Computational Optimization and Applications manuscript No.
(will be inserted by the editor)

Numerical enclosures of the optimal cost of the Kantorovitch's mass transportation problem.

Nicolas Delanoue · Mehdi Lhommeau · Philippe Lucidarme

Received: November 4th 2014 / Accepted: date

Abstract The problem of optimal transportation was formalized by the French mathematician Gaspard Monge in 1781. Since Kantorovitch, this (generalized) problem is formulated with measure theory. Based on Interval Arithmetic, we propose a guaranteed discretization of the Kantorovitch's mass transportation problem. Our discretization is spatial: supports of the two mass densities are partitioned into finite families. The problem is relaxed to a finite dimensional linear programming problem whose optimum is a lower bound to the optimum of the initial one. Based on Kantorovitch duality and Interval Arithmetic, a method to obtain an upper bound to the optimum is also provided. Preliminary results show that good approximations are obtained.

Keywords Optimal Transportation · Interval Arithmetic · Continuous programming · Optimization

1 Introduction

Optimal Transportation is a mathematical research topic which started with Monge theory "des remblais et déblais" in 1781. In the 40's, Kantorovitch [11] gave the modern formulation of this problem. This problem is to minimize the transport cost between two mass densities μ and ν. Without loss of generality,

Nicolas Delanoue
LARIS, 62 avenue Notre Dame du Lac, 49000 Angers, France.
E-mail: nicolas.delanoue@univ-angers.fr

Mehdi Lhommeau
LARIS, 62 avenue Notre Dame du Lac, 49000 Angers, France.
E-mail: mehdi.lhommeau@univ-angers.fr

Philippe Lucidarme
LARIS, 62 avenue Notre Dame du Lac, 49000 Angers, France.
E-mail: philippe.lucidarme@univ-angers.fr

the total mass to be moved can be supposed to be equal to 1. Therefore, the transportation problem is stated, in modern literature, as follow:

$$\mathcal{T}(\mu,\nu) = \min_{\pi \in \Gamma(\mu,\nu)} \int_{X\times Y} c(x,y) d\pi(x,y), \tag{1}$$

where $\Gamma(\mu,\nu)$ denotes the collection of all probability measures on $X \times Y$ with marginals μ on X and ν on Y. This problem is an infinite dimensional linear programming problem with convex constraints. Following the seminal discoveries of Brenier in the 90's [5,6], Optimal Transportation has received renewed attention from mathematical analyst and the Fields Medal awarded in 2010 to C. Villani [28] (see also [9] for a survey).

Our main contribution is to provide a bounded approximation of optimum $\mathcal{T}$. This approximation is given in a guaranteed way. That means, the approximation is such that:

$$\underline{\mathcal{T}} \leq \mathcal{T} \leq \overline{\mathcal{T}},$$

where $\underline{\mathcal{T}}$ (resp. $\overline{\mathcal{T}}$) is the lower bound (resp. upper bound) of optimum $\mathcal{T}$. The main idea to obtain this guaranteed approximation can be sketched, with some simplifications, as follows. The lower bound $\underline{\mathcal{T}}$ is obtained by a spatial discretization of μ and ν, the integral (1) is then replaced by a sum. The proof of these results is presented in Section 3. Needed enclosures on c, μ and ν can be computed thanks to Interval Arithmetic which is presented in Section 2. For the upper bound $\overline{\mathcal{T}}$, this is the Kantorovitch duality which is carried. This part is detailed in Section 4. In each of these sections, an example illustrates our approach. Our approach is quite similar to [23] but their method supposes to be able to compute exactly minimum and maximum values of the function c over subsets of $X \times Y$.

Our finite dimensional relaxation scheme is different to the one presented in [25,26]. In [25], measures are approximated by discrete measures (weighted sum of Dirac). As in [25], our method generates a sequence of finite-dimensional linear programs such that the optimum $\mathcal{T}$ is the limit of the optimal values of these programs. Note also that there exists an approach based on the gradient descent method [8,7].

Otherwise, our contribution can be compared with obtained results by applying the methodology developed by J.B Lasserre [13] (See Proposition 7.7 p. 177). Indeed, his approach based on moments creates a sequence of semidefinite programs. The computed sequence $\{\mathcal{T}_i\}_{i\in\mathbb{N}}$ of optimums of these semidefinite programs is nonincreasing and converges to $\mathcal{T}$. The approximation scheme based on moments can be seen as a spectral decomposition whereas our approach is spatial. Moreover, our method is not limited to cost functions c that are polynomial and to measures with basic semi-algebraic support. Very few other authors have solved numerically the mass transfer problem, we can cite Anderson [24], Benamou [4] and Mérigot [17].

To our knowledge, no other method is able to generate guaranteed bounds of $\mathcal{T}$ when c is not supposed to be polynomial.

2 Interval Arithmetic

This section introduces notations and definitions related to interval analysis. An interval $[\underline{x}, \overline{x}]$ is a set which can be written as $\{x \in \mathbb{R}^n, \underline{x} \leq x \leq \overline{x}\}$ with $\underline{x}$ and $\overline{x}$ in $\mathbb{R}^n$. Here the relation $\leq$ has to be understood component-wise. Note that this definition implies that intervals are bounded. The set intervals is usually denoted by $\mathbb{IR}^n$.

Definition 1 A function $[f] : \mathbb{IR}^n \to \mathbb{IR}^m$ is said to be an inclusion function of $f : \mathbb{R}^n \to \mathbb{R}^m$ if $\forall [x] \in \mathbb{IR}^n, f([x]) \subset [f]([x])$ (where $f([x]) = \{f(x) | x \in [x]\}$).

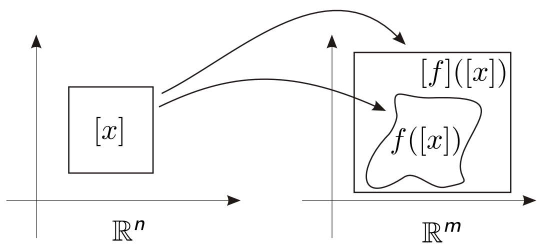

Fig. 1 Illustration of inclusion function.

Interval Arithmetic [14] provides an effective method to build inclusion functions. In [18], Neumaier proves that it is always possible to find an inclusion function $[f]$ when f is defined by an arithmetical expression. This possibility to enclose the image of an interval $[x]$ under f is powerful. Indeed, let us suppose that $0 \notin [f]([x])$, one can conclude that $\forall x \in [x], f(x) \neq 0$. Since Moore's works [14] [15] that introduced Interval Arithmetic, many algorithms have been developed in different areas, for example in global optimization [21], non-linear dynamical systems [16] ...

As interval analysis provides rigorous methods, these algorithms can prove mathematical assertion. For instance, in 2003, Hales launched the "Flyspeck project" ("Formal Proof of Kepler") in an attempt to use computers to automatically verify every step of the proof (partially based on interval analysis) of the Kepler's conjecture. Another important example is a generalization of the Newton method called Interval Newton method. This method can be applied to find all zeros of a given differentiable map $f : \mathbb{R}^n \to \mathbb{R}^n$. The interval Newton method creates a sequence of intervals containing zeros of f and has very interesting properties: combined with Brouwer fixed point theorem, it can prove existence and uniqueness of a zero of f [3], [19]. About dynamical systems, Warwick Tucker proved that the Lorentz equations support a strange attractor with an approach partially based on Interval Arithmetic [22].

In this paper, Interval Analysis is used to generate bounds of the cost function c in (1). Moreover, it is also possible to use this Arithmetic to compute guaranteed enclosures for definite integrals. In Subsection 2.1, Interval Arithmetic is clearly defined. Subsection 2.2 provides a method, based on Interval Arithmetic, to compute bounds of a definite integral.

2.1 Guaranteed bounds with Interval Arithmetic

Definition 2 (Interval Arithmetic) Let us consider two real intervals $[x] = [\underline{x}, \overline{x}]$ and $[y] = [\underline{y}, \overline{y}]$, *i.e.* $[x], [y] \in \mathbb{IR}$. The four basic interval operations are defined, as in [12], by:

$$\begin{array}{l} [x] + [y] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}], \\ [x] - [y] = [\underline{x} - \overline{y}, \overline{x} - \underline{y}], \\ [x] \times [y] = [\min\{\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}\}\ ,\ \max\{\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}\}], \\ [x] \div [y] = [x] \times \left[\frac{1}{\overline{y}}, \frac{1}{\underline{y}}\right],\ \text{if } \underline{y}\overline{y} > 0. \end{array} \tag{2}$$

Proposition 1 *The four basic interval operations (2) are inclusion function of addition, subtraction, multiplication and division defined on reals.*

Proof Trivial. □

Interval Arithmetic also define negative, square and power functions as follows:

$$-[x] = [-\overline{x}, -\underline{x}], \tag{3}$$

$$[x]^n = \begin{cases} [1] & \text{if n=0}, \\ [\underline{x}^n, \overline{x}^n] & \text{if n is odd or } \underline{x} \geq 0, \\ [\overline{x}^n, \underline{x}^n] & \text{if n is even and } \overline{x} \leq 0, \\ [0, \max\{\underline{x}^n, \overline{x}^n\}] & \text{otherwise.} \end{cases} \tag{4}$$

The inclusion functions are stable by composition. Therefore, as soon a function is given by its expression and we have inclusion function for their atoms, one can automatically generate an inclusion function.

Proposition 2 *If f and g are functions with inclusion function $[f]$ and $[g]$, then $[f] \circ [g]$ is an inclusion function of $f \circ g$.*

Proof Let $[x] \in \mathbb{IR}^n$, then $g([x]) \subset [g]([x])$. So $f \circ g([x]) \subset [f] \circ [g]([x])$. □

Example 1 Let us assume that $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \frac{3}{2}(1 - 2x + x^2).$$

The function $[f] : \mathbb{IR} \to \mathbb{IR}$ defined by

$$[f]([\underline{x}, \overline{x}]) = \left[\frac{3}{2}, \frac{3}{2}\right] \times \left([1, 1] - [2, 2] \times [\underline{x}, \overline{x}] + [\underline{x}, \overline{x}]^2\right),$$

is an inclusion function for f.

Moreover, if $\underline{x}$ is supposed to be non negative,

$$[f]([\underline{x}, \overline{x}]) = \left[\frac{3}{2} - 3\overline{x} + \frac{3}{2}\underline{x}^2\ ,\ \frac{3}{2} - 3\underline{x} + \frac{3}{2}\overline{x}^2\right].$$

Remark 1 Please note that the real numbers $\frac{3}{2}$, 1 and 2 were converted to intervals of width zero and all real operation were changed to interval ones. In practice, using floating point arithmetic, if a real number α is not representable, then this real number is converted to an interval $[\underline{\alpha}, \overline{\alpha}]$ with representable bounds containing α. In a way, one can avoid the rounded errors of the floating point arithmetic, and finally obtain guaranteed bounds.

2.2 Guaranteed approximation to a definite integral

In this subsection, we propose a method to rigorously enclose a definite integral. The approach and the convergence are similar to the classical rectangle method.

Definition 3 Let X be a subset of $\mathbb{R}^n$ and μ be a probability measure. A finite collection $\{X_i\}_i$ of subsets of $\mathbb{R}^n$ is said to be a $\mu-$paving if $X \subset \cup_i X_i$ and $i \neq j \Rightarrow \mu(X_i \cap X_j) = 0$.

Proposition 3 *Let $f : X \to \mathbb{R}$ be in $L^1(\lambda)$ and $\{X_i\}_{i \in I}$ a $\lambda-$paving of X, then*

$$\int_X f(x) d\lambda(x) \in \sum_{i \in I} [f](X_i)\ \lambda(X_i), \tag{5}$$

where λ denotes the Lebesgue measure.

Proof Let X_i be an element of the paving. Since $[f]$ is an inclusion function for f, we can write

$$\forall x \in X_i, \underline{f}(X_i) \leq f(x) \leq \overline{f}(X_i)$$

with $[\underline{f}(X_i), \overline{f}(X_i)] = [f](X_i)$. Therefore, one has

$$\int_{X_i} \underline{f}(X_i) d\lambda(x) \leq \int_{X_i} f(x) d\lambda(x) \leq \int_{X_i} \overline{f}(X_i) d\lambda(x).$$

Equivalently,

$$\underline{f}(X_i) \int_{X_i} d\lambda(x) \leq \int_{X_i} f(x) d\lambda(x) \leq \overline{f}(X_i) \int_{X_i} d\lambda(x).$$

Since $\{X_i\}_i$ is an λ-paving, we end up with

$$\sum_{i \in I} \underline{f}(X_i)\ \lambda(X_i) \leq \int_X f(x) d\lambda(x) \leq \sum_{i \in I} \overline{f}(X_i)\ \lambda(X_i).$$

□

Figure 2 shows an example of function from $\mathbb{R}$ to $\mathbb{R}$. The definite integral, illustrated by the algebraic area, is between sum of areas of the rectangles on left and on right.

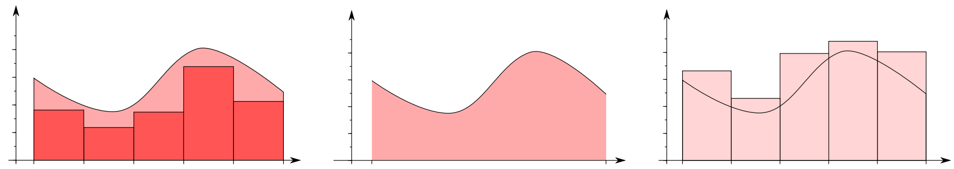

Fig. 2 Illustration of Proposition 3. The definite integral $\int_a^b f(x)dx$ (center) is between $\sum_{i\in I} \underline{f}(X_i)\ \lambda(X_i)$ (left) and $\sum_{i\in I} \overline{f}(X_i)\ \lambda(X_i)$ (right).

Example 2 Let us consider the same function f and inclusion function $[f]$ as in example 1 and define a regular paving $\{X_i\}$ defined by $X_i = [\frac{i-1}{n}, \frac{i}{n}]$. The paving $\{X_i\}$ is a $\lambda-$paving of $[0, 1]$ since $\forall i \in \{1, \dots, n\}, \lambda(\{\frac{i}{n}\}) = 0$. According to Proposition 3, one can conclude that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{3}{2} - 3\frac{i}{n} + \frac{3}{2}\left(\frac{i-1}{n}\right)^2 \leq \int_0^1 f(x)d\lambda(x) \leq \frac{1}{n}\sum_{i=1}^{n} \frac{3}{2} - 3\frac{i-1}{n} + \frac{3}{2}\left(\frac{i}{n}\right)^2.$$

Proposition 4 proves that the method given by Proposition 3 is convergent. To speak about convergence, one needs to topologize the set of intervals. This can be done by the distance d defined by

$$[x], [y] \in \mathbb{IR}^n, d([x], [y]) = \max\{\|\underline{x} - \underline{y}\|, \|\overline{x} - \overline{y}\|\}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^n$. This distance is equivalent to the classical Hausdorff distance defined on compact subsets of $\mathbb{R}^n$. With this topology, the notion of *continuous inclusion function* is now well defined. Finally, with $[x]$ an element of $\mathbb{IR}^n$, let us also define the diameter of $[x]$ by $\Delta([x]) = \max\{\|x - y\| \mid x, y \in [x]\}$.

Proposition 4 (Convergence) *Let X be an element of $\mathbb{IR}^n$, and let $[f]$ be a continuous inclusion function of f satisfying $\forall x \in X, [f](\{x\}) = \{f(x)\}$. Let $\{X_i\}_{i\in I}$ be a $\lambda-$paving of X and $h = \max_{i\in I}\{\Delta(X_i)\}$ then*

$$\lim_{h\to 0} \sum_{i\in I} [f](X_i)\ \lambda(X_i) = \int_X f(x)d\lambda(x).$$

Proof Since X is compact and $[f]$ continuous, by Heine-Cantor Theorem $[f]$ is uniformly continuous. That is to say:

$$\forall\varepsilon > 0\ \exists\delta > 0\ \forall A, B \subset X, d(A, B) < \delta \Rightarrow d([f](A), [f](B)) < \varepsilon.$$

In particular, with $B = \{a\} \subset A$ one has

$$\forall\varepsilon > 0\ \exists\delta > 0\ \forall A \subset X, d(A, \{a\}) < \delta \Rightarrow d([f](A), f(\{a\})) < \varepsilon.$$

Therefore,

$$\forall\varepsilon > 0\ \exists\delta > 0\ \forall A \subset X, \Delta(A) < \delta \Rightarrow \Delta([f](A)) < \varepsilon.$$

Let $\{X_i\}$ be a λ-paving of X such that $\forall i \in I, \Delta(X_i) < h$,

$$\forall \varepsilon > 0\ \exists \delta > 0, h < \delta \Rightarrow \sum_{i\in I}(\overline{f}(X_i) - \underline{f}(X_i))\lambda(X_i) < \varepsilon\lambda(X).$$

That is to say, $\sum_{i\in I}(\overline{f}(X_i) - \underline{f}(X_i))\lambda(X_i)$ converges to 0 as h tends to 0. In other words, the two bounds of enclosure (5) converge to the same value. We end up with

$$\lim_{h\to 0}\sum_{i\in I}[f](X_i)\ \lambda(X_i) = \int_X f(x)d\lambda(x).$$

□

3 Lower bound of the optimal value

This section contains one of the two main contributions of this article. The following theorem lets us generate a finite dimensional linear programming problem where the optimal value is a lower bound for the optimal value of (1). In Subsection 3.1, the theorem and its proof are given. Subsection 3.2 illustrates this theorem with an example. Finally, assuming continuous hypothesis about inclusion function and compactness, the method can be proved to be convergent. In other words, one can generate a sequence of lower bounds which converges to the optimal value. This result is provided in Subsection 3.3.

3.1 Finite dimensional relaxation

Theorem 1 (Relaxation) *Let μ and ν (with support X and Y) be absolutely continuous measures with respect to Lebesgue measure λ. If $\{X_i\}_{1\leq i\leq n}$ and $\{Y_j\}_{1\leq j\leq m}$ be $\lambda-$pavings of X and Y. Suppose that $\mu(X_i) \in [\underline{\mu}_i, \overline{\mu}_i]$, $\nu(Y_j) \in [\underline{\nu}_j, \overline{\nu}_j]$, and there exist reals $c_{ij} \in \mathbb{R}$ such that $\forall x, y \in X_i \times Y_j, \underline{c}_{ij} \leq c(x, y)$ and define $\underline{\mathcal{T}}$ by*

$$\begin{aligned}
\underline{\mathcal{T}} = \inf_{\pi_{ij}\in\mathbb{R}^n\otimes\mathbb{R}^m} \quad & \sum_{i,j}\underline{c}_{ij}\pi_{ij} \\
\text{subject to} \quad & \forall i,\ \underline{\mu}_i \leq \sum_j \pi_{ij} \leq \overline{\mu}_i, \\
& \forall j,\ \underline{\nu}_j \leq \sum_i \pi_{ij} \leq \overline{\nu}_j, \\
& \forall i, \forall j,\ \pi_{ij} \geq 0,
\end{aligned} \tag{6}$$

then $\underline{\mathcal{T}} \leq \mathcal{T}(\mu, \nu)$.

Remark 2 The finite dimensional linear programming problem (6) has $n \times m$ variables which are denoted by π_{if}. Indeed, the tensor product $\mathbb{R}^n \otimes \mathbb{R}^m$ is isomorphic to $\mathbb{R}^{n\times m}$.

Proof The first part of the proof is to prove that the cost function (1) is greater than the cost of (6). Let π be a transference plan from μ to ν then, the collection $\{X_i \times Y_j\}_{\{(i,j)\in I\times J\}}$ is a π-paving. Therefore, one has:

$$\sum_{i,j} \int_{X_i\times Y_j} c(x,y)d\pi(x,y) = \int_{X\times Y} c(x,y)d\pi(x,y).$$

Since $\forall x, y \in X_i \times Y_j, \underline{c}_{ij} \leq c(x,y)$, one can conclude that

$$\underline{c}_{ij}\pi(X_i \times Y_j) \leq \int_{X_i\times Y_j} c(x,y)d\pi(x,y).$$

Therefore:

$$\sum_{ij} \underline{c}_{ij}\pi_{ij} \leq \int_{X\times Y} c(x,y)d\pi(x,y)$$

with $\pi_{ij} = \pi(X_i \times Y_j)$.

The second part of the proof consists to prove that constraints of (1) implies constraints of (6). Let π be a probability measure on $X \times Y$ with μ and ν as marginals. For each i, one has

$$\sum_j \pi(X_i \times Y_j) = \mu(X_i).$$

Since $\mu(X_i)$ is supposed to belong to $[\underline{\mu}_i, \overline{\mu}_i]$, one has:

$$\forall i,\ \underline{\mu}_i \leq \sum_j \pi_{ij} \leq \overline{\mu}_i$$

with $\pi_{ij} = \pi(X_i \times Y_j)$. Finally, as π is a probability measure over $X \times Y$, then $\pi_{ij} \geq 0$. □

3.2 An academic example

Let us consider the two probability measures μ and ν with support $[0,1]$ defined by $\mu = \mathrm{d}x$ and $\nu = \frac{3}{2}(1-y^2)\mathrm{d}y$. Those two probability measures are represented by Figure 3. For this optimization problem, we consider that c is the square of the Euclidean distance $c(x,y) = (x-y)^2$.

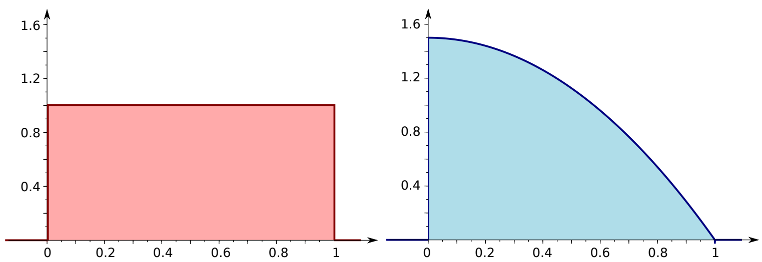

Fig. 3 Probability measures μ and ν.

Let us divide the interval $[0,1]$ with a regular paving with $n = m = 6$ elements

$$X_i = \left[\frac{i-1}{n}, \frac{i}{n}\right],\ Y_j = \left[\frac{j-1}{m}, \frac{j}{m}\right]$$

with $i, j \in \{1, \ldots, n\}$. Our approach generates the following finite dimensional linear programming:

$$\begin{aligned} \underline{\mathcal{T}} = \min_{\pi_{ij} \in \mathbb{R}^n \otimes \mathbb{R}^m} \quad & \sum_{i,j} \underline{c}_{ij} \pi_{ij} \\ \text{subject to} \quad & \forall i,\ \underline{\mu}_i \leq \sum_j \pi_{ij} \leq \overline{\mu}_i, \\ & \forall j,\ \underline{\nu}_j \leq \sum_i \pi_{ij} \leq \overline{\nu}_j, \\ & \forall i, \forall j,\ \pi_{ij} \geq 0, \end{aligned} \tag{7}$$

with the following data:

$$\left(\underline{c}_{ij}\right)_{1 \leq i,j \leq n} = \begin{pmatrix} 0.000 & 0.000 & 0.027 & 0.111 & 0.250 & 0.444 \\ 0.000 & 0.000 & 0.000 & 0.027 & 0.111 & 0.250 \\ 0.027 & 0.000 & 0.000 & 0.000 & 0.027 & 0.111 \\ 0.111 & 0.027 & 0.000 & 0.000 & 0.000 & 0.027 \\ 0.250 & 0.111 & 0.027 & 0.000 & 0.000 & 0.000 \\ 0.444 & 0.250 & 0.111 & 0.027 & 0.000 & 0.000 \end{pmatrix}.$$

$$\begin{aligned} \underline{\mu} &= (0.166, 0.166, 0.166, 0.166, 0.166, 0.166), \\ \overline{\mu} &= (0.167, 0.167, 0.167, 0.167, 0.167, 0.167), \\ \underline{\nu} &= (0.247, 0.233, 0.205, 0.164, 0.108, 0.039), \\ \overline{\nu} &= (0.248, 0.234, 0.207, 0.165, 0.109, 0.040). \end{aligned}$$

Lower bounds $\underline{c}_{ij}$ have been computed using interval arithmetic as shown in Subsection 2.1. Data $\underline{\nu}, \underline{\mu}, \overline{\nu}$ and $\overline{\mu}$ have been generated using the method to compute a guaranteed approximation to a definite integral as presented in Subsection 2.2.

This linear programming problem has been solved using the GLPK solver [2] with library gmp [1] to avoid rounded errors coming from floating-point number representation. According to Theorem 1, we have a proof that the solution to the infinite dimensional linear problem considering in this example satisfies

$$1.085 \times 10^{-3} \leq \mathcal{T}.$$

The same approach can be used to generate better lower bounds to the problem by taking bigger n. Figure 4 shows lower bounds computed for different values of n.

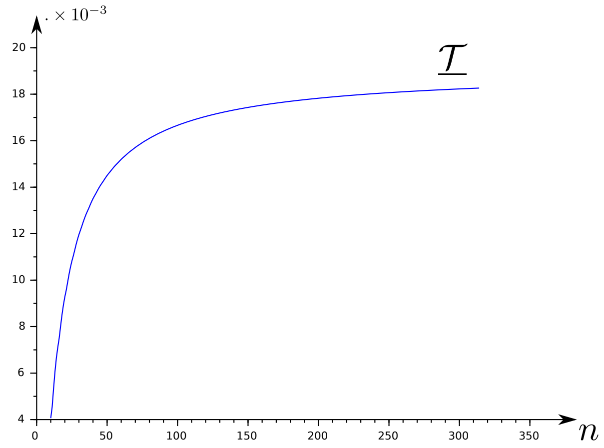

Fig. 4 Computed lower bounds on $\mathcal{T}$ with respect to n (with $m = n$).

For this example, regular pavings are chosen with the same cardinality, i.e. $m = n$. Note that generated linear problems are of dimension n^2 and composed with $2n$ linear constraints.

3.3 Convergence

Proposition 5 (Convergence) *Let X, Y be elements of $\mathbb{IR}^k$, $\mu = f d\lambda$, $\nu = g d\lambda$ be non negative measures with supports in X and Y. Let us suppose that $[f]$, $[g]$, $[c] = [\underline{c}, \overline{c}]$ are continuous inclusion functions of f, g, c satisfying $\forall x \in X \forall y \in Y, [f](\{x\}) = \{f(x)\}, [g](\{y\}) = \{g(y)\}$ and $[c](\{(x,y)\}) = \{c(x,y)\}$. Let $\{X_i\}_{i\in I}$ and $\{Y_j\}_{j\in J}$ be a $\lambda-$pavings of X and Y. Let us denote by $h = \max\{\max\{\Delta(X_i) \mid i\}, \max\{\Delta(Y_j) \mid j\}\}$, if $\underline{\mathcal{T}}$ is the optimal solution of (6) then*

$$\lim_{h\to 0} \underline{\mathcal{T}} = \mathcal{T}$$

Proof Let us denote by $\tilde{\mathcal{T}}$ the following optimal value:

$$\begin{aligned} \tilde{\mathcal{T}} = \inf_{\tilde{\pi}_{ij} \in \mathbb{R}^n \otimes \mathbb{R}^m} \quad & \sum_{i,j} \tilde{c}_{ij} \tilde{\pi}_{ij} \\ \text{subject to} \quad & \forall i,\ \sum_j \tilde{\pi}_{ij} = \mu(X_i), \\ & \forall j,\ \sum_i \tilde{\pi}_{ij} = \nu(Y_j), \\ & \forall i, \forall j,\ \tilde{\pi}_{ij} \geq 0, \end{aligned} \tag{8}$$

where $\tilde{c}_{ij} = \sup\{c(x,y) \mid (x,y) \in X_i \times Y_j\}$.

The main idea of the proof consists to demonstrate the following two assertions. First, inequalities $\underline{\mathcal{T}} \le \mathcal{T} \le \tilde{\mathcal{T}}$ are true and, second, the quantity $\tilde{\mathcal{T}} - \underline{\mathcal{T}}$ converges to 0 as h tends to 0.

Inequality $\underline{\mathcal{T}} \le \mathcal{T}$ was proven in Theorem 1. Let us now show the inequality $\mathcal{T} \le \tilde{\mathcal{T}}$. Let $\tilde{\pi}^* = (\tilde{\pi}^*_{ij})$ be an optimal solution to problem (8). From $\tilde{\pi}^*$, let us define a measure $\pi \in \Gamma(\mu,\nu)$ satisfying $\forall i \forall j, \pi(X_i \times Y_j) = \tilde{\pi}^*_{i,j}$. One has

$$\int_{X\times Y} c(x,y) d\pi(x,y) = \sum_{i,j} \int_{X_i \times Y_j} c(x,y) d\pi(x,y).$$

Since $\forall (x,y) \in X_i \times Y_j, c(x,y) \le \tilde{c}_{ij}$, one can conclude:

$$\int_{X\times Y} c(x,y) d\pi(x,y) \le \sum_{i,j} \int_{X_i \times Y_j} \tilde{c}_{ij} d\pi(x,y) = \sum_{i,j} \tilde{c}_{ij} \tilde{\pi}^*_{i,j} = \tilde{\mathcal{T}}$$

with $\pi \in \Gamma(\mu,\nu)$. In other words, from any optimal solution of (8), it is possible to create a feasible solution of (1). Moreover, this feasible solution to (1) is smaller than $\tilde{\mathcal{T}}$. In particular, with $\mathcal{T}$ the optimal value of (1), one can conclude $\mathcal{T} \le \tilde{\mathcal{T}}$.

Let us now prove that $\tilde{\mathcal{T}} - \underline{\mathcal{T}}$ converges to 0 as h tends to 0. The main idea is to prove that both $\tilde{\mathcal{T}} - \underset{\sim}{\mathcal{T}}$ and $\underset{\sim}{\mathcal{T}} - \underline{\mathcal{T}}$ converge to 0 as h tends to 0 where $\underset{\sim}{\mathcal{T}}$ is the optimal value to the following finite dimensional linear problem:

$$\begin{aligned} \underset{\sim}{\mathcal{T}} = \inf_{\pi_{ij} \in \mathbb{R}^n \otimes \mathbb{R}^m} \quad & \sum_{i,j} \underline{c}_{ij} \pi_{ij} \\ \text{subject to} \quad & \forall i,\ \sum_j \pi_{ij} = \mu(X_i), \\ & \forall j,\ \sum_i \pi_{ij} = \nu(Y_j), \\ & \forall i, \forall j,\ \pi_{ij} \ge 0. \end{aligned} \tag{9}$$

One has:

$$\underline{\mathcal{T}} \le \underset{\sim}{\mathcal{T}} \le \mathcal{T} \le \tilde{\mathcal{T}}.$$

Following the same idea of proof of Proposition 4, by Heine-Cantor Theorem, since $[c]$ is uniformly continuous, thus

$$\forall \varepsilon > 0\ \exists \delta > 0, h < \delta \Rightarrow (\tilde{c}_{ij} - \underline{c}_{ij}) < \varepsilon.$$

Therefore, $\forall (\pi_{ij}) \in \mathbb{R}^n \otimes \mathbb{R}^m$ satisfying

$$\begin{cases} \forall i,\ \sum_j \pi_{ij} = \mu(X_i), \\ \forall j,\ \sum_i \pi_{ij} = \nu(Y_j), \end{cases}$$

one has

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \sum_{ij} (\tilde{c}_{ij} - \underline{c}_{ij})\pi_{ij} < \varepsilon \mu(X). \tag{10}$$

In particular, with (π_{ij}) an optimal solution to (9), one can conclude

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \sum_{ij} \tilde{c}_{ij}\pi_{ij} - \sum_{ij} \underline{c}_{ij}\pi_{ij} < \varepsilon \mu(X). \tag{11}$$

i.e.

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \sum_{ij} \tilde{c}_{ij}\pi_{ij} - \underset{\sim}{\mathcal{T}} < \varepsilon \mu(X).$$

i.e.

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \sum_{ij} \tilde{c}_{ij}\pi_{ij} < \varepsilon \mu(X) + \underset{\sim}{\mathcal{T}}.$$

As $\tilde{\mathcal{T}} \leq \sum_{ij} \tilde{c}_{ij}\pi_{ij}$, one has

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \tilde{\mathcal{T}} < \varepsilon \mu(X) + \underset{\sim}{\mathcal{T}}.$$

Therefore,

$$\forall \varepsilon > 0 \ \exists \delta > 0, h < \delta \Rightarrow \tilde{\mathcal{T}} - \underset{\sim}{\mathcal{T}} < \varepsilon \mu(X).$$

Let us now prove that $\underset{\sim}{\mathcal{T}} - \underline{\mathcal{T}}$ converges to 0 as h tends to 0. The real numbers $\underline{\mathcal{T}}$ and $\underset{\sim}{\mathcal{T}}$ are optimal values of the two finite dimensional linear programming problems (6) and (9). Using matrix notation, we have:

$$\begin{aligned}
\underset{\sim}{\mathcal{T}} &= \inf\{\underset{\sim}{c}^T \pi \mid \quad b = A\pi = b, \quad \pi \geq 0\}, \\
\underline{\mathcal{T}} &= \inf\{\underset{\sim}{c}^T \pi \mid b - \underline{\epsilon} \leq A\pi \leq b + \overline{\epsilon}, \pi \geq 0\}
\end{aligned}$$

with

$$A = \underbrace{\left(\begin{array}{ccc}
\underbrace{1 \cdots 1}_{m} & & \\
& \ddots & \\
& & \underbrace{1 \cdots 1}_{m} \\
\hline
1 & & 1 \\
\ddots & \cdots & \ddots \\
\underbrace{\quad 1}_{m} & & \underbrace{\quad 1}_{m}
\end{array} \right)}_{n \text{ times.}}, \pi = \begin{pmatrix} \pi_{11} \\ \vdots \\ \pi_{1m} \\ \pi_{21} \\ \vdots \\ \pi_{2m} \\ \vdots \\ \pi_{m1} \\ \vdots \\ \pi_{nm} \end{pmatrix}, b = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \\ \nu_1 \\ \vdots \\ \nu_m \end{pmatrix},$$

and

$$\underline{\epsilon} = \begin{pmatrix} \mu_1 - \underline{\mu_1} \\ \vdots \\ \mu_n - \underline{\mu_n} \\ \nu_1 - \underline{\nu}_1 \\ \vdots \\ \nu_m - \underline{\nu}_m \end{pmatrix}, \overline{\epsilon} = \begin{pmatrix} \overline{\mu_1} - \mu_1 \\ \vdots \\ \overline{\mu_n} - \mu_n \\ \overline{\nu}_1 - \nu_1 \\ \vdots \\ \overline{\nu}_m - \nu_m \end{pmatrix}.$$

Therefore, those two linear programs have the same cost function c, and constraint matrix A. Since $[f]$ and $[g]$ are supposed to satisfy, $\forall x \in X, [f](\{x\}) = \{f(x)\}$ and $\forall y \in Y, [g](\{y\}) = \{g(y)\}$, according to Proposition 4, $\underline{\epsilon}$ and $\overline{\epsilon}$ can be as small as we need. So $\underline{\mathcal{T}}$ may be seen as the optimal value of a perturbed linear program of (9). Using bounds provided by [20], we can prove that $\underline{\mathcal{T}}$ converges to $\underset{\sim}{\mathcal{T}}$ as long as $\underline{\epsilon}$ and $\overline{\epsilon}$ converges to 0. □

4 Upper bound of the optimal value

In this section, Kantorovitch duality is used to propose an upper bound to the optimal value. Kantorovitch Theorem states the following:

Theorem 2 (Kantorovitch duality) *Let μ and ν be probability measures and let $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a lower semi-continous cost function. Define*

$$\begin{aligned} \mathcal{T}'(\mu,\nu) = \sup_{\varphi\in\mathcal{C}(X),\psi\in\mathcal{C}(Y)} \quad & \int_X \varphi d\mu + \int_Y \psi d\nu, \\ \textit{subject to} \quad & \varphi(x) + \psi(y) \leq c(x,y), \end{aligned}$$

then

$$\mathcal{T}(\mu,\nu) = \mathcal{T}'(\mu,\nu).$$

With $\mathcal{C}(X)$ the set of continuous real bounded functions over X.

The proof of this Theorem can be found in [27].

4.1 Finite dimensional relaxation

Theorem 3 (Upper relaxation) *Let μ and ν (with compact supports X and Y) be absolutely continuous measures with respect to Lebesgue measure. Let $\{X_i\}_i$ and $\{Y_j\}_j$ be $\lambda-$pavings of X and Y. Suppose that $\mu(X_i) \leq \overline{\mu}_i$, $\nu(Y_j) \leq \overline{\nu}_j$, and $\forall (x,y) \in X_i \times Y_j, c(x,y) \leq \overline{c}_{ij}$,*

$$\begin{aligned} \overline{\mathcal{T}} = \sup_{(\varphi_i)\in\mathbb{R}^n,(\psi_j)\in\mathbb{R}^m} \quad & \sum_i \varphi_i \overline{\mu}_i + \sum_j \psi_j \overline{\nu}_i \\ \textit{subject to} \quad & \varphi_i + \psi_j \leq \overline{c}_{ij}, \end{aligned} \tag{12}$$

then

$$\mathcal{T} \leq \overline{\mathcal{T}}.$$

Proof Let (φ, ψ) be in $\mathcal{C}(X) \times \mathcal{C}(Y)$. Since each X_i and Y_j is compact, there exist real numbers $\varphi_i = \sup_{x \in X_i} \varphi_i(x)$, $\psi_i = \sup_{y \in Y_j} \psi_i(y)$ such that

$$\int_{X_i} \varphi(x) d\mu \leq \varphi_i \; \mu(X_i) \text{ and } \int_{Y_j} \psi(x) d\nu \leq \psi_j \; \nu(Y_j).$$

As $\mu(X_i)$ is supposed to be smaller than $\overline{\mu}_i$, one has:

$$\int_X \varphi d\mu + \int_Y \psi d\nu \leq \sum_i \varphi_i \overline{\mu}_i + \sum_j \psi_j \overline{\nu}_i.$$

The second part of the proof consists to demonstrate that condition $\forall (x, y), \varphi(x) + \psi(y) \leq c(x, y)$ implies $\forall (i, j), \varphi_i + \psi_j \leq \overline{c}_{ij}$. Let us suppose that $\forall (x, y), \varphi(x) + \psi(y) \leq c(x, y)$, then

$$\forall (x, y) \in X_i \times Y_j, \varphi(x) + \psi(y) \leq c(x, y) \leq \overline{c}_{ij}.$$

As X_i and Y_j are supposed to be compact, there exists $(\xi, \zeta) \in X_i \times Y_j$ such that

$$\varphi(\xi) = \sup_{x \in X_i} \varphi(x) = \varphi_i, \text{ and } \psi(\zeta) = \sup_{y \in Y_j} \psi(x) = \psi_j.$$

We can write

$$\varphi_i + \psi_j \leq \overline{c}_{ij}.$$

□

4.2 An academic example

Here, we consider the same example as in Subsection 3.2. Let us consider a regular paving of the unit interval $[0, 1]$. Theorem 3 generates, with $m = n = 6$, the following finite dimensional linear programming problem:

$$\begin{array}{lll} \overline{\mathcal{T}} = & \sup\limits_{(\varphi_i) \in \mathbb{R}^n, (\psi_j) \in \mathbb{R}^m} & \sum\limits_i \varphi_i \overline{\mu}_i + \sum\limits_j \psi_j \overline{\nu}_i \\ & \text{subject to} & \varphi_i + \psi_j \leq \overline{c}_{ij} \end{array} \tag{13}$$

with the following data:

$$\overline{\mu} = (0.167, 0.167, 0.167, 0.167, 0.167, 0.168),$$

$$\overline{\nu} = (0.248, 0.234, 0.207, 0.165, 0.109, 0.040),$$

and

$$(\overline{c}_{ij})_{1\leq i,j\leq n} = \begin{pmatrix} 0.028 & 0.112 & 0.250 & 0.445 & 0.695 & 1.000 \\ 0.112 & 0.028 & 0.112 & 0.250 & 0.445 & 0.695 \\ 0.250 & 0.112 & 0.028 & 0.112 & 0.250 & 0.445 \\ 0.445 & 0.250 & 0.112 & 0.028 & 0.112 & 0.250 \\ 0.695 & 0.445 & 0.250 & 0.112 & 0.028 & 0.112 \\ 1.000 & 0.695 & 0.445 & 0.250 & 0.112 & 0.028 \end{pmatrix}.$$

As for example in Subsection 3.2, reals $\overline{c}_{ij}$ have been computed using interval arithmetic. This finite dimensional linear programming problem admits $\overline{\mathcal{T}} = 0.091$ as optimal value. According to Theorem 3, we can write

$$\mathcal{T} \leq 91 \times 10^{-3}.$$

This upper bound can be improved by increasing m and n. Figure 5 shows this upper bound decreases with respect to n (with $m = n$).

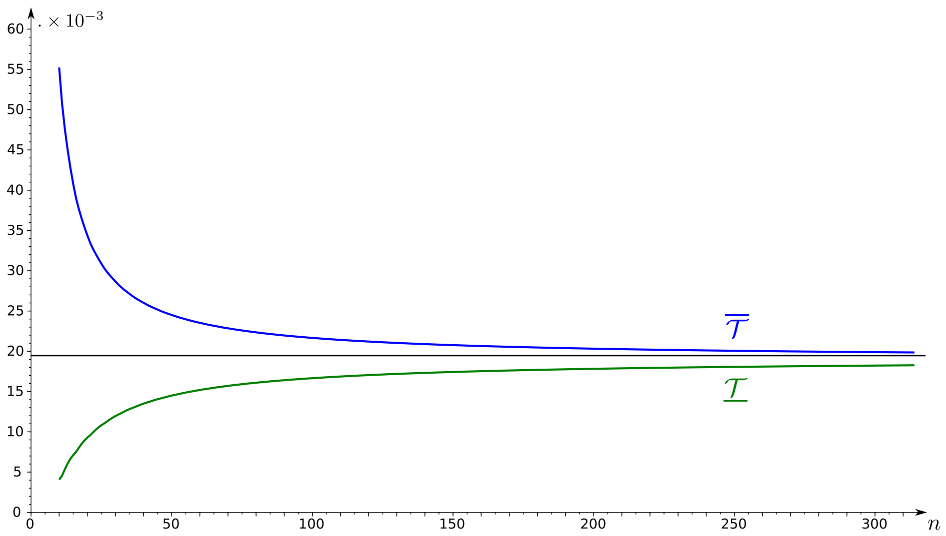

Fig. 5 Computed lower and upper bounds on $\mathcal{T}$ with respect to n (with $m = n$).

In general, the finite dimensional linear programming problem is of dimension $m + n$ with mn constraints.

Remark 3 The vectors $\overline{\mu}$ and $\overline{\nu}$ must be generated with attention. Indeed, if $\sum_i \overline{\mu}_i \neq \sum_j \overline{\nu}_j$, then solution to (13) may be unbounded. One can remark that $\overline{\mu}_6$ has been chosen to 0.168 instead of 0.167 in order to guarantee $\sum_i \overline{\mu}_i = \sum_j \overline{\nu}_j$.

Remark 4 Since the cost c is convex and μ and ν has supports on the real line. We can prove that

$$\mathcal{T} = \int_0^1 \left|G^{-1}(t) - F^{-1}(t)\right|^2 dt,$$

where F^{-1} and G^{-1} are respectively the generalized inverses of the cumulative distribution functions of f and g:

$$F(x) = \int_{-\infty}^{x} d\mu, \text{and } G(x) = \int_{-\infty}^{x} d\nu.$$

The proof of this result can be found in [27]. In our case, one has

$$\begin{aligned} F(x) &= x, && \forall x \in [0,1], \\ G(x) &= \frac{3}{2}x - \frac{1}{2}x^3, && \forall x \in [0,1], \end{aligned}$$

and

$$\begin{aligned} F^{-1}(t) &= t, && \forall t \in [0,1], \\ G^{-1}(t) &= 2\cos\left(\frac{\arccos(-t)}{3} + \frac{4\pi}{3}\right), && \forall t \in [0,1]. \end{aligned}$$

Finally, using a standard method for numerical integration, we obtain

$$\mathcal{T} = \int_0^1 \left|G^{-1}(t) - F^{-1}(t)\right|^2 dt \simeq 19,0476 \times 10^{-3}.$$

This value is represented by the black line in Figure 5 and is, of course, between our computed bounds.

4.3 Numerical examples

In this subsection, we first consider the following problem

$$\mathcal{T}(\mu,\nu) = \min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} |x - y| d\pi(x,y), \tag{14}$$

where two probability measures μ and ν with support $[0,1] = X = Y$ defined by $\mu = \mathrm{d}x$ and $\nu = \frac{3}{2}(1 - y^2)\mathrm{d}y$. We consider regular subdivision of the unit interval with n intervals. Figure 6 shows the computed lower and upper bounds.

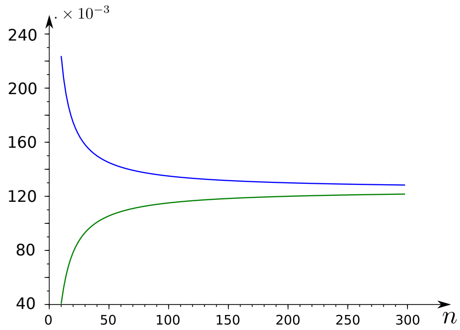

Fig. 6 Computed lower and upper bounds on $\mathcal{T}$ with respect to n (with $m = n$).

Finally, as mentioned in the introduction, our approach is not limited to polynomial cost and to measures with basic semi-algebraic support. Let us consider

$$\mathcal{T}(\mu, \nu) = \min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} (x - y)^2 d\pi(x, y), \tag{15}$$

where two probability measures μ and ν with support $[0, 1] = X = Y$ defined by $\mu = \frac{3}{2}(1 - x^2)\mathrm{d}x$ and $\nu = 2\sin^2(\pi y)\mathrm{d}y$. The two measures are represented on Figure 7.

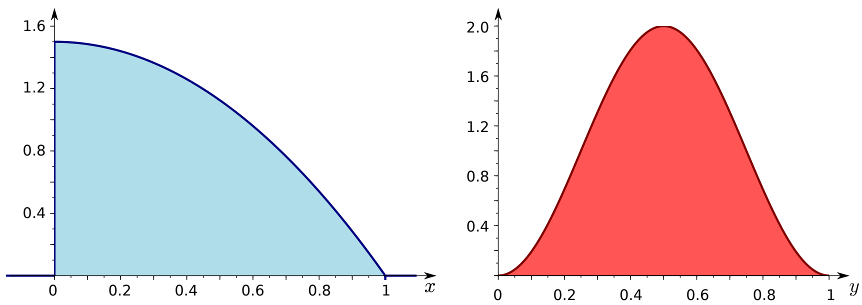

Fig. 7 Probability measures μ and ν.

Computed lower and upper bounds on $\mathcal{T}$ are shown on Figure 8.

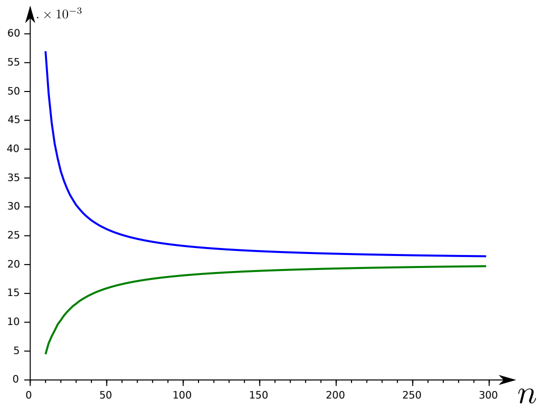

Fig. 8 Computed lower and upper bounds on $\mathcal{T}$ with respect to n with $\mu = \frac{3}{2}(1 - x^2)\mathrm{d}x$, $\nu = 2\sin^2(\pi y)\mathrm{d}y$ and $c(x, y) = (x - y)^2$.

5 Conclusion

In this paper, two rigorous spatial discretizations has been proposed to enclose the optimum of the Kantorovitch mass transportation problem. Our approach is general in the sense that only existence of inclusion function of c in (1) is required. The functionality of the designed algorithms, implemented in C++, was confirmed on some illustrating examples. The source code is available on the webpage of the author. We hope that obtained bounds could be used in a constraint propagation method to compute bounds of the solution [10].

References

1. The GNU Multiple Precision Arithmetic Library `https://gmplib.org/`
2. GLPK: GNU Linear Programming Kit. `http://www.gnu.org/software/glpk/`.
3. G. Alefeld, *Inclusion methods for systems of nonlinear equations - the interval Newton method and modifications.*, In "Topics in Validated Computation". Proceedings of the IMACS-GAMM International Workshop on Validated Computation, Oldenburg, Germany, August 30 - September 3, 7-26, 1993. Editor: J. Herzberger. Elsevier Amsterdam 1994.
4. J.D. Benamou, *Numerical resolution of an "unbalanced" mass transport problem.* ESAIM Mathematical Modelling and Numerical Analysis Vol. 37, no.5, pp. 851-868 (2003).
5. Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math. 305 (1987), no. 19, 805-808 (French, with English summary). MR 923203 (89b:58226)
6. Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions* Communications on Pure and Applied Mathematics 44(4) 375-417 (1991)
7. D. Bosc, *Numerical approximation of optimal transport maps.* SSRN (2010).
8. R. Chartrand, K. R. Vixie, B. Wohlberg and E. M. Bollt, *A gradient descent solution to the Monge-Kantorovich problem*, Appl. Math. Sci., vol. 3, no. 22, pp. 1071–1080, 2009.
9. L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, in Current developments in mathematics, 1997 (Cambridge, MA), Int. Press, Boston, MA, 65-126, 1999.

10. L. Jaulin, M. Kieffer, O. Didrit and E. Walter, *Applied Interval Analysis with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, 2001
11. L. V. Kantorovitch, *On the translocation of masses*, Doklady. Akad. Nauk SSSR, 37, No. 7-8, 227-229, 1942.
12. R. B. Kearfott. *Interval computations: Introduction, uses, and resources.* Euromath Bulletin , 2(1):95-112, 1996.
13. J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2009.
14. R. E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1996
15. R. E. Moore, *Practical aspects of interval computation*, Appl. Math., 13, pages 52-92
16. Rudolf J. Lohner, *Enclosing the solutions of ordinary initial and boundary value problems*, Computer Arithmetic, 255–286, Teubner, Stuttgart, 1987.
17. Q. Mérigot, *A multiscale approach to optimal transport.* Computer Graphics Forum, Vol 30, Issue 5, 1583-1592 (2011).
18. A. Neumaier, *Interval Methods for Systems of Equations* Encyclopedia of Mathematics and its Applications 37, Cambridge Univ. Press, Cambridge 1990
19. N. Revol, *Interval Newton iteration in multiple precision for the univariate case Numerical Algorithms,* vol 34, no 2, pp 417–426, 2003
20. Stephen M. Robinson, *Bounds for error in the solution set of a perturbed linear program* In Linear Algebra and its Applications Vol. 6, pp 69-81, 1973.
21. L.G. Casado, J.A. Martinez, I. Garcia and YA.D. Sergeyev, *New Interval Analysis Support Functions Using Gradient Information in a Global Minimization Algorithm*, In Journal of Global Optimization 25: 345-362, 2003.
22. W. Tucker, *The Lorenz attractor exists*, C. R. Acad. Sci. Math. t. 328, Série I, 1197-1202, 1999.
23. S. Dubuc, I. Kagabo *Numerical solutions of the mass transfer problem*, RAIRO - Operations Research Volume: 40, Issue: 1, page 1-17, 2006.
24. E. J. Anderson, A. B. Philpott *Duality and an Algorithm for a Class of Continuous Transportation Problems*, Mathematics of Operations Research, Vol. 9, No. 2, pp. 222-231, 1984.
25. J. Gabriel, B. Escobedo-Trujillo, *An approximation scheme for the mass transfer Problem* Morfismos, Vol. 10, no.2, pp. 1-13, 2006.
26. J. Gabriel, J. González-Hernández, R. López-Martínez, *Numerical approximations to the mass transfer problem on compact spaces.* IMA Journal of Numerical Analysis 30, 1121-1136 (2010).
27. C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, 2003.
28. C. Villani, *Optimal transport: old and new*, Springer Verlag, 2008.