

A METAHEURISTIC FOR A NUMERICAL APPROXIMATION TO THE MASS TRANSFER PROBLEM

MARTHA L. AVENDAÑO-GARRIDO ^{a,*}, JOSÉ R. GABRIEL-ARGÜELLES ^a,
LIGIA QUINTANA-TORRES ^a, EFRÉN MEZURA-MONTES ^b

^aFaculty of Mathematics

University of Veracruz, Circuito Gonzalo Aguirre Beltrán S/N, Zona Universitaria, 91090, Xalapa, Veracruz, Mexico
e-mail: {maravendano, jgabriel, liquintana}@uv.mx

^bArtificial Intelligence Research Centre

University of Veracruz, Sebastián Camacho 5, Col. Centro, 91000, Xalapa, Veracruz, Mexico
e-mail: emezura@uv.mx

This work presents an improvement of the approximation scheme for the Monge–Kantorovich (MK) mass transfer problem on compact spaces, which is studied by Gabriel *et al.* (2010), whose scheme discretizes the MK problem, reduced to solve a sequence of finite transport problems. The improvement presented in this work uses a metaheuristic algorithm inspired by scatter search in order to reduce the dimensionality of each transport problem. The new scheme solves a sequence of linear programming problems similar to the transport ones but with a lower dimension. The proposed metaheuristic is supported by a convergence theorem. Finally, examples with an exact solution are used to illustrate the performance of our proposal.

Keywords: Monge–Kantorovich mass transfer problem, finite dimensional linear programming, transport problem, metaheuristic algorithm, scatter search.

1. Introduction

In 1781 the French mathematician Gaspard Monge posed the mass transfer problem (Monge, 1781). He thought, intuitively, to find an optimal transportation plan to move earth from an embankment to a hole. The cost function considered was a distance. Monge was looking for functions that assign each particle of the embankment to its corresponding position in the hole.

In 1942 the Russian mathematician Leonid V. Kantorovich posed the problem of translocation of masses (Kantorovich, 2006a), which consists in minimizing the translocation work to move an initial mass distribution to a final mass distribution. He considered this problem in compact metric spaces, Borel sets and a non-negative continuous function. In 1948 Kantorovich realized that when the cost function is a distance, the translocation mass problem is a generalization of the Monge problem (Kantorovich, 2006b). Since then this has been known as the Monge–Kantorovich (MK) mass transfer problem.

The MK problem became the prototype of many problems in probability theory, differential geometry, linear programming, stochastic control, information theory, cybernetic and matrix theory (see Rachev and Rüschendorf, 1998). In particular, Kantorovich’s metric is defined by an MK problem. This metric is used in several areas of abstract mathematics, e.g., measure theory, ergodic theory, functional analysis, statistics and others (see Rachev and Rüschendorf, 1998). Also, it is used in several applications, for example, probability metrics (Rachev, 1991), control of cancer radiotherapy (Hanin *et al.*, 1993), image registration and warping (Haker *et al.*, 2004), or limit theorems and recursive stochastic equations (Rachev and Rüschendorf, 1998), among other things.

Several authors have studied approximations schemes for the MK problem. For example, Anderson and Nash (1987) or Anderson and Philpott (1984) studied an algorithm in which the underlying spaces are the unit interval $[0, 1]$. In the work of Hernández-Lerma and Lasserre (1998), a general approximation scheme based

*Corresponding author

on infinite dimensional linear programming problems is given, which can be applied to the MK problem. Gabriel *et al.* (2010) and González-Hernández *et al.* (2006) proposed a general approximation scheme for the MK problem on compact and Polish spaces, respectively. Other schemes for the MK problem were studied by Benamou and Brenier (2000), Caffarelli *et al.* (2002), Benamou (2003), Guittet (2003), and recently by Bosc (2010) and Mèrigot (2011). This work presents an improvement to the approximation scheme for the MK problem on compact spaces given by Gabriel *et al.* (2010).

The improvement proposed relies on the usage of a metaheuristic algorithm inspired by the scatter search (SS) algorithm (see Martí *et al.*, 2006). At the beginning, the SS was considered a new evolutionary method because it combines two or more solutions in an initial set, called the reference set, to create new solutions that improve the original ones. In this sense, it is said that SS is an evolutionary method. However, it breaks the premise that evolutionary approaches must be based on randomization. In contrast, SS adopts a deterministic approach to select two or more elements of the reference set to create new solutions. Such a way of work based on deterministic decisions has distinguished SS from other evolutionary methods. Furthermore, it has showed advantages to solve a variety of complex optimization problems (see Laguna *et al.*, 2014). Unlike other metaheuristic algorithms which are trajectory based (e.g., simulated annealing), SS is based on a population of solutions and a deterministic recombination process to generate new ones. It is also based on the idea that potential solutions of a given problem have different important elements. Therefore, a suitable way to combine them is very convenient so as to improve the quality of the solutions within an interactive process.

The scheme proposed in this work solves a linear programming (LP) problem similar to the transport (T) problem but with a lower dimension. The LP problems are constructed using a metaheuristic algorithm inspired by SS.

The remainder of this paper is organized as follows. In Section 2, we present the scatter search algorithm. In Section 3, we study the MK problem and the numerical approximation proposed by Gabriel *et al.* (2010). In Section 4, we present an algorithm, inspired by scatter search, to approximate the MK problem through a sequence of LP problems in order to decrease the number of variables of the transport problems and, as a consequence, to decrease the execution time as well. In Section 5, we prove our main results, i.e., the sequence of solutions of the LP problems that approximate the MK problem converges to the optimal value of the MK problem. In Section 6, we use three examples in order to illustrate the performance of the proposed scheme.

Finally, Section 7 summarizes our conclusions.

2. Scatter search: An evolutionary method

The evolutionary method known as scatter search (SS) was introduced in the 1960s by Fred Glover, who published the first description of the method in 1977, establishing its main features. In the work of Glover (1998), SS is detailed. The SS method follows the steps presented in Algorithm 1.

Algorithm 1. Scatter search.

Step 1. Generate an initial set of different solutions and improve them by using a suitable heuristic process. The best solutions out of the initial set will be called the reference solution set.

Step 2. Create new solutions by combining subsets of the reference solutions in a structured way.

Step 3. Apply the heuristic adopted in Step 1 to improve the new solutions generated in Step 2.

Step 4. The subset of best solutions obtained in Step 3 are added to the current reference set.

Step 5. Repeat Steps 2, 3 and 4 until the reference set does not change. To promote diversity in the reference set, Step 1 can be called again. The whole process ends when a termination condition (e.g., a number of iterations) is reached.

The proposed scheme in this work is inspired by SS. However, it has some differences, what we are going to show further in this paper.

3. Monge–Kantorovich problem and its numerical approximation

In the MK problem the following data are given:

1. two metric spaces X and Y , endowed with the corresponding Borel σ -algebras $\mathbb{B}(X)$ and $\mathbb{B}(Y)$, respectively;
2. a measurable cost function $c : X \times Y \rightarrow \mathbb{R}$;
3. a probability measure ν_1 on X and a probability measure ν_2 on Y .

Let $M(X \times Y)$ be the linear space of finite signed measures on $\mathbb{B}(X \times Y)$ endowed with the topology of weak convergence, and let $M^+(X \times Y)$ be the convex cone of non-negative measures in $M(X \times Y)$.

If μ is in $M(X \times Y)$, we denote by $\Pi_1\mu$ and $\Pi_2\mu$ the marginals (or projections) of μ on X and Y , respectively. That is, for all $A \in \mathbb{B}(X)$ and $B \in \mathbb{B}(Y)$ we have

$$\begin{aligned}\Pi_1\mu(A) &:= \mu(A \times Y), \\ \Pi_2\mu(B) &:= \mu(X \times B).\end{aligned}$$

Then the MK problem can be stated as follows:

$$\begin{aligned} \text{MK :} \quad & \text{minimize} && \langle \mu, c \rangle := \int_{X \times Y} c d\mu \quad (1) \\ & \text{subject to} && \Pi_1 \mu = \nu_1, \quad (2) \\ & && \Pi_2 \mu = \nu_2, \quad (3) \\ & && \mu \in M^+(X \times Y). \quad (4) \end{aligned}$$

A measure μ in $M(X \times Y)$ is a feasible solution for the MK problem if it satisfies (2), (3) and (4), and $\langle \mu, c \rangle$ is finite. The MK problem is called *consistent* if the set of feasible solutions is non-empty, in which case its optimum value is defined as

$$\text{inf(MK)} := \inf\{\langle \mu, c \rangle \mid \mu \in \mathcal{F}\},$$

where

$$\mathcal{F} := \{\mu \mid \mu \text{ is a feasible solution to the MK problem}\}.$$

The MK problem is said to be solvable if there is a feasible solution $\hat{\mu}$ that attains the optimum value. In this case, $\hat{\mu}$ is called an optimal solution to the MK problem, and the value inf(MK) is written as $\text{min(MK)} = \langle \hat{\mu}, c \rangle$.

If there exists a Borel measurable function $f : X \rightarrow Y$ such that

$$\nu_2(B) = \nu_1(f^{-1}(B))$$

for all $B \in \mathbb{B}(Y)$, then it is called the optimal coupling for the MK problem. The former function is a transport map or Monge's solution.

Assumption 1. The following assumptions are required:

- (a) The spaces X and Y are compact metric spaces.
- (b) The cost function $c(x, y)$ is continuous.
- (c) The optimal solution for the MK problem exists and it is unique.

Remark 1. The product measure $\mu := \nu_1 \times \nu_2$ is a feasible solution for the MK problem. Therefore it is consistent. Furthermore, in this case the MK problem is solvable (see Anderson and Nash, 1987, Theorem 5.2; Hernaández-Lerma and Gabriel, 2002, Remark 2.5). Moreover, uniqueness conditions for the MK problem are established by Levin (2006, Theorems 2.1 and 2.2.).

We shall need the next three propositions and a theorem proved by Gabriel *et al.* (2010).

Proposition 1. *There exist two probability measure sequences $\{\nu_1^n\}$ on $\mathbb{B}(X)$ and $\{\nu_2^n\}$ on $\mathbb{B}(Y)$, both with finite supports, such that $\{\nu_1^n\}$ weakly converges to ν_1 and $\{\nu_2^n\}$ weakly converges to ν_2 .*

Remark 2. By the proof of Proposition 3.1 of Gabriel *et al.* (2010), for a given sequence of positive numbers $\{r_n\}$ such that $r_n \downarrow 0$, the two probability

measure sequences $\{\nu_1^n\}$ on $\mathbb{B}(X)$ and $\{\nu_2^n\}$ on $\mathbb{B}(Y)$ have supports on finite sets contained in X^n and Y^n , respectively, where $\cup_n X^n$ and $\cup_n Y^n$ are denumerable dense sets in X and Y , respectively. We consider the set $S^n = X^n \times Y^n$. Note that the elements of S^n are pairs (x, y) with $x \in X$ and $y \in Y$.

Now, a sequence of MK problems is introduced. For each positive integer n , the following MK problem is considered:

$$\begin{aligned} \text{MK}^n : \quad & \text{minimize} && \langle \mu, c \rangle \\ & \text{subject to} && \Pi_1 \mu = \nu_1^n, \\ & && \Pi_2 \mu = \nu_2^n, \\ & && \mu \in M^+(X \times Y). \end{aligned}$$

Proposition 2. *If μ is a feasible solution to the MK problem, then there exists a probability measure sequence $\{\mu^n\}$ on $\mathbb{B}(X \times Y)$, with marginals $\Pi_1 \mu^n = \nu_1^n$ and $\Pi_2 \mu^n = \nu_2^n$, such that $\{\mu^n\}$ weakly converges to μ .*

Proposition 3. *If μ is a feasible solution for MK^n , then μ has a finite support denoted by $\text{supp}(\mu)$ and, moreover,*

$$\text{supp}(\mu) \subset S^n.$$

Using the above propositions, the MK^n problem can be discretized by a T^n problem in the following way:

$$\begin{aligned} T^n : \quad & \text{minimize} && \sum_{(x,y) \in S^n} c_{xy} \lambda_{xy}^n \\ & \text{subject to} && \sum_{y \in Y^n} \lambda_{xy}^n = a_x^n \quad \forall x \in X^n, \\ & && \sum_{x \in X^n} \lambda_{xy}^n = b_y^n \quad \forall y \in Y^n, \\ & && \lambda_{xy}^n \geq 0 \quad \forall (x, y) \in S^n, \end{aligned}$$

where

$$c_{xy} = c(x, y), \quad a_x^n = \nu_1^n(\{x\}), \quad b_y^n = \nu_2^n(\{y\})$$

with $x \in X^n$ and $y \in Y^n$.

Remark 3. Note that

$$\sum_{x \in X^n} a_x^n = \sum_{y \in Y^n} b_y^n = 1,$$

and under this condition the T^n problem has an optimal solution (see Bazaraa *et al.*, 2010).

Now, if $\{\hat{\lambda}_{xy}^n\}$ is an optimal solution to a T^n problem, the measure is defined as

$$\hat{\mu}^n(\cdot) = \sum_{(x,y) \in S^n} \hat{\lambda}_{xy}^n \delta_{(x,y)}(\cdot),$$

where $\delta_{(x,y)}$ denotes the Dirac measure concentrated at (x, y) in $X \times Y$.

Then we have the following convergence theorem.

Theorem 1. *There exist a subsequence $\{\widehat{\mu}^{n_i}\}$ of $\{\widehat{\mu}^n\}$ and a probability measure $\widehat{\mu}$, such that the following holds:*

1. $\{\widehat{\mu}^{n_i}\}$ weakly converges to $\widehat{\mu}$,
2. $\widehat{\mu}$ is an optimal solution to the MK problem, and
3. $\lim_{n \rightarrow \infty} \langle \widehat{\mu}^n, c \rangle = \min(\text{MK})$.

A bound to the convergence order of this scheme is given by

$$|\langle \widehat{\mu}, c \rangle - \langle \widehat{\mu}^n, c \rangle| \leq \varepsilon_n,$$

where

$$\varepsilon_n = \sup \{ |c(x, y) - c(z, w)| \},$$

with (x, y) and (z, w) in $X \times Y$, $d(x, z) < r_n$ and $d(y, w) < r_n$ (see Gabriel et al., 2010, Theorem 4.5).

In practice, for each n , the previous approximation scheme can be implemented by solving the T^n problem. Nevertheless, as n increases, the computational time becomes very high and it is expensive to calculate the optimal solution to the T^n problem. Thus, it is necessary to design a more efficient solution scheme.

4. Proposed scheme

In order to decrease the execution time to solve the T^n problem, we will implement an approximation scheme inspired by the SS method, which reduces the number of variables of the T^n problem. In this way, we can obtain much better approximations in less execution time, as will be shown further.

The traditional SS generates a set of solutions; they are improved and recombined until competitive solutions are found. Our proposal takes that idea, but it is not applied to solutions, but to sets of points that generate linear programming problems; i.e., it takes a set of points obtained after solving a T problem. After that, the set of points is improved and combined with another set of points to generate a new linear programming problem and so on, until a competitive solution of the linear programming problem is obtained. In the next section such a scheme is detailed.

We construct a sequence of LP problems using Algorithm 2. Next, we give the detailed steps.

Step 1.

- (a) Take appropriate $n \in \mathbb{N}$.
- (b) Determine X^n and Y^n , where $\cup_n X^n$ and $\cup_n Y^n$ are dense denumerable sets in X and Y , respectively, for r_n given in Remark 2.
- (c) Define $S^n = X^n \times Y^n$.

Algorithm 2. Proposed algorithm.

Step 1. For a specific n , construct and resolve the T^n problem. Take as a reference set all the points, where $\widehat{\mu}^n$ is different from zero. (Inspired by Step 1 of SS.)

Step 2. Using the points found in Step 1, add additional points based in a deterministic method. (Inspired by Step 2 of SS.)

Step 3. Construct and resolve the LP^{n+1} problem. Take as a new reference set all the points, where $\widehat{\mu}^{n+1}$ is different from zero. (Inspired by a combination of Steps 3 and 4 of SS.)

Step 4. Repeat Steps 2 and 3 until reaching a specified iteration limit. (Similar to Step 5 of SS.)

(d) Construct and solve the T^n problem:

$$\begin{aligned} T^n : \quad & \text{minimize} && \sum_{(x,y) \in S^n} c_{xy} \lambda_{xy}^n \\ & \text{subject to} && \sum_{y \in Y^n} \lambda_{xy}^n = a_x^n \quad \forall x \in X^n, \\ & && \sum_{x \in X^n} \lambda_{xy}^n = b_y^n \quad \forall y \in Y^n, \\ & && \lambda_{xy}^n \geq 0 \quad \forall (x, y) \in S^n, \end{aligned}$$

where

$$\begin{aligned} c_{xy} &= c(x, y), \\ a_x^n &= \nu_1^n(\{x\}), \\ b_y^n &= \nu_2^n(\{y\}) \end{aligned}$$

with $x \in X^n$ and $y \in Y^n$.

(e) Define

$$\widehat{\mu}^n(\cdot) = \sum_{(x,y) \in S^n} \widehat{\lambda}_{xy}^n \delta_{(x,y)}(\cdot),$$

where $\{\widehat{\lambda}_{xy}^n\}$ is the optimal solution of the T^n problem.

(f) Compute

$$R^n = \{(x, y) | (x, y) \in S^n \text{ and } \widehat{\mu}^n(x, y) \neq 0\}.$$

(Note that $R^n \subset S^n$ and $\widehat{\lambda}_{xy}^n \neq 0$ whenever $\widehat{\mu}^n(x, y) \neq 0$.)

Step 2.

(a) For each $(x, y) \in R^n$, there exist points (x^*, y^*) such that

$$x^* = \arg \min_x d(x, x')$$

for all $(x', y') \in R^n$ with $x \neq x'$, where

$$\arg \min_a f(a) := \{a | \forall b, f(a) \leq f(b)\}.$$

If $d(y, y^*) > r_n$, take the set

$$H_{(x,y)}^n = \{x, x^*\} \times (B_r(y) \cap B_r(y^*)) \cap Y^n,$$

where $r = d(y, y^*)$ and $B_r(z)$ is the open neighborhood with center in z and radius r .

(b) Define the set

$$\dot{R}^n = \left(\bigcup_{(x,y)} H_{(x,y)}^n \right) \cup R^n.$$

(Note that $\dot{R}^n \subset S^n$.)

(c) Obtain X^{n+1} and Y^{n+1} where $\cup_{n+1} X^{n+1}$ and $\cup_{n+1} Y^{n+1}$ are dense denumerable sets in X and Y respectively, for r_{n+1} given by Remark 2.

(d) Define $S^{n+1} = X^{n+1} \times Y^{n+1}$.

(e) For each $(x, y) \in \dot{R}^n$, take

$$N_{(x,y)}^n = B_{r_{n+1}}(x, y) \cap S^{n+1}.$$

(f) Define the set

$$\tilde{R}^{n+1} = \bigcup_{(x,y) \in \dot{R}^n} N_{(x,y)}^n.$$

(Note that $\tilde{R}^{n+1} \subset S^{n+1}$.)

Step 3.

(a) Construct and solve the following LP^{n+1} problem:

$$\begin{aligned} LP^{n+1} : \text{minimize} \quad & \sum_{(x,y) \in \tilde{R}^{n+1}} c_{xy} \lambda_{xy}^{n+1} \\ \text{subject to} \quad & \sum_{y:(x,y) \in \tilde{R}^{n+1}} \lambda_{xy}^{n+1} = a_x^{n+1} \\ & \forall x \text{ s. t. } (x, y) \in \tilde{R}^{n+1}, \\ & \sum_{x:(x,y) \in \tilde{R}^{n+1}} \lambda_{xy}^{n+1} = b_y^{n+1} \\ & \forall y \text{ s. t. } (x, y) \in \tilde{R}^{n+1}, \\ & \lambda_{xy}^{n+1} \geq 0 \quad \forall (x, y) \in \tilde{R}^{n+1}, \end{aligned}$$

where $c_{xy} = c(x, y)$, $a_x^{n+1} = \nu_1^{n+1}(\{x\})$ and $b_y^{n+1} = \nu_2^{n+1}(\{y\})$ with $(x, y) \in \tilde{R}^{n+1}$.

(b) Define

$$\hat{\mu}^{n+1}(\cdot) = \sum_{(x,y) \in \tilde{R}^{n+1}} \hat{\lambda}_{xy}^{n+1} \delta_{(x,y)}(\cdot), \quad (5)$$

where $\{\hat{\mu}^{n+1}\}$ is the optimal solution of the LP^{n+1} problem.

(c) Take

$$R^{n+1} = \left\{ (x, y) \mid (x, y) \in \tilde{R}^{n+1}, \hat{\mu}^{n+1}(x, y) \neq 0 \right\}.$$

(Note that $R^{n+1} \subset S^{n+1}$.)

Step 4. Repeat Steps 2 and 3 until a stop condition is reached.

Remark 4. Let be $(\tilde{R}^{n+1})^c = S^{n+1} - \tilde{R}^{n+1}$. If the variables λ_{xy}^{n+1} with (x, y) in $(\tilde{R}^{n+1})^c$ are the non basic variables of the T^{n+1} problem, then the set of the extreme points of the T^{n+1} problem coincides with the set of the extreme points of the LP^{n+1} problem (see Bazaraa *et al.*, 2010, Chapter 3).

5. Convergence theorem

The sequence of the LP^{n+1} problems generated by the previous algorithm converges due to the following theorem.

Theorem 2. Let T^{n+1} be the T problem that discretizes the MK problem and let LP^{n+1} be the LP problem generated using Algorithm 2. Then

$$\min(T^{n+1}) = \min(LP^{n+1}).$$

Proof. We define the sets

$$\mathcal{F}_T = \{ \{ \lambda_{xy} \} \mid \{ \lambda_{xy} \} \text{ is a feasible solution to the } T^{n+1} \text{ problem} \}$$

and

$$\mathcal{F}_{LP} = \{ \{ \lambda_{xy} \} \mid \{ \lambda_{xy} \} \text{ is a feasible solution to the } LP^{n+1} \text{ problem} \}.$$

By Remark 4, we have

$$\mathcal{F}_T = \mathcal{F}_{LP}$$

and

$$\min(T^{n+1}) = \min(LP^{n+1}). \quad \blacksquare$$

Corollary 1. Let $\{\hat{\mu}^{n+1}\}$ be a sequence of measures given in Eqn. (5). Then the sequence $\{\langle \hat{\mu}^{n+1}, c \rangle\}$ converges to $\min(\text{MK})$.

Proof. It is a consequence of Theorem 3.6 proved by Gabriel *et al.* (2010) and the previous theorem. \blacksquare

Remark 5. The order of convergence of the proposed scheme is given by

$$|\langle \hat{\mu}, c \rangle - \langle \hat{\mu}^n, c \rangle| \leq \varepsilon_n.$$

6. Numerical examples

In this section, we illustrate the proposed scheme using three cost functions. The first one was previously studied by Gabriel *et al.* (2010) as well as Anderson and Philpott (1984), the second one by Gabriel *et al.* (2010), and the third one by Anderson and Nash (1987). These examples are used because they have an exact solution and they can be compared with the approximated solutions obtained by our proposal. The examples are designed with two goals: the first is to graphically present the sets generated using the proposed algorithm and the second is to compare the number of variables of the original T problem with the number of variables of the LP problem generated under the new scheme proposed in this work.

In the examples, we consider $X = Y = [0, 1]$ with the usual topology and $\nu_1 = \nu_2$ as the Lebesgue measure. These examples satisfy Assumption 1 since the cost functions are polynomial functions and X and Y are compact metric spaces.

We take the sets

$$X^n = Y^n = \left\{ \frac{k}{2^n} \mid 0 \leq k \leq 2^n \right\}.$$

The free software R was used to optimize the respective LP problems.

6.1. Graphical illustration of the proposed scheme.

First, we graphically illustrate the sets given by the algorithm proposed in Section 4 in order to show its performance at each iteration. We start the algorithm with $n = 3$. Despite the fact that $n = 3$ is a small value, it is suitable to show the sets S^n, \tilde{R}^n and R^{n+1} . It is necessary to emphasize that if we take n larger than 3, it is difficult to visualize the sets.

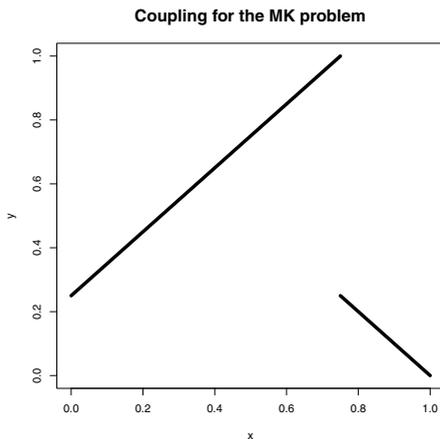


Fig. 1. Optimal coupling using the cost function $c(x, y) = x^2y - xy^2$.

Example 1. Let $c(x, y) = x^2y - xy^2$ be the cost function. Gabriel *et al.* (2010) as well as Anderson and Philpott (1984) showed that its optimal coupling is given by

$$\text{Graph}(f) = \{(t, f(t)) \mid t \in [0, 1]\},$$

where

$$f(t) = \begin{cases} \frac{1}{4} + t & \text{for } t \in [0, \frac{3}{4}), \\ 1 - t & \text{for } t \in [\frac{3}{4}, 1] \end{cases}$$

(see Fig. 1).

In the first image of Fig 2, we show the points in the set S^3 (Step 1(c)), the points in the set R^3 (Step 1(f)) and the points in the set \tilde{R}^4 (Step 2(f)). In the second and third images of Fig. 2, we show the next two iterations. In the other images of Fig. 2, we only show the points in sets R^6, R^7 and R^8 because they are too close and it is difficult to visualize them. We can observe that the points in R^8 are over the graph of the optimal coupling shown in Fig. 1. ♦

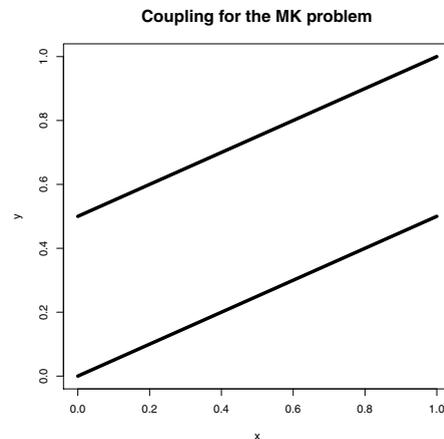


Fig. 3. Coupling using the cost function $c(x, y) = (2y - x - 1)^2(2y - x)^2$.

Example 2. Let $c(x, y) = (2y - x - 1)^2(2y - x)^2$ be the cost function. Its coupling is studied by Gabriel *et al.* (2010) and it is given by

$$\text{Graph}(f_1) \cup \text{Graph}(f_2),$$

where $f_1 : [0, 1] \rightarrow [0, 1]$ and $f_2 : [0, 1] \rightarrow [0, 1]$ with

$$f_1(t) = \frac{t}{2}, \quad f_2(t) = \frac{t+1}{2}$$

(see Fig. 3).

In the first panel of Fig. 4, we show the points in the set S^3 (Step 1(c)), the points in the set R^3 (Step 1(f)) and the points in the set \tilde{R}^4 (Step 2(f)). In the second and third

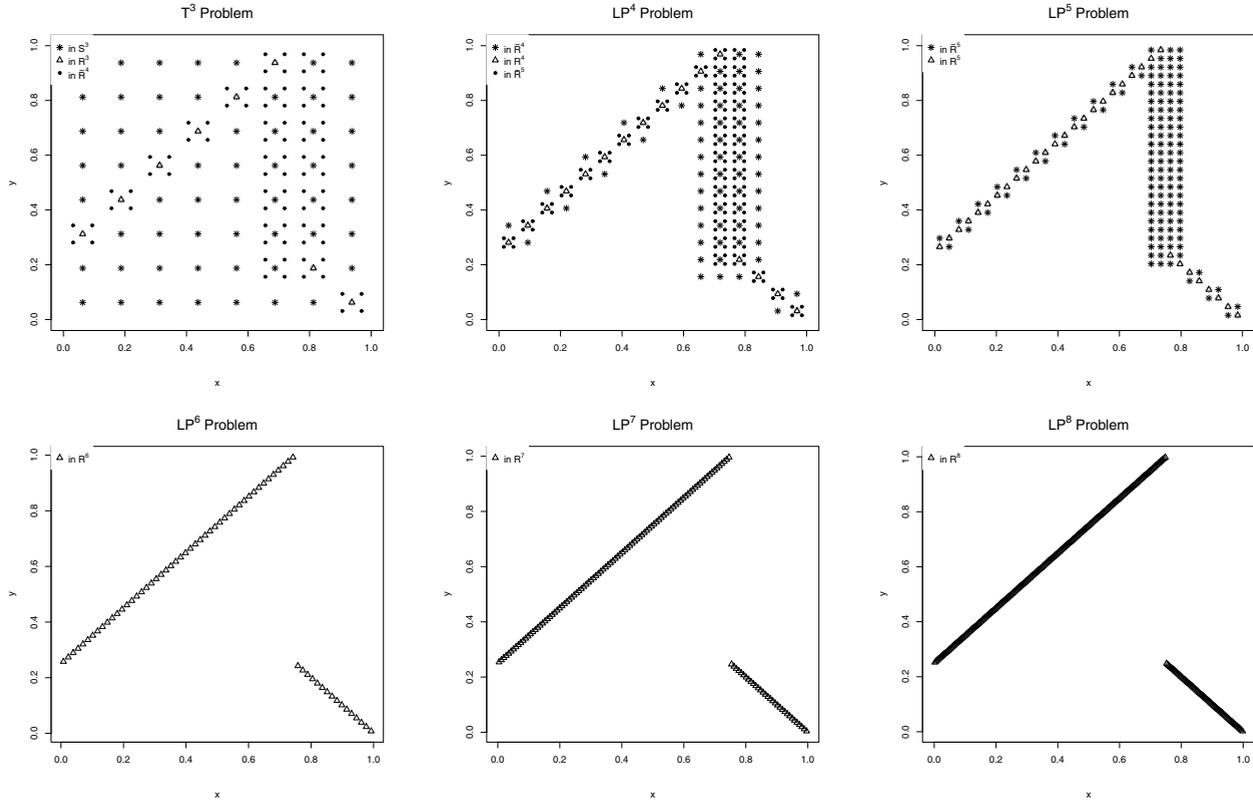


Fig. 2. Solution of the cost function $c(x, y) = x^2y - xy^2$.

panels of Fig. 4, we show the next two iterations. In the other panels of Fig. 4, we show those points in the sets R^6 , R^7 and R^8 , respectively. We can observe, as in the previous example, that the points in R^8 are over the graph of the optimal solution shown in Fig. 3. ♦

6.2. Comparison of the proposed algorithm against the previous scheme. Now, we compare the scheme proposed in the algorithm of Section 4 with the one by Gabriel *et al.* (2010). By Corollary 1 and Remark 5, the results and the bound of the error are the same; however, we make the comparison based on the number of variables of the T^n and LP^n problems.

Note that $\text{card}(X^n) = \text{card}(Y^n) = 2^n$. Then the T^n problem has $(2^n)(2^n)$ variables. Thus, solving the T^n problem produces a very high computational cost when n increases.

In these examples, we start with $n = 8$, because it is more appropriate to obtain better approximations and it is not necessary visualize the sets \tilde{R}^n .

Example 3. (Continuation of Example 1) Anderson and Philpott (1984) proved that the optimal value for the MK problem given in Example 1 is $-\frac{9}{256} \approx -0.03515625$. In Table 1 we show the number of elements in \tilde{R}^n that are

taken to construct the LP^n problem. We can see that the number of variables of the LP^n problem is remarkably lower than the number of variables of the T^n problem. The results for the T^{10} and T^{11} problems are not reported because of the high cost of their implementation. The results of the LP^8 problem are not presented because they require the solution of the T^7 problem.

Table 1. Comparison using the cost function $c(x, y) = x^2y - xy^2$.

Problem	No. of variables	Error	Value
T^8	65536	9.5×10^{-7}	-0.0351553
LP^8	—	—	—
T^9	262144	2.4×10^{-7}	-0.0351560
LP^9	2560	2.4×10^{-7}	-0.0351560
T^{10}	1048576	—	—
LP^{10}	5120	6.0×10^{-8}	-0.03515619
T^{11}	4194304	—	—
LP^{11}	10240	1.5×10^{-8}	-0.03515624

Example 4. Let $c(x, y) = 4x^2y - xy^2$ be the cost function. This function was studied by Anderson and Nash (1987), and its optimal value is $\frac{107}{432} \approx 0.2476852$

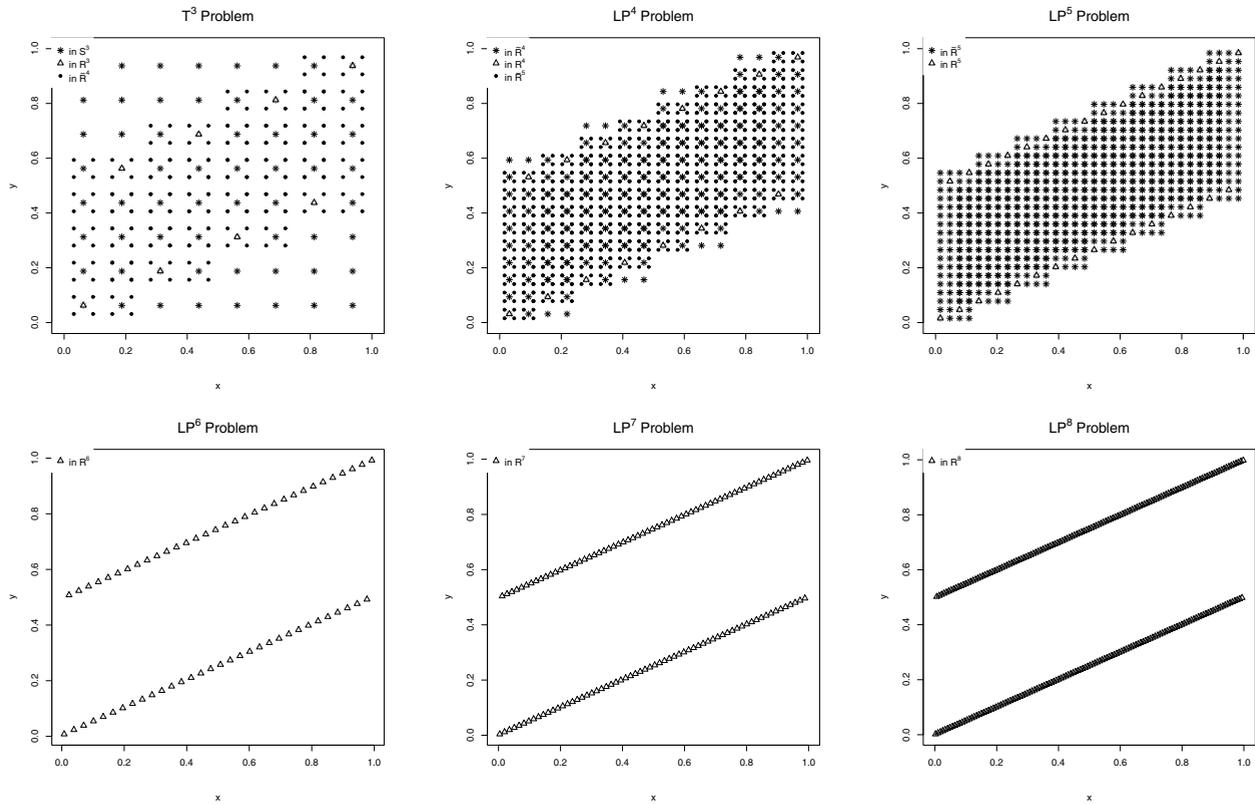


Fig. 4. Solution of the cost function $c(x, y) = (2y - x - 1)^2(2y - x)^2$.

while its optimal solution has support on the set

$$\text{Graph}(f) = \{(t, f(t)) | t \in [0, 1]\}$$

with

$$f(t) = \begin{cases} \frac{2}{3} + \frac{4}{3}t & \text{for } [0, \frac{1}{4}), \\ 2 - 4t & \text{for } [\frac{1}{4}, \frac{1}{3}), \\ 1 - t & \text{for } [\frac{1}{3}, 1]. \end{cases}$$

It is shown in Fig. 5(a).

In Table 2, we show the number of variables in the T^n and LP^n problems. We can note again that the dimension of the LP^n problem is considerably lower than that of the T^n problem. The results for the T^{11} problem are not reported due to the related high computational cost. Finally, in Fig. 5(b) we can observe that the points in R^{11} are in the support of the optimal coupling for the MK problem. The results of the LP^8 problem are not presented because they require the solution of the T^7 problem.



7. Conclusion

An approximation scheme for the MK problem was studied by Gabriel *et al.* (2010), who approximated the

Table 2. Comparison using the cost function $c(x, y) = 4x^2y - xy^2$.

Problem	No. of variables	Error	Value
T^8	65536	2.5×10^{-6}	0.2476877
LP^8	—	—	—
T^9	262144	6.4×10^{-7}	0.2476858
LP^9	2040	6.4×10^{-7}	0.2476858
T^{10}	1048576	1.6×10^{-7}	0.2476853
LP^{10}	4080	1.6×10^{-7}	0.2476853
T^{11}	4194304	—	—
LP^{11}	8160	4.2×10^{-8}	0.2476852

solution to the MK problem through the solution to the T problems. This approximation is natural. However, the LP problems generated by the T problems can be very large and computationally expensive. In this work, such a previous scheme is improved using a metaheuristic inspired by scatter search. The LP problems generated by the proposed metaheuristic have the same solution, but they are significantly less expensive than the T problem. Such performance implies that it is possible to obtain better approximations in less time. The algorithm was implemented using the free software R and tested with

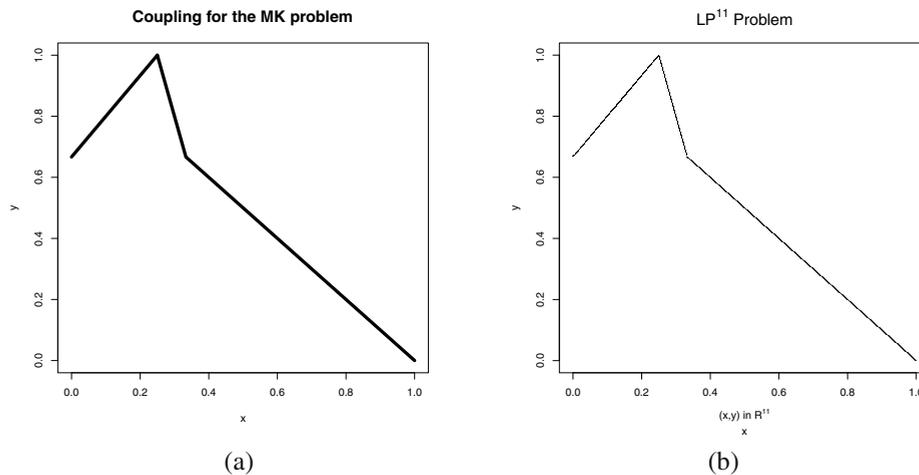


Fig. 5. Coupling (a) and approximated solution (b) to the cost function $c(x, y) = 4x^2y - xy^2$.

some examples previously reported in the specialized literature.

References

- Anderson, E. and Nash, P. (1987). *Linear Programming in Infinite-dimensional Spaces*, Wiley, New York, NY.
- Anderson, E. and Philpott, A. (1984). Duality and an algorithm for a class of continuous transportation problems, *Mathematics of Operations Research* **9**(2): 222–231.
- Bazaraa, M.S., Jarvis, J.J. and Sherali, H.D. (2010). *Linear Programming and Network Flows*, Wiley-Interscience, Hoboken, NJ.
- Benamou, J. (2003). Numerical resolution of an unbalanced mass transport problem, *ESAIM Mathematical Modelling and Numerical Analysis* **37**(5): 851–868.
- Benamou, J. and Brenier, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem, *Numerische Mathematik* **84**(3): 375–393.
- Bosc, D. (2010). Numerical approximation of optimal transport maps, *SSRN Electronic Journal*, DOI: 10.2139/ssrn.1730684.
- Caffarelli, L., Feldman, M. and McCann, R. (2002). Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs, *Journal of the American Mathematical Society* **15**(1): 1–26.
- Gabriel, J., González-Hernández, J. and López-Martínez, R. (2010). Numerical approximations to the mass transfer problem on compact spaces, *IMA Journal of Numerical Analysis* **30**(4): 1121–1136.
- Glover, F. (1998). A template for scatter search and path relinking, in J.-K. Hao *et al.* (Eds.), *Artificial Evolution*, Lecture Notes in Computer Science, Vol. 1363, Springer, Berlin/Heidelberg, pp. 1–51.
- González-Hernández, J., Gabriel, J. and Hernández-Lerma, O. (2006). On solutions to the mass transfer problem, *SIAM Journal on Optimization* **17**(2): 485–499.
- Guittet, K. (2003). On the time-continuous mass transport problem and its approximation by augmented Lagrangian techniques, *SIAM Journal on Numerical Analysis* **41**(1): 382–399.
- Haker, S., Zhu, L., Tannenbaum, A. and Angenent, S. (2004). Optimal mass transport for registration and warping, *International Journal of Computer Vision* **63**(3): 225–240.
- Hanin, L., Rachev, S. and Yakovlev, A. (1993). On the optimal control of cancer radiotherapy for non-homogeneous cell population, *Advances in Applied Probability* **25**(1): 1–23.
- Hernández-Lerma, O. and Gabriel, J. (2002). Strong duality of the Monge–Kantorovich mass transfer problem in metric spaces, *Mathematische Zeitschrift* **239**(3): 579–591.
- Hernández-Lerma, O. and Lasserre, J. (1998). Approximation schemes for infinite linear programs, *SIAM Journal on Optimization* **8**(4): 973–988.
- Kantorovich, L. (2006a). On a problem of Monge, *Journal of Mathematical Sciences* **133**(4): 225–226.
- Kantorovich, L. (2006b). On the translocation of masses, *Journal of Mathematical Sciences* **133**(4): 1381–1382.
- Laguna, M., Gortázar, F., Gallego, M., Duarte, A. and Martí, R. (2014). A black-box scatter search for optimization problems with integer variables, *Journal of Global Optimization* **58**(3): 497–516.
- Levin, V. (2006). Optimality conditions and exact solutions to the two-dimensional Monge–Kantorovich problem, *Journal of Mathematical Sciences* **133**(4): 1456–1463.
- Martí, R., Laguna, M. and Glover, F. (2006). Principles of scatter search, *European Journal of Operational Research* **169**(2): 359–372.
- Mèrigot, Q. (2011). A multiscale approach to optimal transport, *Computer Graphics Forum* **30**(5): 1583–1592.
- Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*, De l’Imprimerie Royale, Paris.

Rachev, S. (1991). *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York, NY.

Rachev, S. and Rüschendorf, L. (1998). *Mass Transportation Problems*, Vol. I, Springer, New York, NY.



Martha L. Avendaño-Garrido was born in Xalapa, Mexico, in 1980. She received her diploma degree in mathematics from Universidad Veracruzana (Mexico) in 2004, her M.Sc. degree in industrial mathematics and computer science from Centro de Investigación en Matemáticas (Mexico), and her Ph.D. degree in mathematics from Universidad Complutense de Madrid (Spain) in 2013. She is a full professor at the Faculty of Mathematics, Universidad Veracruzana.

She is interested in optimization, applied mathematics and computer science.



José R. Gabriel-Argüelles was born in Banderilla, Veracruz, Mexico, in 1969. He received his Ph.D degree from the Department of Mathematics of CINVESTAV, IPN (Mexico), in 2000. He is a full professor at the Faculty of Mathematics, Universidad Veracruzana (Mexico). He has published 12 research papers in the areas of infinite linear programming, the Monge–Kantorovich mass transfer problem and schemes of approximation.



Efrén Mezura-Montes was born in Xalapa, Mexico, in 1973. He received his diploma degree in computer systems engineering from the University of the Americas, Puebla, in 1997, his M.Sc. degree in artificial intelligence from the University of Veracruz in 2001, and his Ph.D. degree in computer science from the Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV-IPN) in 2004. He has published more than 100 papers in peer-reviewed journals and conferences. His research interests are in the design, study and application of nature-inspired meta-heuristic algorithms to solve complex optimization problems. Dr. Mezura-Montes is currently the head of the Ph.D. graduate program in artificial intelligence of the Artificial Intelligence Research Center at the University of Veracruz. He is also a member of the IEEE Computational Intelligence Society Evolutionary Computation Technical Committee and the IEEE Systems, Man and Cybernetics Society Soft Computing Technical Committee. Dr. Mezura Montes is also a member of the Mexican National Researchers System Level 1 and a regular member of the Mexican Computing Academy.



Ligia Quintana-Torres was born in Mexico City in 1972. She received her diploma degree in mathematics from Universidad Veracruzana in 1995, studied for a Master's in artificial intelligence and received a Ph.D. degree in mathematics from Universidad Veracruzana in 2016. She is interested in optimization and mathematics applied to computer science.

Received: 9 January 2016

Revised: 6 June 2016

Accepted: 10 August 2016