J Math Imaging Vis (2011) 41:147–167
DOI 10.1007/s10851-011-0284-0

Transportation Distances on the Circle

Julien Rabin · Julie Delon · Yann Gousseau

Published online: 6 May 2011

Abstract This paper is devoted to the study of the Monge-Kantorovich theory of optimal mass transport, in the special case of *one-dimensional and circular distributions*. More precisely, we study the Monge-Kantorovich problem between discrete distributions on the unit circle S^1, in the case where the ground distance between two points x and y is defined as $h(d(x, y))$, where d is the geodesic distance on the circle and h a *convex and increasing function*. This study complements previous results in the literature, holding only for a ground distance equal to the geodesic distance d. We first prove that computing a Monge-Kantorovich distance between two given sets of pairwise different points boils down to cut the circle at a well chosen point and to compute the same distance on the real line. This result is then used to obtain a dissimilarity measure between 1-D and circular discrete histograms. In a last part, a study is conducted to compare the advantages and drawbacks of transportation distances relying on convex or concave cost functions, and of the classical L^1 distance. Simple retrieval experiments based on the hue component of color images are shown to illustrate the interest of circular distances. The framework is eventually applied to the problem of color transfer between images.

Keywords Optimal mass transportation theory · Earth Mover's Distance · Circular histograms

1 Introduction

The theory of optimal transportation was first introduced by Monge [27] in its *Mémoire sur la théorie des déblais et des remblais* (1781) and rediscovered by Kantorovich [19] in the late '30s. The Monge-Kantorovich problem can be described in the following way. Given two probability distributions f and g on X and c a nonnegative measurable cost function on $X \times X$, the aim is to find the optimal transportation cost

$$\mathrm{MK}_c(f, g) := \inf_{\pi \in \Pi(f,g)} \iint_{x,y} c(x, y)\, d\pi(x, y) \tag{1}$$

where $\Pi(f, g)$ is the set of probability measures on $X \times X$ with marginals f and g (such measures are called transportation plans). The existence, uniqueness and behavior of optimal transportation plans has been thoroughly studied in the last decades [1, 13, 24, 25, 44, 45].

This framework is nowadays widely used in many fields of research, such as cosmology [10], meteorology [7], fluid mechanics or electromagnetic (see [1] for a complete review).

The use of the Monge-Kantorovich framework in image processing and computer vision has been popularized by Rubner et al. [40] for image retrieval and texture classification with the introduction of the so-called Earth Mover's Distance (EMD). Although the definition of the EMD is

This work has been supported by the French National Research Agency (project OTARIE) under grant BLAN07-2_183172.

J. Rabin (✉)
CMLA, ENS de Cachan, 61, Avenue du Président Wilson, 94230 Cachan, France
e-mail: julien.rabin@cmla.ens-cachan.fr

J. Delon · Y. Gousseau
CNRS LTCI, Télécom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France

J. Delon
e-mail: julie.delon@telecom-paristech.fr

Y. Gousseau
e-mail: yann.gousseau@telecom-paristech.fr

slightly different from the original Monge-Kantorovich formulation, these are equivalent when considering distributions having the same total weight. In the following years and up to now, a large body of works has relied on the use of such distances for image retrieval, see e.g. [12, 15, 20, 23, 34, 48]. This extensive use of transportation distances is largely due to their robustness when comparing histograms or discrete distributions. For the same reason, these distances are also successfully used to compare local features between images, see [21, 33–36]. Other uses of transportation distances for images include: image registration [16], image morphing [49] or junction and edge detection [39].

A strong limitation of transportation distances is their computational cost. Standard approaches quickly become intractable when dealing with a large amount of data in dimensions more than two. Indeed, the simplex algorithm, interior point methods or the Hungarian algorithm all have a complexity of at least $O(N^3)$ (N being the size of the data, either the number of samples or the number of histogram bins). Therefore, several works have proposed to speed up the computation or the approximation of optimal transport, in particular in the field of image processing, where the amount of data is often massive, see [11, 17, 21, 42]. One particular case in which the computation is elementary and fast is the case of *one-dimensional histograms*, for which it is well known that optimal transport, in the case of a *convex cost function*, is equivalent to the pointwise difference between cumulative distribution functions [44]. A question that arises is then the possibility to perform such simple and efficient computations in the case of *circular* histograms, i.e. histograms in which the first and last bins are neighbors.

Indeed, circular histograms are especially important in image processing and computer vision. First, the local geometry is often efficiently coded by the distribution of gradient orientations. Such representations offer the advantage of being robust to various perturbations, including noise and illumination changes. This is in particular the case for the well known SIFT [22] descriptor and its numerous variants. In such a situation, the comparison of local features reduces to the comparison of one-dimensional *circular* histograms. Other local features involving circular histograms include the so-called Shape Context [3]. Second, the color content of an image can be accounted for by its *hue*, in color spaces such as HSV or LCH. In such cases again, information is coded in the form of circular one-dimensional histograms. Several works in the field of computer vision have explicitly addressed the use of transportation distances in the case of circular histograms, either using thresholded concave cost functions [33, 34] or L^1 cost functions [33, 35, 36].

Contributions The first contribution of this paper is to give a general formulation of the optimal transportation cost when the ground cost is a convex function of the Euclidean distance on the circle. This formulation gives a practical way to compute distances in linear time in this case. This study extends previous results holding for L^1 cost functions, either between sets of points [4, 5, 43, 46, 47] or between discrete histograms [33–36]. Second, we provide various experiments dealing with image manipulation or retrieval, for which the interest of circular transportation distances is shown. We conclude with a discussion (that actually applies to both circular and non-circular cases) on the respective interest of transportation distances with either convex or concave cost functions and the classical L^1 bin-to-bin distance.[1] It is shown that the choice between these three family of distances should essentially be driven by the type of perturbation the histograms are likely to suffer from.

Outline The paper is organized as follows. In Sect. 2, the optimal transportation flow of the Monge-Kantorovich problem is investigated in the circular case for any convex cost function. The definition of this problem being recalled, a new formula is introduced and a sketch of the proof is proposed (details of the proof are provided in the Appendix).

In Sect. 3, several experiments are performed to analyze the practical interest of such a dissimilarity measure for various cost functions described in Sect. 3.1. First, a comparison of time complexity with other algorithms proposed in literature is given in Sect. 3.2. Then, in Sect. 3.3, a discussion about the robustness and the limitations of Monge-Kantorovich distances for histogram comparison is given, along with some simple experiments of image retrieval. Eventually, an application to hue transfer between images is proposed in Sect. 3.4.

2 The Monge-Kantorovich Transportation Problem on the Circle

In this section, we present some results on the Monge-Kantorovich transportation cost between two circular histograms. In particular, we give an analytic formulation of this cost when the ground cost between points on the circle can be written as an increasing and convex function of the Euclidean distance along the circle.

2.1 Definitions

Consider two discrete and positive distributions

$$f = \sum_{i=1}^{N} f[i]\delta_{x_i} \quad \text{and} \quad g = \sum_{j=1}^{M} g[j]\delta_{y_j}, \tag{2}$$

[1] The term "bin-to-bin" coins the fact that with such a distance, only bins having the same index are compared.

where $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_M\}$ are two sets of points on a subset Ω of $\mathbb{R}^K$. Assume that these distributions are normalized in the sense that $\sum_{i=1}^N f[i] = \sum_{j=1}^M g[j] = 1$. Let $c : \Omega \times \Omega \mapsto \mathbb{R}^+$ be a nonnegative cost function (called *ground cost*), the quantity

$$\mathrm{MK}_c(f,g) := \min_{(\alpha_{i,j}) \in \mathcal{M}} \sum_{i=1}^N \sum_{j=1}^M \alpha_{i,j} c(x_i, y_j), \quad \text{with} \tag{3}$$

$$\mathcal{M} = \Big\{ (\alpha_{i,j}) \in \mathbb{R}^N \times \mathbb{R}^M ; \alpha_{i,j} \geq 0, \sum_i \alpha_{i,j} = g[j], \sum_j \alpha_{i,j} = f[i] \Big\}, \tag{4}$$

is called the *optimal transportation cost* between f and g for the ground cost c. Matrices $(\alpha_{i,j})$ in $\mathcal{M}$ are called *transport plans* between f and g. If $(\alpha_{i,j})$ is optimal for (3), we say that $(\alpha_{i,j})$ is an *optimal transport plan*.

Let d be a distance on Ω and assume that the ground cost can be written $c(x,y) = d(x,y)^\lambda$, with the convention $d(x,y)^0 = \mathbb{1}_{x \neq y}$. It can be shown [44] that

- when $\lambda \in [0, 1[$, MK_{d^λ} is a *distance*[2] between probability distributions;
- when $\lambda \in [1, \infty[$, $(\mathrm{MK}_{d^\lambda}(f,g))^{\frac{1}{\lambda}}$ is also a distance between probability distributions.

These distances are called *Monge-Kantorovich distances*, or Wasserstein distances. For $\lambda = 1$, MK_d is also known as the Kantorovich-Rubinstein distance, or in computer vision as the *Earth Mover's Distance* (EMD), as introduced by Rubner in [40].[3]

Computing optimal transportation costs is generally time consuming. The main exception is the case of the real line: if $\Omega = \mathbb{R}$, and if the cost c is a convex and increasing function of the Euclidean distance $|x-y|$, then the optimal transport plan between f and g is the monotone rearrangement of f onto g, which sends the mass starting from the left. This result is usually false if c is not a convex function of the Euclidean distance on the line.

In the following, we take interest in the case where Ω is a circle S^1 of perimeter 1, and where c is an increasing function of the geodesic distance d along the circle. In particular, we will see that the previous result on the line can be adapted in this case. As mentioned in the introduction, the interest for the circle S^1 is motivated by its common use to describe images, e.g. to represent color or orientation.

[2]Throughout this paper, a "distance" function is a function that satisfies the three following properties: positive definiteness, symmetry and triangle inequality.

[3]Observe that the EMD has been introduced by Rubner [40] for unbalanced problems, for which f and g do not necessarily have the same total weight.

2.2 Optimal Transportation for Convex Functions of the Distance

The main result of this section is an analytic formulation of the optimal transportation cost between the discrete distributions f and g on the circle S^1 when the ground cost c can be written $c(x,y) = h(d(x,y))$, with $h : \mathbb{R} \to \mathbb{R}^+$ an increasing and convex function and d the geodesic distance along the circle. This formulation generalizes to convex cost functions different results presented in [4, 5, 33–36, 43, 46, 47]. In the following, we use the same notations for points on the circle and their coordinates along the circle, regarded as variables taking their values on the reduced interval $[0, 1[$ (*modulo* 1). It follows that d can be written

$$d(x,y) = \min(|x-y|, 1-|x-y|). \tag{5}$$

The distributions f and g on S^1 can be seen equivalently as periodic distributions of period 1 on $\mathbb{R}$.

Let us define the cumulative distribution function of f on $[0, 1[$ as

$$\forall y \in [0,1[, \quad F(y) = \sum_{i=1}^N f[i] \cdot \mathbb{1}_{\{x_i \in [0,y[\}}. \tag{6}$$

F is increasing and left continuous, and can be extended on the whole real line with the convention $F(y+1) = F(y) + 1$. This boils down to consider f as a periodic distribution on $\mathbb{R}$. We define also the pseudo-inverse of F as $F^{-1}(y) = \inf\{t;\ F(t) > y\}$. The interest of these definitions lies in the next result.

2.2.1 An Analytic Formulation of Optimal Transportation on the Circle

Theorem 1 *Assume that d is given by (5) and that the ground cost c can be written $c(x,y) = h(d(x,y))$, with $h : \mathbb{R} \to \mathbb{R}^+$ an increasing and convex function. Let f and g be two discrete probability distributions on the circle, with cumulative distribution functions F and G, and let G^α denote the function $G - \alpha$. Then*

$$\mathrm{MK}_c(f,g) = \inf_{\alpha \in \mathbb{R}} \int_0^1 h(|F^{-1} - (G^\alpha)^{-1}|). \tag{7}$$

Idea of the proof This result is a generalization of the real line case, where it is well known [44] that the global transportation cost between two probability distributions f and g can be written

$$\mathrm{MK}_c(f,g) = \int_0^1 h(|F^{-1} - G^{-1}|). \tag{8}$$

Figure 1 illustrates this formula for the special case $c(x,y) = |x-y|$. A proof of (7) in a continuous setting has been

proposed very recently in [8], where it is shown that this equation holds for any couple of probability distributions. However, this proof involves some complex notions of measure theory which are not needed in the discrete setting. For the sake of completeness and simplicity, we provide in the Appendix a simpler proof of these theorem in the case of discrete distributions. The proof first focus on the case where f and g can be written as sums $f = \frac{1}{P}\sum_{k=1}^{P}\delta_{x_k}$, and $g = \frac{1}{P}\sum_{k=1}^{P}\delta_{y_k}$, where $\{x_1,\ldots,x_P\}$ and $\{y_1,\ldots,y_P\}$ are discrete sets of points on the unit circle S. When the points are all pairwise different, we show that the circle can always be "cut" at some point, such that computing the optimal transport between f and g boils down to compute an optimal transport between two distributions on the real line (see Fig. 2). This result is proven first for strictly convex functions h and for any optimal transport plan, then for any convex function h and a well chosen optimal plan. Once the problem has been reduced to the real line, formula (7) follows from the fact that the optimal transport on $\mathbb{R}$ is given by the ordering of the points. The generalization of this formula to any kind of discrete distribution results from the continuity of the global transport cost $\mathrm{MK}_c(f,g)$ in the values of the masses and their positions on the circle.

In practice, formula (7) can be computed for any convex function h at a precision ε with a complexity in $O((N+M)\log\frac{1}{\varepsilon})$ [8], where N and M are the respective numbers of points in the distributions f and g (i.e. the number of masses in the distributions).

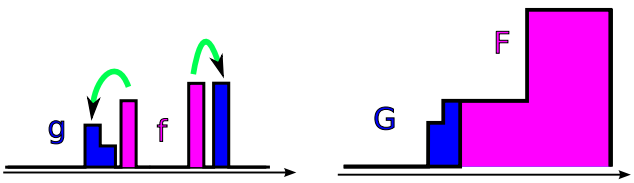


Fig. 1 Using $c(x,y) = |x-y|$ on the real line, the optimal transportation plan between two discrete histograms f and g is the L^1 distance of the difference between the cumulative histograms F and G (formula (8))

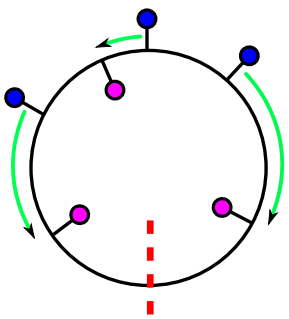
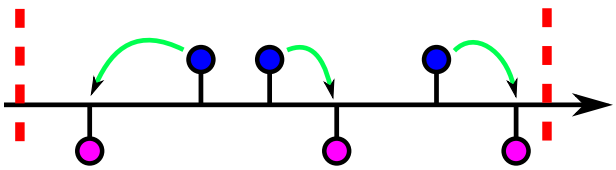


Fig. 2 When the distributions are sums of unitary different masses, and when the ground cost is a nonnegative, increasing and convex function of the distance along the circle, there is a (non-necessarily unique) "cut" on the circle such that the optimal transportation on S^1 boil down to optimal transportation on the real line



2.2.2 The Case $c(x,y) = d(x,y)$

If h is a power function $x \mapsto x^\lambda$, with $\lambda \geq 1$, Theorem 1 gives us a way to compute Monge-Kantorovich distances between f and g:

$$(\mathrm{MK}_{d^\lambda}(f,g))^{\frac{1}{\lambda}} = \left(\inf_{\alpha\in\mathbb{R}}\int_0^1 |F^{-1} - (G^\alpha)^{-1}|^\lambda\right)^{\frac{1}{\lambda}}. \tag{9}$$

Observe that the infimum of (9) is obtained for a value $\hat{\alpha}$ which depends on λ. Moreover, the optimal cut on the circle $\hat{\alpha}$ is shown in the Appendix (see (27)) to belong to the set of N values $\{F(x) - G(x), x \in [0,1[\}$.

In the case $\lambda = 1$ (i.e. when the ground cost c is the distance d along the circle), this result can be rewritten

$$\mathrm{MK}_d(f,g) = \inf_{\alpha\in\mathbb{R}} \|F - G - \alpha\|_1 = \inf_{\alpha\in\mathbb{R}}\int_0^1 |F - G - \alpha|, \tag{10}$$

where we denote by $\|.\|_1$ the L^1 norm on $[0,1[$.

In practice, since $F-G$ is piecewise constant for discrete distributions, the infimum of (10) can be computed easily by computing the weighted median of the (finite number of) values $F(t) - G(t)$ when $t \in [0,1[$, the weights being the lengths of the intervals on which $F - G$ is constant. In practice, this yields a $O(N)$ exact algorithm to compare two normalized distributions of N points with different masses using linear time weighted median algorithms (see [14] for a review).

To the best of our knowledge, formulation (10) has been known since [46], where it is proved for sets of points with unitary masses on the circle. A similar result is shown in [4, 5] for the Kantorovich-Rubinstein problem, which is known to be equivalent (see [44], chapter 1) to the Monge-Kantorovich problem when the cost $c(x,y)$ is a distance, which is true for $\lambda = 1$ (but false for $\lambda > 1$). All of these papers notice that computing (10) can be done with a $O(N)$ algorithm.

In the case of circular histograms, i.e. discrete circular distributions living on a uniform grid of N bins, bins 0 and $N-1$ are neighbors. If the cost c is $c(i,j) = \min(|i -$

$j|, N-|i-j|)$ along the circle, formula (10) can be rewritten

$$\mathrm{MK}_d(f,g) = \inf_{\alpha} \sum_{i=0}^{N-1} |F[i] - G[i] - \alpha| = \|F - G - \mu\|_1, \tag{11}$$

where $\|.\|_1$ is the discrete L^1 norm and where μ is the median of the set of values $\{F[i] - G[i], 0 \le i \le N-1\}$. This median value can be computed in linear time. The corresponding linear algorithm is presented in particular in [33], where the distance between histograms is designed by the acronym $\mathrm{EMD}_{\mathrm{MOD}}$.

An equivalent distance is proposed independently of [33] in [35, 36] and designed by the acronym CEMD, for Circular Earth Mover's Distance. Indeed, it is easily checked that the distance defined by formula (11) is equivalent to CEMD, defined as

$$\mathrm{CEMD}(f,g) = \min_{k \in \{0,\ldots,N-1\}} \|F_k - G_k\|_1, \tag{12}$$

where $F_k[i]$ is defined as $F[i] - F[k]$ if $i \in \{k, \ldots, N-1\}$ and $F[i] - F[k] + 1$ if $i \in \{0, \ldots, k-1\}$ (the definition being similar for G_k by replacing f by g). In other words, the distance $\mathrm{MK}_d(f,g)$ is also the minimum in k of the L^1 distance between F_k and G_k, the cumulative histograms of f and g starting at the k^{th} quantization bin.

2.3 Optimal Transportation for Concave Functions of the Distance

In practice, it may be useful to choose the ground cost c as a nonnegative, concave and increasing function h of the ground distance d. For instance, for the task of image retrieval, several authors [15, 39, 40] claim that good results can be achieved with a function

$$h(t) = 1 - e^{-\tau t}. \tag{13}$$

Notice that if h is increasing, concave and such that $h(0) = 0$, it is easy to show that $c = h(d)$ is also a distance, and thus MK_c is also a distance between probability distributions. Another property of concave costs is that they do not move the mass which is shared by the distributions [44]: if f and g are histograms, the problem is reduced to the transport of $(f-g).\mathbb{1}_{f-g\ge 0}$ onto $(g-f).\mathbb{1}_{f-g<0}$, which have disjoint supports.

However, in the case of such concave functions h, Theorem 1 does not apply, and there is no general and fast algorithm to compute corresponding optimal transportation costs, either on the real line, or on the circle. In most cases, we are reduced to use linear programming, i.e. simplex or interior point algorithms, which are known to have at best a $O(N^{2.5}\log(NC_{\max}))$ complexity to compare two histograms on N bins [2, 29] (where $C_{\max}$ is the maximal cost between two bins). We describe in the following some special cases of concave function h for which this complexity can be reduced.

2.3.1 L^1 as a Monge-Kantorovich Distance

If the distributions f and g are discrete histograms on N bins, and if $h(t) = \mathbb{1}_{t\neq 0}$ (one everywhere except at 0), then the Monge-Kantorovich distance between f and g is [44]

$$\mathrm{MK}_{\mathbb{1}_{d\neq 0}}(f,g) = \frac{1}{2}\sum_{i=1}^{N} |f[i] - g[i]| = \frac{1}{2}\|f-g\|_1. \tag{14}$$

Indeed, since h is a concave function, an optimal plan $(\alpha_{i,j})$ between f and g do not move the mass shared by the distributions, which means that $\alpha_{i,i} = \min(f[i], g[i])$. Now, observing that $\sum_j \alpha_{i,j} = f[i]$ for $(\alpha_{i,j}) \in \mathcal{M}$,

$$\begin{aligned}
\mathrm{MK}_{\mathbb{1}_{d\neq 0}}(f,g) &= \min_{(\alpha_{i,j})\in\mathcal{M}} \sum_{i,j} \alpha_{i,j} d(i,j) \\
&= \min_{(\alpha_{i,j})\in\mathcal{M}} \sum_i \sum_{j\neq i} \alpha_{i,j} \\
&= \sum_i (f[i] - \alpha_{i,i}) \\
&= \sum_i (f[i] - \min(f[i], g[i])) \\
&= \sum_i (f[i] - g[i]) \mathbb{1}_{f[i]\ge g[i]} = \frac{1}{2}\|f-g\|_1.
\end{aligned}$$

In other words, the L^1 distance between two normalized histograms f and g is proportional to a Monge-Kantorovich distance for the concave function $\mathbb{1}_{t\neq 0}$.

2.3.2 Thresholded Distances

In [33, 34], Pele and Werman consider thresholded ground distances, using $h(t) = \min(t, T)$, with T a given threshold. Up to a multiplicative factor, this function h can almost be seen as a discrete version of (13), where τ is chosen proportional to T. They show that in this case, the computation of the optimal cost can be solved by a min-cost-max-flow algorithm, whose complexity is smaller than classical linear programming algorithms. Observe that $T = 1$ corresponds to the case studied in Sect. 2.3.1. This thresholded distance is used in [33] with $T = 2$ to compare unbalanced circular histograms (in the sense that their total masses are not necessarily equal). In order to take into account the positive difference Δ of total mass between two non normalized histograms f and g, a new point is added to the smallest

histogram f, with weight Δ. The ground distance between this point and all other points is set to a constant β times the maximum ground distance. This variant of the Earth Mover's Distance is denoted by $\widehat{\text{EMD}}$ in [33]. In what follows, we show the connection between this distance and the classical L^1 bin-to-bin distance.

Using the thresholded distance $\min(d, 2)$, the cost $c(i, j)$ can take only three values: 0 if $i = j$, 1 if i and j are neighbors, and 2 in other cases. If we note $\alpha_{i,j}$ the quantity of mass going from bin i of f to bin j of g in the optimal transport plan, then

$$\widehat{\text{EMD}}(f, g) = \sum_{i=1}^{N} \left\{ \left(\sum_{|j-i|\geq 2} 2\alpha_{i,j} \right) + \alpha_{i,i+1} + \alpha_{i,i-1} \right\} + 2\beta\Delta. \quad (15)$$

Now, the thresholded distance is a concave function of d, which implies that all the shared mass remains in place: for a given i, $\sum_{j\neq i} \alpha_{i,j} = (f[i] - g[i])\mathbb{1}_{f[i]\geq g[i]}$. This implies that

$$\widehat{\text{EMD}}(f, g) = 2\sum_{i=1}^{N} (f[i] - g[i])\mathbb{1}_{f[i]\geq g[i]} - \sum_{i=1}^{N} (\alpha_{i,i+1} + \alpha_{i,i-1}) + 2\beta\Delta. \quad (16)$$

Observe that for unbalanced histograms f and g, $\|f - g\|_1 = 2\sum_{i=1}^{N} (f[i] - g[i])\mathbb{1}_{f[i]\geq g[i]} + \Delta$. It follows that

$$\widehat{\text{EMD}}(f, g) = \|f - g\|_1 - \sum_{i=1}^{N} (\alpha_{i,i+1} + \alpha_{i,i-1}) + \Delta(2\beta - 1). \quad (17)$$

If $\beta = 0.5$ or if f and g have the same total mass, then

$$\widehat{\text{EMD}}(f, g) = \text{MK}_{\min(d,2)}(f, g) = \|f - g\|_1 - \sum_{i=1}^{N} (\alpha_{i,i+1} + \alpha_{i,i-1}). \quad (18)$$

Now, notice that $\alpha_{i,i+1}$ is different from 0 only if $f[i] \geq g[i]$ (otherwise, all the mass in i stay in place) and $f[i + 1] < g[i + 1]$ (otherwise the mass $g[i + 1]$ is already "filled" by a part of $f[i + 1]$). In other words, the only points where the quantities $\alpha_{i,i-1}$ or $\alpha_{i,i+1}$ are different from 0 are the points where the densities of f and g are crossing.

It follows that, when the histograms are normalized, the thresholded distance $\widehat{\text{EMD}} = \text{MK}_{\min(d,2)}$ is often very close to L^1, in particular when N is large and when histograms are crossing at only a few places, as we will see in the experiments of Sect. 3.3.

In order to allow larger ground displacements, the use of values of T larger than 2 is proposed in [34]. The authors of this paper propose an algorithm that can handle any cost function and multi-dimensional histograms. This is made at the price of a non-linear time complexity and necessitates a compromise in the tuning of T (smaller values yield faster computations). We will come back on the use of such concave cost functions in the experimental section.

In the next section, we investigate the practical interest of circular Monge-Kantorovich distances over the classical L^1 bin-to-bin distance.

3 Experiments

This section is devoted to an experimental analysis of the previous optimal transportation framework. At this point it should be underlined that the following experimental study is proposed to illustrate the differences between various cost functions, and is not meant to demonstrate the superiority of a particular one. On the contrary, it will be shown in the experiments that performance may vary considerably depending on the dominant perturbation on histograms and the ground distance chosen.

In the following, we first introduce (Sect. 3.1) the cost functions compared in this section. We then investigate the practical time complexity involved by the computation of the Monge-Kantorovich measures according to the ground distance chosen (Sect. 3.2). A synthetic experiment is proposed in Sect. 3.3 to compare the relative robustness of convex and concave cost functions to various perturbations on histograms, followed by some illustrations on small examples from color image retrieval. An application to the hue transfer between images is finally studied in Sect. 3.4.

3.1 Cost Functions

The following cost functions are displayed on Fig. 3:

- Monge-Kantorovich measures with convex ground functions
  - MK_{d^p}, the Monge-Kantorovich measure[4] using d^p ground cost with $p = 1$, 2 and 3. The case $p = 1$ is equivalent to the Circular Earth Mover Distance (CEMD) defined in formula (11);
  - $\text{MK}_{H\tau}$, the Monge-Kantorovich measure using a Huber cost function with parameter $\tau = 10$:

$$c(x, y) = \begin{cases} d(x, y)^2 & \text{if } d(x, y) \leq \tau \\ \tau(2d(x, y) - \tau) & \text{otherwise} \end{cases}. \quad (19)$$

[4] We refer to MK_{d^p} transportation cost as "measure", since a normalization is required (see (9)) to get a distance.

Table 1 Comparison of average computation times for various Monge-Kantorovich measures between circular and one-dimensional histograms with $N \in \{10, 100, 1000\}$ bins

Distance	L^1	CEMD (MK_d)	MK_{d^2}	MK_{d^2}	MK_{d^2}	MK_{T10}	MK_{T2}
Formulation	Eq. (14)	Eq. (11)	Eq. (10)	[41]	[30]	[30]	[31]
			$N = 10$ bins				
Average CPU time (in seconds)	5×10^{-8}	3×10^{-7}	1.6×10^{-6}	2.7×10^{-5}	1.1×10^{-4}	1.1×10^{-4}[a]	1.45×10^{-7}
CPU time ratio (w.r.t L^1)	1	6	32	1740	2200	2200	3
			$N = 100$ bins				
Average CPU time (in seconds)	6×10^{-7}	5×10^{-6}	1.9×10^{-5}	4.6×10^{-3}	2.5×10^{-2}	3.8×10^{-3}	1.2×10^{-6}
CPU time ratio (w.r.t L^1)	1	8.3	32	7.7×10^3	4.2×10^4	6.3×10^3	2
			$N = 1000$ bins				
Average CPU time (in seconds)	7×10^{-6}	6×10^{-5}	3.4×10^{-4}	3.2	$\simeq$100	0.37	1.2×10^{-5}
CPU time ratio (w.r.t L^1)	1	8.6	49	4.6×10^5	1.4×10^7	5.3×10^4	1.7

[a]Note that with $N = 10$ bins histograms, the distance MK_{T10} is equivalent to MK_d

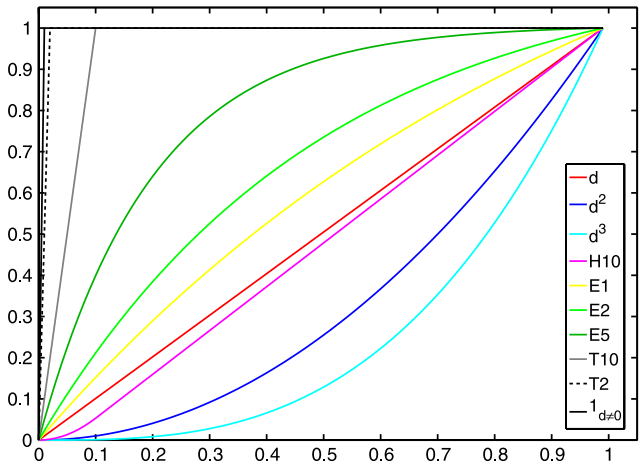


Fig. 3 Convex and concave cost functions used in this experimental section: Minkowski distance functions (referred to as d^λ, with $\lambda = 1, 2$ and 3), Huber cost function ($H10$), exponential cost function ($E\tau$, with $\tau = 1, 2$ and 5) and truncated distances ($T\tau$, with $\tau = 2$ and 10)

- Monge-Kantorovich distances with concave ground functions
  - $\mathrm{MK}_{T\tau}$ defined when using a thresholded cost function (see Sect. 2.3.2), as introduced in [33, 34], that is $c(x, y) = \min\{d(x, y), \tau\}$ with parameter $\tau = 2$ and 10;
  - the L^1 bin-to-bin distance, which can be seen as the Monge-Kantorovich distance with $\mathbb{1}_{d\neq 0}$ as ground cost (i.e. equivalent to MK_{T1}, see Sect. 2.3.1);
  - $\mathrm{MK}_{E\tau}$ defined from formula (3) when using the exponential cost function (as used for instance in [15, 39, 40]), that is $c(x, y) = 1 - \exp\{-\tau \cdot d(x, y)\}$ with parameter $\tau = 1,\ 2$ and 5.

In the following paragraph, we provide a comparative study of computation times of the different distances involved in this experimental study.

3.2 Implementation and Computation Time

An interesting aspect of the proposed closed-form solution (7) for computing Monge-Kantorovich transportation costs with convex ground distances is that it can be easily computed at a relatively small computational expense when compared with the classical L^1 distance.

In Table 1 are provided average computation times for the comparison of one-dimensional, normalized and circular histograms using Monge-Kantorovich measures with the various ground cost functions described in the previous paragraph. These average computation times have been estimated over 1000 runs on a laptop with Intel Core T9600 @ 2.80 GHz and 4 GB RAM, using C programming implementation, between pairs of histograms with $N = \{10, 100, 1000\}$ bins. The list of the different Monge-Kantorovich transportation costs and the corresponding algorithms is given below:

- The linear time L^1 distance (equivalent to MK_{T1} and $\mathrm{MK}_{\mathbb{1}_{d\neq 0}}$);
- The linear time CEMD distance (equivalent to MK_d) using formula (11);
- The MK_{d^2} distance using formula (10);
- The MK_{d^2} distance using the code of [41] based on a simplex algorithm;

- The MK_{d^2} distance using the code of [30] based on a min-cost-max-flow algorithm;
- The MK_{T10} distance using the code of [30] based on a min-cost-max-flow algorithm;
- The linear time MK_{T2} distance using the code of [31].

Firstly, observe that the MK_d distance (CEMD in formula (11)) only requires the extra computation of a median value[5] in comparison with the L^1 distance, which results in an order of magnitude increase in execution time. The complexity for computing the MK_{d^2} distance then differs in that it also requires the construction of pseudo-inverses to find the optimal cut on the circle (see formulas (10) and (7)) which again results in an order of magnitude increase in execution time. In these experiments, pseudo-inverses are estimated using the same quantization N as in the original histograms. Note that the computation times of Monge-Kantorovich measures MK_{d^3} and MK_{H10}, with respectively d^3 and Huber convex cost functions, have been omitted for the sake of simplicity, since those are almost the same than the computation time of MK_{d^2}.

Secondly, observe that the optimal transportation flow corresponding to the Monge-Kantorovich distance with non-convex ground costs can not be directly computed (see Sect. 2.3). The solving of such linear programming problem requires the use of algorithmic approaches such as the simplex algorithm, the interior point method or the min-cost-max-flow algorithm. Even if such methods are more generic (algorithms proposed by Rubner et al. [41] and Pele et al. [30] can both deal with multi-dimensional histograms and with any ground costs), it should be noted that it implies much more computation time (several orders of magnitude) and also memory limitations. Note that the computation times of Monge-Kantorovich measures $\mathrm{MK}_{E\tau}$, with exponential concave cost function, have been omitted for the sake of simplicity, since those are almost the same than the computation time of MK_{d^2} when using the simplex code of [41] or the min-cost-max-flow code of [30].

Observe that for the special case of the MK_{T2} distance, the first code of Pele et al. [31] is in linear time. As previously discussed in Sect. 2.3.2, experimental study will show (Sect. 3.3) that its performance is almost the same as the L^1 distance, which therefore strongly limits its practical interest when the number of histogram bins increases.

3.3 Comparative Analysis of Transportation Distances for Histogram Comparison

Now that we have discussed the advantage of the proposed framework according to time complexity considerations, we intend to investigate its practical interest for histogram comparison with some small-scale experiments. More precisely, the following comparative analysis is aimed at identifying situations where convex cost functions outperform classical concave cost functions, and situations in which concave costs are more interesting.

3.3.1 Synthetic Experiments

This paragraph provides a general discussion on the relative advantages of Monge-Kantorovich measures using convex cost functions, those using concave cost functions, and the L^1 bin-to-bin distance. The discussion is not specific to the circular case and will be made from non-circular examples.

Class of perturbations Several experiments (including the ones to be presented further) have led us to observe that mainly two different kinds of perturbation on histogram (illustrated in Fig. 4) are involved when using Monge-Kantorovich measures with convex or concave cost functions:

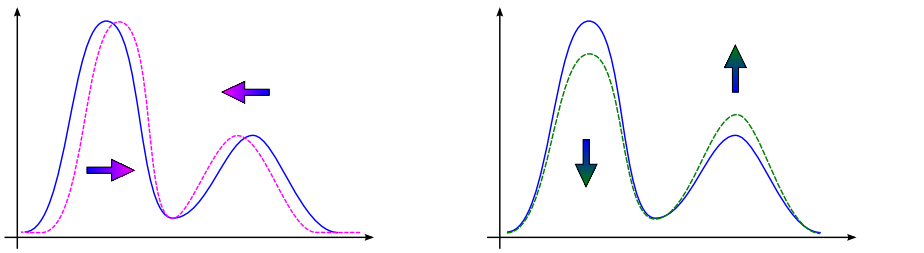
- *Shifts variability:* This happens when the modes of a histogram are shifted (see Fig. 4(a) for an illustration). For instance, this phenomenon appears with geometrical changes when dealing with orientation histograms;
- *Weights variability:* This describes the fact that the relative weights of the modes in a histogram changes (see Fig. 4(b) for an illustration). For example, this phenomenon may happen in a color histogram of an image when the sizes of object change.

Of course other phenomena can affect the performance of the Monge-Kantorovich measures (such as the quantization, the sampling process, or the presence of outlier noise). However, these perturbations cause nearly the same performance degradation, whatever the cost used. On the contrary, as it will be demonstrated in the following experiments, the shift and weight variabilities have a different impact on the Monge-Kantorovich measures' efficiency depending on the cost functions chosen.

Experimental settings We consider the Monge-Kantorovich transportation costs defined in Sect. 3.1 (Fig. 3): MK_{d^p}, with $p = 1, 2, 3$, $\mathrm{MK}_{H\tau}$, with $\tau = 10$, $\mathrm{MK}_{T\tau}$, with $\tau = 2, 10$, L^1 (equivalent to MK_{T1}) and $\mathrm{MK}_{E\tau}$, with $\tau = 1,\ 2,$ and 5[6] is computed with the linear time algorithm of Pele [31]. Recall that among these distances, only L^1, MK_d and MK_{T2} can be computed in linear time. Observe also that, these distances in a sense produce a complete range

[5]Here computed with the linear time partial quicksort algorithm described in [14].

[6]The Earth Mover's Distances with exponential and truncated cost functions have been computed using the codes kindly provided by Y. Rubner [41] and O. Pele [30]. The special case of MK_{T2}.



(a) Illustration of histogram shift (b) Illustration of histogram weight variability

Fig. 4 Illustration of the two classes of perturbations involved in histogram comparison: shift variability (*left*) and weight variability (*right*)

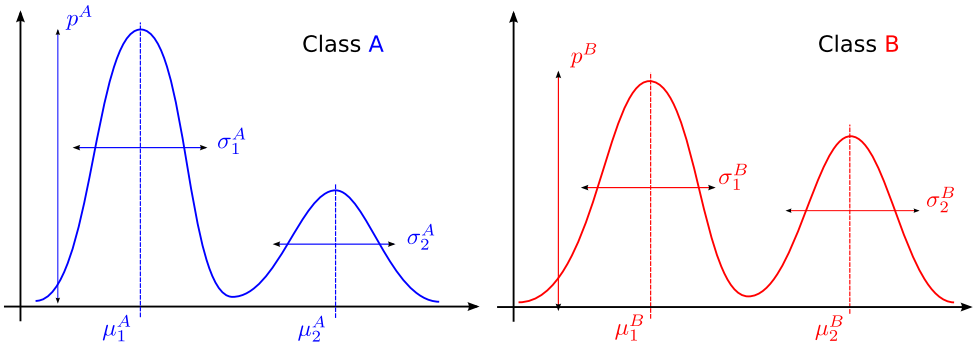


Fig. 5 The two classes A and B are defined as a Gaussian mixture model. For each class, the two Gaussian distributions are defined with 4 parameters (means and standard-deviations), plus a weighting parameter p

of alternatives between bin-to-bin distances (such as L^1) and Monge-Kantorovich associated with highly convex cost functions (e.g. MK_{d^3}). This fact will be quite clear in the following synthetic experiments.

These experiments consist, in order to study the assets of the various distances, to perform simple retrieval experiments from synthetic histograms (mixture of two Gaussians) in the presence of two types of aforementioned perturbations: shifts in the positions of modes on the one hand, and variation in the weight of modes on the other hand (see Fig. 4).

We assume that elements to be compared belong to two classes A and B, and that each element is represented by one N-bins histogram. We model the histograms as the mixture of two Gaussians. Writing $c \in \{A, B\}$ for the class, these two Gaussians have weights p^c and $(1-p^c)$, means μ_1^c and μ_2^c, and standard deviations σ_1^c and σ_2^c (see Fig. 5). In the following experiments, parameters are set as follows

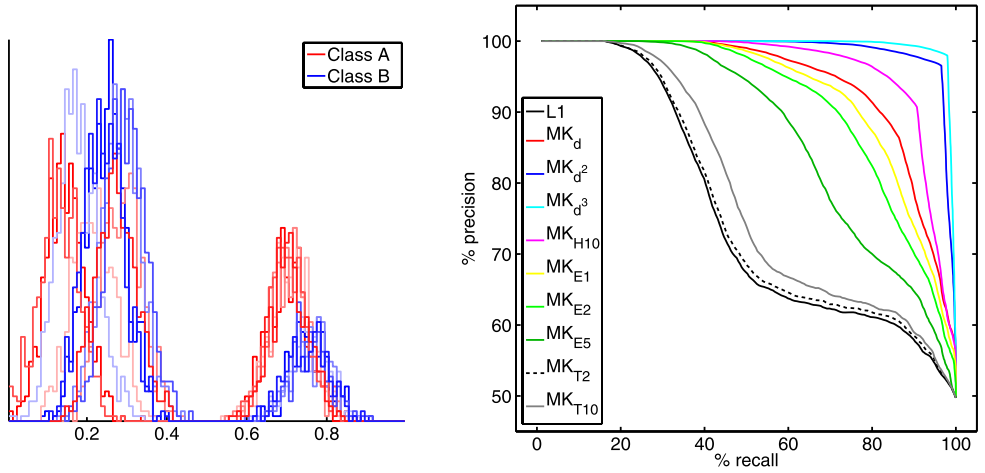
- *Histogram construction* Quantization of histograms: $N = 100$ bins; Number of samples for Gaussian mixture data generation: 1,000 samples in $[0, 1]$; Number of histograms per class: 1,000 histograms.
- *Gaussian mixture parameters* Weights: $p^A = 0.6$ and $p^B = 0.8$; Means: $\mu_2^A = \mu_2^B = 0.2$ and $\mu_2^A = \mu_2^B = 0.7$; Standard-deviations: $\sigma_1^A = \sigma_1^B = \sigma_2^A = \sigma_2^B = 0.05$.

In the following experiments, the robustness of transportation distances for histogram comparison are evaluated by displaying classical performance curves. For a given dissimilarity measure D, a retrieval performance curve is obtained by using each histogram of the dataset as a query and finding the r most similar histograms for D. For each value of r, we hence compute

- the *recall*, which is defined as the average, when the query spans the database, of the ratio between the number of correctly retrieved histograms among r and the size of the query class;
- the *precision*, which represents the average on the whole database of the rate of true positives among the r most-similar histograms.

The curves (r, recall) and (recall, precision) are drawn in the following experiments for various optimal transportation measures with respect to the two different kinds of variability that are simulated with the Gaussian mixture model.

Histogram shift We introduce random shifts in the histogram by modeling the means μ_1^c as random variables. We choose $\mu_1^A = 0.2 + \epsilon_\mu$, where ϵ_μ is uniformly drawn in $[-0.1; 0.1]$. Some of such generated histograms are superposed in Fig. 6(a). The precision-recall curves resulting from this two-class retrieval problem are plotted for different dissimilarity measures in Fig. 6(b). One first observes that distances MK_{d^p}, relying on convex cost functions, give the best results, the larger p the better. Second, it can be seen that transportation distances with concave cost function yields less efficiency. First are distances relying on an exponential cost. Eventually using transportation distances with truncated L^1 distances provides poor results, similar to those obtained with the L^1 distance. This fact is in agreement with the analysis made in Sect. 2.3.2.

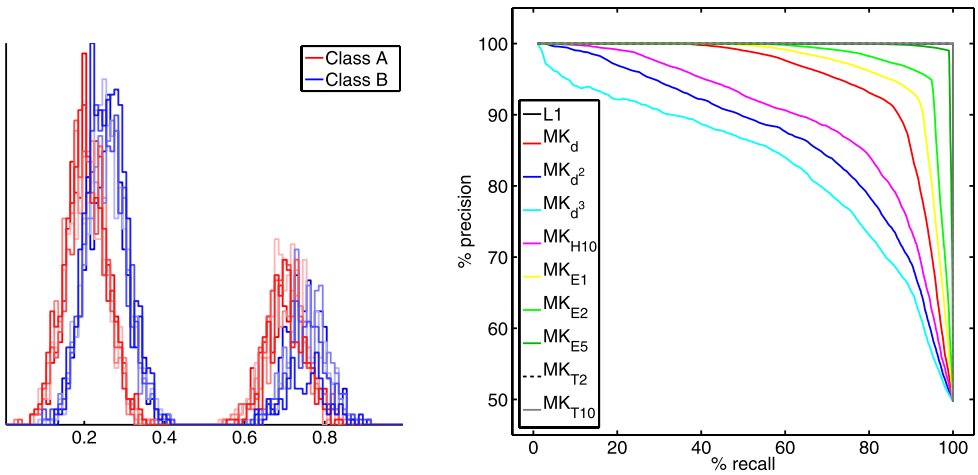


(a) Histogram shift variability experiment

(b) Average precision-recall curve

Fig. 6 Two-class retrieval problem with intraclass shift variability. The effect of the perturbation on histograms is shown in (**a**). The Precision-Recall curves are displayed in (**b**) for several transportation measures obtained with either convex or concave ground cost: $\mathrm{MK}_{T\tau}$ refers to as the transportation distance with truncated cost function according to the threshold $\tau \in \{2, 10\}$; $\mathrm{MK}_{\exp\tau}$ corresponds to the transportation distance with exponential cost function using parameter $\tau \in \{1, 2, 5\}$; MK_{d^p} is the Monge-Kantorovich measure with $p \in \{1, 2, 3\}$ and MK_{H10} is the Monge-Kantorovich measure with Huber cost function. In addition the curve obtained with L^1 metric, which is equivalent to MK_{T1} (see Sect. 2.3.1) is also shown

Fig. 7 Two-class retrieval problem with intraclass weight variability. The effect of the perturbation on histograms is shown in (**a**). The Precision-Recall curves are displayed in (**b**), plotted for different ground costs



(a) Histogram weight variability experiment

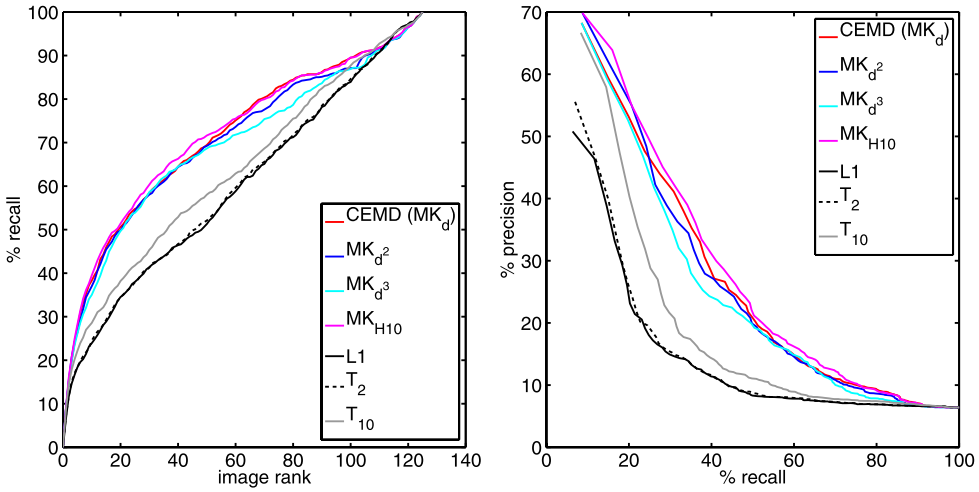
(b) Average precision-recall curve

Histogram weight variability In the second experience, intraclass weight variability are now simulated by modeling weights as random variables: $p_1^A = 0.6 + \epsilon_p$, where ϵ_p is uniformly drawn from $[-0.1; 0.1]$. Some of such generated histograms are superposed in Fig. 7(a). The precision-recall curves resulting from this two-class retrieval problem are plotted for different dissimilarity measures in Fig. 7(b). One observes that with this kind of perturbation, transportation distances with L^1 cost function are less robust than the L^1 distance. This time, it can be seen that distances with concave cost function yield better retrieval performance. Using thresholded cost functions again provides results that are very similar to those obtained with the L^1 distance. In the meantime, distances relying on exponential cost functions are still half-way between convex cost functions and thresholded cost functions.

It therefore appears that higher robustness to one type of perturbation yields poorer robustness to the other type. There is a logical tradeoff between robustness to shifts and weight variability. In this context, and given that it may be computed in linear time, the MK_1 distance appears as a good compromise in term of computational cost and robustness to the two kinds of variability considered here.

In order to illustrate the previous results on real datasets, two image retrieval experiments are shown in the next paragraph.

Fig. 8 Example of a category of 9 pictures extracted from the image database used for image retrieval (results are shown in Fig. 9). These photographs represent the same scene under various illumination conditions and camera settings



(a) Average recall rate according to the rank

(b) Average precision-recall curve

Fig. 9 (Color online) Retrieval on a color image database. Average performance curves are displayed for several similarity measures: *black curves* corresponds to the L^1 bin-to-bin distance; the Monge-Kantorovich distance MK_d (equivalent to the CEMD distance) is displayed in *red*, MK_{d^2} is plotted in *blue* and MK_{d^3} in *cyan*; the Monge-Kantorovich measure MK_{H10} using Huber ground cost is shown in *magenta*; the Monge-Kantorovich distances with truncated cost functions MK_{T2} and MK_{T10} are displayed respectively in *black dashed line* and *gray continuous line*. In this case, convex ground costs yield better performance than concave ones

3.3.2 Toy-Examples on Color Image Retrieval

In [36], it has been already demonstrated that the Circular Earth Mover's Distance (or CEMD, formula (12)) is far more robust than classical bin-to-in distances (L^1 and L^2 metric, χ^2 distance, etc.) to compare SIFT descriptors. In particular, it is underlined that this cross-bin distance is more adapted to two kinds of perturbations (quantization and histogram shifts).

In this section, we illustrate the previous experimental observations through image retrieval experiments. These are based on circular histogram comparison from real and small-scale data. We emphasize here that these experiments do not pretend to achieve state-of-the art results, and are just an experimental illustration of the previous study.

First experiment on color image retrieval For the task of color image retrieval, numerous studies have shown that the Earth Mover's Distance (defined in Sect. 2.1) often achieves better retrieval performance than bin-to-bin distances [9, 12, 15, 20, 23, 34, 39, 40, 48]. In order to illustrate the advantages of the Monge-Kantorovich distances for circular data and in the same context, we rely on hue distributions to perform simple image retrieval experiments on a color image dataset. The dataset[7] contains 14 categories of 9 pictures of the same object, with various camera settings (sensitivity, with or without flash, white balance reference, exposure time, etc.). Nine pictures of the same category are shown as an example in Fig. 8. Each of the $P = 14 \times 9 = 126$ images of the dataset is described by a hue (channel H of the HSV representation) distribution, built on $N = 360$ bins.

In order to show the possible interest of using transportation distances with convex costs, we compare several Monge-Kantorovich measures (see Sect. 3.1) for which recall and precision curves are displayed on Fig. 9.

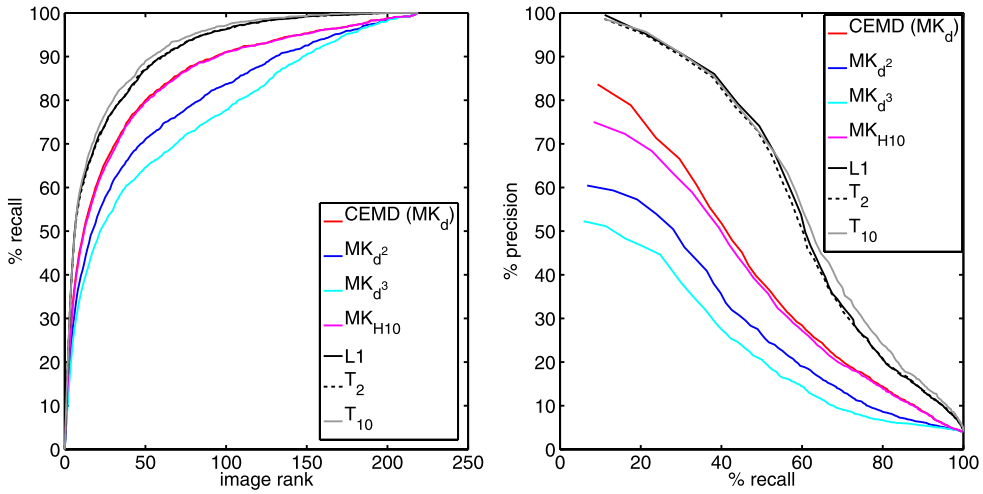
In this experiment, the results of CEMD—equivalent to MK_d—(in red curve), like other Monge-Kantorovich measures with convex ground costs (i.e. MK_{d^2}, MK_{d^3}, and MK_{H10} in blue, cyan and magenta curves), clearly outperform those of L^1 (black curve)—and also the related Monge-Kantorovich distances with truncated ground costs (i.e. MK_{T2}, and MK_{T10} displayed respectively in gray and dashed line). As in the previous synthetic experiment, one can guess that the general superiority of Monge-Kantorovich measures over bin-to-bin distances like L^1 are

[7] The image dataset is available at the following address: http://perso.telecom-paristech/~rabin/database/.

(a) Image dataset for second retrieval experiment

(b) Example of color temperature modification

Fig. 10 22 pictures used for the image retrieval under white balance perturbation (results are shown in Fig. 11)



(a) Average recall rate

(b) Average precision-recall curve

Fig. 11 (Color online) Retrieval results with data corrupted by white balance modification (color temperature correction). Average performance curves are displayed for several similarity measures: the L^1 bin-to-bin distance (in *black*); the Monge-Kantorovich measures MK_{d^p} (for $p = 1, 2$, and 3 respectively displayed in *red*, *blue* and *cyan*) and MK_{H10} (in *magenta*); the Monge-Kantorovich distances with truncated cost functions MK_{T2} and MK_{T10}, shown respectively in *black dashed line* and *gray continuous line*. In contrast with the results of Fig. 9 concave costs yield better performance in this case

due to their natural robustness to shifts in the distributions. Nevertheless, it also becomes clear at this point that the choice of the ground cost is crucial regarding the performance.

Second experiment on color image retrieval In this paragraph, we build another experimental setting showing that convex ground costs do not systematically outperform concave ones for hue comparison, and that the results of the previous experiment should be taken cautiously, as discussed in Sect. 3.3.1.

A small dataset[8] of 22 photographs is shown in Fig. 10(a). For each picture of this dataset, synthetic modifications are proceeded in order to simulate *white balance* correction with a "color temperature" varying from 4400 to 6200 K (an example is given in Fig. 10(b)). Results of retrieval are shown in Fig. 11.

[8]The image dataset is available at the following address: http://perso.telecom-paristech/~rabin/database/.

Under color temperature modification (Fig. 11), one observes the following result: L^1 distance together with MK_{T2} and MK_{T10} provide better retrieval scores than Monge-Kantorovich measures with convex costs (CEMD, MK_{d^2}, MK_{d^3} and MK_{H10}). An examination of the results has led us to observe that, in such a case, the intraclass variability results this time mainly from *differences of weights* of dominant modes in histograms (see Fig. 4(b) for an illustration).

In the next section, we illustrate another interest aspect of the proposed circular Monge-Kantorovich cost to perform color transfer between pair of images.

3.4 Application to Hue Transfer Between Images

Aside from the time complexity and histogram comparison studies of previous sections, another powerful aspect of the Monge-Kantorovich optimal transport is discussed here: the use of the optimal transport flow when considering periodic data. The aim of this section is to illustrate the interest of the proposed framework to transfer a hue distribution from one image to another.

We first give a brief recall about color transfer, which has been extensively studied in image processing, in particular within the optimal transport framework. An extension of this methodology to hue distribution using circular Monge-Kantorovich measures is then studied.

Color transfer First, let us recall that histogram equalization and more generally histogram specifications are merely particular cases of optimal transportation on the real line. Indeed, if u is a discrete image and h_u its gray level distribution, histogram specification consists in finding the optimal transport plan between h_u and a target discrete probability distribution h_t (one speaks of histogram equalization when h_t is a constant distribution). If one considers a cost c equal to the Euclidean distance on the line, then, as explained in Sect. 2.1, the solution of this problem consists in a monotone rearrangement. This rearrangement is obtained by applying the function $H_t^{-1} \circ H_u$ to u, where H_u (resp. H_t) is the cumulative distribution function of h_u (resp. h_t) and H_t^{-1} represent the pseudo-inverse of H_t (see definition in Sect. 2.2, after formula (6)). If u is a color image, such contrast adjustments can be applied to its "intensity" channel (e.g. the channel "Value" in the HSV representation).

Within the Monge-Kantorovich framework, this methodology naturally extends to color histograms (e.g. 3-D histograms) when considering color transfer between two images (see for instance [28]). Nevertheless, because of the time complexity of the aforementioned exact algorithms (i.e. the simplex based algorithm of [40] and the min-cost-max-flow algorithm of [34]) for high-dimensional histograms (2^{24} bins for standard images with 3 channels of 8 bits), only approximate methods can be practically performed, such as [32] and [38]. In order to cope with this dimensionality problem without any approximation, we propose in the following paragraph to restrict the color transfer to the hue component of the HSV representation to perform fast color transfer.

Hue transfer Thanks to formula (11) or (12) (Sect. 2.2.2), one can extend the previous framework to hue distributions, which are circular distributions. Following (11), the optimal mapping between the hue distribution h_u of an image u and the target hue distribution h_t is obtained as $(H_t - \alpha)^{-1} \circ H_u$, where α is the median of the values $\{H_u[i] - H_t[i]\}$. Figure 12 illustrates such transfer of hue on a pair of images.

It can be seen that taking into account the circularity of hue values with CEMD (Fig. 12(b)) results in a much more appealing result than with a non-circular mapping using EMD (Fig. 12(c)). Indeed, in the HSV representation, the origin of the hue is defined as the separation between red and purple values, which boils down to setting the maximum transport cost between those two colors when considering the optimal transport on the real line. This results, in the hue transfer example of Fig. 12(c) using EMD, in the partial correspondence between the light green background and the red foreground on the one hand, and between the dark green background and the blue foreground in the other hand, instead of matching green colors together as with CEMD. For this example, defining hue transfer from optimal transport on the circle with other convex ground cost such as L^p with $p > 1$ gives results similar to CEMD (another example illustrating this fact is shown in the following).

It should be noted that such approaches based on optimal transport are well known to produce some unpleasant artefacts (spatial irregularities) and thus often require some post-processing regularization techniques (for a detailed survey on color transfer between images, we refer the reader to [32, 37]). Using convex ground costs is hence more appropriate than concave costs since it has the interesting property that hue ordering on the circle is preserved, and thus reduces artefacts. On the contrary, using concave costs yields optimal transport flows on the circle which may reverse local ordering. As a result, it produces strong artifacts, since similar colors can have very different mappings. To experiment hue transfer with concave cost functions, we first compute the histogram of hue of each image with a quantization step of 256 bins before computing the optimal flow between the two histograms. This quantization step circumvents the problem of computation time for large histograms mentioned in Sect. 3.2.

Examples of hue transfers using convex and concave cost functions are shown in Figs. 13 and 14. First, as claimed before, observe that optimal transport with very convex cost, such as the $\mathrm{MK}_{d^{10}}$ measure on Fig. 14(a), yields similar hue transfer than with CEMD (Fig. 13(b)). Moreover, it can be seen that, as expected, using concave costs (MK_{E5} measure in Fig. 14(c)) gives very different visual results in comparison with convex costs. Observe that using thresholded cost functions is also not appropriate when considering transportation flow since very different flows will corresponds the same optimal transport cost (see the hue transfer obtained by MK_{T10} distance on Fig. 14(b)).

Eventually, artefacts still arise with convex costs (see for instance the feather on the hat of Lena on Figs. 13(b) and 14(a)) so that an interesting approach would combine the proposed method for hue transfer with the regularization method proposed in [37], but such a study is beyond the scope of this paper.

(a) Original images before hue transfer with optimal transport.

(b) Hue transfer with optimal transport using geodesic on the circle as ground cost (CEMD).

(c) Hue transfer with optimal transport on the real line (EMD).

Fig. 12 (Color online) Hue transfer between images. *First row*: original images. *Second row*: the hue channel of each image has been modified by applying the circular optimal transportation flow of CEMD between the hue channels (see text for details), while other channels ("saturation" and "value" in the HSV representation) are kept unchanged. *Third row*: result of hue transfer when using optimal transportation flow on the real line, i.e. without taking into account the circularity of the definition of the hue

4 Conclusion

In this paper, we have proposed a new formulation (and the proof of this formulation) for Monge-Kantorovich measures on the circle MK_c, when the ground cost c is a convex and increasing function of the geodesic distance on the circle. In the particular case where the cost function is the geodesic distance on the circle d, it has been shown that the transportation distance MK_d between circular histograms (also referred to as CEMD, standing for Circular Earth Mover's Distance) can be deduced by a very simple formula (11) which is computed in linear time.

A comparative analysis of transportation measures with different cost functions has been proposed, considering two types of perturbations which arise with histogram repre-

(a) Original images before hue transfer with optimal transport.

(b) Hue transfer with optimal transport using geodesic on the circle as ground cost (CEMD).

(c) Hue transfer with optimal transport on the real line (EMD).

Fig. 13 (Color online) Other example of hue transfer between images. *First row*: original images. *Second row*: the hue channel of each image has been modified by applying the circular optimal transportation flow of CEMD between the hue channels (see text for details), while other channels ("saturation" and "value" in the HSV representation) are kept unchanged. *Third row:* result of hue transfer when using optimal transportation flow on the real line, i.e. without taking into account the circularity of the definition of the hue

(a) Hue transfer with optimal transport using the convex ground cost d^{10} ($\mathrm{MK}_{d^{10}}$).

(b) Hue transfer with optimal transport using the concave ground cost $min(d, T)$ with $T = 10$ (MK_{T10}).

(c) Hue transfer with optimal transport using exponential ground cost (MK_{E5}).

Fig. 14 (Color online) Hue transfer between images of Fig. 13(a) with other ground costs. *First row*: result of hue transfer when using optimal transportation flow of $\mathrm{MK}_{d^{10}}$ with very convex ground cost. *Second row*: result of hue transfer when using optimal transportation flow of MK_{T10} using truncated ground cost. *Third row*: result of hue transfer when using optimal transportation flow of MK_{E5} with very concave ground cost

sentation: shift and weight changes of dominant modes. We have demonstrated that there is a trade-off between these two phenomena when using either convex or concave cost functions for histogram comparison. Eventually, the proposed CEMD dissimilarity measure offers an interesting compromise between these two choices, while being easy to use.

Eventually, an application of the proposed framework to color transfer has been studied. Other applications could also benefit from the transportation cost MK_c, such as shape recognition based on circular descriptors (see e.g. character recognition with orientation histogram [6], and curvature based descriptor along closed contour [18, 26]).

Acknowledgements Julie Delon acknowledges the support of the *French Agence Nationale de la Recherche* (ANR), under grant BLAN07-2_183172, Optimal transport: Theory and applications to cosmological reconstruction and image processing (OTARIE), and would like to thank J. Salomon and A. Sobolevski for fruitful discussions.

Appendix: Proof of Theorem 1

This appendix provides a complete proof of Theorem 1 in the case where f and g are discrete distributions (as written in (2)). We first prove this theorem for distributions composed of unitary masses, and conclude thanks to continuity arguments.

A.1 Introduction

Consider two discrete sets of points $\{x_1, \dots, x_P\}$ and $\{y_1, \dots, y_P\}$ on the unit circle S, and the corresponding discrete distributions

$$f = \frac{1}{P}\sum_{k=1}^{P} \delta_{x_k}, \quad \text{and} \quad g = \frac{1}{P}\sum_{k=1}^{P} \delta_{y_k}, \tag{20}$$

where the notations x_k, y_k are used equally for points on the unit circle or for their coordinates in $[0, 1[$. Let d be the geodesic distance along the circle (given by (5)) and assume that c can be written $c(x, y) = h(d(x, y))$ with h a nonnegative, increasing and convex function. It is well known (this is a consequence of Birkhoff's theorem, see for example the introduction of [44]) that the optimal transportation cost between f and g equals

$$\mathrm{MK}_c(f, g) = \min_{\sigma \in \Sigma_P} W_\sigma^c(f, g), \quad \text{with}$$
$$W_\sigma^c(f, g) := \frac{1}{P}\sum_k c(x_k, y_{\sigma(k)}) = \frac{1}{P}\sum_k h(d(x_k, y_{\sigma(k)})), \tag{21}$$

where Σ_P is the set of permutations of $\{1, \dots, P\}$. In other words, finding the optimal transportation between f and g boils down to find the optimal permutation σ between the points $\{x_k\}$ and $\{y_j\}$.

A.1.1 Paths

If x and y are two different points of S^1, we note $\gamma(x, y)$ *the geodesic path linking x and y on S^1* (the path is supposed open: it does not contain x and y). This path is always unique except in the case where x and y are in opposite positions on the circle. In this case, we choose $\gamma(x, y)$ as the path going from x to y in the trigonometric direction. A path $\gamma(x, y)$ is said to be *positive* if it goes from x to y in the trigonometric direction. If the path goes from x to y in the opposite direction , it is said to be *negative*.

A.1.2 Cumulative Distribution Functions

The cumulative distribution function of f has been defined in (6). Now, on $[0, 1[$ seen as a the unit circle S, no strict order can be defined between points, which means that we can define as many cumulative distribution functions as there are starting points on the circle. If x is a point in $[0, 1[$, the x-cumulative distribution function F_x of f can be defined by choosing x as the reference point on the circle S^1 and by summing the mass in the trigonometric order from this new reference point:

$$\forall y \in \mathbb{R}, \quad F_x(y) = F(x + y) - F(x). \tag{22}$$

A.2 Preliminary Results

In the following, we prove that if f and g can be written as in (20), if the points $x_1, \dots, x_P$ and $y_1, \dots, y_P$ are pairwise different, and if σ is an optimal permutation for (21), there is always a point on the circle which is not contained in any optimal path of σ. This result is proven first for strictly convex functions h and for any optimal permutation σ, then for convex functions h and a well chosen optimal permutation.

Proposition 1 *Assume that h is strictly convex. Let $x_1, \dots, x_P$ and $y_1, \dots, y_P$ be P points in $[0, 1[$, all pairwise different. Then for each permutation σ of Σ_P which minimizes (21), there exists $k \in \{1, \dots, P\}$ such that for all $l \neq k$, $x_k \notin \gamma(x_l, y_{\sigma(l)})$.*

The proof of this proposition needs the following lemma, which describes some properties of the geodesic paths $\gamma(x_l, y_{\sigma(l)})$ obtained when σ is a minimizer of (21) and h is strictly convex.

Lemma 1 *Assume that h is strictly convex. Let σ be a minimizer of (21) and let $\gamma_l = \gamma(x_l, y_{\sigma(l)})$ and $\gamma_k = \gamma(x_k, y_{\sigma(k)})$ (with $l \neq k$) be two geodesic paths for the assignment defined by σ. Assume also that $x_l \neq x_k$ and $y_{\sigma(l)} \neq y_{\sigma(k)}$. Then, one of the following holds*:

- $\gamma_l \cap \gamma_k = \emptyset$;

- *$\gamma_l \cap \gamma_k \neq \emptyset$ and in this case γ_l and γ_k have the same direction (both positive or both negative) and neither of them is contained in the other.*

Proof Assume that $\gamma_l \cap \gamma_k \neq \emptyset$. If $\gamma_l \cap \gamma_k$ is equal to $\gamma(x_l, x_k)$, then, since h is an increasing function of d, $c(x_l, y_{\sigma(l)}) > c(x_k, y_{\sigma(l)})$ and $c(x_k, y_{\sigma(k)}) > c(x_l, y_{\sigma(k)})$, which contradicts the optimality of σ. The same conclusion holds if $\gamma_l \cap \gamma_k$ is equal to $\gamma(y_{\sigma(l)}, y_{\sigma(k)})$. Moreover, if for example the path γ_l is included in γ_k, then the strict convexity of the function h implies

$$c(x_l, y_{\sigma(l)}) + c(x_k, y_{\sigma(k)}) > c(x_l, y_{\sigma(k)}) + c(x_k, y_{\sigma(l)}),$$

which also contradicts the optimality of σ. Thus, $\gamma_l \cap \gamma_k$ is equal to $\gamma(x_l, y_{\sigma(k)})$ or to $\gamma(x_k, y_{\sigma(l)})$ and it follows that γ_k and γ_l are either both positive or both negative. □

Proof of Proposition 1 Let σ be a minimizer of (21). In the following, we will denote by γ_l the geodesic path $\gamma(x_l, y_{\sigma(l)})$. We can assume without loss of generality that the points $x_1, \dots, x_P$ are in trigonometric order on the circle.

Assume that for each $l \in \{1, \dots, P\}$, there exists $q(l) \neq l$ such that x_l belongs to the open path $\gamma_{q(l)}$. Then, for each l, we have $\gamma_{q(l)} \cap \gamma_l \neq \emptyset$, which means that the geodesic paths $\gamma_{q(l)}$ and γ_l are either both positive or both negative (from Lemma 1). Assume for instance that they are both positive and let us show that in this case $x_l \in \gamma_{l-1}$ (with $l-1 = P$ if $l = 0$). If $q(l) = l-1$, there is nothing to prove. If $q(l) \neq l-1$, it means in particular that $x_{q(l)}, x_{l-1}, x_l$ are in trigonometric order on the circle. Since $\gamma_{q(l)}$ is a positive path starting from $x_{q(l)}$ and containing x_l, it follows that $\gamma_{q(l)}$ contains x_{l-1} (recall that the points are assumed to be pairwise different, in particular $x_{l-1} \neq x_{q(l)}$). Thus $\gamma_{l-1} \cap \gamma_{q(l)} \neq \emptyset$, which implies that γ_{l-1} is positive. Now, x_l must be in γ_{l-1}, otherwise we would have $\gamma_{l-1} \subset \gamma_{q(l)}$, which contradicts Lemma 1. Thus, if the paths $\gamma_{q(l)}$ and γ_l are both positive, $x_l \in \gamma_{l-1}$.

In the same way, if $\gamma_{q(l)}$ and γ_l are both negative, then $x_l \in \gamma_{l+1}$. In any case, for each $l \in \{1, \dots, P\}$, $x_l \in \gamma_{l-1} \cup \gamma_{l+1}$ (with the obvious convention $\gamma_{P+1} = \gamma_1$, $\gamma_0 = \gamma_P$).

Now, suppose that for a given $k \in \{1, \dots, P\}$, x_k is in γ_{k-1}. Then, γ_{k-1} and γ_k have the same direction. From Lemma 1, it follows that x_{k-1} cannot be contained in γ_k. Since we know that $x_{k-1} \in \gamma_{k-2} \cup \gamma_k$, x_{k-1} must be in γ_{k-2}. Recursively, for each $l \in \{1, \dots, P\}$, $x_l \in \gamma_{l-1}$. It follows that for each $l \in \{1, \dots, P\}$, $d(x_l, y_{\sigma(l-1)}) < d(x_{l-1}, y_{\sigma(l-1)})$, and since h is increasing

$$\sum_{l=1}^{P} c(x_l, y_{\sigma(l)}) > \sum_{l=1}^{P} c(x_{l+1}, y_{\sigma(l)}), \tag{23}$$

which contradicts the fact that σ is a minimizer of (21). We come to the same conclusion if for a given $k \in \{1, \dots, P\}$, x_k is in γ_{k+1}. □

The same result can be proven for any convex function h with the difference that it is only satisfied for a good choice of the permutation σ which minimizes (21), and not for all of these permutations. This result can be seen as a limit version of Proposition 1.

Corollary 1 *Assume that h is convex. Let $x_1, \dots, x_P$ and $y_1, \dots, y_P$ be P points in $[0, 1[$. Assume that all these points are pairwise different. Then there exists a permutation σ of Σ_P which minimizes (21) and a point $x_k \in \{x_1, \dots, x_P\}$ such that for all $l \neq k$, $x_k \notin \gamma(x_l, y_{\sigma(l)})$.*

Proof We know that for any strictly convex function h, if σ_h minimizes the cost $\sigma \mapsto W_\sigma^c(f, g)$, there exists $k \in \{1, \dots, P\}$ such that for all $l \neq k$, $x_k \notin \gamma_l = \gamma(x_l, y_{\sigma_h(l)})$.

Now, assume that h is convex (not strictly). One can always find a sequence (h^n) of increasing and strictly convex functions such that h^n converges pointwise towards h when $n \to \infty$. If σ and the points $x_1, \dots, x_P, y_1, \dots, y_P$ are fixed, then the finite sum $W_\sigma^n(f, g) := \frac{1}{P} \sum_k h^n(d(x_k, y_{\sigma(k)}))$ tends towards $W_\sigma(f, g) = \frac{1}{P} \sum_k h(d(x_k, y_{\sigma(k)}))$ when $n \to \infty$. Thus, for each $\varepsilon > 0$, there exists an integer N, such that for all $n \geq N$, $|W_\sigma^n(f, g) - W_\sigma(f, g)| \leq \varepsilon$. Since Σ_P is a finite set, we can chose N large enough such that this property holds for every σ in Σ_P. We can also chose N such that $|\min_\sigma W_\sigma(f, g) - \min_\sigma W_\sigma^n(f, g)| \leq \varepsilon$. Now, if $n \geq N$ and if σ^* is an optimal permutation for $W_\sigma^n(f, g)$, it follows that

$$\begin{aligned} &|\min_\sigma W_\sigma(f, g) - W_{\sigma^*}(f, g)| \\ &\quad \leq |\min_\sigma W_\sigma(f, g) - \min_\sigma W_\sigma^n(f, g)| \\ &\qquad + |W_{\sigma^*}^n(f, g) - W_{\sigma^*}(f, g)| \\ &\quad \leq 2\varepsilon. \end{aligned}$$

Since Σ_P is a finite set, the fact that this distance can be made arbitrarily small implies that when n is large enough, a minimizer σ^* of $W_\sigma^n(f, g)$ is also a minimizer of $W_\sigma(f, g)$. This proves that there exists at least one minimizer σ of $\sigma \mapsto W_\sigma(f, g)$ such that $x_k \notin \gamma(x_l, y_{\sigma_\lambda(l)})$ for some $k \in \{1, \dots, P\}$ and all $l \neq k$. □

A.3 Proof of Theorem 1

Proof of Theorem 1 Let us begin with the case where f and g can be written as sums of unitary masses (see (20)), and where $x_1, \dots, x_P$ and $y_1, \dots, y_P$ are pairwise different. Proposition 1 and Corollary 1 show that if the ground cost c can be written $c(x, y) = h(d(x, y))$ with h a positive, convex and increasing function, we can choose some optimal

permutation σ for which there is some point x_k which is not contained in any path of σ (recall that paths are defined as open: they do not contain their boundaries). Since all points are supposed pairwise different, the only path meeting all the neighborhoods of x_k is γ_k. It follows that there exists some open set on one side of x_k and not containing x_k which does not cross any path of the optimal permutation σ. The middle x of this open set is not contained in any path of σ. We can thus cut the circle S^1 at x and reduce the transportation problem on the circle to the transportation problem on the real line. The optimal permutation σ is thus given by the sorting of the points (formula (72) in [44]), taking x as the reference point on the circle. This means that when points are pairwise different, we have

$$\mathrm{MK}_c(f,g) = \inf_{x\in S^1} \int_0^1 h(|F_x^{-1} - G_x^{-1}|), \tag{24}$$

where F_x^{-1} and G_x^{-1} are the pseudo-inverses (pseudo-inverses are defined in Sect. 2.2) of the increasing functions F_x and G_x defined in (22).

Now, observe that F_x and G_x are horizontal translations of $F - F(x)$ and $G - G(x)$ by the same vector x. In consequence,

$$\int_0^1 h(|F_x^{-1} - G_x^{-1}|) = \int_0^1 h(|(F - F(x))^{-1} - (G - G(x))^{-1}|). \tag{25}$$

If we notice that for all α, $(F-\alpha)^{-1}(t) = F^{-1}(t+\alpha)$, this can be rewritten

$$\begin{aligned}\int_0^1 h(|F_x^{-1} - G_x^{-1}|) &= \int_0^1 h(|F^{-1}(t+F(x)) - (G)^{-1}(t+G(x))|)dt \\ &= \int_{-F(x)}^{1-F(x)} h(|F^{-1}(t) - (G)^{-1}(t+G(x)-F(x))|)dt.\end{aligned}$$

Since F and G have been defined on $\mathbb{R}$ such that for all y, $F(y+1) = F(y)+1$ and $G(y+1) = G(y)+1$, the bounds of this integral can be replaced by any bounds $(t, t+1)$, and in particular by the bounds $(0,1)$. By using the fact that $G^{-1}(t+\alpha) = (G-\alpha)^{-1}(t)$, it follows that the whole integral equals

$$\int_0^1 h(|F^{-1} - (G + F(x) - G(x))^{-1}|). \tag{26}$$

Finally,

$$\mathrm{MK}_c(f,g) = \inf_{x\in S^1} \int_0^1 h(|(F)^{-1} - (G + F(x) - G(x))^{-1}|). \tag{27}$$

In order to conclude, notice that the function $\varphi : \alpha \mapsto \int_0^1 h(|(F)^{-1} - (G+\alpha)^{-1}|)$ is continuous ($h : \mathbb{R} \to \mathbb{R}^+$ is continuous since it is convex) and coercive ($\varphi(\alpha) \to +\infty$ when $|\alpha| \to +\infty$). It follows that φ reaches its minimum at a point $\alpha_0 \in \mathbb{R}$. In addition, the fact that F and G are piecewise constant implies that φ is piecewise affine, with discontinuities of φ' at points $F(x) - G(x)$. Thus,

$$\mathrm{MK}_c(f,g) = \inf_{\alpha\in\mathbb{R}} \int_0^1 h(|(F)^{-1} - (G+\alpha)^{-1}|). \tag{28}$$

The previous result can be generalized to the case where the points x_i, y_j may coincide just by remarking that both quantities in (28) are continuous in the positions of these points. In consequence, the result holds for distributions with rational masses.

In order to generalize the result to any couple of discrete probability distributions, observe that the right term in (28) is continuous in the values of the masses $f[i]$ and $g[j]$. As for the continuity of $\mathrm{MK}_c(f,g)$, assume that a mass ε of the distribution f is transferred from the point x_{i_0} to the point x_{i_1} in f, and let us call the new distribution f^ε. If (α) is an optimal transport plan between f and g, let j_0 be an index such that $\alpha_{i_0,j_0} \geq \varepsilon$. A transport plan (α') between f^ε and g can be defined as

- $\alpha'_{i_0,j_0} = \alpha_{i_0,j_0} - \varepsilon$,
- $\alpha'_{i_1,j_0} = \alpha_{i_1,j_0} + \varepsilon$,
- $\alpha'_{i,j} = \alpha_{i,j}$ for $(i,j) \neq (i_0,j_0), (i_1,j_0)$.

The corresponding transportation cost between f^ε and g is then lower than $\mathrm{MK}_c(f,g) + \varepsilon h(\frac{1}{2})$, which implies that $\mathrm{MK}_c(f^\varepsilon, g) \leq \mathrm{MK}_c(f,g) + \varepsilon h(\frac{1}{2})$. Conversely, we can show that $\mathrm{MK}_c(f,g) \leq \mathrm{MK}_c(f^\varepsilon, g) + \varepsilon h(\frac{1}{2})$. □

References

1. Ambrosio, L., Caffarelli, L.A., Brenier, Y., Buttazzo, G., Villani, C.: Optimal Transportation and Applications. Lecture Notes in Mathematics, vol. 1813. Springer, Berlin (2003)
2. Burkard, R., Dell'Amico, M., Martello, S.: Assignment Problems. SIAM, Philadelphia (2009)
3. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 509–522 (2002)
4. Cabrelli, C.A., Molter, U.M.: The Kantorovich metric for probability measures on the circle. J. Comput. Appl. Math. **57**(3), 345–361 (1995)
5. Cabrelli, C.A., Molter, U.M.: A linear time algorithm for a matching problem on the circle. Inf. Process. Lett. **66**(3), 161–164 (1998)
6. Cha, S.-H., Srihari, S.N.: On measuring the distance between histograms. Pattern Recognit. **35**(6), 1355–1370 (2002)
7. Cullen, M.J.P.: A Mathematical Theory of Large-Scale Atmospheric-Ocean Flow. Imperial College Press, London (2006)
8. Delon, J., Salomon, J., Sobolevskii, A.: Fast transport optimization for Monge costs on the circle. SIAM J. Appl. Math. **70**(7), 2239–2258 (2010)

9. Dvir, G.: Context-based image modelling. In: Proceedings of the 2002 IEEE International Conference on Pattern Recognition (ICPR), vol. 4, p. 40162. IEEE Comput. Soc., Los Alamitos (2002)
10. Frisch, U., Matarrese, S., Mohayaee, R., Sobolevski, A.: A reconstruction of the initial conditions of the universe by optimal mass transportation. Nature (2002)
11. Grauman, K., Darrell, T.J.: Fast contour matching using approximate earth mover's distance. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), pp. 220–227 (2004)
12. Greenspan, H., Dvir, G., Rubner, Y.: Region correspondence for image matching via emd flow. In: CBAIVL '00: Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'00), p. 27. IEEE Computer Society, Washington, DC (2000)
13. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. **177**(2), 113–161 (1996)
14. Gurwitz, C.: Weighted median algorithms for L1 approximation. BIT Numer. Math. **30**(2), 301–310 (1990)
15. Hurtut, T., Gousseau, Y., Schmitt, F.: Adaptive image retrieval based on the spatial organization of colors. Comput. Vis. Image Underst. **112**(2), 101–113 (2008)
16. Haker, S., Zhu, L., Tannenbaum, A., Angenent, S.: Optimal mass transport for registration and warping. Int. J. Comput. Vis. **60**(3), 225–240 (2004)
17. Indyk, P., Thaper, N.: Fast image retrieval via embeddings. In: 3rd International Workshop on Statistical and Computational Theories of Vision, Nice, France (2003)
18. Jalba, A.C., Wilkinson, M.H.F., Roerdink, J.B.T.M.: Shape representation and recognition through morphological curvature scale spaces. IEEE Trans. Image Process. **15**(2), 331–341 (2006)
19. Kantorovich, L.: On the transfer of masses. Dokl. Akad. Nauk **37**(2), 227–229 (1942) (in Russian)
20. Lv, Q., Charikar, M., Li, K.: Image similarity search with compact data structures. In: CIKM '04: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, pp. 208–217. ACM, New York (2004)
21. Ling, H., Okada, K.: An efficient Earth Mover's distance algorithm for robust histogram comparison. IEEE Trans. Pattern Anal. Mach. Intell. **29**(5), 840–853 (2007)
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
23. Liu, Y., Zhang, D., Lu, G., Ma, W.-Y.: Region-based image retrieval with high-level semantic color names. In: MMM '05: Proceedings of the 11th International Multimedia Modelling Conference, Washington, DC, USA, pp. 180–187. IEEE Comput. Soc., Los Alamitos (2005)
24. McCann, R.J.: Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. **80**(2), 309–323 (1995)
25. McCann, R.J.: Exact solutions to the transportation problem on the line. In: Proceedings: Mathematical, Physical and Engineering Sciences, pp. 1341–1380 (1999)
26. Mokhtarian, F.: Silhouette-based occluded object recognition through curvature scale space. Mach. Vis. Appl. **10**(3), 87–97 (1997)
27. Monge, G.: Mémoire sur la théorie des déblais et des remblais. In: Histoire de l'Académie Royale des Sciences (1781)
28. Morovic, J., Sun, P.L.: Accurate 3d image colour histogram transformation. Pattern Recognit. Lett. **24**(11), 1725–1735 (2003)
29. Orlin, J.: A faster strongly polynomial minimum cost flow algorithm. In: STOC (1988)
30. Pele, O.: Source code for EMD. http://www.cs.huji.ac.il/~ofirpele/FastEMD/code/
31. Pele, O.: Source code for MK_{T2}: http://www.cs.huji.ac.il/~ofirpele/SiftDist/code/
32. Pitié, F., Kokaram, A., Dahyot, R.: Automated colour grading using colour distribution transfer. Comput. Vision Image Underst., February 2007
33. Pele, O., Werman, M.: A linear time histogram metric for improved sift matching. In: ECCV08 (2008)
34. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: ICCV (2009)
35. Rabin, J., Delon, J., Gousseau, Y.: Circular Earth Mover's Distance for the comparison of local features. In: Proceedings of the IEEE International Conference on Pattern Recognition (ICPR). IEEE Computer Society, Los Alamitos (2008)
36. Rabin, J., Delon, J., Gousseau, Y.: A statistical approach to the matching of local features. SIAM J. Imaging Sci. (2009)
37. Rabin, J., Delon, J., Gousseau, Y.: Regularization of transportation maps for color and contrast transfer. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). IEEE Computer Society, Los Alamitos (2010)
38. Rabin, J., Peyré, G., Cohen, L.D.: Geodesic shape retrieval via optimal mass transport. In: Proceedings of the European Conference on Computer Vision (ECCV'10) (2010)
39. Ruzon, M.A., Tomasi, C.: Edge, junction, and corner detection using color distributions. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1281–1295 (2001)
40. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)
41. Rubner, Y.: Source code for EMD. http://robotics.stanford.edu/~rubner/
42. Shirdhonkar, S., Jacobs, D.W.: Approximate earth mover's distance in linear time. In: CVPR08, pp. 1–8 (2008)
43. Shen, H.C., Wong, A.K.C.: Generalized texture representation and metric. Comput. Vis. Graph. Image Process. **23**(2), 187–206 (1983)
44. Villani, C.: Topics in Optimal Transportation. Am. Math. Soc., Providence (2003)
45. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2008)
46. Werman, M., Peleg, S., Melter, R., Kong, T.Y.: Bipartite graph matching for points on a line or a circle. J. Algorithms **7**(2), 277–284 (1986)
47. Werman, M., Peleg, S., Rosenfeld, A.: A distance metric for multidimensional histograms. Comput. Vis. Graph. Image Process. **32**(3), 328–336 (1985)
48. Zheng, Q.-F., Wang, W.-Q., Gao, W.: Effective and efficient object-based image retrieval using visual phrases. In: MULTIMEDIA '06: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 77–80. ACM, New York (2006)
49. Zhu, L., Yang, Y., Haker, S., Tannenbaum, A.: An image morphing technique based on optimal mass preserving mapping. IEEE Trans. Image Process. **16**(6), 1481–1495 (2007)

Julien Rabin graduated from the École Normale Supérieure de Cachan, France, in 2006, and received a Ph.D. degree in Signal and Image Processing from Télécom ParisTech, Paris, France, in 2009. After working as a postdoctoral researcher at the University Paris Dauphine in 2010, France, he joined the École Normale Supérieure de Cachan for a postdoctoral position. His research interests are image enhancement, optimal transportation, object and shape recognition.

Julie Delon graduated from the École Normale Supérieure de Cachan, France, in 2001 and received a Ph.D. degree in applied mathematics in 2004. She has been a CNRS researcher at Télécom ParisTech, Paris, France since 2005. Her research interests include the mathematical modeling of images, optimal transportation, film restoration and image processing.

Yann Gousseau graduated from the École Centrale de Paris, France, in 1995, and received a Ph.D. degree in applied mathematics from the University Paris Dauphine in 2000. After working as a postdoctoral researcher at Brown University, Providence, USA, and at the École Normale Supérieure de Cachan, France, he joined Télécom ParisTech, Paris, France, where he has been an Associate Professor since 2001. His research interests include the mathematical modeling of natural images, image analysis, computer vision and image processing.