Capacity Constrained Entropic Optimal Transport, Sinkhorn Saturated Domain Out-Summation and Vanishing Temperature

Jean-David Benamou, Mélanie Martinet

CAPACITY CONSTRAINED ENTROPIC OPTIMAL TRANSPORT, SINKHORN SATURATED DOMAIN OUT-SUMMATION AND VANISHING TEMPERATURE

JEAN-DAVID BENAMOU AND MÉLANIE MARTINET

ABSTRACT. We propose a new method to reduce the computational cost of the Entropic Optimal Transport in the vanishing temperature (ε) limit. As in [Schmitzer, 2016], the method relies on a Sinkhorn continuation "ε-scaling" approach; but instead of truncating out the small values of the Kernel, we rely on the exact "out-summation" of saturated domains for a modified constrained Entropic Optimal problem. The constraint depends on an additional scalar parameter λ. In pratice $\lambda = \varepsilon$ also vanishes and the constraint disappear. Using [Berman, 2017], the convergence of the (ε, λ) continuation method based on this modified problem is established. We then show that the saturated domain can be over estimated from the previous larger (ε, λ). On the saturated zone the solution is constant and known and the domain can be "out-summed" (removed) from Sinkhorn algorithm. The computational and cost and memory foot print can be estimated. The complexity decreases with the dimension and should be close to linear for dimension 3. We also give preliminary 1-D numerical experiments.

CONTENTS

1. INTRODUCTION

Solving numerically the OT problem is important in many fields of applied mathematics. The interested reader is refered to the recent books [Santambrogio, 2015] [Peyré and Cuturi, 2018] [Galichon, 2016]. Optimal Transport is an optimization problem over mapping between two probability measures. The cost to be minimized gives a distance on the space of measure (also known as Earth Dovers

Distance) "lifting" the geometric properties of measure support distance. It is now a standard tool in signal processing and Machine Learning [Kolouri et al., 2016]. The optimal mapping is also used, for instance to register landmark sample points between "shapes" encoded in Euclidean or higher dimensionnal feature spaces.

Entropic regularization [Cuturi, 2013] is a standard and widely used to approximation of Optimal transport. In this setting the optimization variable is a Transport Plan, a probability measure on the space of all possible coupling between the source and target measure supports. Its popularity is linked to its companion "Sinkhorn" algorithm providing a flexible and robust computing tool. The penalization parameter ε in front of the entropic regularization is also called "temperature". The computational and applicative limits of Sinkhorn algorithm are linked to ε. On one hand the computational complexity increases like $1/\varepsilon$, on the other hand the optimal transportation mapping bias increases strongly with ε. This is problematic for applications like the already mentionned registration but also for the reflector OT problem [Benamou et al., 2020] for example.

In the case of the classical quadratic cost and for dimension up to three, non-entropic Semi-Dicrete [Mérigot, 2011] [Lévy, 2015] or finite difference Monge-Ampère solvers [Benamou and Duval, 2017] can be used. The implementation of these methods involve some additional difficulties.

Choosing to stick with Entropic Optimal Transport and in order to mitigate the bias, [Feydy et al., 2018] suggests to modify further the optimization by removing the diagonal entropic cost to send the measure onto themselves. It can be connected to the recent theoretical effort to understand the linearization of Entropic OT around $\varepsilon = 0$ [Conforti and Tamanini, 2019] [Pal, 2019].

A different path detailed in section 1.5 is followed by [Schmitzer, 2016], see also [Feydy, 2020]. It tries to leverage the increasing sparsity of the support of the Entropic transport plan as it converges to the graph of the optimal transport map when $\varepsilon \to 0$. As sketched in the abstract, we propose in this paper an algorithm based on this approach with a theoretical background .

1.1. **Optimal Transportation (OT).** 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 $c : X \times Y \to \mathbf{R}$ and two borel probability measures $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(Y)$. Then, the Kantorovich problem in primal and dual forms (1.1) has solutions.*

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

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 :

$$C = \{(f,g) \in \mathcal{C}(X) \times \mathcal{C}(Y),\ f \oplus g \leq c\}, \tag{1.2}$$

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

Complementary slackness ensures that $\gamma^\star$, the optimal transport plan, has support where the constraint on the optimal dual potentials saturates, i.e. on $\{(x,y) \in X \times Y,\ f^\star(x) + g^\star(y) = c(x,y)\}$ that is on the set $\{(x, y_x^\star),\ x \in X\}$ where

$$y_x^\star = arg \inf_{y \in Y} c(x,y) - g^\star(y) \tag{1.3}$$

The solution of (1.3) depends on the choice and regularity of c; the L^2 cost $c = \frac{1}{2}\|x - y\|^2$ and more generally $c = h(x-y)$, h convex, have been intensively studied since [Brenier, 1991]. In those case (see [Santambrogio, 2015] for a comprehensive mathematical presentation), $x \mapsto y_x^\star$ is a map

and $\gamma^\star$ is concentrated on its graph denoted $\Gamma^\star \in X \times Y$.

This paper relies on the convergence result established in [Berman, 2017] (see theorem 4) and will adress the following setting : we will work on the d-dimensionnal torus $X = Y = \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ endowed with the usual periodic distance

$$c = \inf_{k\in\mathbb{Z}^d} \frac{1}{2}\|x + k - y\|^2. \tag{1.4}$$

See remark 5 on the class of costs we could envision extending the method.

We will keep the generic notations X, Y and c. The probability measures μ and ν have positive and $\mathcal{C}^{2,\alpha}$ continuous densities with respect to the Lebesgue measure on X. We will also need the following result (adapted to our neeeds and notations) on the optimal transport map :

Lemma 2 (Lemma 2.3 and 2.4 [Berman, 2017])**.** *Assume f is C^2-smooth and stricly quasi-convex. Then for any fixed $x \in X$, the unique infimum in (1.3) is attained at $y_x^\star = x - \nabla f(x)$. The map $x \mapsto y_x^\star$ is a C^1 diffeomorphism of X (and the inverse can be computed with g using the symmetric map). Moreover, the function $x \mapsto c(x, y)$ is smooth on some neighborhood of $y_x^\star$ in X and its Hessian is equal to the identity there.*

Remark 1 (Regularity)**.** *The hypothesis on c and (μ, ν) are sufficient to get the C^2-smooth and stricly quasi-convex hypothesis on g.*

1.2. **Entropic Optimal Transport (EOT).** The Kantorovich formulation (1.1) is a linear program. It can be solved numerically using standard LP solvers. The main drawback of this method is it's high dimensionality, a discretization of the space X with N points gives a linear problem with $N \times N$ unknowns and $2\,N$ constraints. The theoretical LP solvers complexity is cubic in the number of unknowns is therefore out of reach for reasonable discretizations (typically $N > 100$).

A review of the different methods which can be applied to this OT problem and its many generalization can be found in the recent book [Peyré and Cuturi, 2018]. Amongst them, a popular alternative is based on a strict convexification of the OT problem based on "Entropic Regularization". It has been introduced for OT computations in [Cuturi, 2013] but can be traced back to the Schroedinger problem (see [Léonard, 2013]).

The "entropic regularization" of the Kantorovich problem (1) is based on the following Kull-Back Leibler divergence or "relative entropy" (KL) penalization :

$$\begin{aligned} OT_\varepsilon(\mu,\nu) := \quad & \min_{\gamma_\varepsilon\in\Pi(\mu,\nu)} \langle c, \gamma_\varepsilon\rangle_{X\times Y} + \varepsilon\ \mathrm{KL}(\gamma_\varepsilon \,|\, \mu\otimes\nu) = \\ & \max_{f_\varepsilon, g_\varepsilon} \langle f, \mu\rangle_X + \langle g_\varepsilon, \nu\rangle_Y - \varepsilon\, \langle \exp(\tfrac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon)) \odot K_\varepsilon - 1, \mu\otimes\nu\rangle_{X\times Y} \end{aligned} \tag{1.5}$$

where $\odot$ is the element-wise multiplication and $\varepsilon > 0$ a small "temperature" parameter,

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

and the kernel

$$K_\varepsilon := \exp(-\frac{1}{\varepsilon}c) \tag{1.6}$$

depends explicitly on the cost c.

The dual problem is unconstrained and the primal-dual optimality conditions given by

$$\gamma_\varepsilon^\star = \exp(\frac{1}{\varepsilon}(f_\varepsilon^\star \oplus g_\varepsilon^\star)) \odot K_\varepsilon \odot \mu\otimes\nu, \tag{1.7}$$

The optimal entropic plan $\gamma_\varepsilon^\star$ is therefore a diagonal scaling by the Kantorovich potentials of K_ε It is diffuse, or not concentrated on a map.

1.3. **Sinkhorn algorithm.** Numerical solutions are produced under a space discretization : we set $X_N = \{x_i\}_{i=1..N}$, $Y_N = \{y_j\}_{j=1..N}$, $c_N = \{c(x_i, y_j)\}_{i,j=1..N}$ and

$$\mu_N = \sum_{i=1}^{N} \frac{\mu(x_i)}{\sum_j \mu(x_j)} \delta_{x_i}, \qquad \nu_N = \sum_{j=1}^{N} \frac{\nu(y_j)}{\sum_j \nu(y_j)} \delta_{y_j} \tag{1.8}$$

Replacing (X, Y, c, μ, ν) by $(X_N, Y_N, c_N, \mu_N, \nu_N)$ provides a natural discretization of of the OT problem. With our notation the dual optimal transport problem takes the same form :

$$OT_{\varepsilon,N}(\mu_N, \nu_N) := \max_{f_\varepsilon, g_\varepsilon} J(f_\varepsilon, g_\varepsilon) = \langle f_\varepsilon, \mu_N \rangle_{X_N} + \langle g_\varepsilon, \nu_N \rangle_{Y_N} - \varepsilon \left\langle \exp(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon)) \odot K_\varepsilon - 1, \mu_N \otimes \nu_N \right\rangle_{X_N \times Y_N}. \tag{1.9}$$

where by abuse of notation, we keep $(f_\varepsilon, g_\varepsilon)$ for discrete vectors in $\mathbb{R}^N$ and $K_\varepsilon := \exp(-\frac{1}{\varepsilon} c_N) \in \mathbb{R}_+^{N \times N}$. As in (1.7), the solution of the discrete primal problem is the discrete probability measure on $X_N \times Y_N$:

$$\gamma_\varepsilon^\star = \exp(\frac{1}{\varepsilon}(f_\varepsilon^\star \oplus g_\varepsilon^\star)) \odot K_\varepsilon \odot \mu_N \otimes \nu_N. \tag{1.10}$$

We solve (1.9) with Sinkhorn iterative algorithm. It can be interpreted as a block coordinate (f_ε and g_ε) maximization. Initialize with $g_\varepsilon^0 = 0_Y$ and then iterate (in m) :

$$\begin{cases} f_\varepsilon^{m+1} = arg\max_{f_\varepsilon} J(f_\varepsilon, g_\varepsilon^m) \\ g_\varepsilon^{m+1} = arg\max_{g_\varepsilon} J(f_\varepsilon^{m+1}, g_\varepsilon), \end{cases} \tag{1.11}$$

giving

$$\begin{cases} f_\varepsilon^{m+1} = -\varepsilon \log(\langle K_\varepsilon \cdot \exp(\frac{1}{\varepsilon}(g_\varepsilon^m)), \nu_N \rangle_{Y_N}) \\ g_\varepsilon^{m+1} = -\varepsilon \log(\langle K_\varepsilon^\top \cdot \exp(\frac{1}{\varepsilon}(f_\varepsilon^{m+1})), \mu_N \rangle_{X_N}) \end{cases} \tag{1.12}$$

where $\cdot$ is the matrix vector multiplication and $.^\top$ the transpose.

The transportation plan is approximated at each iteration as

$$\gamma_\varepsilon^m = \exp(\frac{1}{\varepsilon}(f_\varepsilon^m \oplus g_\varepsilon^m)) \odot K_\varepsilon \odot \mu_N \otimes \nu_N. \tag{1.13}$$

The following lemma was pointed out to us by F. Collino (we do not know if it was previously known). It will be useful in the implementation of the numerical method in section 4.

Lemma 3 (Log Concavity of $y \mapsto \exp(\frac{1}{\varepsilon}(f(x) + g(y) - c(x,y)$ [Collino, 2020])**.** *Assuming that : (i) $c(x,y) = 1/2\|y - x\|^2$ (non periodic case) and (ii) f and g ar solution of the Sinkhorn equation (1.12) ; then $y \mapsto \exp(\frac{1}{\varepsilon}(f(x) + g(y) - c(x,y)$ is log-concave.*

Proof. The proof consists in replacing f using Sinkhorn equation and then checking that the Hessian is negative definite. □

Remark 2. *Note that this result holds for every Sinkhorn m iterate $(f_\varepsilon^m, g_\varepsilon^m)$. In our case we will restrict to a a period $]y_x^\star - 1/2, y_x^\star + 1/2[$ where the periodic cost reduces to L^2 to enforce the log-concavity.*

Sinkhorn algorithm has been and still is popular : As can be seen in (1.12) the algorithm is cost independant and easy to implement. A second reason is that the entropic regularization can be used to smooth out undesirable discretization scale effects on the transport. For many applications, for example in computer graphics or maching learning, working with a finite and "not too small" ε is fine. Conversely, and this will be the case in this paper, using Sinkhorn to approximate

the solution of (1.1), therefore in the $\varepsilon \to 0$ limit, remains a numerical challenge.

The convergence as $\varepsilon \to 0$ of $OT_\varepsilon(\mu,\nu)$ toward $OT(\mu,\nu)$ is well understood in the continuous setting [Léonard, 2013] as well as the in the discrete setting when N is fixed and independent of ε [Cominetti and Martin, 1994].

In practice however, letting $\varepsilon \to 0$ produces serious numerical difficulties. First, even though Sinkhorn iterative algorithm linearly converges, the number of iterations to do so grows at best like $\mathcal{O}(1/\varepsilon)$. Secondly, even if willing to pay the number of iteration price, the numerical stability depends on the following constraints :

- Recalling the matrix $K_\varepsilon := \exp(-\frac{1}{\varepsilon}c_N)$ appearing in (1.13) and denoting

$$\tau = sup_y c(x^\star_y, y), \tag{1.14}$$

  the maximum transport distance we are trying to calculate, we observe that $\exp(-\frac{1}{\varepsilon}\tau)$ needs to remain positive to allow for the mass to travel from $x^\star_y$ to y. In order to do so it exponentially reaches small values and even 0 numerically when $\varepsilon \to 0$. This leads to computer underflows and overflows.
- The entropic discrete transport plan (1.13) resolves the (entropic approximation of the) transport map on a scale h^2/ε where $h = diam(X)/N$ is the discretization scale. Because of finite precision if h^2/ε is too large, K_ε is the identity matrix and no mass displacement is possible. This will be emphasised in theorem 4 where N and ε need to be dependent parameters.

A good review of existing hacks and methods to mitigate the two problems above can be found in [Chizat et al., 2018] and [Schmitzer, 2016] and in particular two techniques called "ε-scaling" and "kernel truncation" (see section 1.5). This work also adresses this problem using the theorem in the next section.

1.4. **Berman Joint Convergence.** To the best of our knowledge joint convergence in N and ε has only be studied in [Berman, 2017], we reproduce here partially his results. First, a technical condition on the sequence of discretization $(X_N, Y_N, c_N, \mu_N \nu_N)$ called "local density property" (Lemma 3.1 [Berman, 2017]) is necessary : For any given open set U intersecting the support X of μ (same for Y and ν)

$$\liminf_{\varepsilon\to 0} \varepsilon \log(\mu_N(U)) = 0 \tag{1.15}$$

On the flat torus (our setting) this condition is trivially satisfied by a uniform grid. Next, a canonical continuous extension of the discrete potential is constructed by replacing $c_N(x_i, y_j)$ respectively by $c(x, y_j), x \in X$ and $c(x_i, y), y \in X$ in the Sinkhorn equations. Omitting the iteration index :

$$f[g_\varepsilon](x) = -\varepsilon \log(\sum_{j=1..N} \exp(\frac{1}{\varepsilon}(g_\varepsilon(y_j) - c(x, y_j)))\nu_N(y_j)), \ \ \forall x \in X \tag{1.16}$$

and the symmetric formula for $g[f_\varepsilon](y)$, $y \in Y$.

Theorem 4 (Berman joint convergence - corollary 1.3 [Berman, 2017])**.** *We assume μ and ν are in $\mathcal{C}^{2,\alpha}(X)$, $(X = Y)$ for some $\alpha > 0$, and that N and ε are dependent parameters : $N = (1/\varepsilon)^d$ where d is the dimension of the problem. We further assume that the discretisation $(X_N, Y_N, , \mu_N \nu_N)$ satisfies the "density property" above and $(\mu_N, \nu_N) \to (\mu, \nu) \in \mathcal{P}(X)$. 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 $f[g^{k_\varepsilon}_\varepsilon]$, built using the canonical extension (1.16) of the discrete Sinkhorn iterate at $f^{m_\varepsilon}_\varepsilon$, satisfy the estimate*

$$\sup_X |f[g^{m_\varepsilon}_\varepsilon] - f^\star| \leq -C\,\varepsilon \log(\frac{1}{\varepsilon}) \tag{1.17}$$

for some constant C (depending on A), where $f^\star$ is an optimal potential for (1.1).

Moreover, the discrete probability measures $\gamma_\varepsilon^{m_\varepsilon}$ on $X_N \times Y_N$ (defined in (1.13)) converge weakly to the optimal transport plan $\gamma^\star$ solution of (1.1) and satisfies the the following estimate

$$\gamma_\varepsilon^{m_\varepsilon} \leq \mathcal{B}_\varepsilon \odot \mu_N \otimes \nu_N, \tag{1.18}$$

where we define

$$(x,y) \mapsto \mathcal{B}_\varepsilon(x,y) := \frac{p}{\varepsilon^p} \exp(-\frac{c(y,y_x^\star)}{\varepsilon\, p}) \tag{1.19}$$

where c is the periodic cost (1.4) and $x \mapsto y_x^\star$ the OT map (see lemma 2).

Remark 3 (Modifications of [Berman, 2017] around (1.18))**.** *We use slightly different notations, in particular the function $\mathcal{B}_\varepsilon$ will be the cornerstone to the proof of convergence in this paper. The estimate (1.18) is slightly sharper than (1.11) in* [Berman, 2017] *but that follows directly from the proof in section 4.1.1 therein. In particular, it holds in the limit :*

$$\gamma_\varepsilon^\star \leq \mathcal{B}_\varepsilon \odot \mu_N \otimes \nu_N \tag{1.20}$$

Remark 4 (Vertical distance)**.** *We will call $y \to c(y, y_x^\star)$ the "vertical distance", I.e. for x fixed it measure the travel cost to the $y_x^\star$,. It will be useful to think of x as a parameter in (1.19).*

Remark 5 (Domains and Cost)**.** *Theorem 4 holds in particular for the torus and the L^2 periodic cost. It also holds on the sphere for the reflector cost* [Wang, 2004]*. The case of densities μ and ν with compact support is not covered.*

Remark 6 (Transportation plan convergence)**.** *The support of the entropic transportation plan γ_ε (see (1.7)), built trough the interpolation procedure in theorem (4), converges exponentially fast to the graph of the non entropic OT map when $\varepsilon \to 0$. It corresponds to the heuristic justification of ε-scaling given in section 1.5.*

Remark 7 (Constants)**.** *The constants A and more importantly p in theorem 4 depend on the curvature of The Brenier potential $y \mapsto \|y\|^2 - g$ and ultimately on the data (μ, ν). More precisely, they should involve bounds on the Hessians of the densities of μ and ν and a strict positive lower bound on the densities.*

Remark 8 (Reduction of the discretization constraint (section 5.4 [Berman, 2017]))**.** *If (μ, ν) are $\mathbb{C}^\infty(X)$, then theorem 4 holds under the relaxed condition*

$$N \geq C_\delta \left(\frac{1}{\sqrt{\varepsilon}}\right)^{\frac{d}{1+\delta}},\ \delta \in]0, 1/2]. \tag{1.21}$$

The increase in the number of discretization slows significantly as ε decrease. The constant depends on δ. In practice, we used

$$N = \left(\frac{2}{\sqrt{\varepsilon}}\right)^d. \tag{1.22}$$

1.5. **ε-scaling and kernel truncation.** Ignoring for now the space discretisation to simplify the idea. We are given a decreasing sequence (ε_l) such that $\lim_{l\to\infty} \varepsilon_l = 0$, to solve at a sequence of $OT_{\varepsilon_l}(\mu, \nu)$ problems (1.5) where K_{ε_l} is replaced by the "stabilised" kernel :

$$\tilde{K}_{\varepsilon_l} := \exp(\frac{1}{\varepsilon_l}(\sum_{i=0,l-1} (\tilde{f}_{\varepsilon_i}^\star \oplus \tilde{g}_{\varepsilon_i}^\star) - c)) \tag{1.23}$$

where $(\tilde{f}_{\varepsilon_l}^\star, \tilde{g}_{\varepsilon_l}^\star)$ are the optimal potentials at at each l iteration. The initialisation is $\tilde{K}_{\varepsilon_0}^0 = K_{\varepsilon_0}$, that is standard Sinkhorn at $\varepsilon = \varepsilon_0$, and then one computes corrections to the kantorovich potentials solutions of the "standard" entropic problem at the next ε_l level :

$$f_{\varepsilon_l}^\star = \sum_{i=0,l} \tilde{f}_{\varepsilon_l}^\star \qquad g_{\varepsilon_l}^\star = \sum_{i=0,l} \tilde{g}_{\varepsilon_l}^\star$$

Assuming $(f^\star_{\varepsilon_l}, g^\star_{\varepsilon_l}) \to (f^\star, g^\star)$, the solution of the standard Kantorovich problem (1.1), the stabilized Kernel satisfies (see (1.3)) :

$$\lim_{l\to\infty} \{f^\star_{\varepsilon_l} \oplus g^\star_{\varepsilon_l} - c\}(x,y) \leq 0,\ \forall (x,y) \text{ and } \lim_{l\to\infty} \{f^\star_{\varepsilon_l} \oplus g^\star_{\varepsilon_l} - c\}(x, y^\star_x) = 0,\ \forall x. \tag{1.24}$$

This technique indeed stabilises the numerical algorithm as the maximum transport τ (see (1.14)) is replaced by the above quantity. The stabilised Kernel (1.23) approaches 1 on $\Gamma^\star$ the support of the optimal transportation plan, and decreases exponentially elsewhere when $\varepsilon_l \to 0$. For all x and in the l-limit, the function $y \mapsto \tilde{K}_{\varepsilon_l}(x,y)$ is expected to behave as a Gaussian function centered at $y^\star_x$ and standard deviation ε_l. This result is made rigorous and quantitative in theorem 4.

Going back to the computable discrete problem, note that ε-scaling has a very strong impact on the cost of the method. A glance at algorithm (1.12) shows that the number of operations for each Sinkhorn iteration is of order $\mathcal{O}(N^2)$: the multiplication of the $N \times N$ matrix K_ε with a N vector. Because of finite precision and as already mentioned, the stabilised Kernels $\tilde{K}_{\varepsilon_l}$ replacing K_ε become sparse. In addition one can expect that low values at iteration l of the ε-scaling can be used to localise the relevant support of the next stabilised Kernel and therefore truncated out. Sparse Kernels with $\mathcal{O}(N)$ non-zeros elements woull lead to a $\mathcal{O}(N)$ cost for one Sinkhorn iteration. This strategy is pursued in [Schmitzer, 2016] under the name "kernel truncation" where convincing numerical results and partial theoretical results are given. The original Kernel K_ε also enjoys the same sparsification but can only capture maps close to the identity.

The combination of ε-scaling and kernel truncation is therefore fundamental to let ε go to 0.

Note finally that a multilevel truncation algorithm applied directly on the linear programming problem (that is for $\varepsilon = 0$) has been proposed in [Oberman and Ruan, 2015]].

1.6. **Our contribution and strategy.** Our goal is to solve (1.9) for the smallest possible ε and also to approach the each the optimal $\mathcal{O}(N)$ complexity.

We will introduce in section 2 a modification of the Entropic optimal problem which includes a "minimum capacity constraint" $\gamma \geq \lambda\, \mu \otimes \nu$ where $1 \geq \lambda > 0$ is an additional parameter. The constrained optimal plans will be generically denoted $\gamma_{\varepsilon,\lambda}$. The idea behind this modification is that the saturation domain, i.e. where $\gamma^\star_{\varepsilon,\lambda} = \lambda \mu \otimes \nu$ can be exactly "out-summed" in the Sinkhorn algorithm.

We then consider in section 3 a continuation method in ε where $\lambda = \lambda_\varepsilon$ depends on ε and $\lim_{\varepsilon\to 0} \lambda_\varepsilon = 0$. Using theorem 4 we show that it is possible (in infinite precision arithmetics) to generate from $\gamma_{\varepsilon,\lambda_\varepsilon}$ a sequence of discrete solutions (remember N depends on ε) $\gamma^0_{\varepsilon,\lambda_\varepsilon}$ which satisfies the convergence result of theorem 4 to $\gamma^\star$ the continuous solution of problem 1.1. We then show that the associated sequence (in ε) of non-saturated domains is monotone decreasing. One can use the saturation property at ε to predict the saturation domain for $\varepsilon' < \varepsilon$ for a carefully chosen ε dependence of λ_ε. It will also provides a theoretical estimate of the complexity gain obtained by their "out-summation" of the Sinkhorn Algorithm.

Section 4 describe the finite arithmetic precision algorithm we have implemented and presents 1-D numerical results.

The conclusion gives a preliminary assessment of the method, its potential and limitations.

2. Adding the minimum capacity constraint

2.1. **The minimum capacity constrained Entropic OT problem.** From this point on, we will consider only OT discrete problems and will sometimes omit N in the subscript/superscript notations

in order to keep the formulae readable. Please remember that $N = (1/\varepsilon)^d$ or $N = C(1/\sqrt{\varepsilon})^d$ in the smoother case (remark 1.22).

We add a minimum capacity constraint to problem (1.9). More precisely and for $1 \geq \lambda \geq 0$ we consider :

$$OT_{\varepsilon,\lambda}(\mu_N, \nu_N) := \min_{\gamma_{\varepsilon,\lambda} \in \Pi_\lambda(\mu_N,\nu_N)} \langle c, \gamma_{\varepsilon,\lambda} \rangle_{X_N \times Y_N} + \varepsilon \ \mathrm{KL}(\gamma_{\varepsilon,\lambda} \mid \mu_N \otimes \nu_N) \tag{2.1}$$

where

$$\Pi_\lambda(\mu,\nu) := \{\gamma \in \mathcal{P}(X_N \times Y_N), \langle 1_{X_N}, \gamma \rangle_{X_N} = \nu_N \ \langle 1_{X_N}, \gamma \rangle_{Y_N} = \mu_N \text{ and } \gamma \geq \lambda \, \mu_N \otimes \nu_N\}$$

Computing the Fenchel-Rockaffellar dual of (2.1) is a standard duality exercise (a simple extension of the capacity constraint on the marginals established in [Chizat et al., 2018]). Note in particular that $\mu_N \otimes \nu_N \in \Pi_\lambda(\mu_N, \nu_N) \neq \emptyset$ as usual.

(2.2)

$$\max_{f_{\varepsilon,\lambda}, g_{\varepsilon,\lambda}, h_{\varepsilon,\lambda} \geq 0} \langle f_{\varepsilon,\lambda}, \mu_N \rangle_{X_N} + \langle g_{\varepsilon,\lambda}, \nu_N \rangle_{Y_N} + \lambda \, \langle h_{\epsilon,\lambda}, \mu_N \otimes \nu_N \rangle_{X_N \times Y_N} - \varepsilon \, \langle \exp(\frac{1}{\varepsilon}((f_{\varepsilon,\lambda} \oplus g_{\varepsilon,\lambda}) + h_{\varepsilon,\lambda} - c)) - 1, \mu \otimes \nu \rangle_{X_N \times Y_N}$$

Where the positive function $(x,y) \mapsto h_{\varepsilon,\lambda}(x,y)$ is the Lagrange multiplier of the new saturation constraint $\gamma \geq \lambda \, \mu_N \otimes \nu_N$. The primal-dual optimality condition is given by

$$\gamma^\star_{\varepsilon,\lambda} = \exp(\frac{1}{\varepsilon}((f^\star_{\varepsilon,\lambda} \oplus g^\star_{\varepsilon,\lambda}) + h^\star_{\varepsilon,\lambda} - c_N)) \odot \mu_N \otimes \nu_N. \tag{2.3}$$

Still following the general framework in [Chizat et al., 2018], i.e. Sinkhorn algorithm corresponds to alternate maximizations on the dual potentials $(f_{\varepsilon,\lambda}, g_{\varepsilon,\lambda}, h_{\varepsilon,\lambda})$ we obtain the folowing modification on the iterates in (1.12) :

$$\begin{cases} f^{m+1}_{\varepsilon,\lambda} = -\varepsilon \log(\langle \exp(\frac{1}{\varepsilon}(g^m_{\varepsilon,\lambda} + h^m_{\varepsilon,\lambda} - c_N)), \nu_N \rangle_{Y_N}) \\ g^{m+1}_{\varepsilon,\lambda} = -\varepsilon \log(\langle \exp(\frac{1}{\varepsilon}(f^{m+1}_{\varepsilon,\lambda} + h^m_{\varepsilon,\lambda} - c_N)), \mu_N \rangle_{X_N}) \\ \exp(\frac{1}{\varepsilon} h^{m+1}_{\varepsilon,\lambda}) = \max\{1, \frac{\lambda}{\exp(\frac{1}{\varepsilon}(f^{m+1}_{\varepsilon,\lambda} \oplus g^{m+1}_{\varepsilon,\lambda} - c_N))}\} \end{cases} \tag{2.4}$$

The last equation above is a point-wise maximum. The convergence proof of the above iterate to a stationnary point $(f^\star_{\varepsilon,\lambda}, g^\star_{\varepsilon,\lambda}, h^\star_{\varepsilon,\lambda})$ can be easily adapted from to theorem 4.1 in [Chizat et al., 2018] and

$$\gamma^m_{\varepsilon,\lambda} = \exp(\frac{1}{\varepsilon}(f^m_{\varepsilon,\lambda} \oplus g^m_{\varepsilon,\lambda}) + h^m_{\varepsilon,\lambda} - c_N) \odot \mu_N \otimes \nu_N. \tag{2.5}$$

converges in $\mathcal{P}(X \times Y)$ to (2.3).

2.2. **Sinkhorn saturated domain "out-summation".** For the sake of clarity, we may drop in this section the dependence on the Sinkhorn iteration index m. The following definitions will be handy (also in order to construct and analyze our method below).

Definition 1 (Super Level sets)**.** The λ super level set of $y \to \phi(z)$ is $L^+_\lambda(\phi) = \{z, \, \phi(z) > \lambda\}$.

Definition 2 (non-satured transportation plan)**.** We will call "non-satured transportation plan" the probability in $\mathcal{P}(X_N \times Y_N)$ denoted

$$\gamma^{us}_{\varepsilon,\lambda} := \exp(\frac{1}{\varepsilon}(f_{\varepsilon,\lambda} \oplus g_{\varepsilon,\lambda}) - c_N) \odot \mu_N \otimes \nu_N. \tag{2.6}$$

It may (and will) also depend (as mentionned above) on m the Sinkhorn iteration index. We will continue to use $(x,y) \in X \times Y$ as generic notations for the product space (i.e $\gamma^0_{\varepsilon,\lambda}(x,y)$ but also $f_{\varepsilon,\lambda}(x)$ and $g_{\varepsilon,\lambda}(y)$). See figure 1 for a comparison of $\gamma^0_{\varepsilon,\lambda}$ and $\gamma_{\varepsilon,\lambda}$.

Definition 3 (Vertical non-satured domain)**.** We will denote "vertical non-satured domain" , $\forall x$:

$$S^x_{\varepsilon,\lambda} := L^+_\lambda(y \mapsto \exp(\frac{1}{\varepsilon}(f_{\varepsilon,\lambda} \oplus g_{\varepsilon,\lambda}) - c_N)) \tag{2.7}$$

By construction $h_{\varepsilon,\lambda} = 0$ on the non-satured domain (see figure 1). The "horizontal non-satured domain" $S^y_{\varepsilon,\lambda}$ may be defined just interverting x and y. Horizontal and vertical satured domain of course coincide :

$$S_{\varepsilon,\lambda} := \{(x, S^x_{\varepsilon,\lambda}),\, x \in X_N\} = \{(S^y_{\varepsilon,\lambda}, y),\, y \in Y_N\} \tag{2.8}$$

where $S_{\varepsilon,\lambda}$ is the full saturated domain in $X_N \times X_N$. These domains are to be understood either for converged solutions or for a fixed final iteration m.

Using the two definition above and the third equation in (2.4) direct computations give the following trivial properties

$$\forall x,\, \gamma_{\varepsilon,\lambda}(x, y \in S^x_{\varepsilon,\lambda}) = \lambda, \qquad \forall y,\, \gamma_{\varepsilon,\lambda}(x \in S^y_{\varepsilon,\lambda}, y) = \lambda. \tag{2.9}$$

Section 3 introduces a continuation method in (ε, λ) to compute domains containing exactly $(S^{x,y}_{\varepsilon,\lambda})$.

Assuming for now that the saturation domain is given , we are now ready to "out-sum" the satured domain from Sinkhorn algorithm (2.4) :

- The third equation is trivially simplified as :

$$\forall x,\, \exp(\frac{1}{\varepsilon}h_{\varepsilon,\lambda}(x, y \in S^x_{\varepsilon,\lambda})) = 1.$$

- Using (2.9) the first two equations (m index omitted again), we sum out the constant λ values on respectively $X_N \setminus S^x_{\varepsilon,\lambda}$ and $Y_N \setminus S^y_{\varepsilon,\lambda}$:

$$\begin{cases} \forall x,\ f^{m+1}_{\varepsilon,\lambda}(x) = \varepsilon\,(\log(1 - \lambda\langle 1, \nu_N\rangle_{Y_N \setminus S^x_{\varepsilon,\lambda}}) - \log(\langle \exp(\frac{1}{\varepsilon}(g^m_{\varepsilon,\lambda} + h^m_{\varepsilon,\lambda} - c_N)), \nu_N\rangle_{S^x_{\varepsilon,\lambda}})) \\ \forall y,\ g^{m+1}_{\varepsilon,\lambda}(y) = \varepsilon\,(\log(1 - \lambda\langle 1, \mu_N\rangle_{X_N \setminus S^y_{\varepsilon,\lambda}}) - \log(\langle \exp(\frac{1}{\varepsilon}(f^{m+1}_{\varepsilon,\lambda} + h^m_{\varepsilon,\lambda} - c_N)), \mu_N\rangle_{S^y_{\varepsilon,\lambda}})) \end{cases} \tag{2.10}$$

  By construction $h^m_{\varepsilon,\lambda} \equiv 0$ on $S^x_{\varepsilon,\lambda}$ and $S^y_{\varepsilon,\lambda}$. We do not use

Remark 9 (Reduction in complexity)**.** *The first terms in the above simplification are constant. The complexity of each Sinkhorn iteration now is $\mathcal{O}(|S_{\varepsilon,\lambda}|)$ (See (2.8)).*

3. A CONTINUATION METHOD IN (ε, λ)

The parameters N and ε are already dependent either trough $N = (1/\varepsilon)^d$ or $N = C(1/\sqrt{\varepsilon})^d$. In order to avoid overloading the notations, this dependence will not be explicitely used in the text. We will further assume here that $\lambda = \lambda_\varepsilon$ depends on ε (hence also on N) and

$$\lim_{\varepsilon \to 0} \lambda_\varepsilon = 0 \text{ or equivalently } \lim_{N \to +\infty} \lambda_\varepsilon = 0. \tag{3.1}$$

3.1. **Construction of the sequence of plans.** We first define the non-saturated transport plans (2.6) marginals at the limit, that is when (2.4) has converged :

$$\mu^0_N = \langle 1, \gamma^{us}_{\varepsilon,\lambda}\rangle_{Y_N} \qquad \nu^0_N = \langle 1, \gamma^{us}_{\varepsilon,\lambda}\rangle_{X_N} \tag{3.2}$$

By construction (see 2.4) $\exp(\frac{1}{\varepsilon}h_{\varepsilon,\lambda}) \geq 1$, we thereof have the following estimates :

$$\mu_N - \lambda < \mu^0_N < \mu_N \qquad \nu_N - \lambda < \nu^0_N < \nu_N \tag{3.3}$$

The marginals (μ^0_N, ν^0_N) are probability measures in the limit only. They have the same support as (μ_N, ν_N). We immediatly get from (3.3) :

Lemma 5. *Assuming that (X_N, Y_N, μ_N, ν_N) satisfies the "density property" (1.15) and $(\mu_N, \nu_N) \to (\mu, \nu) \in \mathcal{P}(X)$, then so does $(X_N, Y_N, \mu^0_N \nu^0_N)$.*

3.2. **Convergence to $OT(\mu,\nu)$.** We can now define our converging sequence of plans. We denote $(f_\varepsilon^{0,m}, g_\varepsilon^{0,m})$ the sequence of potentials associated to the Sinkhorn resolution (1.12) associated with the resolution of $OT_\varepsilon(\mu_N^0, \nu_N^0)$. They will be useful for the analysis but are never computed. The associated transport plans are given by (see (1.13) :

$$\gamma_\varepsilon^{0,m} = \exp(\frac{1}{\varepsilon}(f_\varepsilon^{0,m} \oplus g_\varepsilon^{0,m} - c_N)) \odot \mu_N^0 \otimes \nu_N^0. \tag{3.4}$$

Using theorem 4 and lemma 5 directly gives (see remark 4.1 in [Berman, 2017] about un-normalized discrete measure) :

Lemma 6. *Setting $m_\varepsilon = [-A\log(\varepsilon)/\varepsilon]$ as in theorem 4, the discrete probability measures $\gamma_\varepsilon^{0,m}$ converges weakly to the optimal transport plan $\gamma^\star$ solution of (1.1) and satisfies the the following estimate*

$$\gamma_\varepsilon^{0,m_\varepsilon} \le \mathcal{B}_\varepsilon \odot \mu_N^0 \otimes \nu_N^0. \tag{3.5}$$

As already mentionned we will never solve $OT_\varepsilon(\mu_N^0, \nu_N^0)$ but instead rely on the $(f_{\varepsilon,\lambda}^m, g_{\varepsilon,\lambda}^m)$, the iterates of the (capacity constrained) Sinkhorn algorithm (2.4) with fixed parameters (ε, λ) (for which Sinkhorm "out-summing" will be possible). They are not the same as $(f_{\varepsilon,\lambda}^{0,m}, g_{\varepsilon,\lambda}^{0,m})$, but will be used as a proxy. We will need, in particular the following lemma :

Lemma 7 (Convergence of the non saturated plan)**.** *We assume (as in theorem 4) that (μ,ν) are bounded below by a strict positive number. Choosing $\lambda_\varepsilon = \varepsilon^{1+\beta}$, $\beta > 0$, there is a constant $C_{\mu,\nu}$ depending only on (μ,ν) such that on the lattice $X_N \times Y_N$:*

$$|\{f_{\varepsilon,\lambda_\varepsilon}^m \oplus g_{\varepsilon,\lambda_\varepsilon}^m\} - \{f_\varepsilon^{0,m} \oplus g^{0,m}\}| \le m\,\varepsilon\,\lambda\,C_{\mu,\nu}, \quad \forall\, m \tag{3.6}$$

which directly gives :

$$\exp(\frac{1}{\varepsilon}(f_{\varepsilon,\lambda_\varepsilon}^{m_\varepsilon} \oplus g_{\varepsilon,\lambda_\varepsilon}^{m_\varepsilon} - c_N) \le \exp(-C_{\mu,\nu}\,\varepsilon^\beta \log(\varepsilon))\ \exp(\frac{1}{\varepsilon}(f_{\varepsilon,\lambda_\varepsilon}^{0,m_\varepsilon} \oplus g_{\varepsilon,\lambda_\varepsilon}^{0,m_\varepsilon} - c_N) \text{ on the lattice } X_N \times Y_N \tag{3.7}$$

(where we absorbed the constant A in $C_{\mu,\nu}$).

Finally $\gamma_{\varepsilon,\lambda_\varepsilon}^{ns,m_\varepsilon}$ (see (2.6)) converges weakly to the optimal transport plan $\gamma^\star$ when $\varepsilon \to 0$.

Proof. The proof of the first statement is done by estimating the discrepancy of $|f_{\varepsilon,\lambda}^m - f_\varepsilon^{0,m}|$ (also for the g(s)) along the Sinkhorn iterations processes (1.12) and (2.4). To simplify the presentation we will drop the $(\varepsilon, \lambda_\varepsilon)$ notations.

First we recall that the "LogSumExp operator" :

$$LSE_\nu(g) := -\varepsilon \log(\langle \exp(\frac{1}{\varepsilon}(g - c)), \nu\rangle_Y)$$

is decreasing in both ν and g but also 1-Lipschitz (see [Vialard, 2019] for a detailed presentation of this result). We therefore have on one hand and $(\forall\, x)$ from (2.4) and (3.3) :

$$f^{m+1} = LSE_{\nu_N}(g^m + h^m) \le LSE_{\nu_N^0}(g^m). \tag{3.8}$$

Using definition 3 for the satured/non-satured domain, we remark that on $S_{\varepsilon,\lambda}^x$, $h^m = 0$ and on the satured domain $Y_N \setminus S_{\varepsilon,\lambda}^x$:

$$\lambda = \exp(\frac{1}{\varepsilon}(f^m \oplus g^m + h^m - c_N)) \ge \exp(\frac{1}{\varepsilon}(f^m \oplus g^m - c_N)) \ge 0$$

we therefore have

$$0 \le \exp(\frac{1}{\varepsilon}(f^m \oplus g^m + h^m - c_N)) - \exp(\frac{1}{\varepsilon}(f^m \oplus g^m - c_N)) \le \lambda$$

Using the above and $\nu_N \leq \nu_N^0 + \lambda$ in the sum we obtain.

(3.9)
$$\begin{aligned}\langle \exp(\tfrac{1}{\varepsilon}(g^m + h^m - c_N)), \nu_N\rangle_{Y_N} &\leq \langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), \nu_N^0 + \lambda\rangle_{Y_N} + \lambda \, \exp(-\tfrac{1}{\varepsilon} f^m) \, \langle 1, \nu_N^0 + \lambda\rangle_{Y_N} \\ &\leq \langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), \nu_N^0\rangle_{Y_N} + \lambda \, R + \lambda^2 \, \exp(-\tfrac{1}{\varepsilon} f^m).\end{aligned}$$

Where the remainded R is given by

$$R = \langle \exp(\frac{1}{\varepsilon}(g^m - c_N)), 1\rangle_{Y_N} + \exp(-\frac{1}{\varepsilon} f^m) \, \langle 1, \nu_N^0\rangle_{Y_N}$$

In order to bound above R we will use the following :

- The mass of ν_0^N (i.e. $\langle 1, \nu_N^0\rangle_{Y_N}$) is bounded over by 1.
- The definition of μ_0^N (3.2) is $\langle \exp(\frac{1}{\varepsilon}(f^m + g^m - c_N)), \mu_N \times \nu_N\rangle_{Y_N} = \mu_N^0$, therefore

$$\begin{aligned}\exp(-\tfrac{1}{\varepsilon} f^m) &= \frac{\mu_N}{\mu_N^0} \, \langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), \nu_N\rangle_{Y_N} \\ &\leq \frac{\mu_N}{\mu_N^0} \, \langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), \nu_N^0 + \lambda\rangle_{Y_N} \\ &\leq \frac{\mu_N}{\mu_N - \lambda} \, \Big(\langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), \nu_N^0\rangle_{Y_N} + \lambda \, \langle \exp(\tfrac{1}{\varepsilon}(g^m - c_N)), 1\rangle_{Y_N} \Big)\end{aligned}$$

- And finally, we will use the bound

$$\langle \exp(\frac{1}{\varepsilon}(g^m - c_N)), 1\rangle_{Y_N} \leq \frac{1}{\min_{Y_N} \nu_N^0} \langle \exp(\frac{1}{\varepsilon}(g^m - c_N)), \nu_N^0\rangle_{Y_N} \leq \frac{1}{\min_{Y_N}(\nu_N - \lambda)} \langle \exp(\frac{1}{\varepsilon}(g^m - c_N)), \nu_N^0\rangle_{Y_N}.$$

We assume that (μ, ν) are bounded below by a strict positive number. Then, for small enough λ, there is a constant $C_{\mu,\nu}$ depending only on μ and ν such that plugging the bullets above in (3.9) we get

$$\langle \exp(\frac{1}{\varepsilon}(g^m + h^m - c_N)), \nu_N\rangle_{Y_N} \leq (1 + \lambda \, C_{\mu,\nu}) \, \langle \exp(\frac{1}{\varepsilon}(g^m - c_N)), \nu_N^0\rangle_{Y_N} \tag{3.10}$$

Applying the $-\varepsilon \log(.)$ function :

$$f^{m+1} \geq LSE_{\nu_N^0}(g^m) - \varepsilon \log(1 + \lambda \, C_{\mu,\nu})) \geq LSE_{\nu_N^0}(g^m) - \varepsilon \, \lambda \, C_{\mu,\nu} \tag{3.11}$$

again for small λ and a constant depending only on (μ, ν).

Combining (3.11) and (3.8) we have :

$$|f^{m+1} - f^{0,m+1}| \leq |LSE_{\nu_N^0}(g^m) - LSE_{\nu_N^0}(g^{0,m})| + \varepsilon \, \lambda \, C_{\mu,\nu} \tag{3.12}$$

The symmetric in (f, g) estimate holds for $|g^{m+1} - g^{0,m+1}|$, hence thanks to the 1-Lipschitz property of $g \mapsto LSE_\nu(g)$ we obtain the estimate (3.6) (both sequences have same intialization) $\forall m$

$$|\{f_{\varepsilon,\lambda_\varepsilon}^m \oplus g_{\varepsilon,\lambda_\varepsilon}^m\} - \{f_\varepsilon^{0,m} \oplus g^{0,m}\} \leq m \, \varepsilon \, \lambda \, C_{\mu,\nu}.$$

Thanks to (3.3) and (3.7) we know that $\gamma_{\varepsilon,\lambda_\varepsilon}^{ns,m_\varepsilon}$ weakly converges to a transport plan concentrated on the support of the graph of the optimal monotone map with marginals (μ, ν), i.e. $\gamma^\star$.

□

Estimate (3.7) will allow us in the next section to approximate the saturation domain (2.7) and perform the saturation domain "out-summing" (2.10).

Remark 10. *The above estimate are probably not sharp, the Sinkhorn iterate are actually contractant in the Hilbert or Oscillation norm. The function $\mathcal{B}_\varepsilon$ converges exponentially fast to the support and the additional coefficient, induced by using $(f_{\varepsilon,\lambda}^m, g_{\varepsilon,\lambda}^m)$ as a proxy for $(f_{\varepsilon,\lambda}^{0,m}, g_{\varepsilon,\lambda}^{0,m})$, can be absorbed in this rate. We used $\beta = 0$ in the numerics.*

Remark 11 (Link with Kernel truncation [Schmitzer, 2016])**.** *A legitimate question is the usefulness of the capacity constraints. The proof above seems to also work by setting to 0 the λ sub-levels sets of γ. Combined with ε-scaling this is vey close to the Kernel truncation method and provides a rule to set the truncation level as well as a convergence result.*

3.3. **Saturated domain localisation.** We start by establishing a few properties and use of the "Berman" function (1.19) :

Proposition 8 (Vertical monotonic properties of (1.19))**.** *We* fix the $x \in X_N$ coordinate and work on one period of the vertical y-axis of the torus centered at $y^\star_x$. We will use $LSE_\varepsilon = \varepsilon^{1+\beta}$, $\beta > 0$ (as in lemma 7). We define, see also figure 3, :

$$A^x_{\varepsilon,\lambda_\varepsilon} := L^+_{\lambda_\varepsilon}(y \mapsto \exp(-C_{\mu,\nu}\,\varepsilon^\beta \log(\varepsilon))\,\mathcal{B}_{\varepsilon,\lambda_\varepsilon}(x,y)), \tag{3.13}$$

the λ_ε super level set of $y \mapsto \mathcal{B}_{\varepsilon,\lambda_\varepsilon}(x,y)$.

(i) For a sufficiently small ε, $A^x_{\varepsilon,\lambda_\varepsilon} = B(y^\star_x, d\mathcal{B}_{\varepsilon,\lambda_\varepsilon})$ is the ball of center $y^\star_x$ and radius

$$d\mathcal{B}_{\varepsilon,\lambda_\varepsilon} = \sqrt{-\,p\,\varepsilon\,\left(\log\frac{\varepsilon^{p+1+\beta}}{p} + C\,\varepsilon^\beta\,\log\varepsilon\right)}.$$

The second term under the square root can be aborbed in the constant p, even for $\beta = 0$, in practice we will use

$$d\mathcal{B}_{\varepsilon,\lambda_\varepsilon} = \sqrt{-\,p\,\varepsilon\,\log\frac{\varepsilon^{p+1}}{p}} \tag{3.14}$$

(ii) The estimates (3.5) and (3.7) gives

$$S^x_{\varepsilon,\lambda_\varepsilon} \subset L^+_{\lambda_\varepsilon}(y \mapsto \exp(-C_{\mu,\nu}\,\varepsilon^\beta \log(\varepsilon))\,\exp(\frac{1}{\varepsilon}(f^{0,m_\varepsilon}_{\varepsilon,\lambda_\varepsilon} \oplus g^{0,m_\varepsilon}_{\varepsilon,\lambda_\varepsilon} - c_N)) \subset A^x_{\varepsilon,\lambda_\varepsilon}. \tag{3.15}$$

Please keep in mind that $S^x_{\varepsilon,\lambda_\varepsilon}$ uses the m_ε iterate of Sinkhorn (2.4).

(iii) The actual computational domain will be defined and denoted by :

$$\mathbb{S}^x_{\varepsilon,\lambda_\varepsilon} := \bigcup_{y \in Co(S^x_{\varepsilon,\lambda_\varepsilon})} B(y, d\mathcal{B}_{\varepsilon,\lambda_\varepsilon}) \tag{3.16}$$

where is the convex hull. of sets. Then $\mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}$ is convex and $A^x_{\varepsilon,\lambda_\varepsilon} \subset \mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}$.

(iv) Setting $\varepsilon' = \alpha\,\varepsilon$, $\alpha < 1$ and assuming $|x - x'| < \varepsilon^{1/2}$ we have, for small ε :

$$A^x_{\varepsilon',\lambda_{\varepsilon'}} \subset A^x_{\varepsilon,\lambda_\varepsilon}. \tag{3.17}$$

(v) $diam(\mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}) \le 3\,d\mathcal{B}_{\varepsilon,\lambda_\varepsilon}$ and $|\{x, \mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}\}_{x\in X_N}| = \mathcal{O}(\log(N)^{d/2}\,N^{3/2})$ if $N = (1/\varepsilon)^d$ as in theorem 4 or $|\{x, \mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}\}_{x\in X_N}| = \mathcal{O}(\log(N)^{d/2}\,N)$ if $N = (1/\sqrt{\varepsilon})^d$ as in remark 8.

Proof. (i) For a small ε, $\mathcal{B}_{\varepsilon,\lambda_\varepsilon}$ behaves as a Gaussian in the considered vertical period of the torus. Definition (1.19) directly gives (i).

(iii) Thanks to Lemma 3 and for a small ε, the intermediate set in (3.15) is convex. The convex hull of $S^x_{\varepsilon,\lambda_\varepsilon}$) is therefore also contained in $A^x_{\varepsilon,\lambda_\varepsilon}$ and (iii) follows (see also remark 12).

(iv) Under the hypothesis of lemma 2, the OT map $x \mapsto y^\star_x$ is C^1, therefore $|y^\star_{x'} - y^\star_x| \le C\,\varepsilon^{1/2}$. We need to check that

$$d\mathcal{B}_{\varepsilon',\lambda_{\varepsilon'}} + C\,\varepsilon^{1/2} \le d\mathcal{B}_{\varepsilon,\lambda_\varepsilon} \tag{3.18}$$

(remember that N depends on ε). Expending (3.18) and absorbing higher order terms in ε into the constant C we find the necessary condition

$$\varepsilon^{(p+1)(1-\alpha)} \le \exp C/\alpha$$

which is satisfied for small ε as $\alpha < 1$.

(v) The first assertion directly follows (ii). The complexity results assume that we are using regular cartesian grid with the corresponding number of points. One just needs to divide $d\mathcal{B}_{\varepsilon,\lambda_\varepsilon}$ by $1/N$, take it to the power d and then multiply by N.

□

We are now ready to formulate the lemma used to localize automatically the non-saturated domain as $(\varepsilon, \lambda_\varepsilon)$ decrease in order to "sum it out" in Sinkhorn for the next smaller $(\varepsilon', \lambda'_{\varepsilon'})$

Lemma 9 (non-saturated domain localisation)**.** *We choose $(\varepsilon, \lambda_\varepsilon)$ and $(\varepsilon', \lambda'_{\varepsilon'})$ as in proposition 8 (iv). We denote N and N' the corresponding number of disscretization points. Then, for $x \in X_N$ and $x' \in X'_N$ such that $|x - x'| < \varepsilon^{1/2}$, proposition 8 (ii) and (iv) directly gives :*

$$S^{x'}_{\varepsilon',\lambda_{\varepsilon'}} \subset A^{x'}_{\varepsilon',\lambda_{\varepsilon'}} \subset A^{x}_{\varepsilon,\lambda_\varepsilon} \subset \mathbb{S}^{x}_{\varepsilon,\lambda_\varepsilon}. \tag{3.19}$$

The saturated domain $S^x_{\varepsilon,\lambda_\varepsilon}$ at $(\varepsilon, \lambda_\varepsilon)$ can be used to construct $\mathbb{S}^x_{\varepsilon,\lambda_\varepsilon}$ with (3.16) which is guaranteed to contain $S^{x'}_{\varepsilon',\lambda_{\varepsilon'}}$ for the next smaller $(\varepsilon', \lambda_{\varepsilon'})$.

Remark 12 (Convexity of $S^x_{\varepsilon,\lambda}$)**.** *For small ε, we can assume that we are restricted to a period of the torus and working with the usual Euclidean distance. The intermediate level set in (3.15) is convex thanks to lemma (3). We do not know the log-concavity result extends to constrained Sinkhorn sequence $(f^m_{\varepsilon,\lambda}, g^m_{\varepsilon,\lambda})$. However $S^x_{\varepsilon,\lambda}$ is exponentially close in ε to a convex set and its convex enveloppe is in any case contained in $A^x_{\varepsilon,\lambda_\varepsilon}$.*

4. Numerics

4.1. **The numerical procedure.** The method will be a continuation method in $(\varepsilon, \lambda_\varepsilon)$ and N as it also depends on ε. They will be chosen according to proposition 8 and remark 8. *In the remaining part of the paper we may sometimes drop the $.^\star$ and $._\lambda$ dependence in the notations.*

Given the m_ε iterate $(f^{m_\varepsilon}_\varepsilon, g^{m_\varepsilon}_\varepsilon, h^{m_\varepsilon}_\varepsilon)$ of (2.2), We detail below how to compute the next solution level at $\varepsilon' = \varepsilon\,\alpha$, $\alpha < 1$ using the "out-summing" Sinkhorn algorithm (2.10). We will use the same notations as in proposition 8. In particular $x_i \in X_N$, $y_j \in Y_N$ will be points on a regular discretization grid and $\lambda = \lambda_\varepsilon = \varepsilon$ (i.e. $\beta = 0$ slightly abusing the hypothesis in proposition 8).

(Step A) *$\forall x \in X_N$, we compute S^x_ε :* Thanks to lemma 3 and remark 2, we know that S^x_ε (which is dominated by A^x_ε) is convex in the period $]y^\star_x - 1/2, y^\star_x - 1/2[$. It will be convenient to work with :

$$P_\varepsilon = \exp(\frac{1}{\varepsilon}(f_\varepsilon \oplus g_\varepsilon) - c_N). \tag{4.1}$$

For $d = 1$ ($d > 1$ is discussed in remark 16) one can therefore use the definition of S^x_ε :

$$S^x_\varepsilon =]\min_{y\in]y^\star_x-1/2,y^\star_x-1/2[}\{P_\varepsilon(x,y) \geq \lambda_\varepsilon\}, \max_{y\in]y^\star_x-1/2,y^\star_x-1/2[}\{P_\varepsilon(x,y) \geq \lambda_\varepsilon\}[\tag{4.2}$$

In practice (4.2) cannot be implemented as the OT map $x \mapsto y^\star_x$ is not known. Instead one compute on an arbitrary period the (unique thanks to the log-concavity again) couple

$$\{y_{j_m}\} = \{y_j,\, P_\varepsilon(x,y_j) \leq \lambda_\varepsilon < P_\varepsilon(x,y_{j+1})\} \qquad \{y_{j_M}\} = \{y_j,\, P_\varepsilon(x,y_j) > \lambda_\varepsilon \geq P_\varepsilon(x,y_{j+1})\}$$

and sort them to determine if S^x_ε contains a periodic boundary or not.

(Step B) *$\forall x \in X_N$, we compute $\mathbb{S}^x_\varepsilon$:* using the definition (3.16) this simply is

$$\mathbb{S}^x_\varepsilon =]y_{j_m} - d\mathcal{B}_{\varepsilon,\lambda}, y_{j_M} + d\mathcal{B}_{\varepsilon,\lambda}[. \tag{4.3}$$

(Step C) *"Out-Summation" at ε' :* $\forall x' \in X'_N$ (the finer grid), we do not know $S^{x'}_{\varepsilon'}$ but thanks to lemma 9 we know it is contained in $\mathbb{S}^{x[x']}_{\varepsilon}$ where

$$x[x'] \in X_N \text{ is he nearest point to } x' \text{ on the coarser grid.}$$

Indeed, the discretization step $1/N$ depends on ε and $|x' - x[x']| < \varepsilon^{1/2}$.

By construction the saturated domain of the solution of $OT_{\varepsilon',\lambda'}(\mu'_N, \nu'_N)$ (2.2) is contained in $\{x', Y_{N'} \setminus \mathbb{S}^{x[x']}_{\varepsilon}\}$, $\forall x'$. It is also contained in $\{X_{N'} \setminus \mathbb{S}^{y[y']}_{\varepsilon}, y'\}$ $\forall y'$, obtained by switching the x and y coordinates. We can now formulate the modified and implementable "out-summed" Sinkhorn algorithm (2.10) :

ITERATE :

(4.4)
$$\begin{cases} \forall x' \in X_{N'},\ f^{m+1}_{\varepsilon'}(x') = \varepsilon' \left(\log(1 - \lambda \langle 1, \nu_{N'} \rangle_{Y_{N'} \setminus \mathbb{S}^{x[x']}_{\varepsilon}}) - \log(\langle \exp(\frac{1}{\varepsilon'}(g^m_{\varepsilon'} + h^m_{\varepsilon'} - c_{N'})), \nu_{N'} \rangle_{\mathbb{S}^{x[x']}_{\varepsilon}})\right) \\ \forall y' \in Y_{N'},\ g^{m+1}_{\varepsilon'}(y') = \varepsilon' \left(\log(1 - \lambda \langle 1, \mu_{N'} \rangle_{X_{N'} \setminus \mathbb{S}^{y[y']}_{\varepsilon}}) - \log(\langle \exp(\frac{1}{\varepsilon'}(f^{m+1}_{\varepsilon'} + h^m_{\varepsilon'} - c_N)), \mu_N \rangle_{\mathbb{S}^{y[y']}_{\varepsilon}})\right) \\ \forall x' \in X_{N'},\ \{\exp(\frac{1}{\varepsilon'} h^{m+1}_{\varepsilon'})\}(x', y' \in \mathbb{S}^{x[x']}_{\varepsilon}) = \max\left(1, \dfrac{\lambda}{\{\exp(\frac{1}{\varepsilon'}(f^{m+1}_{\varepsilon'} + g^{m+1}_{\varepsilon'} - c_N)))\}(x', y' \in \mathbb{S}^{x[x']}_{\varepsilon})}\right). \end{cases}$$

UNTIL : We end the loop, in theory, when $m = m_\varepsilon$ to comply with the hypothesis of lemma 7 which is the foundation of proposition 8 and lemma 9. In practice we found that the error criterium :

(4.5)
$$\begin{aligned} EC_{\varepsilon,\lambda_\varepsilon} &:= \|\langle 1, \gamma^{ns,m}_{\varepsilon,\lambda_\varepsilon} \rangle_{X_{N'}} - \mu_{N'}\|_\infty + \|\langle 1, \gamma^{ns,m}_{\varepsilon,\lambda_\varepsilon} \rangle_{Y_{N'}} - \nu_{N'}\|_\infty \\ &= \|\mu^0_N - \mu_{N'}\|_\infty + \|\nu^0_N - \nu_{N'}\|_\infty < \lambda'_\varepsilon \end{aligned}$$

was also good and resulted in less iterations (see figures (5, 11, 17, 23)). Here $\gamma^{ns,m}_{\varepsilon,\lambda_\varepsilon}$ is constructed according to (2.6) from the mth iterate of (4.4). This condition ensures that it is a solution of the EOT (1.9) with marginals satisfying (3.3).

(Step D) Set $\varepsilon = \varepsilon'$ and $\varepsilon' = \varepsilon\,\alpha$ and the other parameters accordingly, go to (A).

Remark 13 (Computational complexity)**.** *Assuming we we will be repeating the Steps $A \to D$ a finite number of times and based on the number on the number of Sinkhorn iterations $m_\varepsilon = -A\log(\varepsilon)/\varepsilon$ needed at each ε step. We finally get, from 8 (v), a computational cost of $\mathcal{O}(\log(N)^{1+d/2}\, N^{3/2+1/d})$ if $N = (1/\varepsilon)^d$ and $\mathcal{O}(\log(N)^{1+d/2}\, N^{1+2/d})$ if $N = (1/\sqrt{\varepsilon})^d$; N is of course the largest discretization reached (for the smaller ε).*

Remark 14 (Stabilisation property of (4.4))**.** *A nice side effect of the "out-summation" is to avoid overflows in the third equation of (4.4) : We remove (for the next ε' level) the points such that $\exp(\frac{1}{\varepsilon}(f_\varepsilon + g_\varepsilon - c_N)) < \lambda$, so at the next level $(\varepsilon', \lambda') < (\varepsilon, \lambda)$ and for a "reasonable" decrease, it may be smaller but cannot reach very small numbers.*

Remark 15 (ε-scaling and kernel trunction)**.** *The discussion on the unstability induced by finite precision arithmetics in section 1.3 is still relevant. We still need to implement ε-scaling as explained in section 1.5. Kernel truncation however is replaced by the exact λ "out-summation".*

Remark 16 ($d > 1$)**.** *As detailed in remark 13, the memory foot print and computational cost is (or is not far from) linear in the N the number of points to discretize X. We only implemented this algorithm for $d = 1$ but we do not foresee any difficulty for $d > 1$. The only delicate task would be the computation of S^x_ε (Step A), but it is close to convex thanks to lemma 3 (see remark 12) and therefore can be approximated by a cartesian box with the same ideas.*

4.2. **Numerical Results.** All four presented test cases (all figures at the end of the paper) have been computed with the same parameters. The procedure Steps A $\to$ D is run 6 times with $\alpha = 0.2$, the decreasing sequence of ε is $\{0.1\,\alpha^k\}_{k=1..6}$. We always take $\lambda = \varepsilon$ and $N = (2/\sqrt{\varepsilon})^d$. Regarding theorem 4 parameters, we choose arbitrarily $p = A = 1$. Figure 3 shows a log plot of $\mathcal{B}_\varepsilon$ (1.19), and the $\mathcal{B}_\varepsilon > \lambda$ level sets (see (3.9)) for this decreasing sequence of ε.

The numerical method was implemented in Matlab using sparse matrices to represent the saturated domain. The code has not been optimized. The full computation for the 6 levels takes on average a minute on a modern 1 GHz Intel Core i7, 16GB memory laptop.

- Figures (4, 10, 16, 22) show the marginals (μ_N, ν_N) for the successive ε/N discretization. The last case does not satifies the C^2 hypothesis.
- Figures (5, 11, 17, 23) show the Sinkhorn convergence curves for the successive ε/N discretization until (4.5) is satisfied. The $\mathcal{O}(N^2)$ number of iterations is also log plotted in figure 2.
- Figures (6, 12, 18, 24) show the computed transport plans $\gamma^{us}_{\varepsilon,\lambda_\varepsilon}$ (see (2.6) or the successive ε/N discretization. It converges to $\gamma^\star$ the solution of the non entropic $OT(\mu,\nu)$ problem (1.1).
- Figures (7, 13, 19, 25) show the computational domain $\{x, \mathbb{S}^x_\varepsilon\}$, $x \in X_N$ (see (4.3) and therefore the sparsity pattern induced by the "out-summation" of the saturated domain, **nz** is the number of kept (non zero) elements . Axis labels correspond to the cartesian grid discretization indices.
- Figures (8, 14, 20, 26) Show the saturation potential $h_{\varepsilon,\lambda_\varepsilon}$ on $\{x, \mathbb{S}^x_\varepsilon\}$. It is not computed and set to 0 elsewhere but is in theory greater than 1 there. These figures show that the method captures correctly the inclusions (3.19) in lemma 9. The 0 saturated zone $S^x_{\varepsilon',\lambda'}[x']$ is inside $\mathbb{S}^x_{\varepsilon,\lambda}$ for all x'.
- Figures (9, 15, 21, 27) show a log/log plot of $|\{x, \mathbb{S}^x_\varepsilon\}|$ versus N. Despite the (slightly) superlinear complexity found in remark 13 we find a linear behavior.

5. Conclusion

The proposed method comes with theoretical and numerical guarantees. The p parameter in theorem 4 has not however been estimated rigorously (we took $p = 1$). Likewise the C_δ parameter in lemma 8 has been ignored. The comparison of the constrained/unconstrained Sinkhorn iterates in lemma 7 introduces technical complications which can probably be avoided.

It is also unclear if the choice of parameters $\varepsilon = \lambda$ and $\alpha = 0.2$ are optimal. It gave good results with our very crude 1D numerical implementation, even for non-smooth and non lower bounded densitties (μ, ν). It heavily relies on Matlab sparse matrix and cell structures. It can certainly be improved and optimized but already demonstrates that the method is robust and converges as expected.

Using $N = (1/\sqrt{\varepsilon})^d$ (smooth marginals) the memory footprint is linear. The number of operations scales like $N^{1+2/d}$ (remark 13). This is cubic for $d = 1$ instead of $\mathcal{O}(N^4)$ with no "out-summation". But a larger gain expected when d increases.

Testing on higher performance computers would also make sense to check the stability of the (ε, λ) continuation method for smaller values/larger discretizations. A numerical study using the finer $N = 1/\varepsilon$ discretization rule should also be undertaken.

We partially tested the method on the non-periodic l^2 cost and it seems to work with a careful treatment of the saturation at the boundaries. Based on [Berman, 2017], it seems reasonable to conjecture that the method could be applied at least to costs in the form $h(x - y)$, h convex.

AKNOWLEDGEMENTS

We want to thank Robert Berman, Francis Collino, Jean Feydy and François-Xavier Vialard for helpful discussions.

REFERENCES

[Benamou and Duval, 2017] Benamou, J.-D. and Duval, V. (2017). Minimal convex extensions and finite difference discretization of the quadratic monge-kantorovich problem.

[Benamou et al., 2020] Benamou, J.-D., Ijzerman, W. L., and Rukhaia, G. (2020). An Entropic Optimal Transport Numerical Approach to the Reflector Problem. working paper or preprint.

[Berman, 2017] Berman, R. J. (2017). The sinkhorn algorithm, parabolic optimal transport and geometric monge-amp\ere equations. *arXiv preprint arXiv:1712.03082*.

[Brenier, 1991] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417.

[Chizat et al., 2018] Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018). Scaling algorithms for unbalanced transport problems. *to appear in Mathematics of Computation*.

[Collino, 2020] Collino, F. (2020). private communication.

[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, 2020] Feydy, J. (2020). *Geometric data analysis, beyond convolutions*. PhD thesis.

[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.

[Galichon, 2016] Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press, 1 edition.

[Kolouri et al., 2016] Kolouri, S., Park, S., Thorpe, M., Slepčev, D., and Rohde, G. K. (2016). Transport-based analysis, modeling, and learning from signal and data distributions.

[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.

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

[Oberman and Ruan, 2015] Oberman, A. M. and Ruan, Y. (2015). An efficient linear programming method for optimal transportation.

[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*.

[Santambrogio, 2015] Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing.

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

[Vialard, 2019] Vialard, F.-X. (2019). An elementary introduction to entropic regularization and proximal methods for numerical optimal transport. Lecture, hal-02303456.

[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.

FIGURES

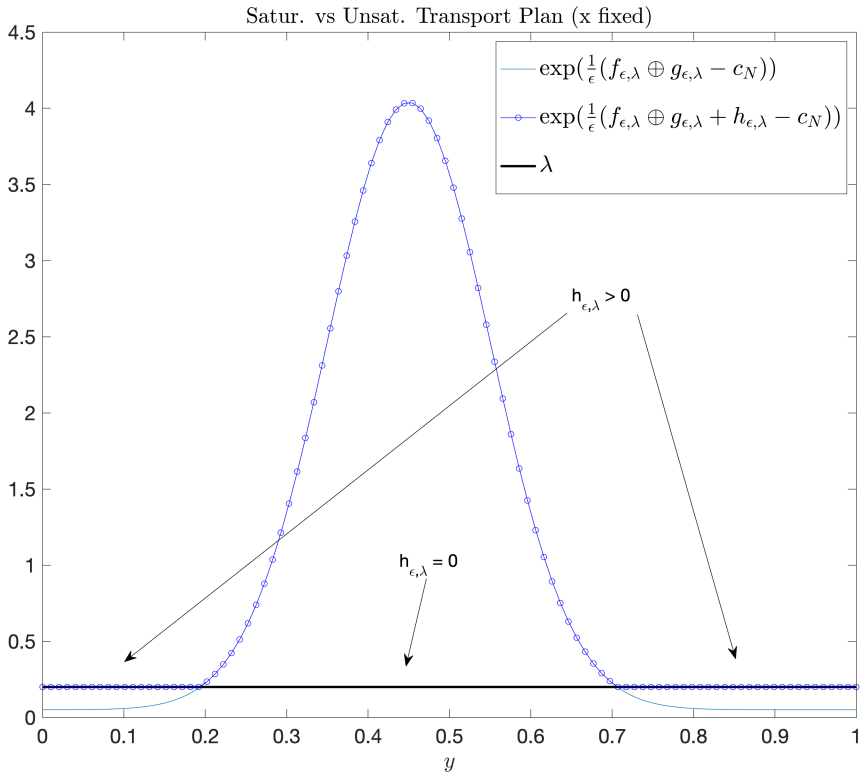

FIGURE 1. Vertical Slice of saturated versus non saturated transport plans.

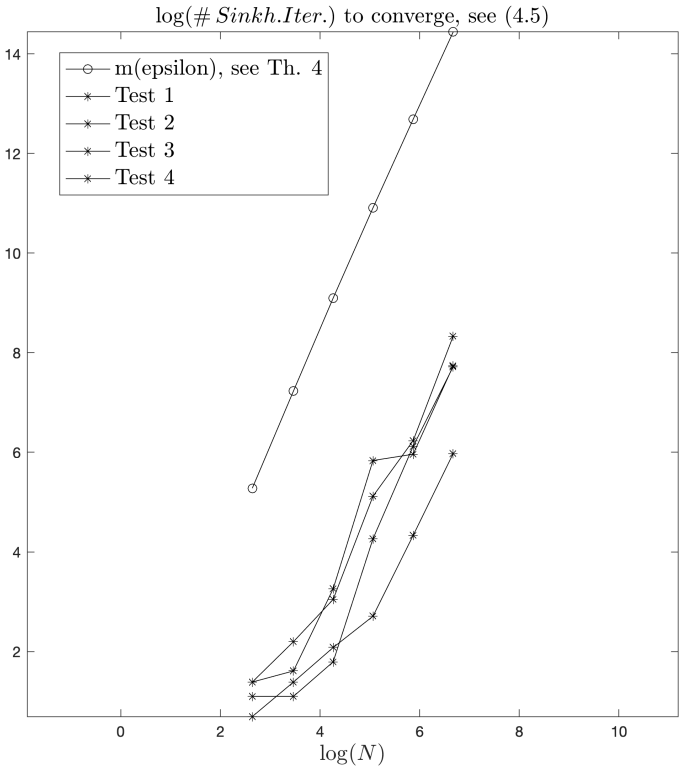

FIGURE 2. Log/Log plot of the number of Sinhorn iterations to reach (4.5) versus N (which depends on $\varepsilon = \lambda$). Also represented is the theoretical number of iterations m_ε in theorem 4 to get joint convergence.

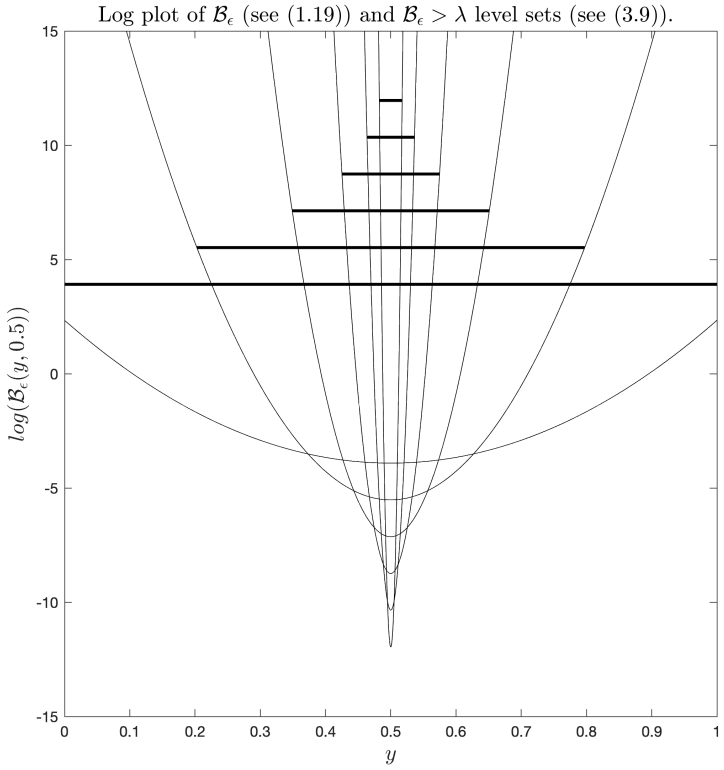

FIGURE 3. Log plot of $\mathcal{B}_\epsilon$ (see (1.19)) and $\mathcal{B}_\epsilon > \lambda$ level sets (see (3.9)), $\varepsilon = \lambda$ decrease as in the numerical tests.

TEST 1

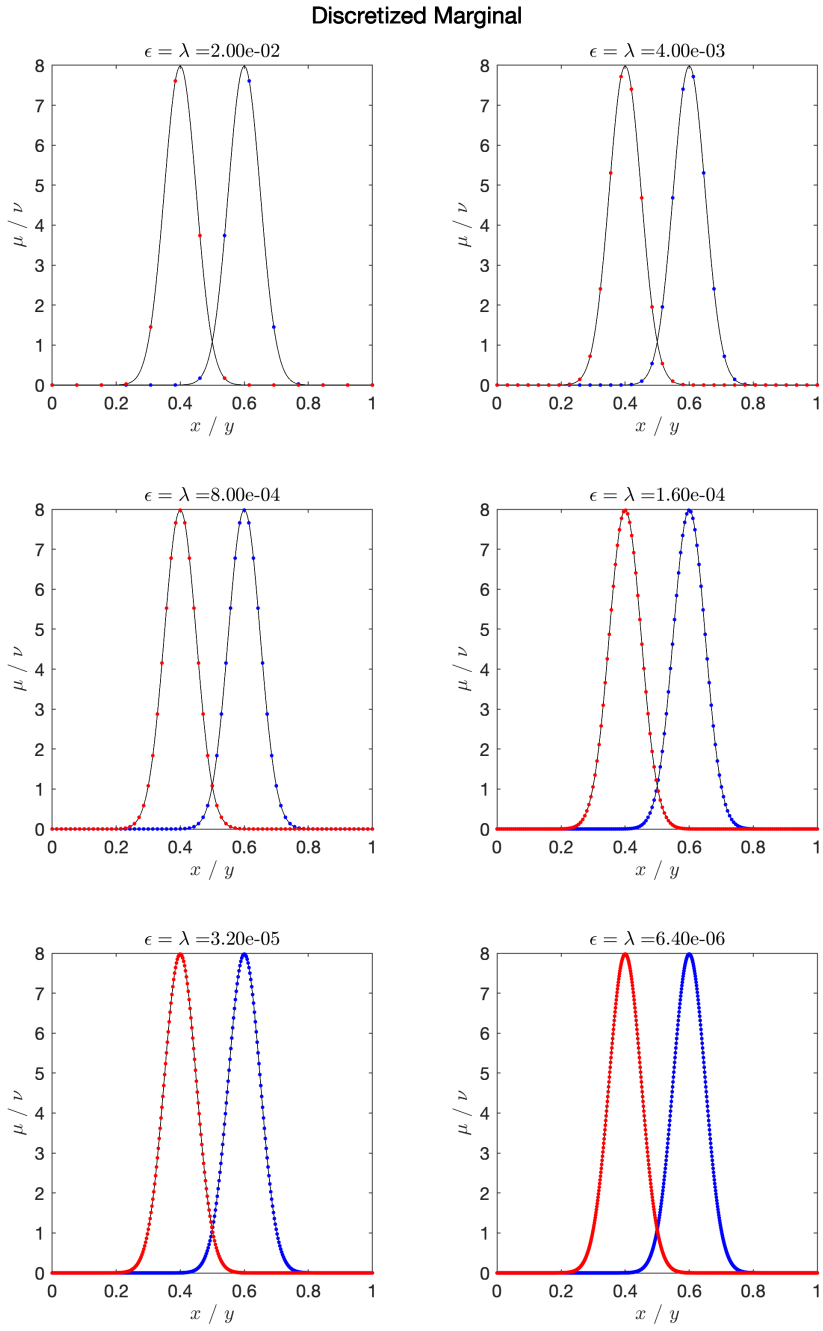

FIGURE 4. Sequence (in ε/N) of (μ_N, ν_N) .

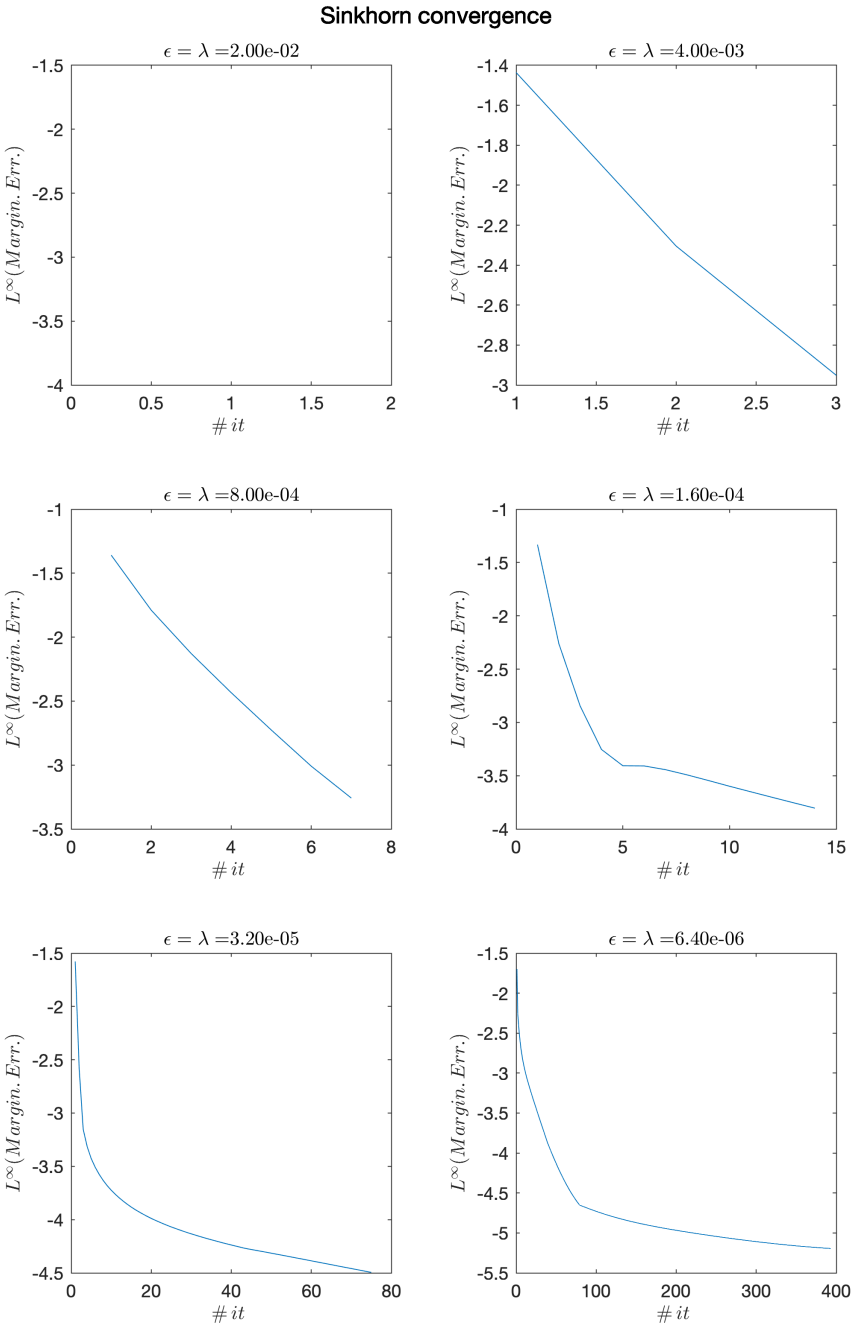

FIGURE 5. Sinkhorn Convergence. Error defined by (4.5). For $\varepsilon = 2.00e-02$, 4.5 is satified after the first iteration.

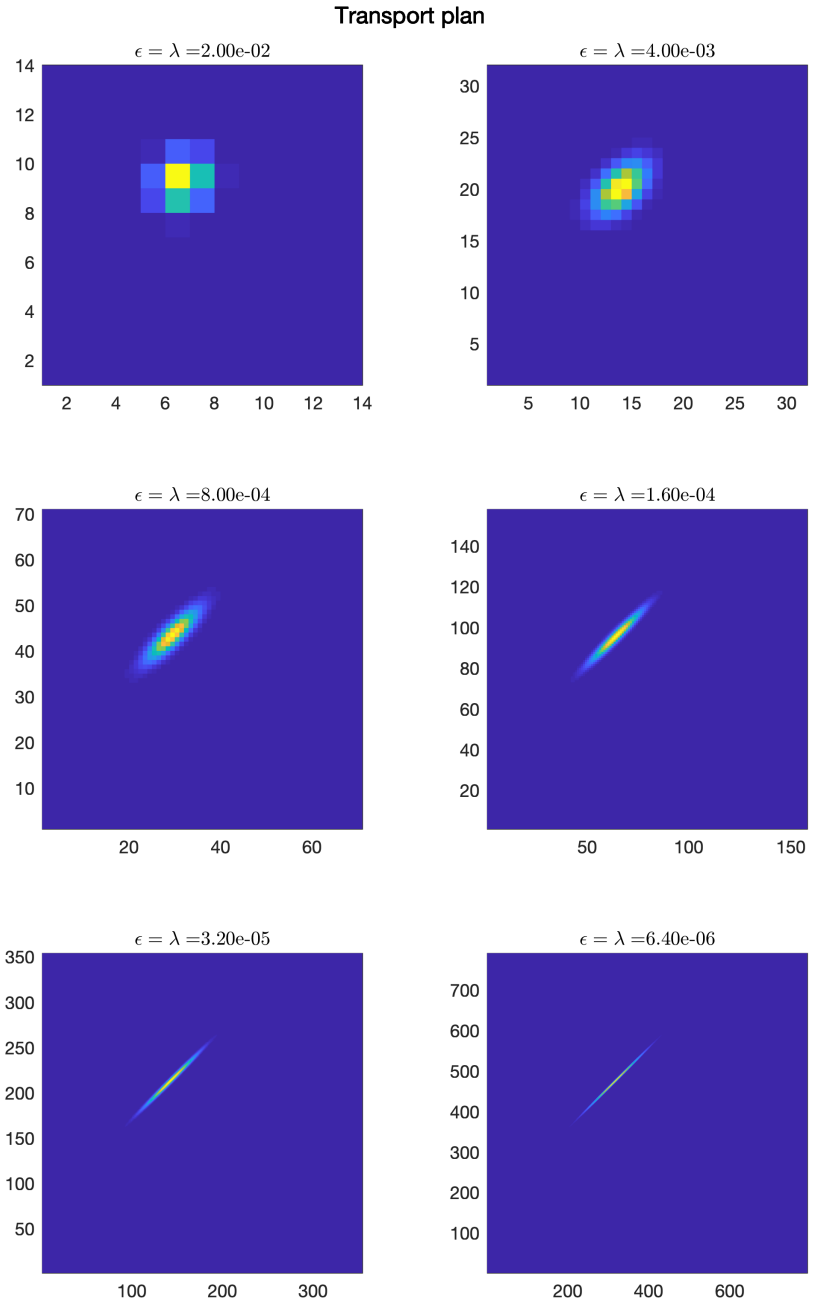

FIGURE 6. Transport Plan $\gamma^{us}_{\varepsilon,\lambda_\varepsilon}$ (see (2.6)).

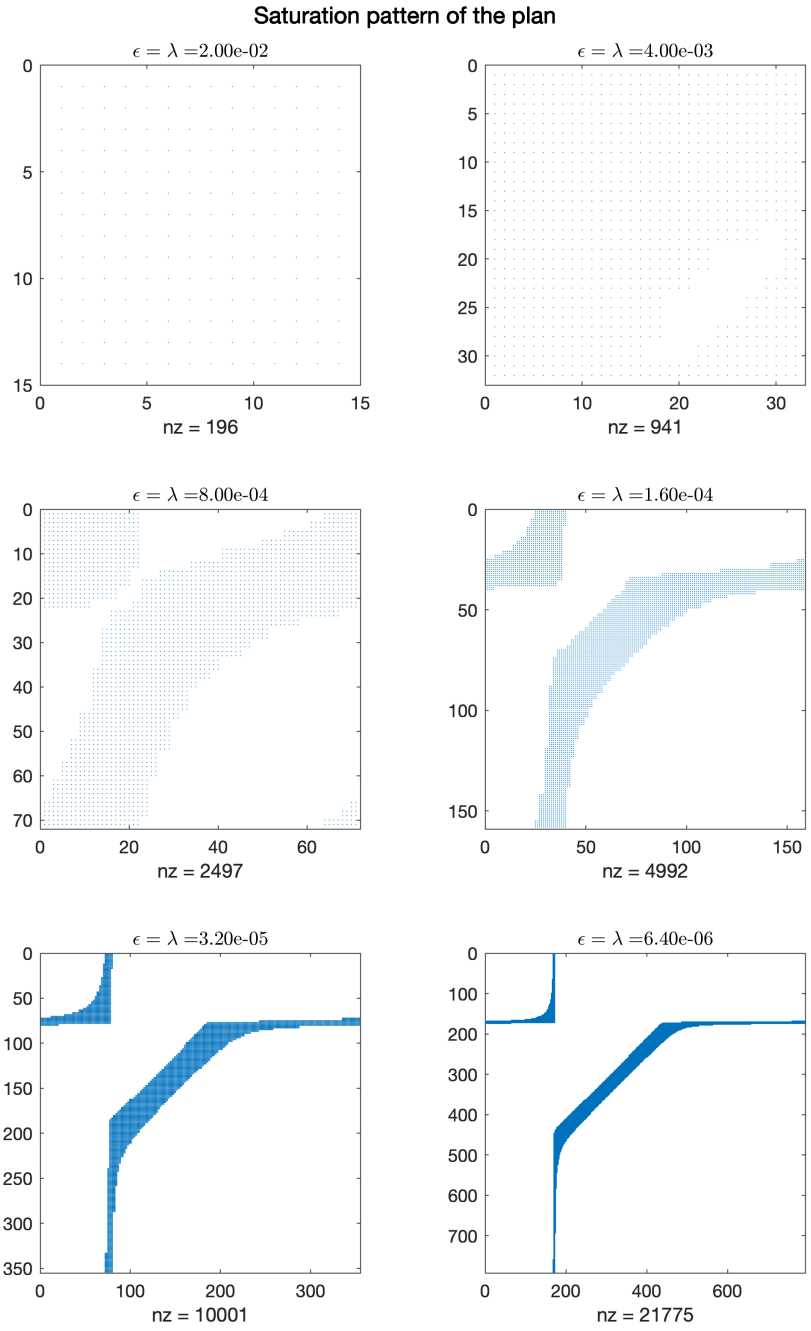

FIGURE 7. Sparsity pattern $\{x, \mathbb{S}_\epsilon^x\}$ (see (4.3), **nz** is the number of kept (non-zero) elements. Axis show the grid indices.

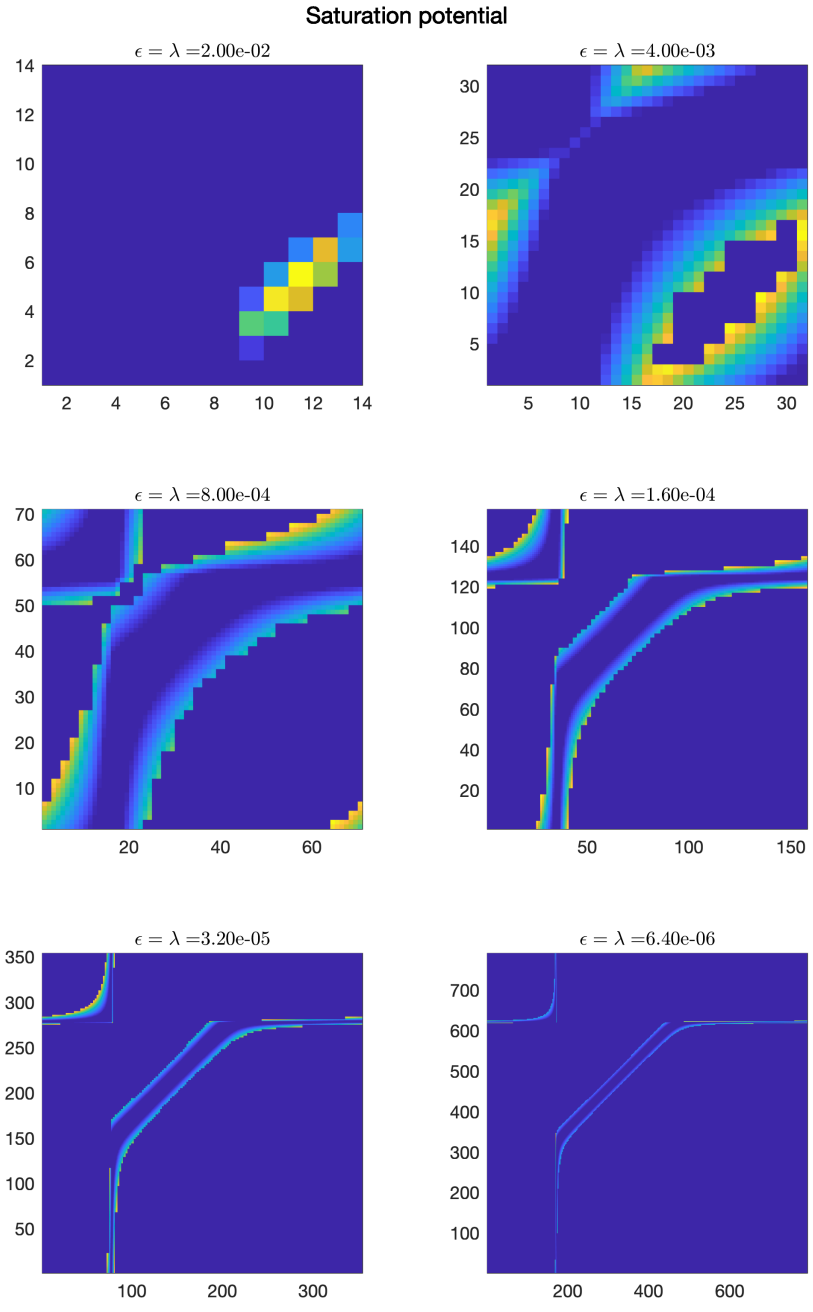

FIGURE 8. Saturation potential $h_{\varepsilon,\lambda_\varepsilon}$ on $\{x, \mathbb{S}_\varepsilon^x\}$ (set to 0 elsewhere as it is not computed but should be > 1) . Axis show the grid indices.

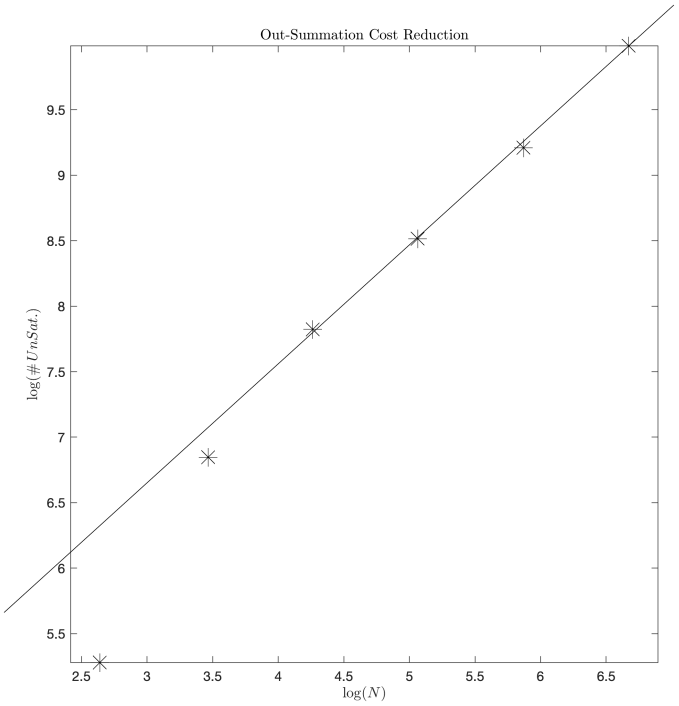

FIGURE 9. Log/Log plot of $|\{x, \mathbb{S}_\varepsilon^x\}|$ versus N plus a linear fit.

TEST 2

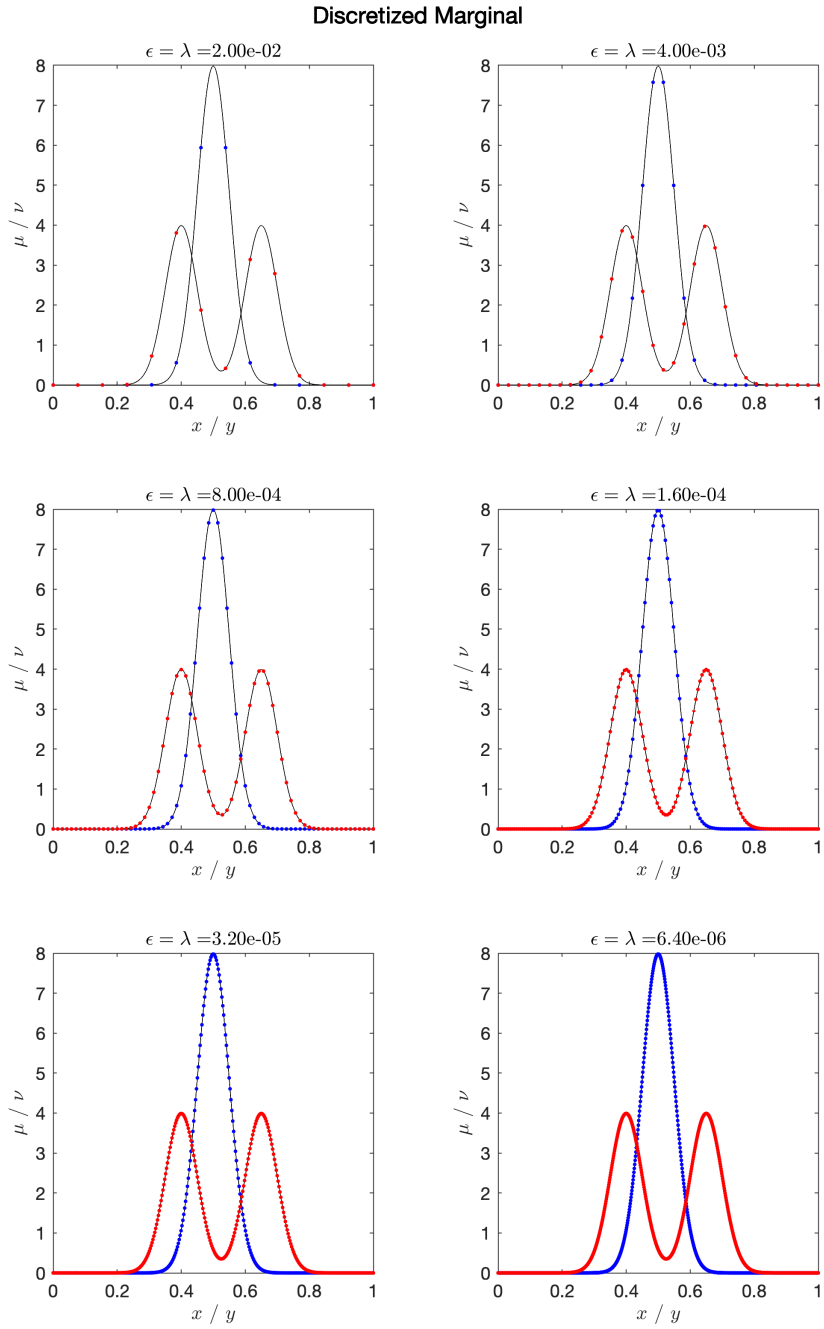

FIGURE 10. Sequence (in ε/N) of (μ_N, ν_N) .

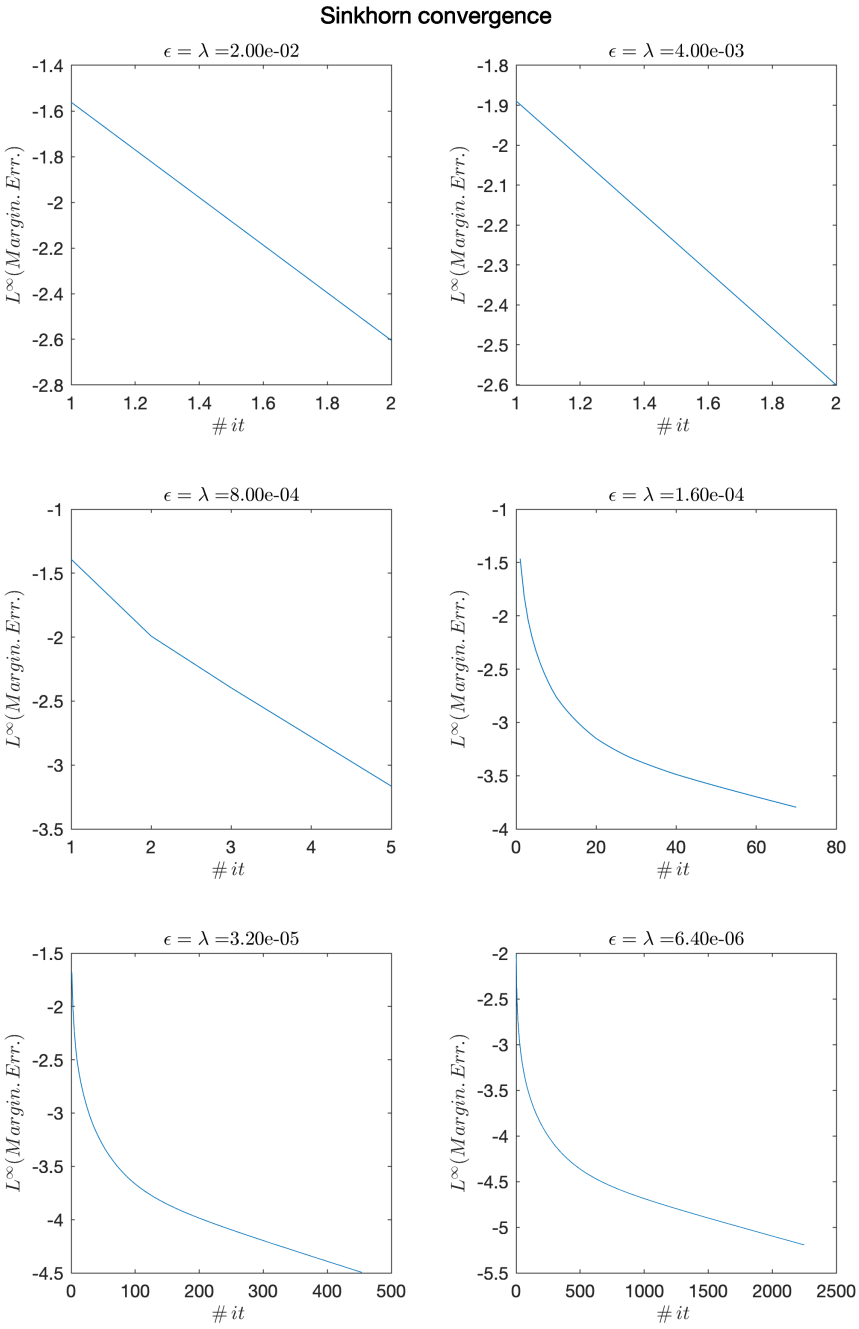

FIGURE 11. Sinkhorn Convergence. Error defined by (4.5)

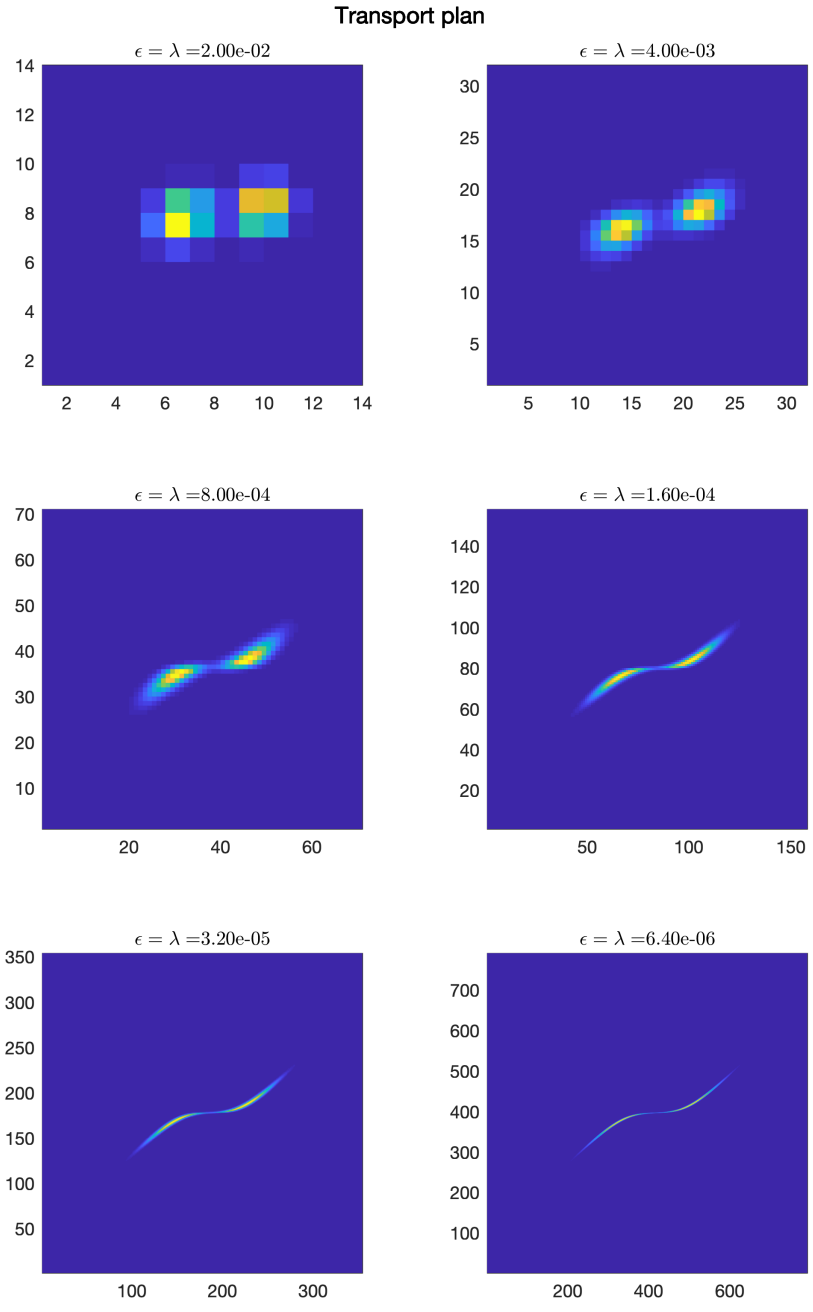

FIGURE 12. Transport Plan $\gamma^{us}_{\varepsilon,\lambda_\varepsilon}$ (see (2.6)).

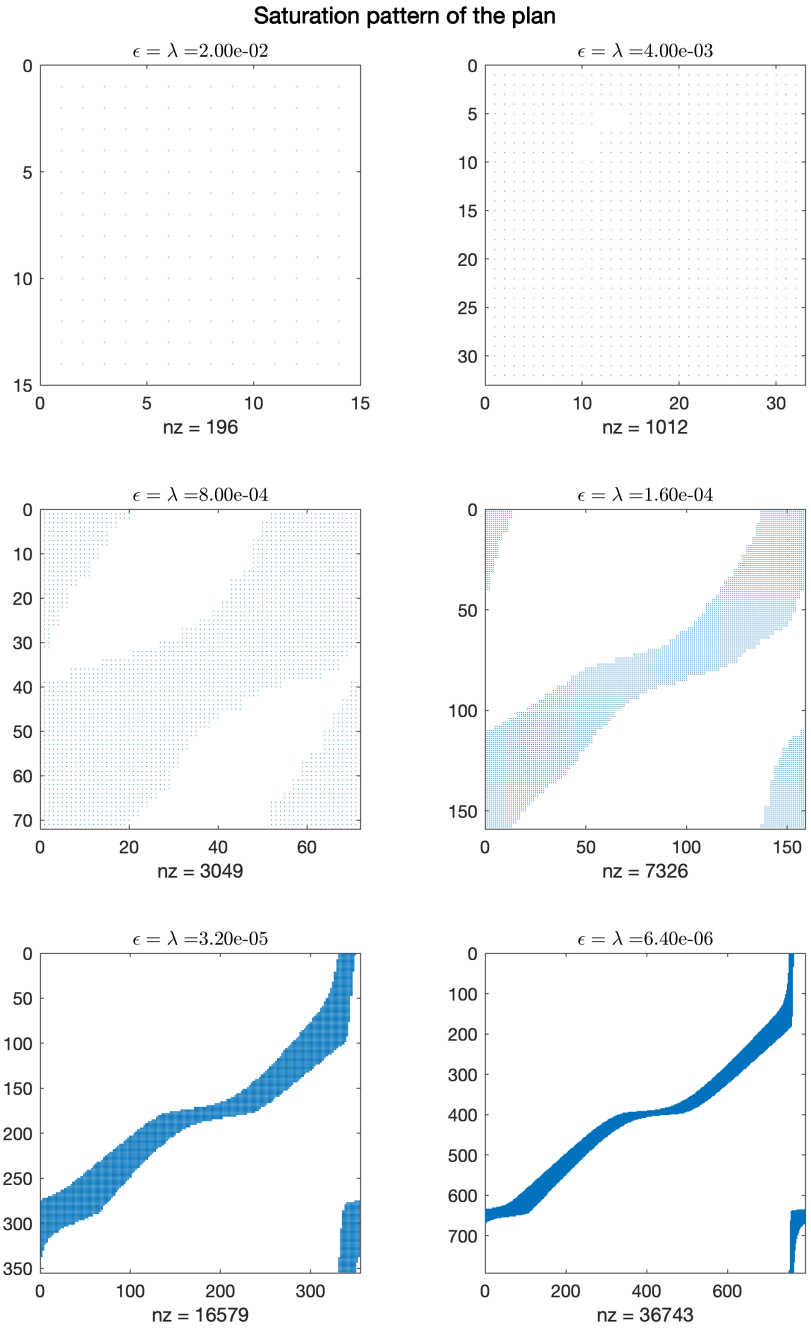

FIGURE 13. Sparsity pattern $\{x, \mathbb{S}_\varepsilon^x\}$ (see (4.3), **nz** is the number of kept (non-zero) elements

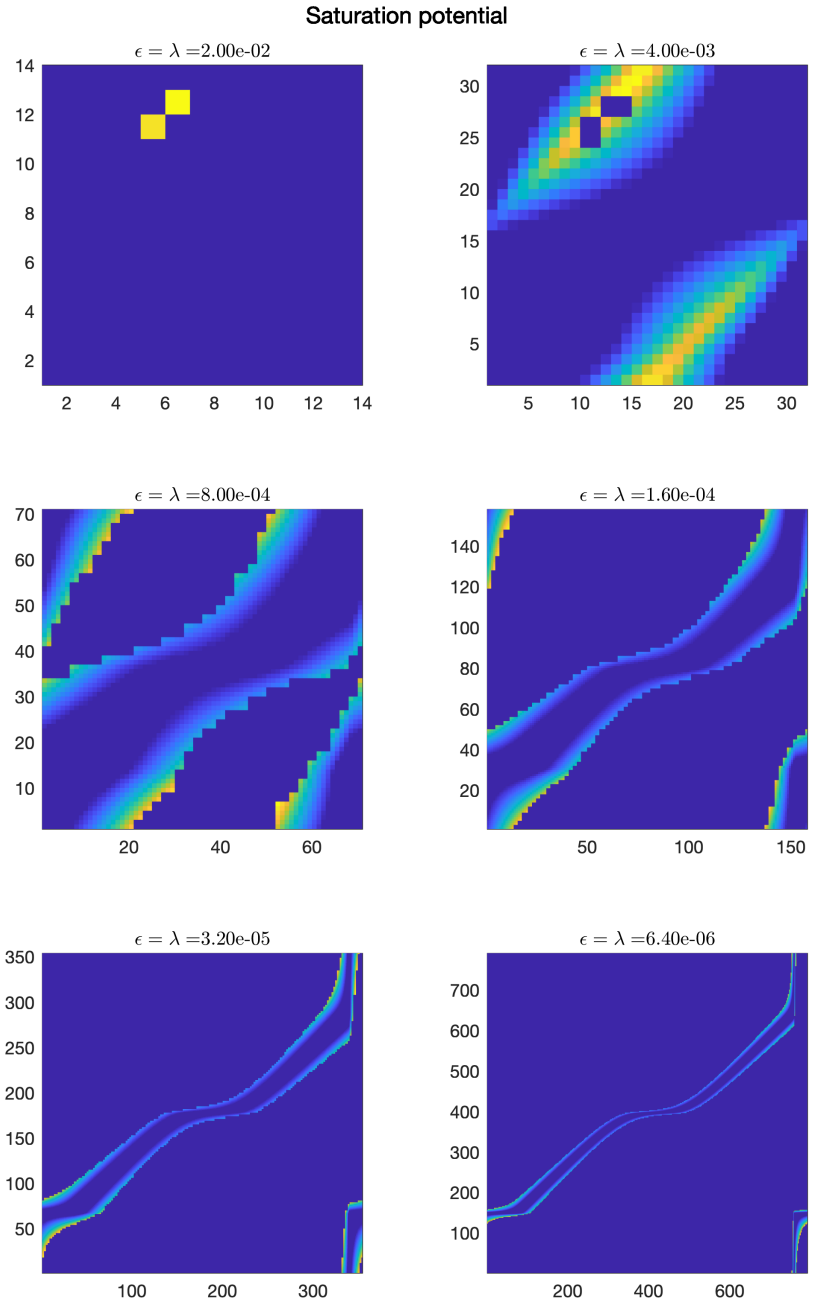

FIGURE 14. Saturation potential $h_{\varepsilon,\lambda_\varepsilon}$ on $\{x, \mathrm{S}_\varepsilon^x\}$ (set to 0 elsewhere as it is not computed but should be > 1) . Axis show the grid indices.

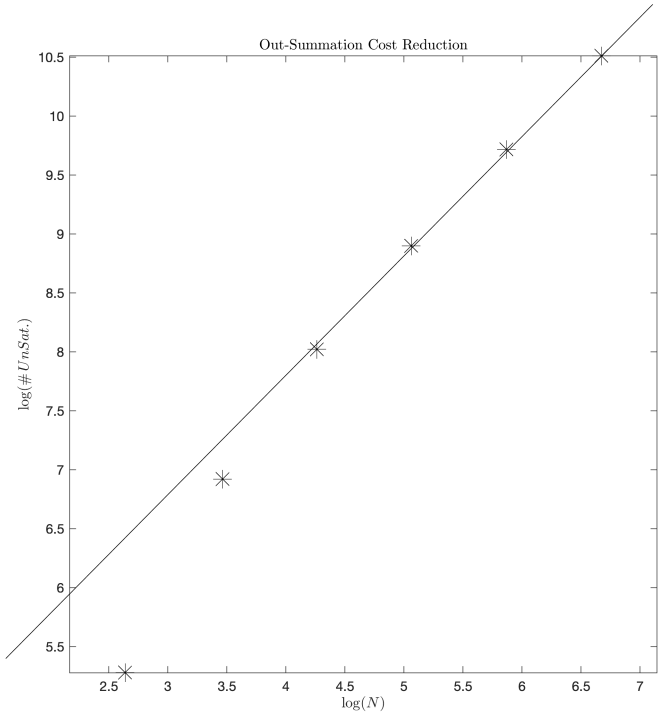

FIGURE 15. Log/Log plot of $|\{x, \mathbb{S}_\varepsilon^x\}|$ versus N plus a linear fit.

TEST 3

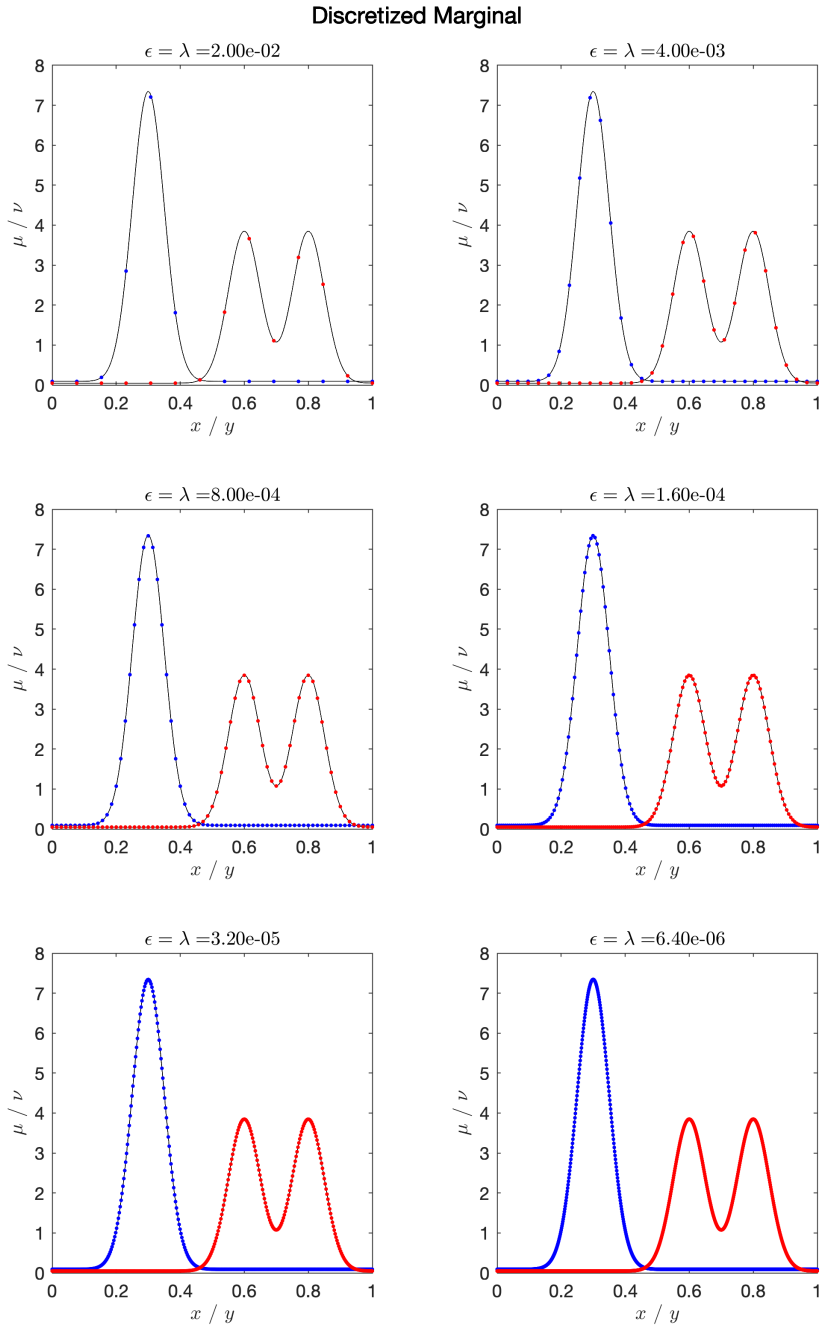

FIGURE 16. Sequence (in ε/N) of (μ_N, ν_N) .

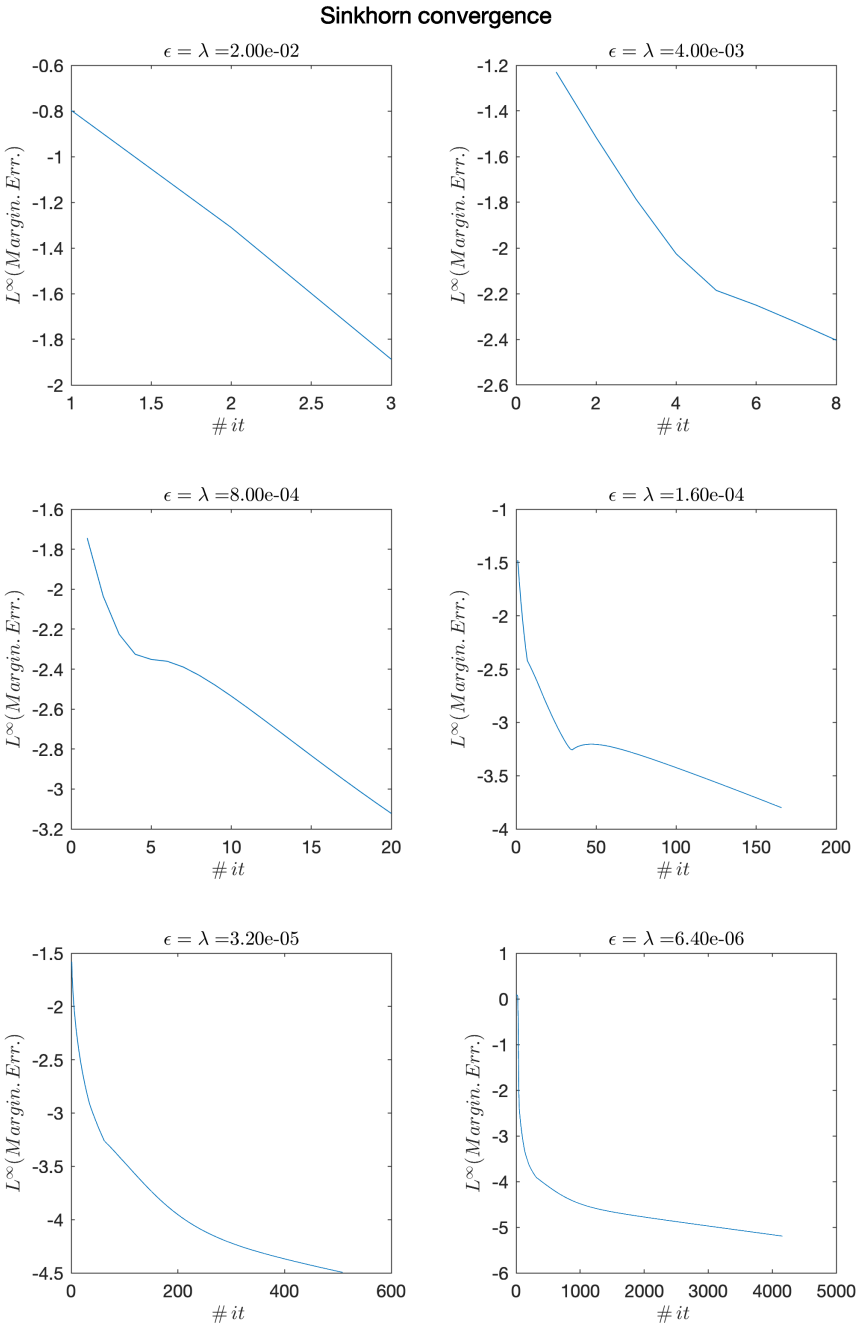

FIGURE 17. Sinkhorn Convergence. Error defined by (4.5)

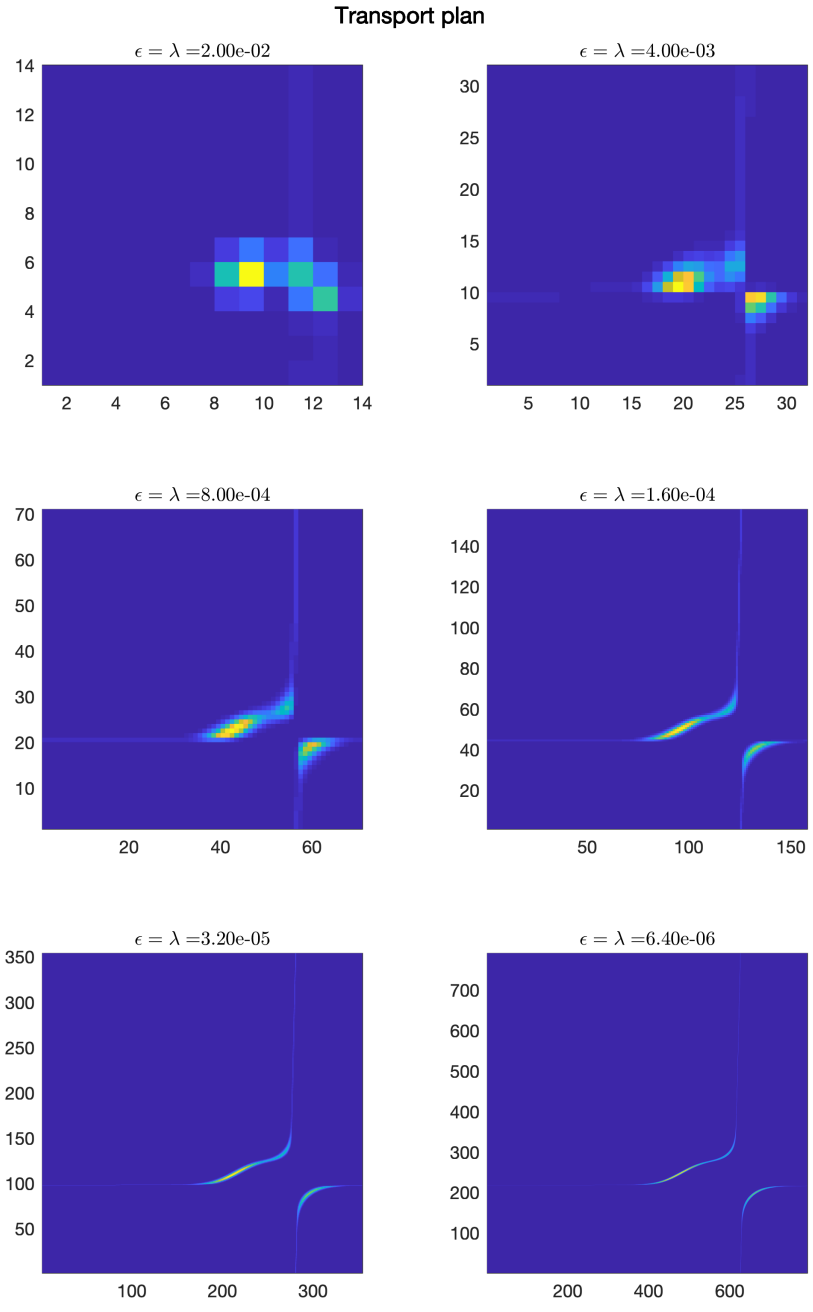

FIGURE 18. Transport Plan $\gamma^{us}_{\varepsilon,\lambda_\varepsilon}$ (see (2.6)).

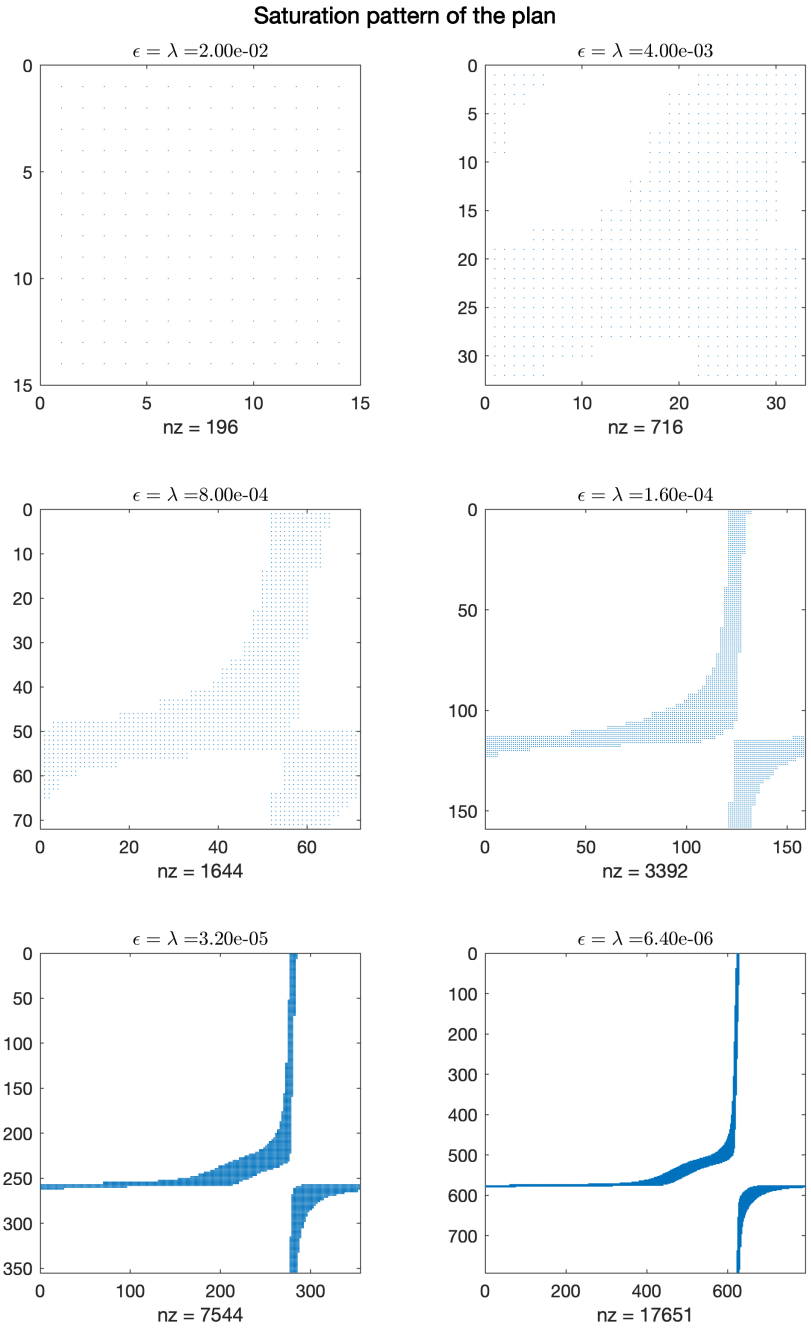

FIGURE 19. Sparsity pattern $\{x, \mathbb{S}_\varepsilon^x\}$ (see (4.3), **nz** is the number of kept (non-zero) elements

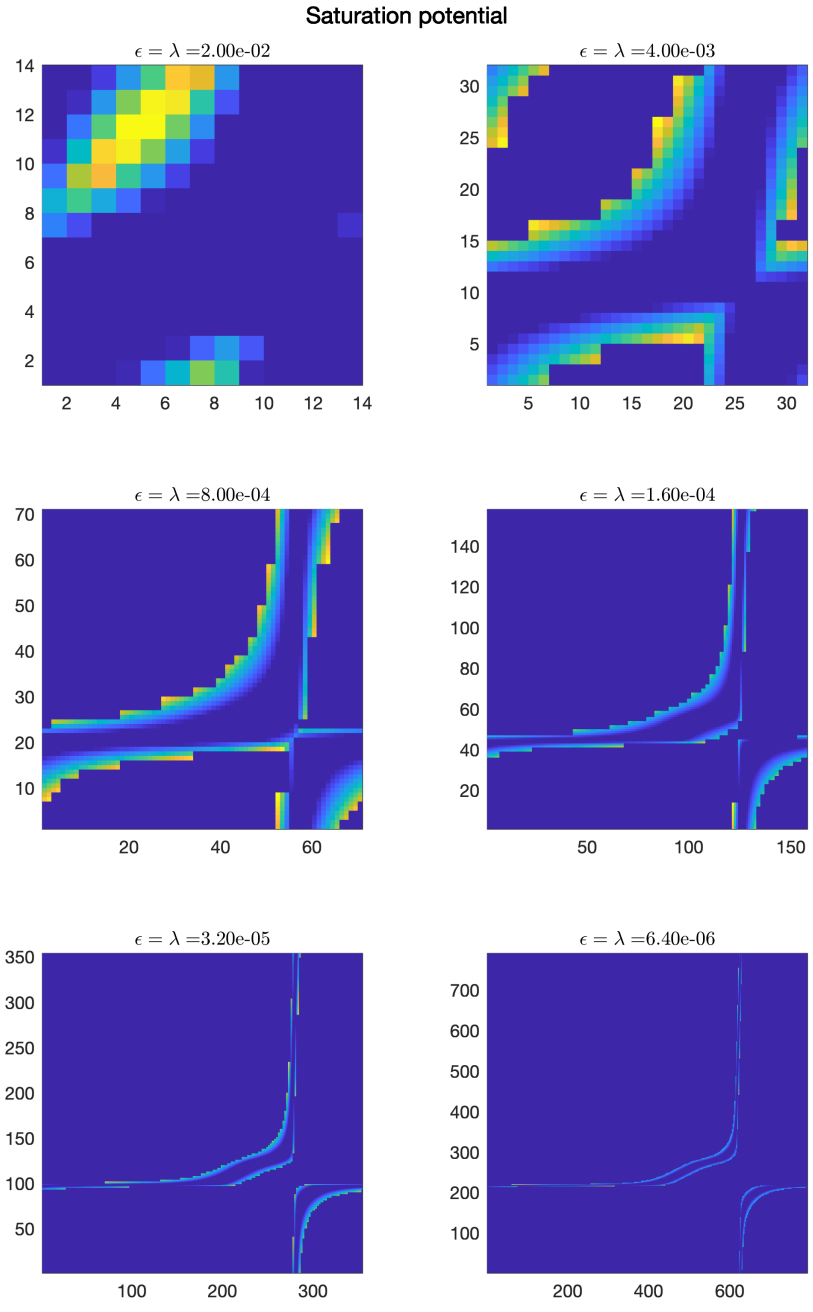

FIGURE 20. Saturation potential $h_{\varepsilon,\lambda_\varepsilon}$ on $\{x, \mathrm{S}_\varepsilon^x\}$ (set to 0 elsewhere as it is not computed but should be > 1) . Axis show the grid indices.

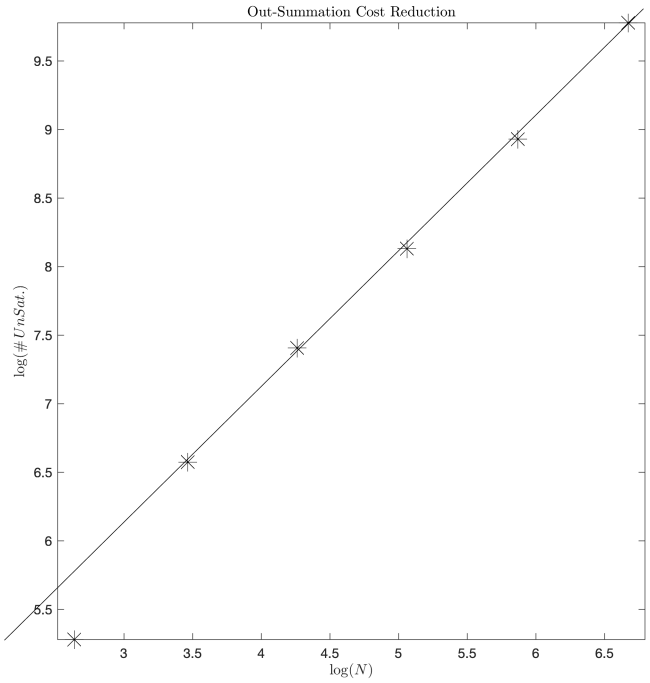

FIGURE 21. Log/Log plot of $|\{x, \mathbb{S}_\varepsilon^x\}|$ versus N plus a linear fit.

TEST 4

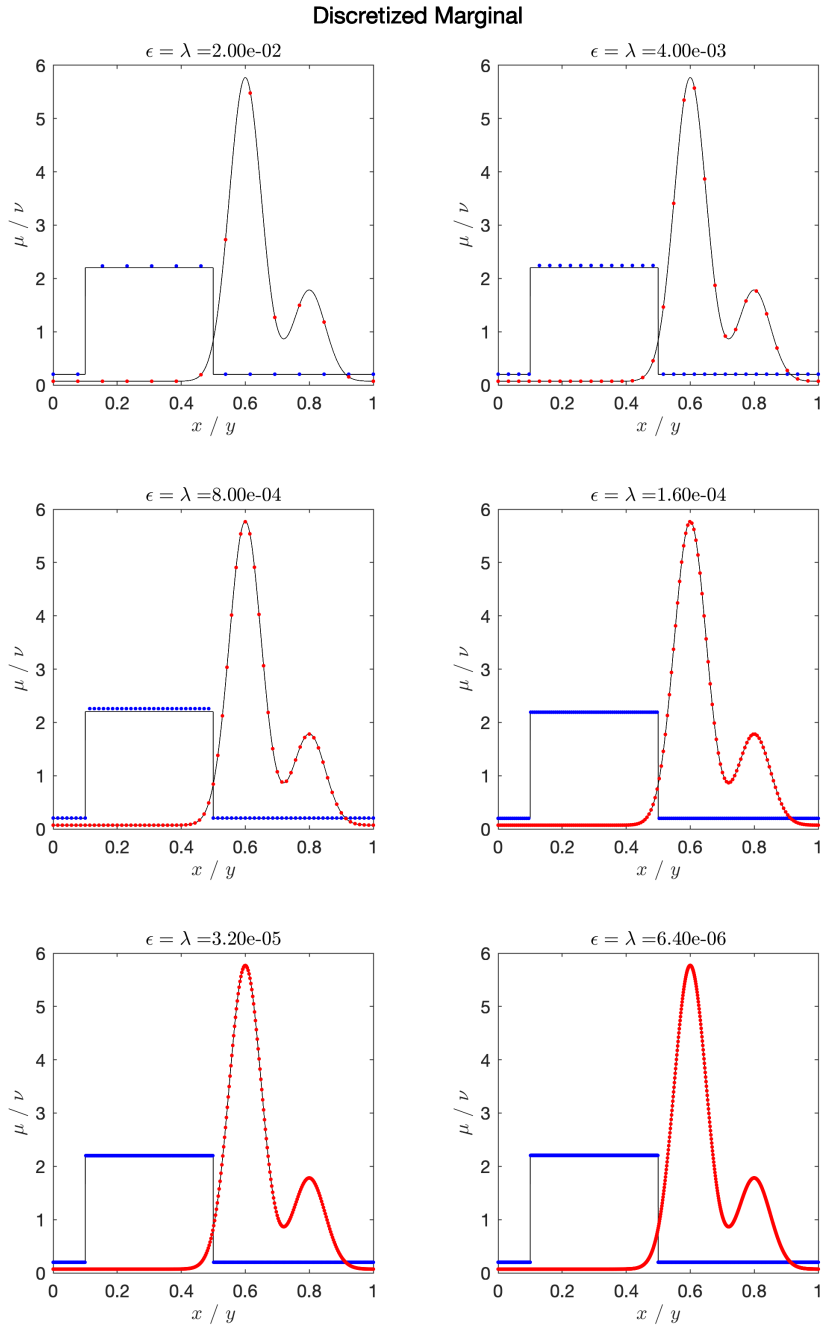

FIGURE 22. Sequence (in ε/N) of (μ_N, ν_N) .

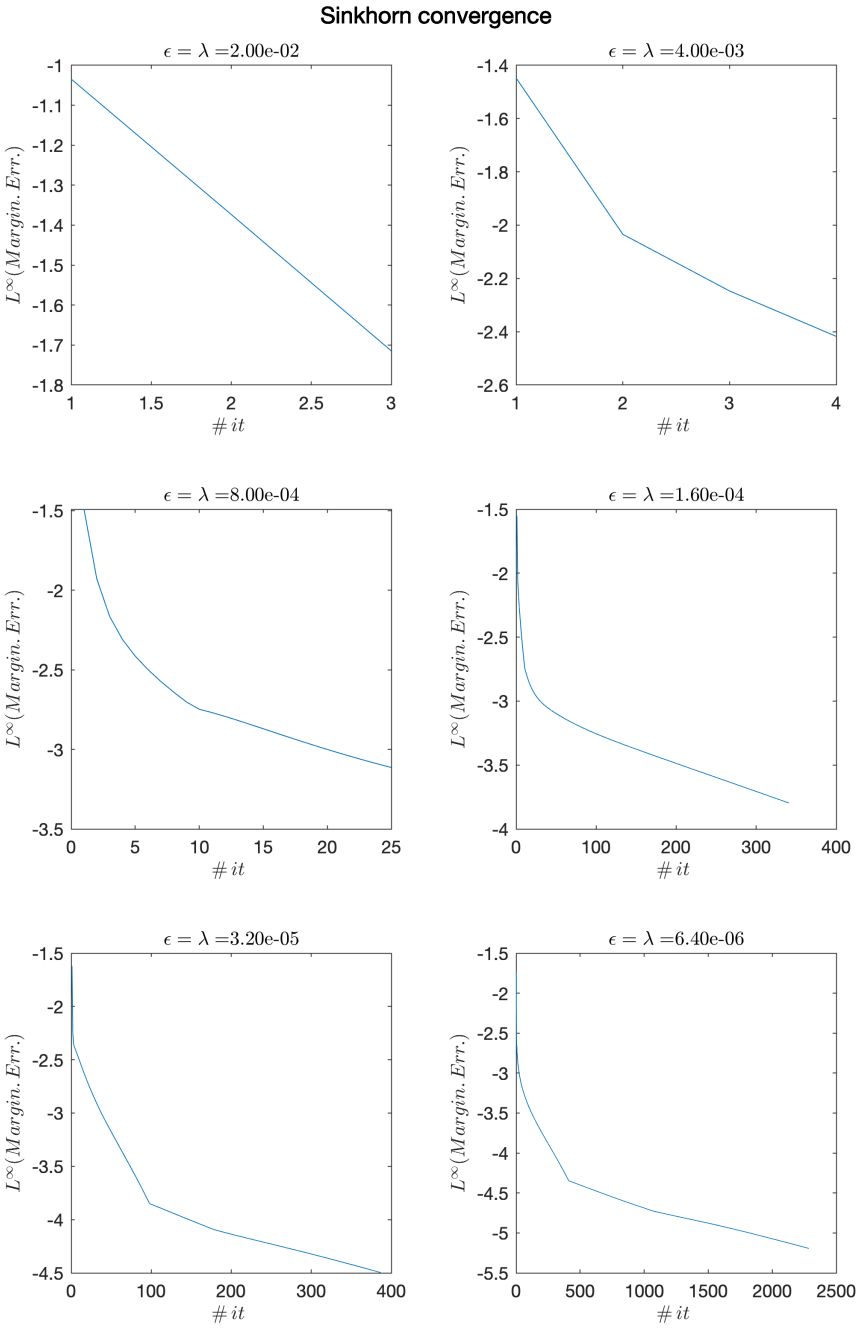

FIGURE 23. Sinkhorn Convergence. Error defined by (4.5)

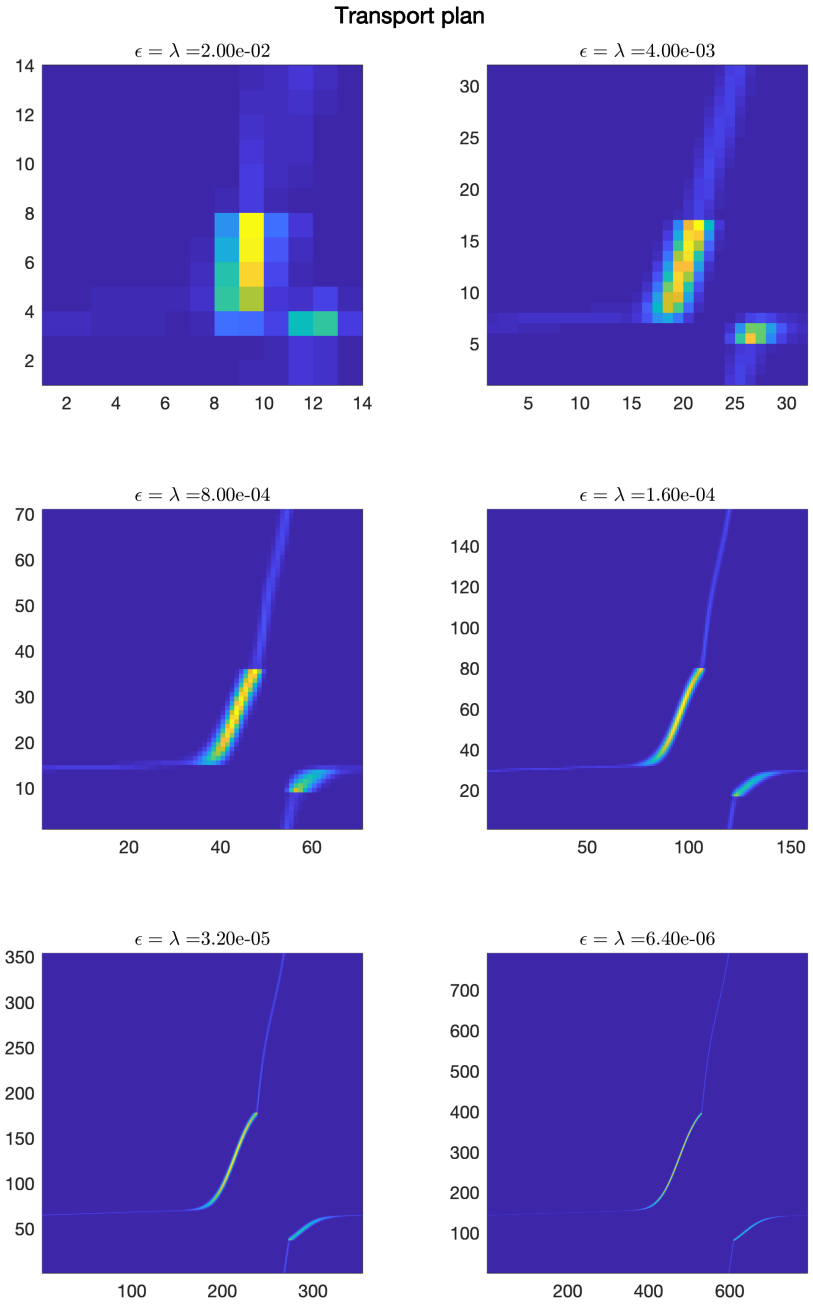

FIGURE 24. Transport Plan $\gamma^{us}_{\varepsilon,\lambda_\varepsilon}$ (see (2.6)).

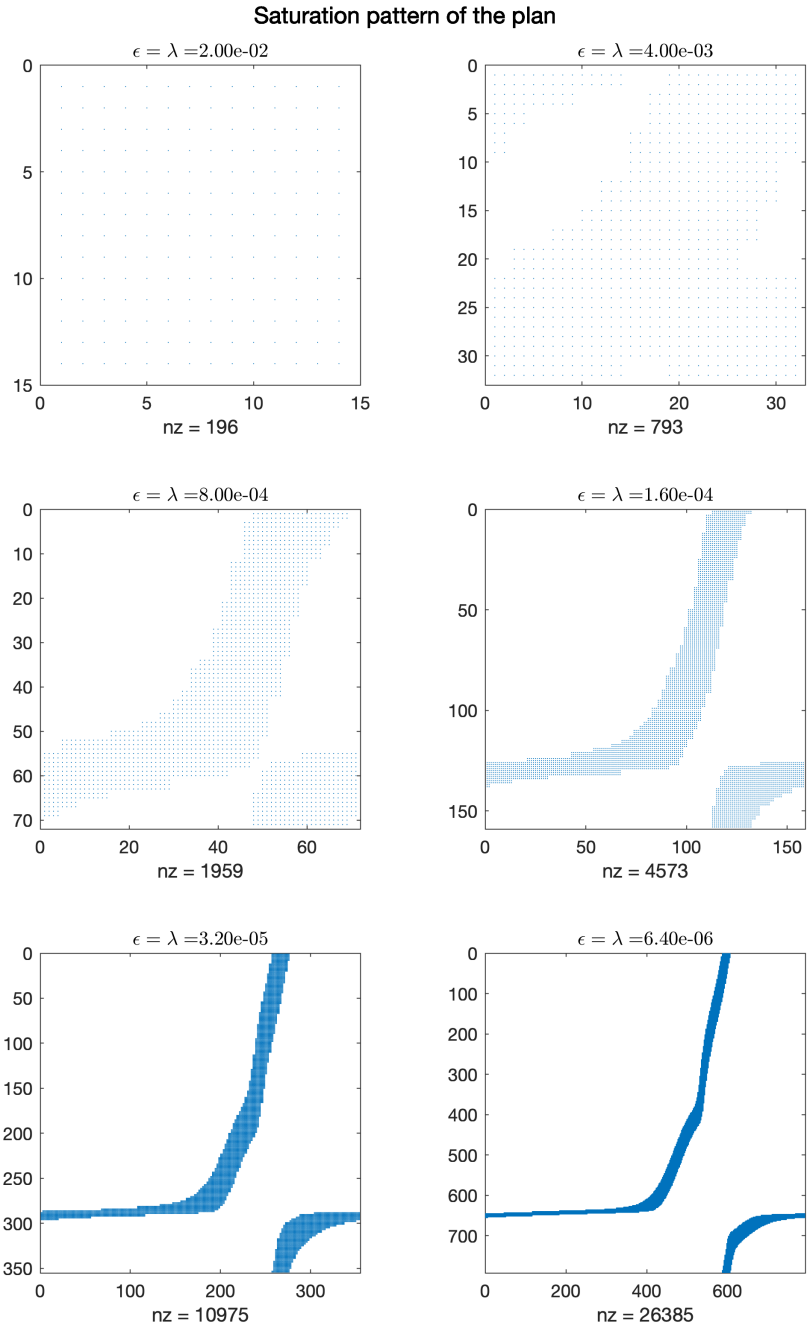

FIGURE 25. Sparsity pattern $\{x, \mathbb{S}_\varepsilon^x\}$ (see (4.3), **nz** is the number of kept (non-zero) elements

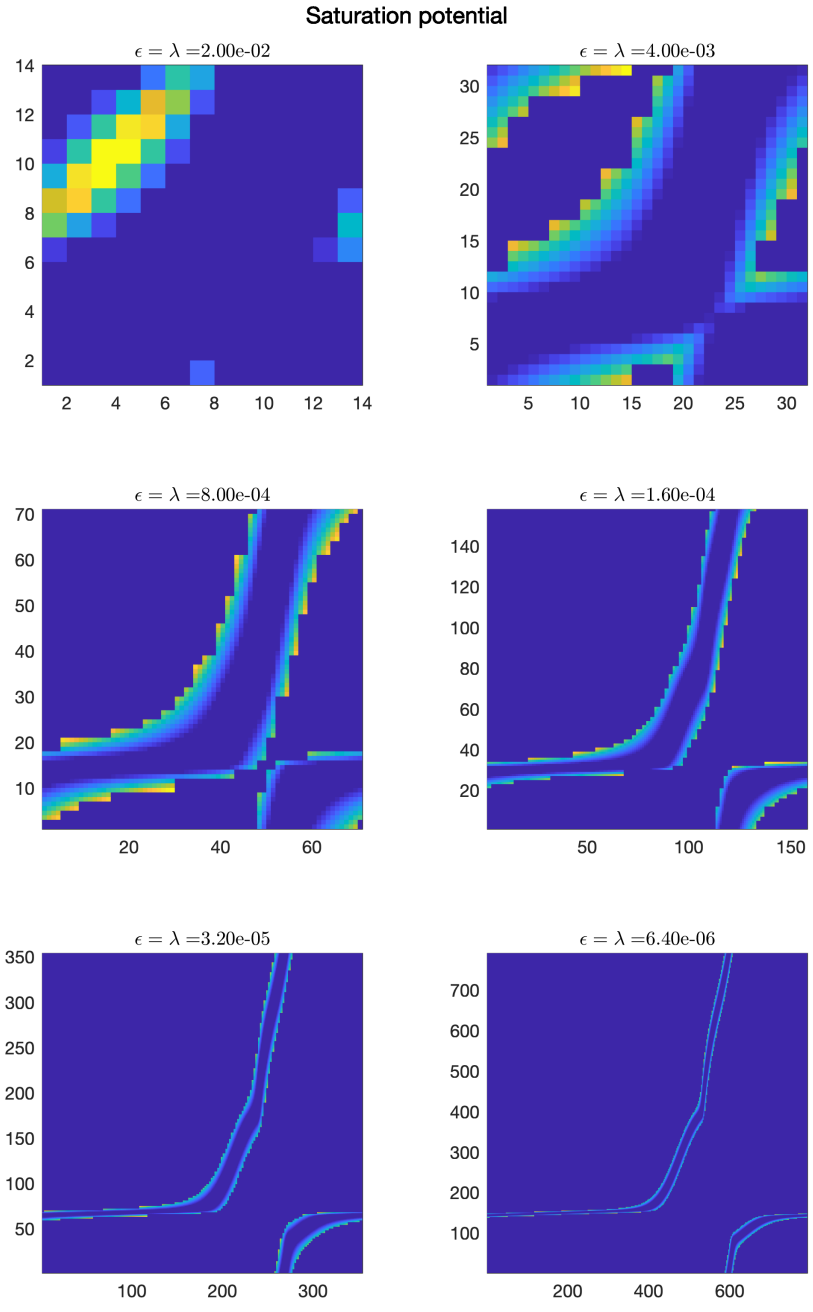

FIGURE 26. Saturation potential $h_{\varepsilon,\lambda_\varepsilon}$ on $\{x, \mathrm{S}_\varepsilon^x\}$ (set to 0 elsewhere as it is not computed but should be > 1) . Axis show the grid indices.

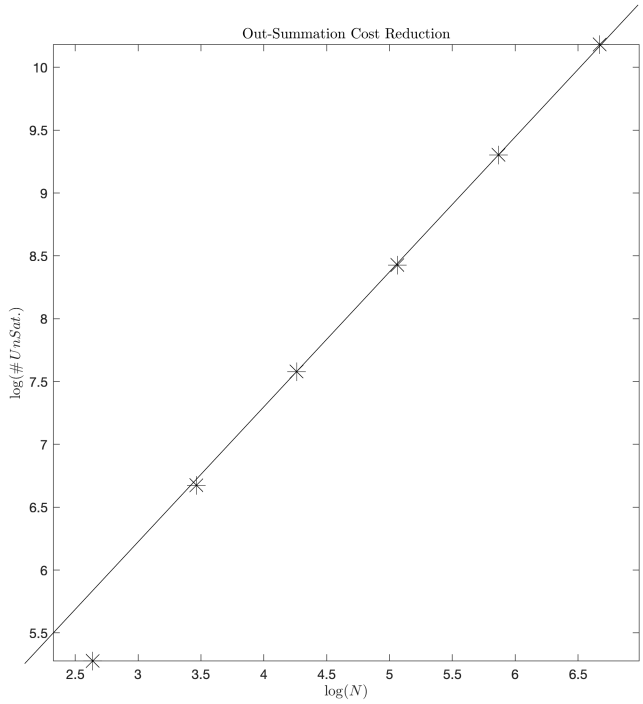

FIGURE 27. Log/Log plot of $|\{x, \mathbb{S}_\varepsilon^x\}|$ versus N plus a linear fit.

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

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: lucas.martinet@inria.fr