

METRIC-PRESERVING REDUCTION OF EARTH MOVER'S DISTANCE

YUICHI TAKANO¹ AND YOSHITSUGU YAMAMOTO¹

ABSTRACT. Earth mover's distance (EMD for short) is a perceptually meaningful dissimilarity measure between histograms. The computation of EMD reduces to a network flow optimization problem; however, it lays a heavy computational burden when the number of locations of histograms is large. In this paper, we address an efficient formulation for computing the exact EMD value. We prove that the EMD problem reduces to a problem with half the number of constraints regardless of the ground distance. We then propose a further reduced formula in which the number of variables is reduced from $O(m^2)$ to $O(m)$ for histograms with m locations when the ground distance is derived from a graph with a homogeneous neighborhood structure. Specifically, EMD problems with L_1 , L_∞ and D -norm ground distances can be reduced in this manner. Some experiments show that the reduction helps compute the EMD efficiently.

1. INTRODUCTION

Earth mover's distance (EMD for short) is a mathematical measure of the dissimilarity between distributions. It was first introduced by Rubner et al. [13] and has been successfully used in classification, image retrieval and multidimensional scaling [14]. It dates back to the work by Werman et al. [16] although the name EMD was not yet used. EMD is defined as the minimal working cost that must be paid to transform one distribution to the other; therefore, it has some desirable properties that other dissimilarity measures fail to have, such as reflecting the perceptual dissimilarity.

The computation of EMD between histograms with m locations reduces to a network flow optimization problem with $O(m^2)$ variables, which grows rather rapidly and makes the computation time-consuming as m grows. One possible remedy is to resort to approximation methods, e.g., [8, 15]. Indyc and Thaper [8] approximated the EMD by L_1 distance between vectors made by summing the weights in rectangular cells with increasing length of sides and proposed a fast image retrieval algorithm. Based on their work, Grauman and Darrell [5, 6] made a contour matching algorithm and proposed a pyramid match kernel. Another approximation method is a wavelet EMD by Shirdhonkar and Jacobs [15]. They consider the dual problem of EMD using the wavelet transform and compute an approximate EMD by a weighted sum of wavelet coefficients. Among other interesting findings are high-dimensional EMD [2, 17], EMD under translations, rigid motions and similarity operations [9].

In contrast to those approximation approaches, Ling and Okada [12] proposed a new efficient formulation EMD- L_1 to compute the exact EMD value. When the ground distance, which is the distance between locations, is L_1 , their formulation is equivalent to the EMD problem. Motivated by their work, we will propose a reduced EMD formulation and prove its equivalence to the original EMD problem via the flow decomposition theorem regardless of the ground distance employed. We also show that the number of variables of the reduced EMD formulation is reduced from $O(m^2)$ to $O(m)$ when the ground distance is derived from a graph with a homogeneous neighborhood structure. This property lightens the computational burden when the ground distance is L_1 , L_∞ or D -norm. In experimental results, we will show that EMD is a perceptually reasonable dissimilarity measure by contrast with Frobenius norm, and the reduced formulation helps compute the EMD efficiently.

¹Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan; e-mail : {takano10, yamamoto}@sk.tsukuba.ac.jp

The rest of this paper is organized as follows: In Section 2, we define the EMD and introduce the reduced EMD. In Section 3, we show the equivalence between the EMD problem and the reduced EMD problem regardless of the ground distance. Based on the homogeneous neighborhood structure of the graph, we show in Section 4 that the number of variables is reduced from $O(m^2)$ to $O(m)$. Section 5 discusses the imbalance case, and in the last section some experimental results are reported.

2. EARTH MOVER'S DISTANCE

Let us consider two histograms $\{p_{(i,j)} \mid (i,j) \in \mathcal{N}\}$ and $\{q_{(i,j)} \mid (i,j) \in \mathcal{N}\}$ defined on the two-dimensional coordinate system $\mathcal{N} := \{(i,j) \mid 1 \leq i \leq m_1, 1 \leq j \leq m_2\}$. A histogram is a mapping from a set of grid locations $(i,j) \in \mathcal{N}$ to the set of non-negative weights $p_{(i,j)}$. Regarding the weight $p_{(i,j)}$ in one histogram as the amount of “dirt” at location (i,j) and the weight $q_{(i,j)}$ in the other as the capacity of “hole” at location (i,j) , EMD is viewed as the minimum transportation cost to fill all the holes by dirt. With the assumption that the total amount of dirt is equal to the total capacity of holes, i.e.,

$$(2.1) \quad \sum_{(i,j) \in \mathcal{N}} p_{(i,j)} = \sum_{(i,j) \in \mathcal{N}} q_{(i,j)},$$

EMD is computed as an optimal value of the following well-known transportation problem of Hitchcock type:

$$(EMD) \quad \left| \begin{array}{ll} \text{minimize} & \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}} d_{(i,j)(k,l)} f_{(i,j)(k,l)} \\ \text{subject to} & \sum_{(k,l) \in \mathcal{N}} f_{(i,j)(k,l)} = p_{(i,j)} \quad \text{for all } (i,j) \in \mathcal{N} \\ & \sum_{(k,l) \in \mathcal{N}} f_{(k,l)(i,j)} = q_{(i,j)} \quad \text{for all } (i,j) \in \mathcal{N} \\ & f_{(i,j)(k,l)} \geq 0 \quad \text{for all } (i,j), (k,l) \in \mathcal{N}, \end{array} \right.$$

where $f_{(i,j)(k,l)}$ is the flow from location (i,j) to location (k,l) . The objective function coefficient $d_{(i,j)(k,l)}$ is a distance between location (i,j) and location (k,l) , and referred to as the *ground distance*. Let $m = m_1 \times m_2$. For $k = 1, 2, \dots, m$ let E_k be the $m \times m$ zero matrix with its k th row replaced by the m -dimensional row vector $\mathbf{e} := (1, 1, \dots, 1)$. Let A denote the $m \times m^2$ matrix $[E_1 \mid E_2 \mid \dots \mid E_m]$ and B denote the matrix $[I \mid I \mid \dots \mid I]$ of the same size, where I is the $m \times m$ identity matrix. By an appropriate definition of row vector \mathbf{d} , column vectors \mathbf{p} and \mathbf{q} , and variable column vector \mathbf{f} , problem (EMD) is rewritten as follows:

$$(EMD) \quad \left| \begin{array}{ll} \text{minimize} & \mathbf{d}\mathbf{f} \\ \text{subject to} & A\mathbf{f} = \mathbf{p} \\ & B\mathbf{f} = \mathbf{q} \\ & \mathbf{f} \geq \mathbf{0}. \end{array} \right.$$

In the sequel we consider

$$(R) \quad \left| \begin{array}{ll} \text{minimize} & \mathbf{d}\mathbf{g} \\ \text{subject to} & (A - B)\mathbf{g} = \mathbf{p} - \mathbf{q} \\ & \mathbf{g} \geq \mathbf{0}, \end{array} \right.$$

which we call problem (R), standing for the *reduced (EMD)*. We denote the optimal value of a problem by $v(\cdot)$.

Lemma 2.1.

$$v(\text{EMD}) \geq v(\text{R}).$$

Proof. Straightforward from the fact that a feasible solution of (EMD) is a feasible solution of (R). \square

3. EQUIVALENCE OF THE TWO PROBLEMS

First note that the matrix $A - B$ is of the form

$$[E_1 - I \mid E_2 - I \mid \dots \mid E_m - I],$$

and that this is the incidence matrix of a complete directed graph without a self loop on node set \mathcal{N} . We denote its arc set by \mathcal{D} . We classify the nodes according to the sign of $p_{(i,j)} - q_{(i,j)}$, namely

$$\begin{aligned} \mathcal{N}_+ &:= \{ (i, j) \in \mathcal{N} \mid p_{(i,j)} - q_{(i,j)} > 0 \} \\ \mathcal{N}_0 &:= \{ (i, j) \in \mathcal{N} \mid p_{(i,j)} - q_{(i,j)} = 0 \} \\ \mathcal{N}_- &:= \{ (i, j) \in \mathcal{N} \mid p_{(i,j)} - q_{(i,j)} < 0 \}. \end{aligned}$$

Following the convention of network flow theory (see for example [1]), we refer to a node in each set as *deficit node*, *balanced node* and *excess node*, respectively. Problem (R) is known as an arc flow formulation of network flow problem and a feasible solution \mathbf{g} of (R) is called an *arc flow*. Another formulation, a path-and-cycle flow formulation, of the network flow problem starts with enumerating all directed paths between any pair of nodes and all directed cycles. The decision variables are the flow value on each path and cycle.

Theorem 3.1 (Theorem 3.5 (Flow Decomposition Theorem), [1]). *Every arc flow can be represented as a path-and-cycle flow (though not necessarily uniquely) such that every directed path with positive flow connects a deficit node to an excess node.*

Let Π and Γ be the set of all directed paths and the set of all directed cycles of the network $(\mathcal{N}, \mathcal{D})$, respectively. Applying the above theorem to problem (R), we obtain the following corollary.

Corollary 3.2. *Let \mathbf{g} be a feasible solution of (R). Then for each directed path $\pi \in \Pi$ there is a non-negative path flow value $f(\pi)$, and for each directed cycle $\gamma \in \Gamma$ there is a non-negative cycle flow value $f(\gamma)$ with the following two properties:*

(1) *For every arc $((i, j)(k, l)) \in \mathcal{D}$ it holds that*

$$(3.1) \quad g_{(i,j)(k,l)} = \sum_{\pi:((i,j)(k,l)) \in \pi \in \Pi} f(\pi) + \sum_{\gamma:((i,j)(k,l)) \in \gamma \in \Gamma} f(\gamma).$$

(2) *$f(\pi)$ is positive only when path π connects a node in \mathcal{N}_+ to a node in \mathcal{N}_- .*

The *arc-path incidence vector* of a directed path π is the vector $\boldsymbol{\delta}(\pi)$ of components

$$\delta_{(i,j)(k,l)}(\pi) := \begin{cases} 1 & \text{when } ((i, j)(k, l)) \in \pi \\ 0 & \text{otherwise.} \end{cases}$$

The *arc-cycle incidence vector* of a directed cycle γ , denoted by $\boldsymbol{\delta}(\gamma)$, is defined in the same way. Then (3.1) is rewritten as

$$\mathbf{g} = \sum_{\pi \in \Pi} f(\pi) \boldsymbol{\delta}(\pi) + \sum_{\gamma \in \Gamma} f(\gamma) \boldsymbol{\delta}(\gamma).$$

Let

$$(3.2) \quad \mathbf{g}' = \sum_{\pi \in \Pi} f(\pi) \delta(\pi).$$

Lemma 3.3. *If \mathbf{g} is a feasible solution of (R), the following statements hold.*

- (1) \mathbf{g}' is a feasible solution of (R),
- (2) $\mathbf{d}\mathbf{g}' \leq \mathbf{d}\mathbf{g}$.

Proof. Straightforward from the fact that $(A-B)\delta(\gamma) = \mathbf{0}$ for every $\gamma \in \Gamma$, $\mathbf{d} \geq \mathbf{0}$ and the construction (3.2) of \mathbf{g}' . \square

Take a pair of nodes $(i, j) \in \mathcal{N}_+$ and $(k, l) \in \mathcal{N}_-$ and let $\Pi((i, j)(k, l))$ be the set of all directed paths connecting (i, j) to (k, l) , i.e., starting at (i, j) and ending at (k, l) . Let \mathbf{g}'' be the vector of components

$$(3.3) \quad g''_{(i,j)(k,l)} := \begin{cases} \sum_{\pi \in \Pi((i,j)(k,l))} f(\pi) & \text{when } (i, j) \in \mathcal{N}_+ \text{ and } (k, l) \in \mathcal{N}_- \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 1 shows the node set \mathcal{N} and two path-flows and a cycle-flow. The broad arrow from (i, j) to (k, l) shows $g''_{(i,j)(k,l)}$.

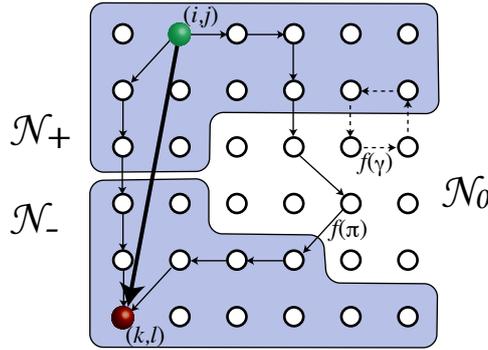


FIGURE 1. Reduction procedure

Lemma 3.4. *If \mathbf{g} is a feasible solution of (R), the following statements hold.*

- (1) \mathbf{g}'' is a feasible solution of (R),
- (2) $g''_{(k,l)(i,j)} = 0$ for all $(i, j) \in \mathcal{N}_+$ and $(k, l) \in \mathcal{N}$,
- (3) $g''_{(i,j)(k,l)} = g''_{(k,l)(i,j)} = 0$ for all $(i, j) \in \mathcal{N}_0$ and $(k, l) \in \mathcal{N}$,
- (4) $g''_{(i,j)(k,l)} = 0$ for all $(i, j) \in \mathcal{N}_-$ and $(k, l) \in \mathcal{N}$, and
- (5) $\mathbf{d}\mathbf{g}'' \leq \mathbf{d}\mathbf{g}'$.

Proof. The first four claims are readily seen by Corollary 3.2 (2) and the construction (3.3) of \mathbf{g}'' . Let $s(\pi)$ and $t(\pi)$ denote the starting node and the terminal node of path π , respectively. The last claim

is seen as follows.

$$\begin{aligned}
\mathbf{d}g' &= \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}} d_{(i,j)(k,l)} g'_{(i,j)(k,l)} \\
&= \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}} d_{(i,j)(k,l)} \sum_{\pi: ((i,j)(k,l)) \in \pi \in \Pi} f(\pi) \\
&= \sum_{\pi \in \Pi} f(\pi) \sum_{((i,j)(k,l)) \in \pi} d_{(i,j)(k,l)} \\
&\geq \sum_{\pi \in \Pi} f(\pi) d_s(\pi) t(\pi) \\
&= \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}} d_{(i,j)(k,l)} \sum_{\pi \in \Pi((i,j)(k,l))} f(\pi) \\
&= \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}} d_{(i,j)(k,l)} g''_{(i,j)(k,l)} \\
&= \mathbf{d}g'',
\end{aligned}$$

where the inequality is due to the triangle inequality of ground distance $d_{(i,j)(k,l)}$. \square

By the above lemma and the equality constraint of (R)

$$\sum_{(k,l) \in \mathcal{N}} g_{(i,j)(k,l)} - \sum_{(k,l) \in \mathcal{N}} g_{(k,l)(i,j)} = p_{(i,j)} - q_{(i,j)}$$

we see

$$(3.4) \quad \sum_{(k,l) \in \mathcal{N}} g''_{(i,j)(k,l)} = p_{(i,j)} - q_{(i,j)} \quad \text{for } (i,j) \in \mathcal{N}_+$$

$$(3.5) \quad \sum_{(k,l) \in \mathcal{N}} g''_{(i,j)(k,l)} = \sum_{(k,l) \in \mathcal{N}} g''_{(k,l)(i,j)} = 0 \quad \text{for } (i,j) \in \mathcal{N}_0$$

$$(3.6) \quad \sum_{(k,l) \in \mathcal{N}} g''_{(k,l)(i,j)} = -p_{(i,j)} + q_{(i,j)} \quad \text{for } (i,j) \in \mathcal{N}_-$$

Finally add $q_{(i,j)}$ flow to $g''_{(i,j)(i,j)}$ for $(i,j) \in \mathcal{N}_+$, $p_{(i,j)}$ flow to $g''_{(i,j)(i,j)}$ for $(i,j) \in \mathcal{N}_-$, and $p_{(i,j)} = q_{(i,j)}$ flow to $g''_{(i,j)(i,j)}$ for $(i,j) \in \mathcal{N}_0$ to make \mathbf{g}''' . Since $d_{(i,j)(i,j)} = 0$, we obtain the following lemma.

Lemma 3.5. *If \mathbf{g} is a feasible solution of (R), the following statements hold.*

- (1) \mathbf{g}''' is a feasible solution of (EMD),
- (2) $\mathbf{d}g''' = \mathbf{d}g''$.

Combining the above lemmas, we have the following inequality.

Lemma 3.6.

$$v(\text{EMD}) \leq v(R).$$

By Lemma 2.1 and 3.6 we see that problem (R) yields the same optimal objective function value as problem (EMD) does.

Theorem 3.7.

$$v(\text{EMD}) = v(R).$$

Note that this equality holds no matter what distance $d_{(i,j)(k,l)}$ is postulated on \mathcal{N} .

4. PROBLEM REDUCTION BASED ON HOMOGENEOUS NEIGHBORHOOD STRUCTURE

Suppose we are given a connected undirected graph, denoted by \mathcal{G} , with node set \mathcal{N} and edge set \mathcal{E} without a self-loop. The edge connecting nodes (i, j) and (k, l) is denoted by $[(i, j)(k, l)]$ and is assigned a positive value $\ell_{[(i, j)(k, l)]}$ called *length*.

For each pair of nodes (i, j) and (k, l) let $d_{(i, j)(k, l)}^\ell$ be the shortest length of paths between the pair. It is known and easily seen that $d_{(i, j)(k, l)}^\ell$ provides a distance defined on \mathcal{N} .

For each node $(i, j) \in \mathcal{N}$ we define

$$(4.1) \quad \mathcal{N}_{\mathcal{G}}(i, j) := \{ (k, l) \in \mathcal{N} \mid [(i, j)(k, l)] \in \mathcal{E} \},$$

and refer to $\mathcal{N}_{\mathcal{G}}(i, j)$ as node (i, j) 's *neighborhood* on \mathcal{G} .

Definition 4.1. Let \mathcal{H} be a finite subset of integer grid points of \mathbb{R}^2 without $(0, 0)$ and $\ell_{(i', j')}^{\mathcal{H}}$ be a positive number for $(i', j') \in \mathcal{H}$. Graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \ell)$ is said to have the homogeneous neighborhood structure of $(\mathcal{H}, \ell^{\mathcal{H}})$ when

- (1) $\mathcal{N}_{\mathcal{G}}(i, j) = \mathcal{N} \cap \{ (i + i', j + j') \mid (i', j') \in \mathcal{H} \}$ for all $(i, j) \in \mathcal{N}$, and
- (2) $\ell_{[(i, j)(k, l)]} = \ell_{(k-i, l-j)}^{\mathcal{H}}$ for all $(k, l) \in \mathcal{N}_{\mathcal{G}}(i, j)$ and $(i, j) \in \mathcal{N}$.

Two graphs together with corresponding homogeneous neighborhood structures are shown in Fig. 2. The distance d^ℓ defined by the upper graph \mathcal{G} , Manhattan graph, with the neighborhood structure $\mathcal{H} = \{(-1, 0), (0, -1), (0, 1), (1, 0)\}$ and

$$(4.2) \quad \ell_{(i', j')}^{\mathcal{H}} = 1 \text{ for all } (i', j') \in \mathcal{H}$$

is the L_1 distance on \mathcal{N} , while the other graph, Union Jack graph, with the neighborhood structure $\mathcal{H} = \{(-1, 0), (-1, -1), (0, -1), (1, -1), (1, 0), (1, 1), (0, 1), (-1, 1)\}$ and $\ell_{(i', j')}^{\mathcal{H}} = 1$ for all $(i', j') \in \mathcal{H}$ defines the L_∞ distance.

Bertsimas et al. [4] proposed the D -norm for $\mathbf{y} \in \mathbb{R}^n$ and $\rho \in [1, n]$ as the optimal value of the linear program

$$\left| \begin{array}{l} \text{maximize} \quad \sum_{j=1}^n u_j |y_j| \\ \text{subject to} \quad \sum_{j=1}^n u_j \leq \rho \\ \quad \quad \quad 0 \leq u_j \leq 1 \quad \text{for } j = 1, \dots, n. \end{array} \right.$$

The Union Jack graph with

$$(4.3) \quad \ell_{(i', j')}^{\mathcal{H}} = \begin{cases} 1 & \text{for } (i', j') \in \{(-1, 0), (0, -1), (1, 0), (0, 1)\} \\ \rho & \text{for } (i', j') \in \{(-1, -1), (1, -1), (1, 1), (-1, 1)\} \end{cases}$$

defines the D -norm, which gives, by setting the parameter ρ appropriately (e.g. $\rho = \sqrt{2}$), an in-between of L_1 and L_2 .

Suppose the ground distance $d_{(i, j)(k, l)}$ among locations of \mathcal{N} is given as the distance $d_{(i, j)(k, l)}^\ell$ for a graph \mathcal{G} with a homogeneous neighborhood structure. Then for two distinct locations (i, j) and (k, l) there is an undirected path of edges $[(i_0, j_0)(i_1, j_1)], [(i_1, j_1)(i_2, j_2)], \dots, [(i_{n-1}, j_{n-1})(i_n, j_n)]$ such that $(i_0, j_0) = (i, j)$, $(i_n, j_n) = (k, l)$,

$$(i_{r+1}, j_{r+1}) \in \mathcal{N}_{\mathcal{G}}(i_r, j_r) \quad \text{for } r = 0, \dots, n-1$$

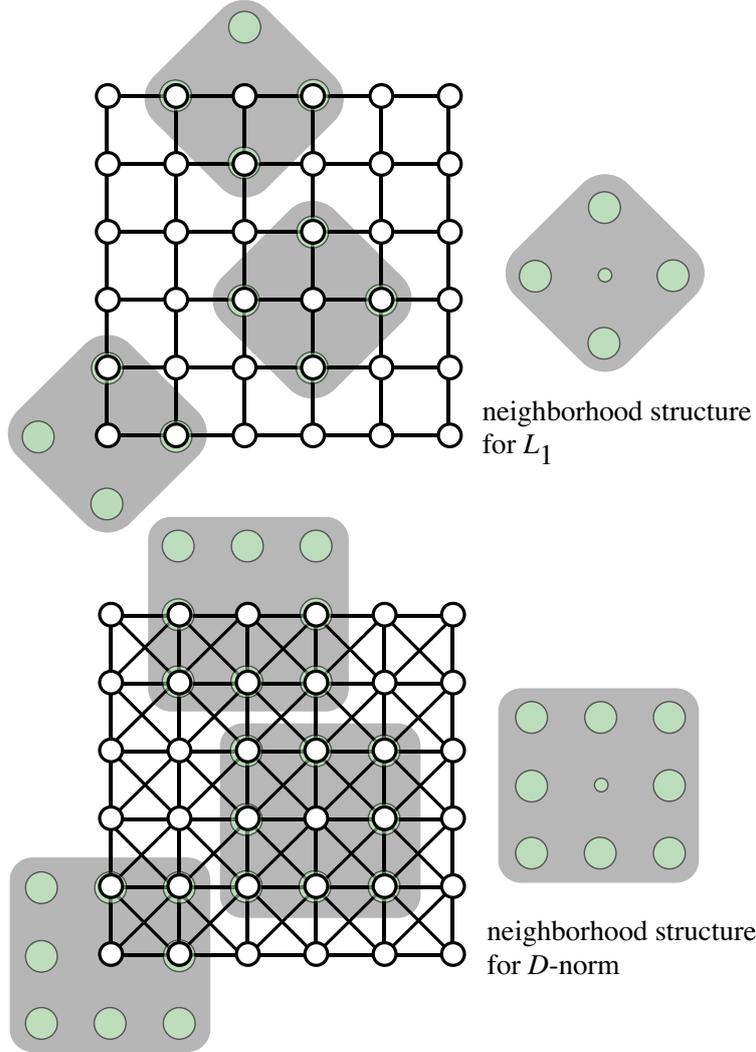


FIGURE 2. Graph and neighborhood structure defining a distance on \mathcal{N}

and

$$(4.4) \quad d_{(i,j)(k,l)} = \sum_{r=0}^{n-1} d_{(i_r, j_r)(i_{r+1}, j_{r+1})} = \sum_{r=0}^{n-1} \ell_{(i_{r+1}-i_r, j_{r+1}-j_r)}^{\mathcal{H}}.$$

Add the constraints

$$g_{(i,j)(k,l)} = 0 \quad \text{for all } (i,j) \in \mathcal{N} \text{ and } (k,l) \notin \mathcal{N}_G(i,j)$$

to problem (R) and denote it by $(\bar{\text{R}})$, i.e.,

$$(\bar{\text{R}}) \quad \left| \begin{array}{l} \text{minimize} \quad \mathbf{d}\mathbf{g} \\ \text{subject to} \quad (\mathbf{A} - \mathbf{B})\mathbf{g} = \mathbf{p} - \mathbf{q} \\ \mathbf{g} \geq \mathbf{0} \\ g_{(i,j)(k,l)} = 0 \quad \text{for all } (i,j) \in \mathcal{N} \text{ and } (k,l) \notin \mathcal{N}_G(i,j), \end{array} \right.$$

or equivalently

$$(\bar{R}) \quad \left\{ \begin{array}{l} \text{minimize} \quad \sum_{(i,j) \in \mathcal{N}} \sum_{(k,l) \in \mathcal{N}_{\mathcal{G}}(i,j)} \ell_{(k-i, l-j)}^{\mathcal{H}} g_{(i,j)(k,l)} \\ \text{subject to} \quad \sum_{(k,l) \in \mathcal{N}_{\mathcal{G}}(i,j)} g_{(i,j)(k,l)} - \sum_{(k,l) \in \mathcal{N}_{\mathcal{G}}(i,j)} g_{(k,l)(i,j)} = p_{(i,j)} - q_{(i,j)} \\ \hspace{15em} \text{for all } (i,j) \in \mathcal{N} \\ g_{(i,j)(k,l)} \geq 0 \quad \text{for all } (i,j) \in \mathcal{N}, (k,l) \in \mathcal{N}_{\mathcal{G}}(i,j). \end{array} \right.$$

We see that problem (\bar{R}) is equivalent to problem (R) .

Lemma 4.2. *Suppose that the graph \mathcal{G} has the homogeneous neighborhood structure $(\mathcal{H}, \ell^{\mathcal{H}})$ and the ground distance $d_{(i,j)(k,l)}$ is given as the shortest length of paths in \mathcal{G} . Then every optimal solution of problem (\bar{R}) is an optimal solution of problem (R) , and*

$$v(\bar{R}) = v(R).$$

Proof. Let $((i,j)(k,l))$ be an arc of \mathcal{D} . Since the ground distance is given as the shortest length of paths in \mathcal{G} , there is a series of arcs $((i_0, j_0)(i_1, j_1)), ((i_1, j_1)(i_2, j_2)), \dots, ((i_{n-1}, j_{n-1})(i_n, j_n))$ such that $(i_0, j_0) = (i, j)$, $(i_n, j_n) = (k, l)$, $(i_{r+1}, j_{r+1}) \in \mathcal{N}_{\mathcal{G}}(i_r, j_r)$ for $r = 0, 1, \dots, n-1$, and also the equality (4.4) holds.

Now suppose we are given a feasible flow \mathbf{g} of problem (R) . The above observation implies that replacing the arc flow of $g_{(i,j)(k,l)}$ on arc $((i,j)(k,l))$ by the path-flow along $((i_0, j_0)(i_1, j_1)), ((i_1, j_1)(i_2, j_2)), \dots, ((i_{n-1}, j_{n-1})(i_n, j_n))$ does not change the objective function value. Repeating this procedure if necessary, we will obtain a feasible flow satisfying the additional equality constraints

$$g_{(i,j)(k,l)} = 0 \text{ for all } (i,j) \in \mathcal{N} \text{ and } (k,l) \notin \mathcal{N}_{\mathcal{G}}(i,j)$$

of (\bar{R}) without changing the objective function value. This completes the proof. \square

Theorem 4.3. *Suppose that the graph \mathcal{G} has the homogeneous neighborhood structure $(\mathcal{H}, \ell^{\mathcal{H}})$ and problem (EMD) employs the shortest length of paths in \mathcal{G} as the ground distance. Then*

$$v(\bar{R}) = v(EMD).$$

Proof. Straightforward from Theorem 3.7 and Lemma 4.2. \square

Let h denote the size of \mathcal{H} , which is four for the Manhattan graph and eight for the Union Jack graph. Then comparing (\bar{R}) with (EMD) , the number of variables reduces from m^2 to mh . This will greatly lighten the computational burden.

5. IMBALANCE CASE

When the equality (2.1) between \mathbf{p} and \mathbf{q} does not hold, a slight adjustment is needed. Suppose $\sum_{(i,j) \in \mathcal{N}} p_{(i,j)} < \sum_{(i,j) \in \mathcal{N}} q_{(i,j)}$. Firstly, add a node, say $(0,0)$, to \mathcal{N} and let

$$p_{(0,0)} := \sum_{(i,j) \in \mathcal{N}} q_{(i,j)} - \sum_{(i,j) \in \mathcal{N}} p_{(i,j)}, \quad q_{(0,0)} := 0,$$

so that the equality $\sum_{(i,j) \in \mathcal{N} \cup \{(0,0)\}} p_{(i,j)} = \sum_{(i,j) \in \mathcal{N} \cup \{(0,0)\}} q_{(i,j)}$ holds. Secondly, define the distance between nodes $(0,0)$ and $(i,j) \in \mathcal{N}$ as

$$d_{(0,0)(i,j)} = d_{(i,j)(0,0)} := L.$$

We see that if L is large enough to meet

$$L \geq \frac{1}{2} \max_{(i,j),(k,l) \in \mathcal{N}} d_{(i,j)(k,l)},$$

d satisfies the triangle inequality and hence is a distance on $\mathcal{N} \cup \{(0, 0)\}$. Therefore the argument in Section 3 applies and we obtain the equivalent reduced problem in variables $g_{(i,j)(k,l)}$ for $(i, j), (k, l) \in \mathcal{N} \cup \{(0, 0)\}$. When the graph \mathcal{G} has the homogeneous neighborhood structure of $(\mathcal{H}, \ell^{\mathcal{H}})$, adding the constraints

$$g_{(i,j)(k,l)} = 0 \quad \text{for all } (i, j) \in \mathcal{N} \text{ and } (k, l) \in \mathcal{N} \setminus \mathcal{N}_{\mathcal{G}}(i, j)$$

does not change the optimal value. We could further add

$$g_{(i,j)(0,0)} = 0 \quad \text{for all } (i, j) \in \mathcal{N}$$

because $(0, 0)$ is a deficit node, i.e., $p_{(0,0)} - q_{(0,0)} > 0$. A parallel argument applies when $\sum_{(i,j) \in \mathcal{N}} p_{(i,j)} > \sum_{(i,j) \in \mathcal{N}} q_{(i,j)}$.

6. EXPERIMENTAL RESULTS

We will report on some experimental results for the handwritten digits. We selected 20 images of 28×28 pixels shown in Fig. 3 out of the MNIST handwritten digit database [10], and applied five different methods to measure the dissimilarity between images.

We normalized the weights of each image or histogram so that they would sum up to 100, and accordingly, the equality (2.1) holds. We compute the dissimilarity between a histogram of group A and a histogram of group B by

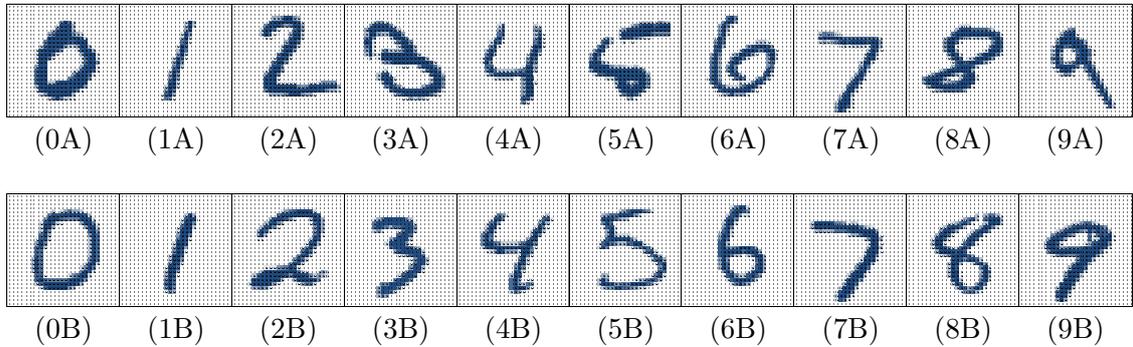


FIGURE 3. 20 images of the MNIST handwritten digit database

TABLE 1. Dissimilarity of the images by computing (a) Frobenius norm

	0B	1B	2B	3B	4B	5B	6B	7B	8B	9B
0A	†14.0	66.3	26.1	*24.5	44.6	34.0	*24.7	21.0	27.0	14.7
1A	91.9	†*0.9	89.1	76.5	100.0	95.8	94.3	98.7	74.1	71.5
2A	35.3	67.9	†31.9	36.1	52.2	36.5	34.7	53.5	41.1	42.6
3A	35.5	63.7	34.7	26.3	*35.1	†*24.0	28.4	41.0	30.5	30.2
4A	41.0	93.4	43.4	57.1	45.2	59.7	52.5	50.2	†38.8	41.7
5A	29.8	60.9	31.4	41.0	44.2	58.8	42.2	46.5	†*21.6	33.4
6A	†*13.5	64.9	36.7	42.0	38.7	44.1	39.5	48.7	27.5	29.8
7A	58.9	79.1	49.1	41.1	43.2	67.4	65.8	†*0.0	55.4	24.1
8A	31.8	60.8	*23.8	37.9	43.2	52.1	36.1	46.1	28.4	†*14.4
9A	57.1	72.1	60.1	39.9	58.7	55.9	32.8	57.0	†26.3	37.6

- (a) Frobenius norm (i.e., $\sqrt{\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (p_{(i,j)} - q_{(i,j)})^2}$),
- (b) (EMD) with L_2 ground distance,
- (c) (\bar{R}) with Manhattan graph and (4.2),
- (d) (\bar{R}) with Union Jack graph and (4.3) with $\rho = 1.3$, and
- (e) (\bar{R}) with Union Jack graph and (4.3) with $\rho = 1$.

All computations are conducted on a personal computer with Core2 CPU (2.66GHz) and 4GB memory. Problems (b), (c), (d) and (e) are solved by using CPLEX 10.1, OPL Studio 5.1.

Tables 1, 2, 3, 4 and 5 show the dissimilarity values of the images, where the minimum of each row is marked with † and the minimum of each column is marked with *, and dissimilarity values are normalized so that the minimum and maximum value become 0 and 100, respectively. Noteworthy points are in order.

Firstly, the pairs (0A)-(0B), (1A)-(1B) and (7A)-(7B) are given remarkably low dissimilarity values by all the methods. This means that these digits are not likely to be misread. On the other hand, the (4A)-(4B) pair is well distinguished by (b), (c), (d) and (e) while the Frobenius norm (a) misreads (4A) as (8B) and (4B) as (3A). Counting how many times the diagonal elements are marked with

TABLE 2. Dissimilarity of the images by computing (b) (EMD) with L_2 ground distance

	0B	1B	2B	3B	4B	5B	6B	7B	8B	9B
0A	†*26.4	60.8	43.3	*36.2	52.8	45.7	*31.5	43.0	39.7	36.2
1A	85.3	†*0.0	77.9	57.1	77.5	90.0	66.4	81.6	60.6	56.3
2A	52.3	78.5	59.7	50.9	78.8	†41.3	42.1	80.4	60.9	78.1
3A	67.3	100.0	81.6	54.0	54.2	†*33.1	47.9	61.9	56.8	71.4
4A	54.5	79.5	60.1	61.5	†*31.3	67.7	44.6	56.1	34.4	46.0
5A	50.2	71.5	48.2	73.4	51.6	83.6	55.7	75.0	†*31.2	56.0
6A	37.3	73.6	53.0	50.6	43.0	53.5	†35.9	58.7	36.8	40.0
7A	60.1	59.9	60.8	37.8	51.5	69.6	75.2	†*14.7	56.9	31.8
8A	47.0	54.5	*41.8	62.5	54.6	77.4	55.0	68.8	41.8	†*30.2
9A	83.8	80.4	90.0	51.7	58.1	62.8	†42.8	75.1	48.2	67.6

TABLE 3. Dissimilarity of the images by computing (c) (\bar{R}) with Manhattan graph and (4.2)

	0B	1B	2B	3B	4B	5B	6B	7B	8B	9B
0A	†*31.9	60.9	48.8	*36.4	55.0	48.0	*38.4	49.1	41.3	41.4
1A	87.8	†*0.0	85.3	60.3	80.8	95.6	66.0	86.5	64.1	64.6
2A	54.0	92.1	66.0	55.6	91.9	†44.0	46.2	90.1	63.9	86.0
3A	77.9	95.2	83.9	55.7	56.9	†*33.4	50.8	70.6	58.6	78.4
4A	56.2	79.7	62.5	61.0	†*35.0	72.3	50.6	60.3	41.0	49.2
5A	56.5	77.4	50.0	83.5	53.9	87.5	58.2	78.8	†*36.0	58.7
6A	40.9	77.0	55.2	57.6	47.1	57.8	40.9	67.2	†39.5	47.5
7A	64.3	64.7	66.5	41.4	57.5	80.5	85.7	†*15.6	57.4	*32.9
8A	54.1	56.4	*45.6	68.0	55.1	84.6	64.9	79.7	46.2	†35.2
9A	93.7	79.8	100.0	48.9	58.3	63.6	†45.0	76.4	53.4	61.7

† and *, we have 6 for (a), 8 for (c), and 9 for (b), (d) and (e). As far as these digit images are concerned, EMD based measures work better than Frobenius norm.

Secondly, we observe that the values of dissimilarity given by (d) are very close to those by (b). By Proposition 3 in [4] the ratio of D -norm with $\rho = 1.3$ to L_2 -norm falls in $[1.3/\sqrt{2}, \sqrt{1+0.3^2}] \approx [0.92, 1.04]$. Union Jack graph with $\rho = 1.3$ fully approximates the L_2 ground distance.

Thirdly, dissimilarity values that are marked with † and/or * are at almost the same places in Tables 2, 3, 4 and 5, however, we observe a contrast when we look closely at the row (6A). Namely, the minimum value of the row (6A) appears in the column (8B) in Table 3, while it appears in the column (6B) in Tables 2, 4 and 5. This shows that our proposed method can be some improvement on Ling and Okada’s method.

For computing the dissimilarity, we have to solve the optimization problems (b), (c), (d) and (e), whereas the computational burden of (a) is by far the lightest. In Table 6, the columns *average*, *min* and *max* show the average time, the minimal time and maximal time for computing the 100 values of

TABLE 4. Dissimilarity of the images by computing (d) (\bar{R}) with Union Jack graph and (4.3) with $\rho = 1.3$

	0B	1B	2B	3B	4B	5B	6B	7B	8B	9B
0A	†*25.6	61.4	42.2	*36.4	52.6	45.2	*30.7	42.5	39.1	35.4
1A	85.1	†*0.0	76.2	57.2	76.7	90.0	66.7	80.0	60.7	54.5
2A	52.7	75.6	59.3	50.5	76.8	†41.2	41.6	79.9	60.8	76.7
3A	65.1	100.0	80.7	55.3	54.8	†*33.6	47.8	61.9	56.7	72.2
4A	54.2	77.2	60.0	61.5	†*31.1	66.8	43.8	56.7	34.2	46.6
5A	50.1	71.1	48.3	72.7	51.3	84.3	56.3	76.4	†*30.4	56.3
6A	37.4	73.1	53.1	50.2	43.0	54.0	†34.5	58.1	36.4	38.7
7A	58.9	58.3	60.3	38.0	51.0	68.0	71.9	†*14.8	56.8	32.3
8A	46.4	55.4	*40.6	62.4	54.8	75.5	54.4	66.3	41.9	†*29.9
9A	83.2	81.3	90.7	51.8	58.5	63.1	†43.0	75.5	48.4	67.7

TABLE 5. Dissimilarity of the images by computing (e) (\bar{R}) with Union Jack graph and (4.3) with $\rho = 1$

	0B	1B	2B	3B	4B	5B	6B	7B	8B	9B
0A	†*23.7	56.9	40.4	*37.6	54.8	44.8	*28.3	40.4	38.5	32.4
1A	83.2	†*0.0	75.1	58.5	74.9	90.6	68.2	77.9	62.3	49.0
2A	54.4	67.4	58.2	49.1	72.5	39.5	†38.9	76.6	61.2	72.1
3A	61.8	100.0	81.9	57.7	55.0	†*34.4	48.3	59.2	59.0	72.9
4A	53.7	78.5	59.5	65.3	†*30.7	65.7	42.4	56.0	32.3	46.7
5A	48.9	70.7	49.4	69.6	51.8	83.4	55.4	77.2	†*29.1	57.6
6A	37.0	69.2	52.2	48.8	42.9	54.3	†32.2	55.8	36.8	35.5
7A	57.4	56.0	57.8	38.0	49.4	63.4	67.1	†*13.7	57.7	33.8
8A	44.3	54.4	*40.2	61.6	57.6	74.5	51.6	61.0	41.3	†*28.9
9A	81.6	86.3	89.7	56.7	61.4	66.1	†43.6	77.3	48.3	74.2

TABLE 6. Computational time and size of (EMD) and (\bar{R})

	CPU time (sec)			#Var	#Const
	average	min	max		
(b)	8.04	6.53	8.67	614,656	1,568
(c)	0.77	0.75	1.09	3,024	784
(d)	0.80	0.75	1.26	5,940	784
(e)	0.81	0.75	1.25	5,940	784

dissimilarity, respectively, and the columns $\#Var$ and $\#Const$ show the number of variables and the number of constraints of each problem, respectively. Because of the remarkable reduction of problem size (see the columns $\#Var$ and $\#Const$), (c), (d) and (e) reduce the computational time sharply in contrast to (b). It is reported in Ling and Okada [12] that the larger the size of histogram grows, the more efficient the reduction becomes. The reduction proposed in this paper would be especially useful when applied to image retrieval systems that need to compute dissimilarity of a large number of pairs of images.

7. CONCLUSION

We have shown that the earth mover’s distance problem reduces to a problem with half the number of constraints regardless of the ground distance. Furthermore, we have proved that a further reduction is possible when the ground distance derives from the graph with a homogeneous neighborhood structure, such as L_1 , L_∞ and D -norm. The preliminary experiment has shown that the reduction helps compute the earth mover’s distance efficiently.

A further direction of this study will be to apply our efficient formulation to *non-negative matrix factorization* (NMF for short). Given a non-negative matrix $M \in \mathbb{R}^{m \times n}$ and a positive integer r less than $\min\{m, n\}$, NMF is to make two non-negative matrices: the basis matrix $U \in \mathbb{R}^{m \times r}$ and the weight matrix $W \in \mathbb{R}^{r \times n}$ such that product UW approximates M . By choosing a measure of dissimilarity $\rho(\cdot)$ the NMF problem is formulated as

$$\begin{cases} \text{minimize} & \rho(M - UW) \\ \text{subject to} & U \in \mathbb{R}_+^{m \times r}, W \in \mathbb{R}_+^{r \times n}. \end{cases}$$

Frobenius norm, Kullback-Leibler divergence and the like are commonly used as ρ . The concern with NMF has been growing since it was used in Lee and Seung [11] for feature extraction and identification. It has a wide range of applications such as image retrieval, text mining and so on, and various NMF algorithms have been proposed (see e.g., Berry et al. [3]). Guillaumet and Vitrià [7] showed in the experimental evaluation that EMD is a good metric when combined with the NMF minimizing Kullback-Leibler divergence. This motivates us to use the earth mover’s distance as the measure of dissimilarity of NMF problems.

Multimedia applications, ranging from magnetic resonance imaging to music recommendation systems, commonly exhibit high-dimensional feature representations [17]. In this paper we have assumed that the location has two coordinates such as (i, j) , however, it can be generalized to a higher dimensional coordinate system with a slight modification.

ACKNOWLEDGMENTS

The authors thank Maiko Shigeno, University of Tsukuba for stimulating discussion, and Junya Gotoh, Chuo University, and Akiko Takeda, Keio University for drawing their attention to D -norm. Also, we would like to thank the anonymous referees. Their suggestions greatly improved the presentation of the manuscript. The first author is supported by the Grant-in-Aid for JSPS Fellows, and the second author is supported by the Grant-in-Aid for Scientific Research (B) 18310101 of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

- [1] R.K. Ahuja, T.L. Magnanti and J.B. Orlin, *Network Flows*. Prentice Hall, 1993.
- [2] A. Andoni, P. Indyk and R. Krauthgamer, "Earth Mover Distance over High-Dimensional Spaces," *Proc. Symposium on Discrete Algorithms*, pp. 343-352, 2008.
- [3] M.W. Berry, M. Browne, A.N. Langville, V.P. Pauca and R.J. Plemmons, "Algorithms and Applications for Approximate Nonnegative Matrix Factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155-173, 2007.
- [4] D. Bertsimas, D. Pachamanova and M. Sim, "Robust Linear Optimization under General Norms," *Operations Research Letters*, vol. 32, no. 6, pp. 510-516, 2004.
- [5] K. Grauman and T. Darrell, "Fast Contour Matching Using Approximate Earth Mover's Distance," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 220-227, 2004.
- [6] K. Grauman and T. Darrell, "The Pyramid Match Kernel: Efficient Learning with Sets of Features," *J. Machine Learning Research*, vol. 8, pp. 725-760, 2007.
- [7] D. Guillamet and J. Vitrià, "Evaluation of Distance Metrics for Recognition Based on Non-negative Matrix Factorization," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1599-1605, 2003.
- [8] P. Indyk and N. Thaper, "Fast Image Retrieval via Embeddings," *Proc. Third Workshop Statistical and Computational Theories of Vision*, 2003.
- [9] O. Klein and R.C. Veltkamp, "Approximation Algorithms for the Earth Mover's Distance under Transformations Using Reference Points," *Proc. 21st European Workshop on Computational Geometry*, 2005.
- [10] Y. LeCun and C. Cortes, "THE MNIST DATABASE of handwritten digits," available at <http://yann.lecun.com/exdb/mnist/>.
- [11] D.D. Lee and H.S. Seung, "Learning the Parts of Objects by Non-negative Matrix Factorization," *Nature*, vol. 401, no. 21, pp. 788-791, 1999.
- [12] H. Ling and K. Okada, "An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 840-853, 2007.
- [13] Y. Rubner, J. Puzicha, C. Tomasi and J.M. Buhmann, "Empirical Evaluation of Dissimilarity Measures for Color and Texture," *Computer Vision and Image Understanding*, vol. 84, pp. 25-43, 2001.
- [14] Y. Rubner, C. Tomasi and L.J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *Int'l J. Computer Vision*, vol. 40, no. 2, pp. 99-121, 2000.
- [15] S. Shirdhonkar and D.W. Jacobs, "Approximate Earth Mover's Distance in Linear Time," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [16] M. Werman, S. Peleg, and A. Rosenfeld, "A Distance Metric for Multidimensional Histograms," *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 328-336, 1985.
- [17] M. Wichterich, I. Assent, P. Kranen and T. Seidl, "Efficient EMD-based Similarity Search in Multimedia Databases via Flexible Dimensionality Reduction," *Proc. ACM SIGMOD Int'l Conf. Management of data*, 2008.