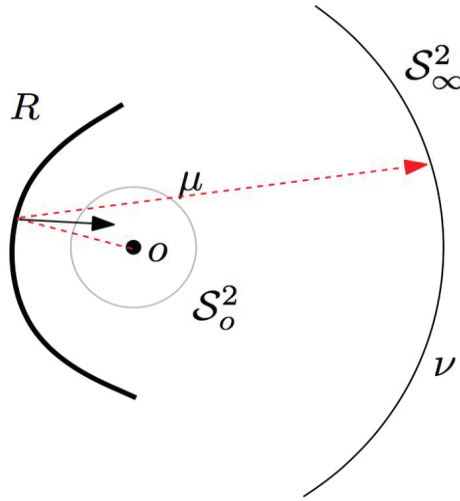


FIG. 1.1. Reflector problem from Point source O to Far Field.

problem in primal and dual forms (1.1) :

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

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, has solutions.

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_Ω is the characteristic function, i.e. a constant 1 on Ω .

Under suitable hypothesis on the ground cost 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 transport plan γ is concentrated on a graph of the OT map $y = T(x)$ implicitly defined by the saturation of the dual constraint:

$$f(x) + g(T(x)) = c(x, T(x)), \mu \text{ a.e.} \tag{1.2}$$

The pair (f, g) are called the Kantorovich potentials and are 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 μ to ν). The pushforward of μ is the measure defined as

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

The following distance property of OT will also be useful when comparing ray traced illuminations:

REMARK 1 (*L^p Wasserstein metric*). *If $X = Y$ and $c := 1/p d^p(x, y)$ such that (X, c) is a compact metric space. Then, the OT cost defines a separable metric on the set of probability measures on X : 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, μ and ν which represent the given illuminance and illumination probability distributions. Then theorem 1 applies to the cost function*

$$c(x, y) = -\log(1 - x \cdot y). \tag{1.4}$$

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 in theorem 1 and the saturation property (1.2) we get

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

Using definition (1.4) for c gives

$$\frac{e^{-g(T(x))}}{1 - x \cdot T(x)} = e^{f(x)} \leq \frac{e^{-g(y)}}{1 - x \cdot y}, \mu \otimes \nu \text{ a.e.} \tag{1.6}$$

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 is given as:

$$R = \{x e^{f(x)} \mid x \in S_0\}. \tag{1.7}$$

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 methods for the Kantorovich problem. We summarize here the general classes of numerical methods, their comparative advantages and disadvantages and whether and they can tackle the reflector cost (1.4).

Linear Programming approach. 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

$$\mu_N = \sum_{i=1}^N p_i \delta_{x_i}, \quad \nu_N = \sum_{j=1}^N q_j \delta_{y_j}, \quad \text{where} \quad \sum_{i=1}^N p_i = \sum_{j=1}^N q_j = 1. \tag{1.8}$$

Of course the number of discrete points for μ and ν may differ, we keep here N for both for simplicity.

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

$$OT_N(p, q) := \min_{\gamma_N \in \Pi(p, q)} \langle c_N, \gamma_N \rangle_{X_N \otimes Y_N} \quad (1.9)$$

where

$$\Pi(p, q) := \{ \gamma_N \in \mathbb{R}_+^{N \times N} \mid \langle 1_{X_N}, \gamma_N \rangle_{Y_N} = p, \langle 1_{Y_N}, \gamma_N \rangle_{X_N} = q \} \quad (1.10)$$

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

Problem (1.9) is a discrete linear programming problem which can be solved numerically using standard linear programming solvers and can be applied to the reflector cost (1.4). The main drawback of this method is its 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 $N > 100$).

Partial Differential approach. At least formally, taking the gradient of (1.2) gives the OT map as a function of f the Kantorovich potential:

$$T(x) = \{y \mapsto \nabla_x c(x, y)\}^{-1}(\nabla f(x)). \quad (1.11)$$

Assuming sufficient smoothness to interpret (1.3) in a pointwise sense and plugging this expression gives a Monge-Ampère PDE for the L^2 cost. This formulation has been studied in detail by Urbas [Urbas, 1997]. The boundary condition is replaced by a state constraint on the gradient of the solution, i.e. the optimal map maps to the support of the target measure. Finite elements and finite differences discretizations of the Monge Ampère operator are available but rather technical see [Neilan et al., 2020] for a review. The discrete non-linear system can be solved using a damped Newton's method. The non standard boundary condition can be tackled as explained in [Benamou et al., 2014]. This approach has an optimal linear complexity for the L^2 cost.

For the reflector cost (1.4) the PDE (see equation (1.5) in [Wang, 2004]) is even more nonlinear 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 and a numerical complexity in $O(N^{\frac{3}{2}})$ where, we recall, N is the number of points discretizing each marginal.

Another heuristic solution method of the Monge-Ampère problem based on a least squares approach is presented in [Romijn et al., 2020] [Romijn et al., 2019]. It can be applied to the reflector cost and the numerical study also demonstrates a $O(N^{\frac{3}{2}})$ complexity.

Semi-discrete approach. Semi-discrete OT takes advantage of the special case where μ has continuous density and the other ν_N is discrete as in (1.8). Then the dual formulation in 1) can be reduced to a semi-dual discrete optimisation problem through the elimination of the constraints replacing f by its c -transform:

$$f(x) = g^c(x) = \min_y (c(x, y) - g(y)) = \min_j (c(x, y_j) - g(y_j)), \quad (1.12)$$

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

It is now well understood that a Newton’s method can be used to solve the SD OT problem under classical regularity hypothesis for the displacement cost [Mérigot, 2011]. The implementation relies on fast (linear time) computations of a tessellation of the target domain called Laguerre cells:

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

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 and leading to additionnal technical difficulties. The resulting method still has linear complexity. Efficient implementations 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.

Entropic approach. The approach, pursued in this paper, is based on a penalization of the OT problem called “entropic regularization”. It was introduced for OT computations in [Cuturi, 2013] (see [Peyré and Cuturi, 2018] for a comprehensive review). This research 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 giving optimal $O(N \log N)$ complexity, this is briefly discussed in remark 4. Secondly, it can be applied to a wide range of costs and does not seem to have yet been investigated yet for the reflector cost. The downside of the entropic regularization with respect to the reflector application is that it computes a biased approximation of the optimal transport distance and map. The bias is controlled by a parameter, usually denoted ε , quantifying the strength of the penalization.

1.3. Outline. In section 2, we detail the entropic approach and recently obtained convergence properties of this approximation when $\varepsilon \rightarrow 0$. We also discuss the entropic 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, choice of discretization and interpolation.

Wasserstein distance is a natural distance to measure discrepancies between measures. In section 4 we therefore propose to evaluate it between the illumination ν and a sampling of its computed approximation. The sampling is obtained via a ray tracing simulations on a continuous interpolation of the reflector. We use this residual error to study the convergence with N and ε . We also show how the continuous approximation of the reflector behaves when the number of traced rays increases.

2. Entropic OT. The “entropic regularization” of the Kantorovich problem (1) is based on the following KullBack-Leibler divergence or “relative entropy” (KL) penalization:

$$\begin{aligned}
 OT_\varepsilon(\mu, \nu) &:= \min_{\gamma_\varepsilon \in \Pi(\mu, \nu)} \langle c, \gamma_\varepsilon \rangle_{X \times Y} + \varepsilon \text{KL}(\gamma_\varepsilon \mid \mu \otimes \nu) \\
 &= \max_{f_\varepsilon, g_\varepsilon} \langle f_\varepsilon, \mu \rangle_X + \langle g_\varepsilon, \nu \rangle_Y - \varepsilon \langle \exp(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c)) - 1, \mu \otimes \nu \rangle_{X \times Y},
 \end{aligned}
 \tag{2.1}$$

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

$$\text{KL}(\gamma \mid \mu \otimes \nu) := \int_{X \times Y} \log\left(\frac{d\gamma}{d\mu \otimes d\nu}\right) d\gamma$$

if γ is absolutely continuous w.r.t to $\mu \otimes \nu$ and $+\infty$ else.

The primal-dual optimality condition is given by

$$\gamma_\varepsilon = \exp\left(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c)\right) \mu \otimes \nu. \quad (2.2)$$

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

2.1. Sinkhorn algorithm. We will restrict from now on to the discrete case (see (1.8)) and the dual optimal transport problem. Numerical solutions of the Entropic OT problem are produced under the discretization (1.8) of this problem, i.e. replacing (X, Y, c, μ, ν) by $(X_N, Y_N, c_N, \mu_N, \nu_N)$ in formulation (2.1):

$$\begin{aligned} OT_{\varepsilon, N} := & \max_{f_\varepsilon, g_\varepsilon} \langle f_\varepsilon, \mu_N \rangle_{X_N} + \langle g_\varepsilon, \nu_N \rangle_{Y_N} \\ & - \varepsilon \langle \exp\left(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon - c_N)\right) - 1, \mu_N \otimes \nu_N \rangle_{X_N \times Y_N}. \end{aligned} \quad (2.3)$$

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

The resulting discrete non-linear system of optimality is solvable using Sinkhorn algorithm. It corresponds to a block coordinate $(f_\varepsilon$ and $g_\varepsilon)$ ascent: Initialize with $g_\varepsilon^0 = 0_{Y_N}$ and then iterate (in k):

$$\begin{aligned} f_\varepsilon^{k+1} &= -\varepsilon \log(\langle \exp\left(\frac{1}{\varepsilon}(g_\varepsilon^k - c_N)\right), \nu_N \rangle_{Y_N}) \\ g_\varepsilon^{k+1} &= -\varepsilon \log(\langle \exp\left(\frac{1}{\varepsilon}(f_\varepsilon^{k+1} - c_N)\right), \mu_N \rangle_{X_N}). \end{aligned} \quad (2.4)$$

Sinkhorn linear convergence in terms of number of iteration for fixed ε has been largely studied (see [Peyré and Cuturi, 2018]) again). For simple costs the rate behaves as $1 - \varepsilon$. The numerical stability depends on typical transport scale of the data $\tau = \sup_{x \in X} c(x, T(x))$ and ε . The computer hits overflows or underflows when the ratio τ/ε is too large. Decreasing ε below a certain threshold is therefore a numerical difficulty. A good review of existing hacks and methods to mitigate this problem can be found in [Schmitzer, 2016] and we rediscuss this in remark 4. The convergence as $\varepsilon \rightarrow 0$ 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 iterations (2.4) are defined on the discrete sets X_N and Y_N but they admit a canonical extension 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:

$$f[g_\varepsilon](x) = -\varepsilon \log\left(\sum_{j=1 \dots N} \exp\left(\frac{1}{\varepsilon}(g_\varepsilon(y_j) - c(x, y_j))\right) \nu_N(y_j)\right), \quad \forall x \in X. \quad (2.5)$$

As our goal is to form the reflector (1.7) from the discrete numerical potential f_ε . We are therefore interested in the convergence of f_ε as $\varepsilon \rightarrow 0$ and $N \rightarrow \infty$. To the best of our knowledge the joint convergence in N and ε has only been studied in [Berman, 2017], we reproduce partially his results:

THEOREM 3 (Berman joint convergence - corollary 1.3 [Berman, 2017]). *We assume μ and ν have $\mathcal{C}^{2, \alpha}$ and positive densities, and that N and ε 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 property”*

(see Remark 2 below) is also necessary. Then there exists a positive constant A_0 such that for any $A > A_0$ the following holds: setting $m_\varepsilon = \lceil -A \log(\varepsilon)/\varepsilon \rceil$ the continuous interpolation provided by $f[g_\varepsilon^{m_\varepsilon}]$, built using the canonical extension (2.5) from the discrete Sinkhorn iterate at $k = m_\varepsilon$, satisfies the estimate

$$\sup_X |f[g_\varepsilon^{m_\varepsilon}] - f| \leq -C\varepsilon \log(\varepsilon), \quad (2.6)$$

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

REMARK 2 (Density property Lemma 3.1 [Berman, 2017]). The “density property” in theorem 3 is defined as follows. For any given open set U intersecting the support X of μ (same for Y and ν)

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

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 μ (1.8) is such that, for U the sequence of approximations $\mu_N(U)$ never converges faster to 0 than ε (remember that $N = (1/\varepsilon)^d$). For $X = \mathbb{R}^d$, uniform grids are fine. See section 3.2 for the reflector case.

REMARK 3 (Domains and Cost). Theorem 3 holds in 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 4 (On the computational efficiency of Sinkhorn). The simplest implementation of Sinkhorn requires $O(N^2)$ operations per iterations. If we use the relation between discretization and entropic parameters in theorem 3, we need at least $O(\sqrt{N} \log(\sqrt{N}))$ iterations to reach $O(\varepsilon \log(\varepsilon))$ precision for the Kantorovich potential. This takes us to a pessimistic $O(N^{\frac{5}{2}} \log(\sqrt{N}))$ far from the optimal complexity of Semi-Discrete OT solvers for example.

This can be largely improved in practice by using a multi-scale method in ε and N . In the small ε limit the entropic plan concentrate on the graph of the optimal transport map. The bandwidth of (2.2) decreases with ε and coarser scales in this parameter can be used to restrict the relevant support on which to perform the summation in the Sinkhorn Iterates. This approach has been proposed and tested in [Schmitzer, 2016] and [Feydy, 2019] and experimentally yield a $O(N \log N)$ cost.

2.2. Entropic bias and Sinkhorn Divergence. From now on, the potentials computed using Sinkhorn iterations (2.4) are denoted by f_{OT_ε} and g_{OT_ε} .

Entropic OT is popular to approximate transport distances between probability measures. The Entropic cost $OT_\varepsilon(\mu, \nu)$ is known to converge exponentially fast to $OT(\mu, \nu) = OT_0(\mu, \nu)$ with ε [Cominetti and Martin, 1994]. Regarding the reflector problem, we are not interested in the transport cost but in f_{OT_ε} itself, the approximation of f . As can be seen formally on the dual formulations of (1) and (2.1) the gradients of the transport and entropic transport costs as a function of measures μ and ν are given by the Kantorovich potentials: $\nabla_{\mu, \nu} OT(\mu, \nu) = (f, g)$ and $\nabla_{\mu, \nu} OT_\varepsilon(\mu, \nu) = (f_{OT_\varepsilon}, g_{OT_\varepsilon})$. The convergence rate (2.6) in theorem 3 with an infinite slope at $\varepsilon = 0$ and the numerical stability limit imposed when decreasing ε in (2.4) does not bode well. Decreasing ε to avoid the pollution of the potential by

the entropy induced bias is a known issue. This problem is discussed in [Feydy et al., 2018] where it is proposed, in order to correct the bias without decreasing ε under the numerical stability limit to add “diagonal terms” to correct the entropic cost:

$$S_\varepsilon(\mu, \nu) = OT_\varepsilon(\mu, \nu) - \frac{1}{2}(OT_\varepsilon(\mu, \mu) + OT_\varepsilon(\nu, \nu)). \tag{2.7}$$

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

$$\begin{aligned} f_{OT_\varepsilon}^{\mu, k+1} &= -\varepsilon \log(\langle \exp(\frac{1}{\varepsilon}(f_{OT_\varepsilon}^{\mu, k} - c_N)), \mu \rangle_X) \\ f_{OT_\varepsilon}^{\nu, k+1} &= -\varepsilon \log(\langle \exp(\frac{1}{\varepsilon}(f_{OT_\varepsilon}^{\nu, k} - c_N)), \nu \rangle_Y). \end{aligned} \tag{2.8}$$

The μ gradient of S_ε , denoted f_{S_ε} may be formed by a simple subtraction. An open question is whether this correction:

$$f_{S_\varepsilon} = f_{OT_\varepsilon} - f_{OT_\varepsilon}^\mu, \tag{2.9}$$

is a better approximation to f than f_{OT_ε} . Numerical simulations of W_2 gradient flows in [Feydy et al., 2018] and, as we will see below, our reflector design simulations do indicate this is the case. Note finally that solving S_ε has the same complexity as OT_ε as the diagonal problems typically converge faster in terms of Sinkhorn iterations.

REMARK 5 (Asymptotics of OT_ε). *The entropic bias may be related formally to asymptotic results on the difference between OT and OT_ε , 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 asymptotic behavior for small ε :*

$$\frac{1}{\varepsilon} \{OT_\varepsilon(\mu, \nu) - OT(\mu, \nu)\} \simeq \frac{1}{2} \left\{ \int_Y \log(\nu) d\nu + \int_X \log(\mu) d\mu \right\}. \tag{2.10}$$

This result is for continuous and compactly supported densities μ and ν with respect to $\mathcal{L}(\mathbb{R}^d)$, the Lebesgue measure.

Using (2.10) and that $OT(\mu, \mu) = OT(\nu, \nu) = 0$, we first remark that the Sinkhorn Divergence correction removes at least the leading terms:

$$\frac{1}{2\varepsilon} \{OT_\varepsilon(\mu, \mu) + OT_\varepsilon(\nu, \nu)\} \simeq \frac{1}{2} \left\{ \int_Y \log(\nu) d\nu + \int_X \log(\mu) d\mu \right\}. \tag{2.11}$$

Formally taking the gradient in μ of this correction we get for small ε :

$$f_{OT_\varepsilon} \simeq f - \frac{\varepsilon}{2} \log(\mu). \tag{2.12}$$

It gives some indication at the leading order of the entropic bias. This is observed in figure 3.1.

Remark finally that the density property discussed in Remark 2 requires for the first order bias terms to converge to 0 when discrete measures μ_N converge to original measure μ .

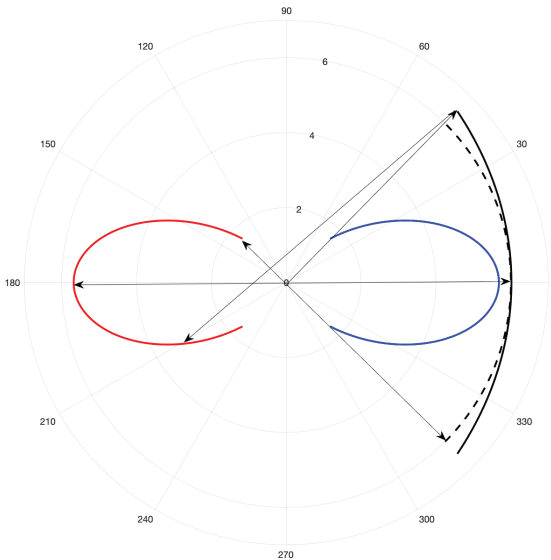


FIG. 3.1. Polar plot of a 1D computation. In blue and red are μ and ν . Solid and dashed black curves are R_ε and R . Also shown are rays traced. The bottom and middle rays are reflected if the exact opposite “identity” direction. The top rays, reflecting on the biased entropic reflector will send light strictly inside the support of ν .

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, each ray is reflected in its opposite direction, i.e. when $\nu(x) = \mu(x + \pi)$, $x \in \mathbb{S}$. 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 $R_\varepsilon = \{x e^{f \circ T_\varepsilon(x)} | x \in X\}$ 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 constants such that the reflectors superimpose at $x = 0$. The illuminance and illumination μ and ν are centered Gaussians and represented respectively in blue and red. This example shows the effect of the entropic bias (2.12). As μ decrease across angles, the reflector moves away from the source. We also trace rays : the bottom and middle rays are reflected if the exact opposite “identity” directions. The top rays, reflecting on the biased entropic reflector will send light strictly inside the support of ν . This shrinking effect of the re-simulation also 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, we propose to use the notion of Sinkhorn divergences (2.7). When the spaces X and Y are different and a for a general cost $c(x, y)$, the natural extension of the diagonal terms is as follows:

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

where $\mu' = \operatorname{argmin}OT(\mu, \cdot)$ and $\nu' = \operatorname{argmin}OT(\cdot, \nu)$. This is consistent with the distance costs $c(x, y) = d_p^p(x, y)$ as $\mu' = \mu$ and $\nu' = \nu$ in this case. Similarly, in the reflector cost case, μ' and ν' correspond to the reflections of respectively μ and ν on the circular reflector mentioned above. Indeed one has $\operatorname{argmin}_{y \in Y} -\log(1 - \langle x, y \rangle) = -x$.

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

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

3.2. Choice of Discretization. In Theorem 3, the requirement that ε is of order $(1/N)^{\frac{1}{d}}$ and that μ_N, ν_N satisfy the density condition (Remark 2) are closely related. Intuitively this condition means that while choosing the discretizations μ_N, ν_N with N points, it is important to make sure that they approximate the corresponding distributions μ, ν with integration error of order $(1/N)^{\frac{1}{d}}$ for functions which do not oscillate on finer scale than $(1/N)^{\frac{1}{d}}$. This density property is satisfied in particular by Quasi Monte-Carlo point clouds ([Berman, 2017]). This discretization is 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 vo-lume 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:

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

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 $s > \frac{d}{p}$:

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

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 μ which is absolutely continuous with respect to the volume form dV with density ρ_μ , the discrete approximation μ_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:

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

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 straightforward 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. We recall that the reflector problem is the inverse problem associated to the following forward problem: Given the source illumination μ 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 μ is reflected from R ? The geometric optics answer to this problem is the simulation of the light propagation known as Ray-Tracing. It can simply be described as follows:

The distribution μ over the sphere is sampled with an empirical measure $\mu^M := \sum_{i=1 \dots M} p_i \delta_{x_i}$. 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 $\nu^M := \sum_{i=1 \dots M} p_i \delta_{y_i}$ 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 sometimes referred to as “re-simulation”. The quality of course also depends on the sampling method and the number of samples M .

REMARK 6 (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 (1.12) and (2.5)).

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-concave interpolation of the first potential f_N is

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

As discussed in section 1.1 (see (1.6)), the continuous reflector $R = \{xe^{\hat{f}(x)} | x \in X\}$ is the envelope of N paraboloids with focal directions $\{y_j\}$. The resulting illumination will always be an N point empirical measure regardless of the number of traced rays.

The second approach, also followed in [Feydy et al., 2018] to compute the continuous gradients of the Kantorovich potentials, is to use either the smoothed transform (2.5) for OT_ϵ or (3.2) for S_ϵ . The Entropic interpolation is also a key ingredient in the proof of theorem 3.

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_ϵ problem, the extension of the first potential $f_{OT_\epsilon, N}$ is

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

Note that (3.7) is the “softmin” version of the minimization (3.6) (its limit when $\varepsilon \rightarrow 0$). It also inherits the smoothness of $x \mapsto c(x, y)$ and is therefore C^∞ .

Finally using the “identity” potentials $(f_{OT_\varepsilon, N}^\mu, f_{OT_\varepsilon, N}^\nu)$ computed when applying Sinkhorn algorithm applied to S_ε , the smooth interpolation is:

$$\tilde{f}_{S_\varepsilon}(x) := \tilde{f}_{OT_\varepsilon} + \varepsilon \log\left(\sum_{j=1\dots N} \exp\left(\frac{1}{\varepsilon}(f_{OT_\varepsilon}^\mu(y_j) - c(x, y_j))\right)\mu'_N(y_j)\right). \quad (3.8)$$

This approach results 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 inputs are analytical descriptions of the illumination/illuminance μ and ν 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 $\{(x_1, x_2) \in \mathbb{R}^2 \mid -0.6 \leq x_1 \leq 0.6, -0.6 \leq x_2 \leq 0.6\}$. Similarly, $Y \subset \mathbb{S}^2$ will be the inverse stereographic projection in the southern hemisphere of the 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 (2.3) will be denoted using the subscript \cdot_{OT_ε} , and those obtained from the minimization of the de-biased functional (3.1) by subscript \cdot_{S_ε} . The discrete Kantorovich potentials are computed for the discretizations (μ_N, ν_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:

$$(x, y, z) \in \mathbb{R}^3 \rightarrow \left(\frac{x}{1+z}, \frac{y}{1+z}\right) \in \mathbb{R}^2.$$

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: ε the entropic regularization, N the discretization of the distributions μ and ν 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 was discussed above and in section 3.3.

We will use the notation $\hat{R}_{OT_\varepsilon}^N, \tilde{R}_{OT_\varepsilon}^N, \hat{R}_{S_\varepsilon}^N, \tilde{R}_{S_\varepsilon}^N$ for the reflectors defined using (1.7) and respectively $\hat{f}_{OT_\varepsilon}^N, \tilde{f}_{OT_\varepsilon}^N, \hat{f}_{S_\varepsilon}^N, \tilde{f}_{S_\varepsilon}^N$.

We will denote by $\nu^M = \hat{R}_{OT_\varepsilon}^N[\mu^M]$ the ray traced sampling with M rays (and a similar notation for the other versions of the reflector). Of course ν^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. The data and numerical results are available on the repository: <https://zenodo.org/record/3888145/files/X6fTkTJKjxU>. 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.

Test Case 1: Square To Circle. The source distribution μ is the uniform distribution over the inverse stereographic projection of a square: $StP(supp(\mu)) = \{(x_1, x_2) \in \mathbb{R}^2 \mid -0.5 \leq x_1 \leq 0.5, -0.5 \leq x_2 \leq 0.5\}$. The target distribution ν is the uniform distribution over the inverse stereographic projection of a disk: $StP(supp(\nu)) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 0.5^2\}$. 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_ε} , \hat{R}_{OT_ε} , \hat{R}_{S_ε} , \hat{R}_{S_ε} and finally also the QMC discretization used for ν_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 ν discretization. All rays hitting one parabola end up at the same position. So up to numerical errors we indeed recover (e), that is ν_N even though $M \gg N$. Also the S_ε 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_ε entropic bias as in section 3.1, which smoothing the density at the boundary of the support would produce, according to formula (2.12) and the effect observed in figure 3.1. As expected, the S_ε (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 ν has a Gaussian density $\rho(x_1, x_2) = e^{-\frac{x_1^2 + x_2^2}{2}}$ over the source domain in Test 1. The source distribution will be the same as in Test Case 1.

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

Test Case 4: Circle To Steep Gaussian. The source distribution μ is the uniform distribution over the inverse stereographic projection of a disk $StP(supp(\mu)) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 0.5^2\}$. The target distribution ν has a Gaussian density $\rho(x_1, x_2) = e^{-16*(x_1^2 + x_2^2)}$ over the source domain in Test 1.

Test Case 5: Circle To Two Steep Gaussians. The source distribution μ is the same as in the fourth. The target distribution ν has a Gaussian density $\rho(x_1, x_2) = e^{-16*((x_1-0.25)^2+(x_2-0.25)^2)} + e^{-16*((x_1-0.25)^2+(x_2+0.25)^2)}$ over the source domain in Test 1.

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 $\hat{f}_{OT_\varepsilon}^N, \hat{f}_{OT_\varepsilon}^N, \hat{f}_{S_\varepsilon}^N, \hat{f}_{S_\varepsilon}^N$ 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 6, 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 ν . For this we will use the standard L^2 Wasserstein distance $W_2(T^{app} \# \mu, \nu)$, or its S_ε entropic counterpart presented in section 2. It is known to provide a smooth distance between empirical distributions [Chizat et al., 2020]. In order to build a numerically computable error estimate we will use the ray traced sample μ^M of the source measure μ .

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

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

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 A giving:

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

where K depends on the data of the problem. The convergence in W^2 norm of such a 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 $\nu_M = \hat{R}_{OT_\varepsilon}^N[\mu^M]$ 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 6):

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

The continuous densities μ and ν 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 μ and ν 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 ε : $W_2(\cdot, \cdot) \simeq S_{\varepsilon=1e-06}(\cdot, \cdot)$, 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 μ and ν are discretized with a finer $M_\infty = 512^2$ points orthogonal grid on the plane.

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

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

For Test cases 4 and 5, we obtained

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

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 ν_M is at best a QMC sampling of ν .
- (2) Decreasing ε improves the OT_ε reflectors.
- (3) Sinkhorn divergence S_ε de-biasing is effective, both in the sense of obtaining lower errors, and being less dependent on the choice of ε .
- (4) As the target is smooth, the entropic interpolation of S_ε 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 $\nu^M = \tilde{R}_{S_\epsilon}^N[\mu^M]$.

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 ν_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 $\nu^M = \tilde{R}_{S_\epsilon}^N[\mu^M]$.
- (2) In logarithmic scales the curves agree with the $M^{-\frac{1}{2}}$ rate predicted by the theory [Fournier and Guillin, 2015].

5. Conclusion. In term of computing the Kantorovich potential f , the PDE or Semi Discrete approaches are the natural choice when c is the Euclidean distance. For the reflector cost however they are rather technical and difficult to implement. The very few references we know off are listed section 1.2.

The entropic regularization method on the other hand is flexible and Sinkhorn algorithm can be implemented very rapidly an efficiently for any cost c as recalled in section 4. The downside is the entropic bias induced by the regularization which is not acceptable for the reflector problem.

We tested a novel approach [Feydy et al., 2018] called Sinkhorn Divergence to correct this bias. Based on theorem 3 [Berman, 2017] we studied numerically the convergence of the corrected method with respect to the entropic regularization and space discretization parameters. Quantitative error measures 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.

Our experiments indicate that the combination of a QMC discretization satisfying the density property (remark 2), the entropic interpolation (2.3) and correcting the Entropic bias with the Sinkhorn divergence technique produce the best results and can be used to design reflectors at least for simple test cases.

Acknowledgements. This work was supported by ROMSOC EID H2020 network under the Marie Curie Grant Agreement No. 765374. We also want to thank Robert Berman, Simone Di Marino, Jean Feydy, Gabriel Peyré and François-Xavier Vialard for helpful discussions.

Figures.

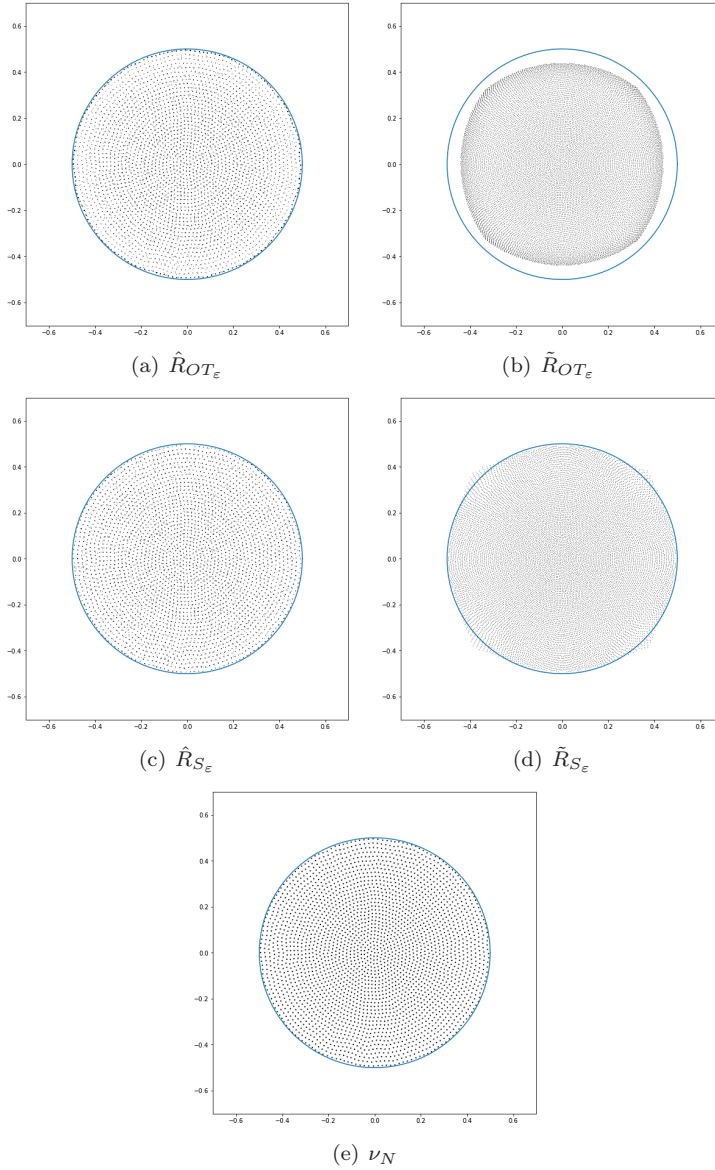


FIG. 5.1. Test Case 1: From left to right and from top to bottom the re-simulations of μ^M ($M=128*128$) using the reflectors \hat{R}_{OT_ε} , $\tilde{R}_{OT_\varepsilon}$, \hat{R}_{S_ε} , $\tilde{R}_{S_\varepsilon}$ (check definition 4 for explanation of the notations) and finally also the QMC discretization used for ν_N , $N = 64 * 64$ points. The regularization parameter was taken to be $\varepsilon = \frac{1}{2*64}$ for all four solutions.

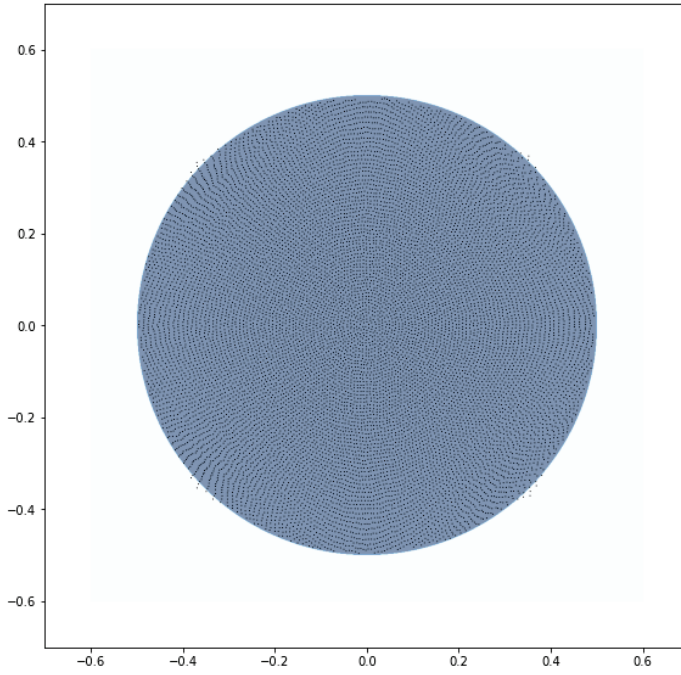


FIG. 5.2. *Test Case 1: Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)*

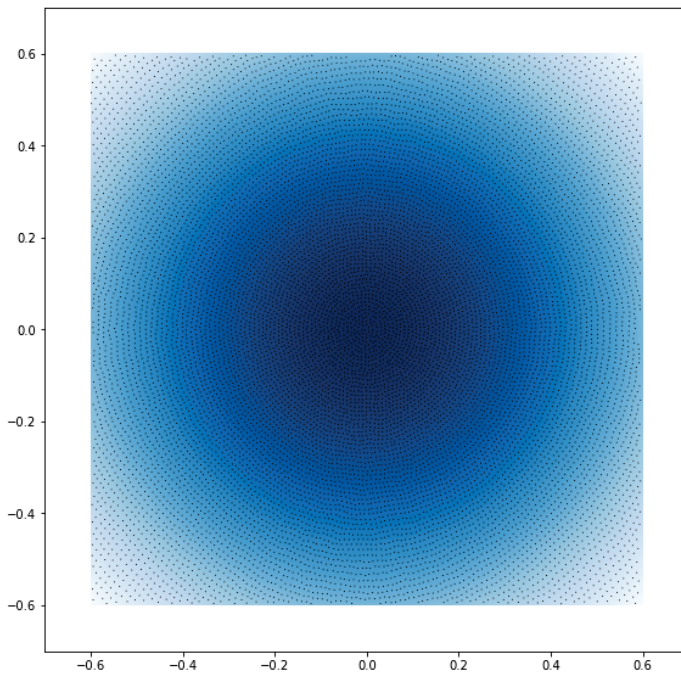


FIG. 5.3. *Test Case 2: Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)*

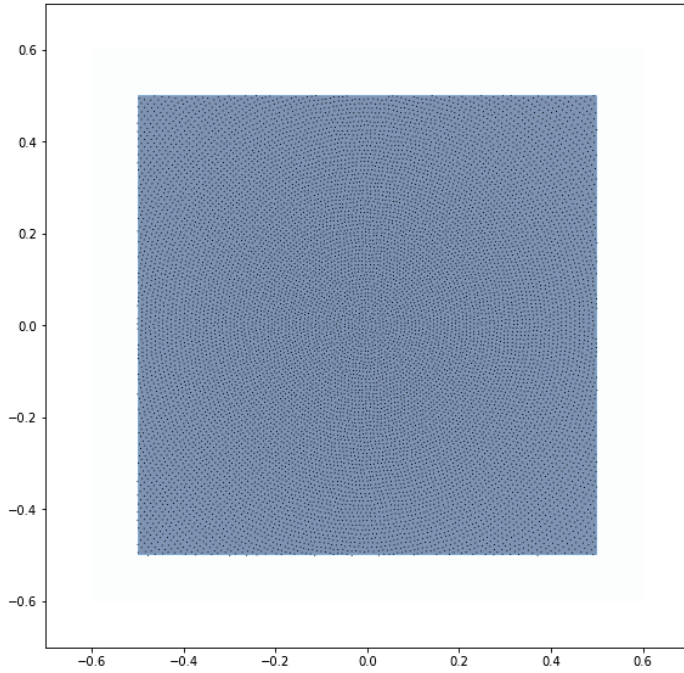


FIG. 5.4. Test Case 3: Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)

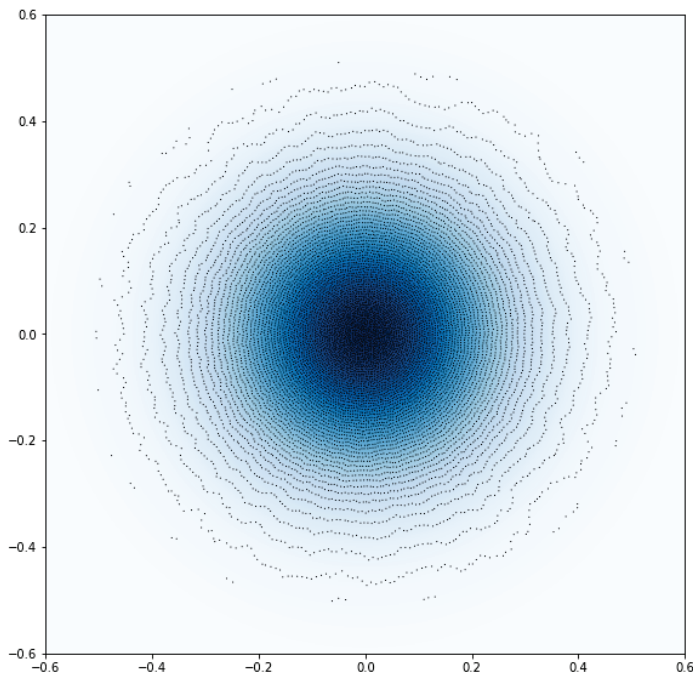


FIG. 5.5. Test Case 4: Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)

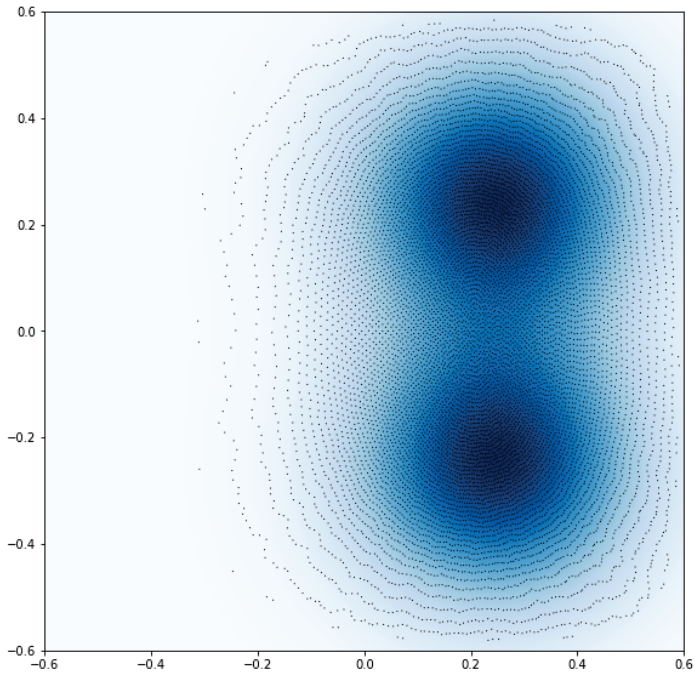


FIG. 5.6. *Test Case 5: Re-simulation of $\tilde{R}_{S_{1/8\sqrt{N}}}$ (point cloud) together with the exact desired distribution (color graph)*

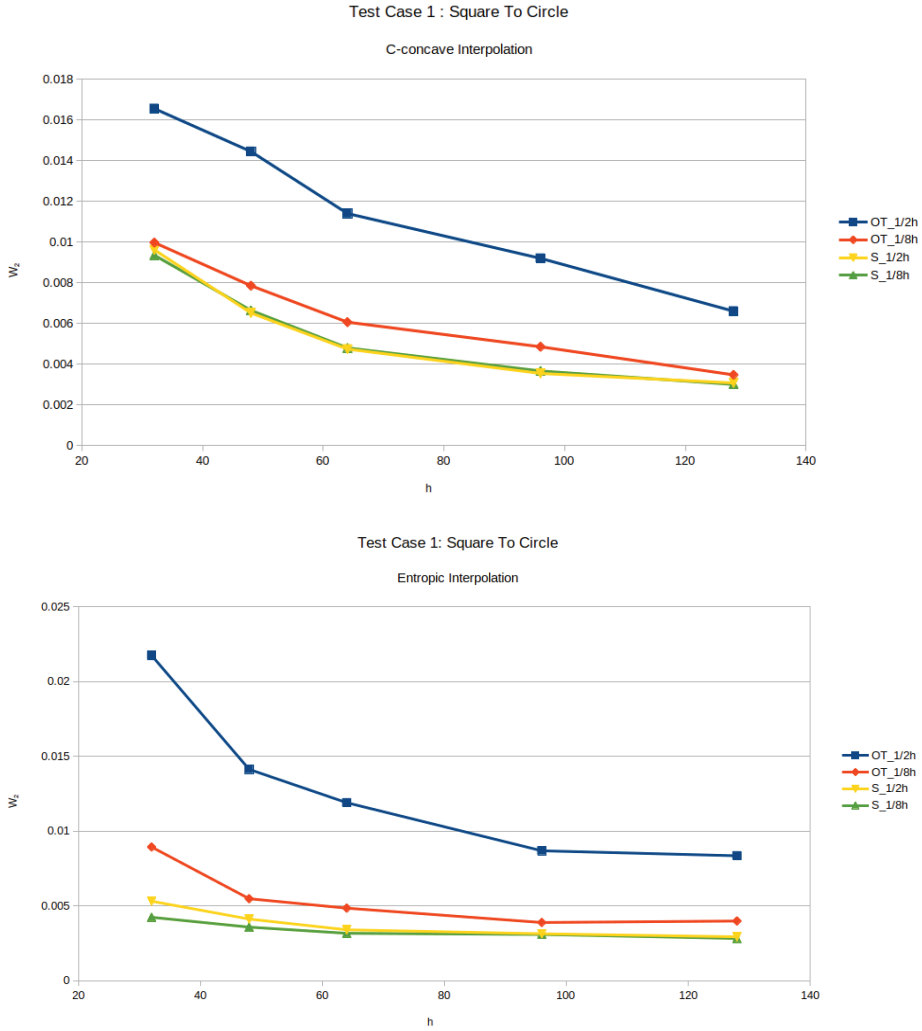


FIG. 5.7. Test Case 1 ($h := \sqrt{N}$): W_2 distance between Re-simulated Push-Forward with 128^2 points and exact target.

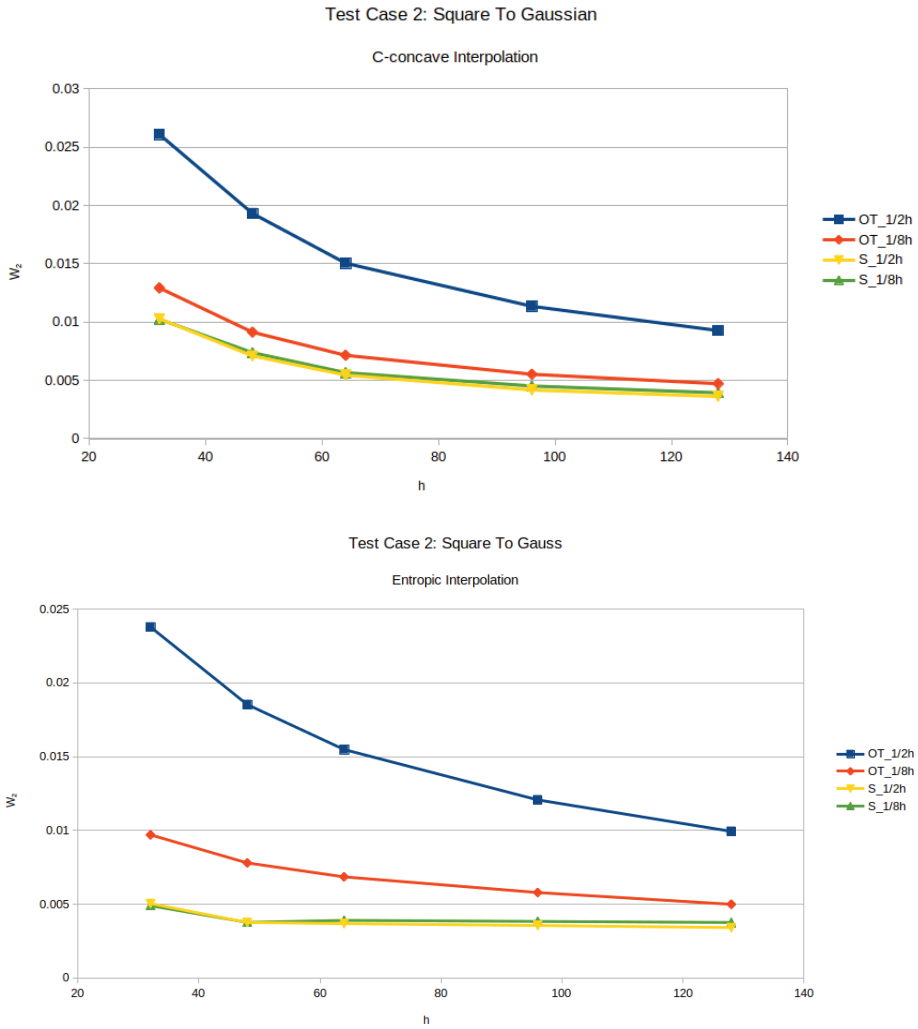


FIG. 5.8. Test Case 2 ($h = \sqrt{N}$): W_2 distance between Re-simulated Push-Forward with 128^2 points and exact target.

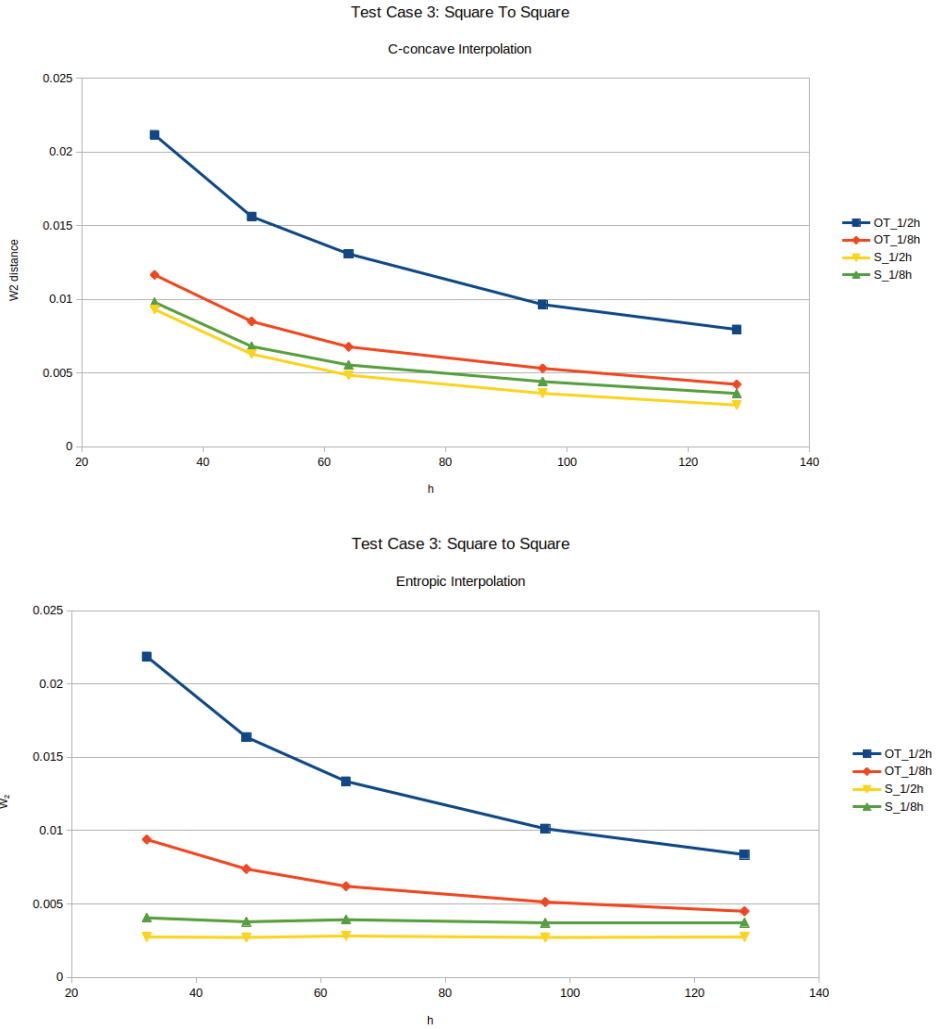


FIG. 5.9. Test Case 3 ($h = \sqrt{N}$): W_2 distance between Re-simulated Push-Forward with 128^2 points and exact target.

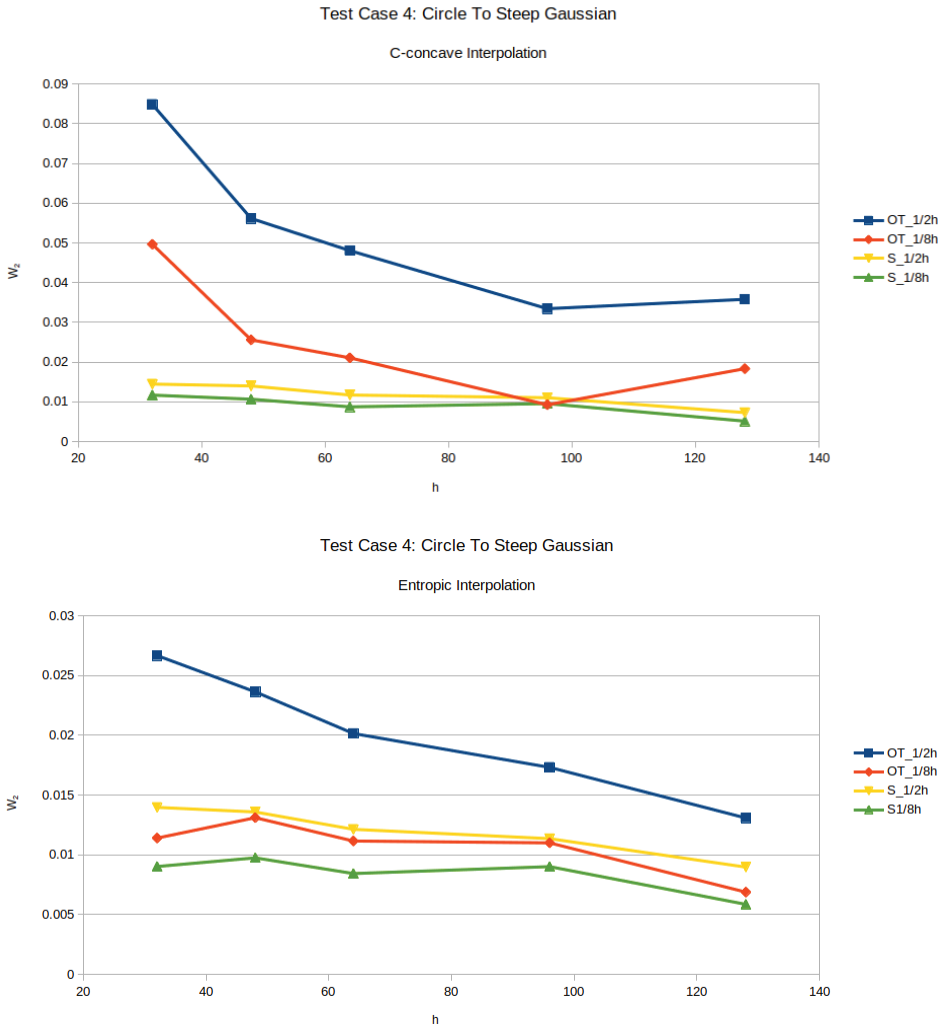


FIG. 5.10. Test Case 4 ($h = \sqrt{N}$): W_2 distance between Re-simulated Push-Forward with 128^2 points and exact target.

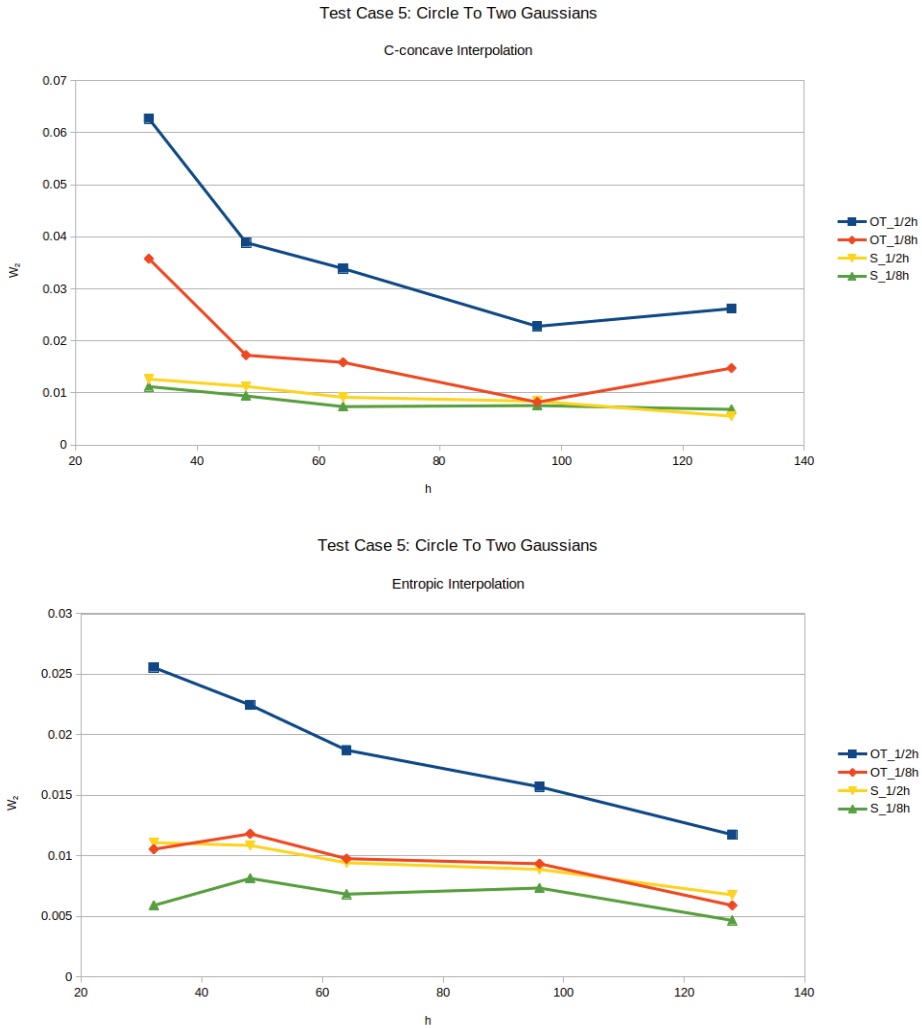


FIG. 5.11. Test Case 5 ($h = \sqrt{N}$): W_2 distance between Re-simulated Push-Forward with 128^2 points and exact target.

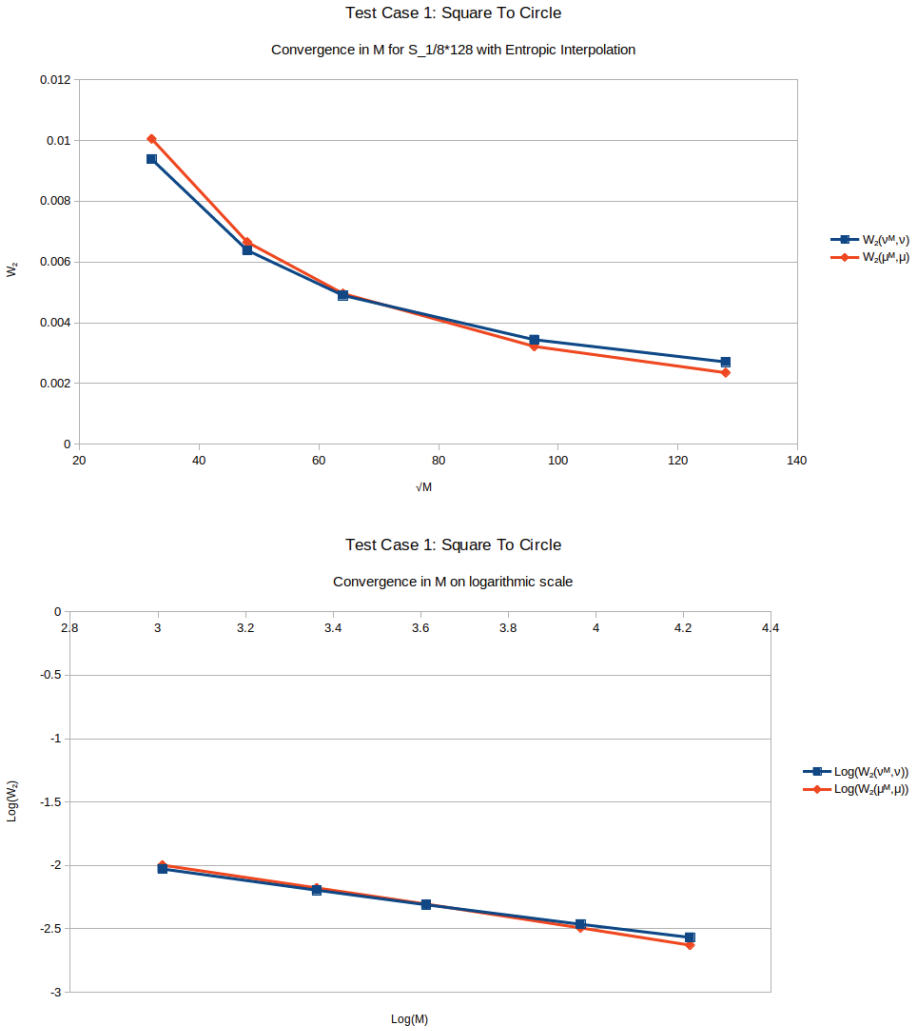


FIG. 5.12. Test Case 1: Convergence of error terms $W_2(\mu^M, \mu)$, $W_2(\nu^M, \nu)$, original (top) and logarithmic (bottom) scales.

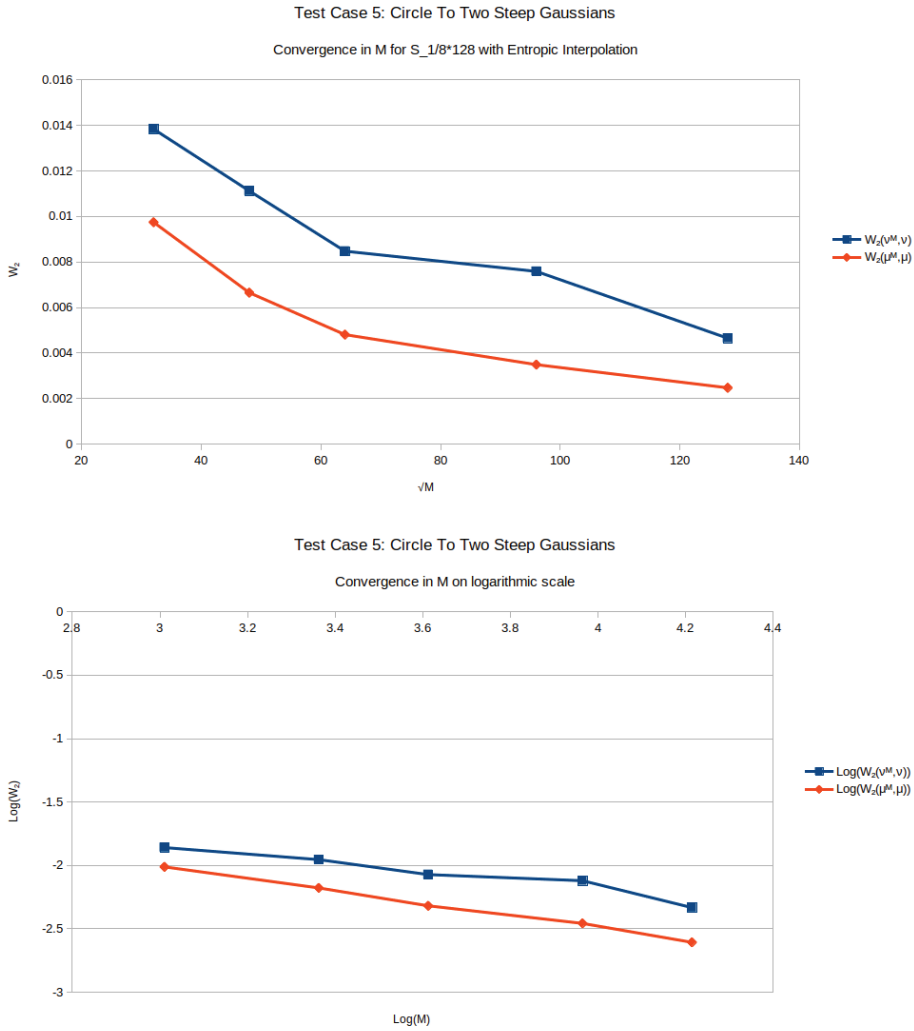


FIG. 5.13. Test Case 5: Convergence of error terms $W_2(\mu^M, \mu)$, $W_2(\nu^M, \nu)$, original (top) and logarithmic (bottom) scales.

Appendix A.

LEMMA 4. *Given two Polish spaces (X, d_X) and (Y, d_Y) , with Borel probability measures μ_1, μ_2 on X and Lipschitz continuous map $T : X \rightarrow Y$ with Lipschitz constant $Lip(T)$, Then*

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

Proof. Let (f, g) be optimal pair of Kantorovich potentials for $W_p(T\#\mu_1, T\#\mu_2)$. Then, for all $x, x' \in X$

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

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:

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

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

When the target support Y is convex and the densities of μ and ν 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

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

where the constant K only depends on the dimension d and the the data μ and ν . See [Philippis and Figalli, 2014] for further refinements and discussions.

REFERENCES

- [Benamou et al., 2014] J.-D. BENAMOU, B. D. FROESE AND A. M. OBERMAN, *Numerical solution of the optimal transportation problem using the monge-ampère equation*, Journal of Computational Physics, 260 (2014), pp. 107–126.
- [Berman, 2017] R. J. BERMAN, *The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampere equations*, ArXiv e-prints, (2017).
- [Brix et al., 2014] K. BRIX, Y. HAFIZOGULLARI AND A. PLATEN, *Solving the monge-ampère equations for the inverse reflector problem*, Mathematical Models and Methods in Applied Sciences, 25 (2014).
- [Brix et al., 2015] K. BRIX, Y. HAFIZOGULLARI AND A. PLATEN, *Designing illumination lenses and mirrors by the numerical solution of monge-ampère equations*, Journal of the Optical Society of America A, 32:11 (2015), 2227.

- [Caffarelli, 1992] L. CAFFARELLI, *The regularity of mappings with a convex potential*, Journal of The American Mathematical Society - J AMER MATH SOC, 5 (1992).
- [Chizat et al., 2020] L. CHIZAT, P. ROUSSILLON, F. LÉGER, F.-X. VIALARD AND G. PEYRÉ, *Faster wasserstein distance estimation with the sinkhorn divergence*, (2020).
- [Cominetti and Martin, 1994] R. COMINETTI AND J. S. MARTIN, *Asymptotic analysis of the exponential penalty trajectory in linear programming*, Mathematical Programming, 67:1 (1994), pp. 169–187.
- [Conforti and Tamanini, 2019] G. CONFORTI AND L. TAMANINI, *A formula for the time derivative of the entropic cost and applications*, (2019).
- [Cuturi, 2013] M. CUTURI, *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*, pp. 2292–2300. Curran Associates, Inc., (2013).
- [Feydy, 2019] J. FEYDY, *GEOMLOSS*, <https://www.kernel-operations.io/geomloss/>, (2019).
- [Feydy et al., 2018] J. FEYDY, T. SÉJOURNÉ, F.-X. VIALARD, S.-I. AMARI, A. TROUVÉ AND G. PEYRÉ, *Interpolating between Optimal Transport and MMD using Sinkhorn Divergences*, working paper or preprint, (2018).
- [Fournier and Guillin, 2015] N. FOURNIER AND A. GUILLIN, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields, 162:3-4 (2015), 707.
- [Genevay et al., 2018] A. GENEVAY, G. PEYRE AND M. CUTURI, *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*, pp. 1608–1617 (2018), Playa Blanca, Lanzarote, Canary Islands. PMLR.
- [Glimm and Oliker, 2003] T. GLIMM AND V. OLIKER, *Optical design of single reflector systems and the monge-kantorovich mass transfer problem*, Journal of Mathematical Sciences, 117 (2003), pp. 4096–4108.
- [Kochengin and Oliker, 2003] S. A. KOCHENGIN AND V. I. OLIKER, *Computational algorithms for constructing reflectors*, Computing and Visualization in Science, 6:1 (2003), pp. 15–21.
- [Leclerc, 2019] H. LECLERC, *Pysdot software* - <https://pypi.org/project/pysdot/>, (2019).
- [Lévy, 2015] B. LÉVY, *A numerical algorithm for L^2 semi-discrete optimal transport in 3D*, ESAIM: Mathematical Modelling and Numerical Analysis, 49:6 (2015), pp. 1693–1715.
- [Léonard, 2013] C. LÉONARD, *A survey of the schrödinger problem and some of its connections with optimal transport*, (2013).
- [Machado Manhães De Castro et al., 2016] P. MACHADO MANHÃES DE CASTRO, Q. MÉRIGOT AND B. THIBERT, *Far-field reflector problem and intersection of paraboloids*, Numerische Mathematik, 134:2 (2016), pp. 389–411.
- [Mérigot, 2011] Q. MÉRIGOT, *A multiscale approach to optimal transport*, Computer Graphics Forum, 30:5 (2011), pp. 1584–1592.
- [Neilan et al., 2020] M. NEILAN, A. J. SALGADO AND W. ZHANG, *Chapter 2 - the monge-ampère equation*, in Bonito, A. and Nochetto, R. H., editors, *Geometric Partial Differential Equations - Part I*, volume 21 of *Handbook of Numerical Analysis*, pp. 105 – 219 (2020). Elsevier.
- [Pal, 2019] S. PAL, *On the difference between entropic cost and the optimal transport cost*, (2019).
- [Peyré and Cuturi, 2018] G. PEYRÉ AND M. CUTURI, *Computational Optimal Transport*, ArXiv e-prints, (2018).
- [Philippis and Figalli, 2014] G. D. PHILIPPIS AND A. FIGALLI, *The monge-ampère equation and its link to optimal transportation*, (2014).
- [Ramdas et al., 2017] A. RAMDAS, N. TRILLOS AND M. CUTURI, *On wasserstein two-sample testing and related families of nonparametric tests*, Entropy, 19:2 (2017), 47.
- [Romijn et al., 2019] L. ROMIJN, J. TEN THIJE BOONKAMP AND W. IJZERMAN, *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, (2019).
- [Romijn et al., 2020] L. ROMIJN, J. TEN THIJE BOONKAMP AND W. IJZERMAN, *Inverse reflector design for a point source and far-field target*, Journal of Computational Physics, 408 (2020).
- [Schmitzer, 2016] B. SCHMITZER, *Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems*, arXiv e-prints, page arXiv:1610.06519, (2016).
- [Urbas, 1997] J. URBAS, *On the second boundary value problem for equations of monge-ampère type*, Journal für die reine und angewandte Mathematik, 487 (1997), pp. 115–124.
- [Villani, 2008] C. VILLANI, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, (2008).
- [Wang, 2004] X.-J. WANG, *On the design of a reflector antenna ii*, Calculus of Variations and Partial

- Differential Equations, 20:3 (2004), pp. 329–341.
- [Womersley, 2017] R. S. WOMERSLEY, *Efficient Spherical Designs with Good Geometric Properties*, ArXiv e-prints, (2017).