Department of Social Systems and Management

Discussion Paper Series

No. 1209

Metric-Preserving Reduction of Earth Mover's Distance and its Application to Non-negative Matrix Factorization

by

Yuichi Takano and Yoshitsugu Yamamoto

May 2008

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

METRIC-PRESERVING REDUCTION OF EARTH MOVER'S DISTANCE AND ITS APPLICATION TO NON-NEGATIVE MATRIX FACTORIZATION

YUICHI TAKANO AND YOSHITSUGU YAMAMOTO

Abstract. We prove that the earth mover's distance problem reduces to a problem with half the number of constraints regardless of the ground distance, and propose a further reduced formulation when the ground distance comes from a graph with a homogeneous neighborhood structure. We also propose to apply our formulation to the non-negative matrix factorization.

1. Introduction

Earth mover's distance (EMD in short) proposed by Rubner et al. [7] is a mathematical measure of the dissimilarity between two distributions. In a recent issue Ling and Okada [6] proposed a new formulation EMD-L_1 to compute EMD when the L_1 ground distance is used. It significantly simplifies the original formulation of EMD. Motivated by their work, we propose in this paper, 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 $O(m)$ for a histogram with m locations when the ground distance is derived from a graph with a homogeneous neighborhood structure. Application to non-negative matrix factorization (NMF) is also described.

2. Earth Mover's Distance

Let us consider two histograms $\{\, p_{(i,j)} \mid 1 \le i \le m_1, 1 \le j \le m_2 \,\}$ and $\{\, q_{(i,j)} \mid 1 \le i \le m_1, 1 \le j \le m_2 \,\}$ defined on the two-dimensional coordinate system. Histogram is a mapping from a set of grid locations (i, j) to the set of non-negative weights $p_{(i,j)}$ or $q_{(i,j)}$, which can be seen a mass of earth (supply) and a collection of holes (demand), respectively. For example, digital imaging can be seen as an histogram if luminosity of each pixel corresponds to the weights. Then, by measuring the least distance to fill the holes with earth, EMD provides the dissimilarity of the two histograms. With the assumption that the total supply and demand are equal, i.e.,

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

where $\mathcal{N} := \{\, (i, j) \mid 1 \le i \le m_1, 1 \le j \le m_2 \,\}$, EMD is computed as an optimal value of the following well-known transportation problem of Hitchcock type:

Date: May 26, 2008.

Discussion Paper No. 1209, University of Tsukuba.

This research is supported in part by the Grant-in-Aid for Scientific Research (B) 18310101 of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

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

where $f_{(i,j)(k,l)}$ is a 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 $k = 1, 2, \ldots, m$ let E_k be the $m \times m$ zero matrix with its kth row replaced by the m-dimensional row vector $\boldsymbol{e} := (1, 1, \ldots, 1)$. Let A denote the $m \times m^2$ matrix $[\, E_1 \,|\, E_2 \,|\, \ldots \,|\, E_m \,]$ and B denote the matrix $[\, I \,|\, I \,|\, \ldots \,|\, I \,]$ of the same size, where I is the $m \times m$ identity matrix. By an appropriate definition of row vector $\boldsymbol{d}$, column vectors $\boldsymbol{p}$ and $\boldsymbol{q}$, and variable column vector $\boldsymbol{f}$, problem (EMD) is rewritten as follows:

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

In the sequel we consider

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

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

Lemma 2.1.

$$v(EMD) \geq v(R).$$

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

3. Equivalence of The Two Problems

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

$$[\, E_1 - I \,|\, E_2 - I \,|\, \ldots \,|\, E_m - I \,],$$

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 $\boldsymbol{g}$ 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 $\boldsymbol{g}$ 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

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

(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 (4.2) is rewritten as

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

Let

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

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

(1) $\boldsymbol{g}'$ is a feasible solution of (R),

(2) $\boldsymbol{d}\boldsymbol{g}' \leq \boldsymbol{d}\boldsymbol{g}$.

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

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 $\boldsymbol{g}''$ be the vector of components

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

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

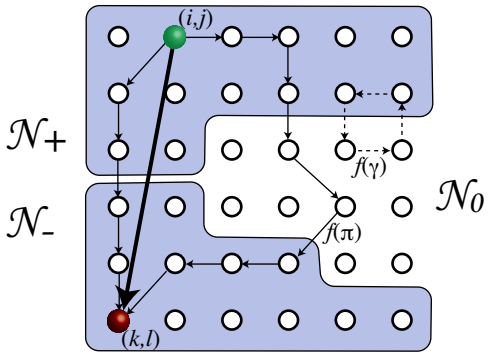

FIGURE 1. Reduction procedure

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

(1) $\boldsymbol{g}''$ 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) $\boldsymbol{d}\boldsymbol{g}'' \leq \boldsymbol{d}\boldsymbol{g}'$.

Proof. The first four claims are readily seen by Corollary 3.2 (2) and the construction (3.3) of $\boldsymbol{g}''$. 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.

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

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

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

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

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

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

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 $\boldsymbol{g}'''$. Since $d_{(i,j)(i,j)} = 0$, we obtain the following lemma.

Lemma 3.5. *If $\boldsymbol{g}$ is a feasible solution of (R), the following statements hold.*
(1) $\boldsymbol{g}'''$ is a feasible solution of (EMD),
(2) $\boldsymbol{d}\boldsymbol{g}''' = \boldsymbol{d}\boldsymbol{g}''$.

Combining the above lemmas, we have the following inequality.

Lemma 3.6.

$$v(EMD) \leq v(R).$$

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.

$$v(EMD) = v(R).$$

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

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

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

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 pints of $\mathbb{R}^2$ without $(0,0)$ and $\ell^{\mathcal{H}}_{(i',j')}$ 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) $\ell_{[(i,j)(k,l)]} = \ell^{\mathcal{H}}_{(k-i,l-j)}$ 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 Figure 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 $\ell^{\mathcal{H}}_{(i',j')} = 1$ for all

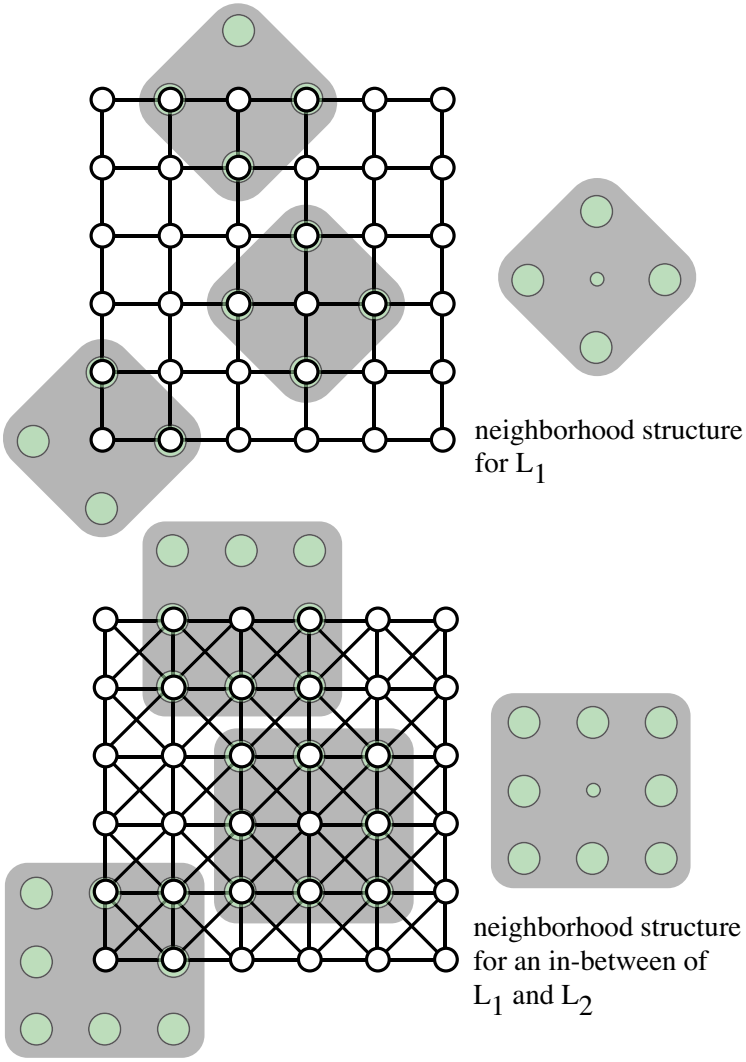

FIGURE 2. Graph and neighborhood structure defining a distance on $\mathcal{N}$

$(i', j') \in \mathcal{H}$ is the L_1 distance on $\mathcal{N}$, while the other graph, Union Jack graph, defines 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 $d^{\ell}_{(i,j)(k,l)}$ 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 $[(i_0, j_0)(i_1, j_1)], [(i_1, j_1)(i_2, j_2)]$, $\ldots, [(i_{n-1}, j_{n-1})(i_n, j_n)]$ such that $(i_0, j_0) = (i, j)$, $(i_n, j_n) = (k, l)$,

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

and

$$d_{(i,j)(k,l)} = \sum_{r=0}^{n-1} d_{(i_r,j_r)(i_{r+1},j_{r+1})} = \sum_{r=0}^{n-1} \ell^{\mathcal{H}}_{(i_{r+1}-i_r, j_{r+1}-j_r)}. \tag{4.2}$$

Add the constraints

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

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

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

or equivalently

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

We see that problem ($\bar{\text{R}}$) 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 $((i_0,j_0)(i_1,j_1)), ((i_1,j_1)(i_2,j_2)), \ldots, ((i_{n-1},j_{n-1})(i_n,j_n))$ 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 $r = 0,1,\ldots,n-1$, and also the equality (4.2) holds.

Now suppose we are given a feasible flow $\boldsymbol{g}$ 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 $((i_0,j_0)(i_1,j_1)), ((i_1,j_1)(i_2,j_2)), \ldots, ((i_{n-1},j_{n-1})(i_n,j_n))$ 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 ($\bar{\text{R}}$) without changing the objective function value. This completes the proof. □

Theorem 4.3. *When 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. □

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

5. non-negative Matrix Factorization

Given a non-negative matrix $M \in \mathbb{R}^{m\times n}$ and a positive integer: the number of basis r less than $\min\{m,n\}$, *non-negative matrix factorization* (NMF in short) 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 UW approximates M. The problem is

$$\text{(NMF)} \qquad \left| \begin{array}{ll} \text{minimize} & \|M - UW\| \\ \text{subject to} & U \in \mathbb{R}_+^{m\times r},\ W \in \mathbb{R}_+^{r\times n}, \end{array} \right.$$

where $+$ denotes the non-negativity of matrices. As the measure $\|\cdot\|$ of the dissimilarity between M and UW, Frobenius norm, Kullback-Leibler divergence and the like are commonly used. Look at the jth column $\boldsymbol{m}_j$ of M, and this reduces to $\boldsymbol{m}_j \approx \sum_{k=1}^{r} w_{kj}\boldsymbol{u}_k$, i.e., the linear combination of columns $\boldsymbol{u}_k$'s of U by non-negative coefficients $w_{k,j}$'s approximates $\boldsymbol{m}_j$.

The concern with NMF has been growing since it was used in Lee and Seung [5]. It is a method for feature extraction and identification, and has a wide range of applications such as image retrieval, text mining and so on. NMF is characterized by non-negativity constraints which enable an only additive combination of the parts in contrast to other methods such as principal component analysis. The NMF algorithms mostly use alternating minimization by fixing U or W, and can be divided into three classes: multiplicative update algorithms, gradient descent algorithms, and alternating least squares algorithms (Berry et al. [2]). The toolbox NMFLAB (Cichocki and Zdunek [3]) of MATLAB consisting of various NMF algorithms is available. Besides, in relation to EDM, Guillamet and Vitrià [4] show in the experimental evaluation that EMD improves the recognition rates. This result stimulates us to apply our formulation to the NMF.

Suppose that we are given U and want to optimize W with respect to the Frobenius norm. Then the problem reduces to the following n problems to determine the jth column $\boldsymbol{w}_j$ of W so that $U\boldsymbol{w}_j$ approximates the jth column $\boldsymbol{m}_j$ of M:

$$\left| \begin{array}{ll} \text{minimize} & \|\boldsymbol{m}_j - U\boldsymbol{w}_j\|_2 \\ \text{subject to} & \boldsymbol{w}_j \in \mathbb{R}_+^r \end{array} \right.$$

for $j = 1, 2, \ldots, n$, where $\|\cdot\|_2$ is the Euclidean norm. We propose to use the earth mover's distance to measure the dissimilarity of $U\boldsymbol{w}_j$ to $\boldsymbol{m}_j$. The ground distance vector $\boldsymbol{d}$ is generally determined from the physical feature of the row indices of M. Then the problem is written as

$$\left| \begin{array}{ll} \text{minimize} & \boldsymbol{d}\boldsymbol{f} \\ \text{subject to} & A\boldsymbol{f} = \boldsymbol{m}_j \\ & B\boldsymbol{f} = U\boldsymbol{w}_j \\ & \boldsymbol{f}, \boldsymbol{w}_j \geq \boldsymbol{0}. \end{array} \right.$$

We have shown that this problem reduces to

$$\left| \begin{array}{ll} \text{minimize} & \boldsymbol{d}\boldsymbol{f} \\ \text{subject to} & (A - B)\boldsymbol{f} = \boldsymbol{m}_j - U\boldsymbol{w}_j \\ & \boldsymbol{f}, \boldsymbol{w}_j \geq \boldsymbol{0}. \end{array} \right.$$

A noteworthy point is that this is a tractable linear optimization problem, for which efficient algorithms such as the interior point algorithm have been developed and a variety of commercial software is available.

6. Conclusion

We have proved 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 proposed a further reduced formulation when the ground distance comes from a graph with a homogeneous neighborhood structure. This generalizes EMD-L_1 in [6], and will help compute the earth mover's distance efficiently. 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 without a slightest modification. We have also proposed to apply our formulation to the non-negative matrix factorization. Results of the computational experiment will be reported shortly.

References

[1] R.K. Ahuja, T.L. Magnanti and J.B. Orlin, *Network Flows, Theory, Algorithms, and Applications*, Prentice-Hall, Englewood Cliffs, 1993.

[2] M.W. Berry, M. Browne, A.N. Langville, V.P Pauca and R.J. Plemmons, "Algorithms and applications for approximate non-negative matrix factorization," *Computational Statistics & Data Analysis*, **52**, 155-173, 2007.

[3] A. Cichocki and R. Zdunek, "NMFLAB for signal processing," available at http://www.bsp.brain.riken.jp/ICALAB/nmflab.html, 2006.

[4] D. Guillamet and J. Vitrià, "Evaluation of distance metrics for recognition based on non-negative matrix factorization," *Pattern Recognition Letters*, **24**, 1599-1605, 2003.

[5] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, **401**, 788-791, 1999.

[6] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison," *IEEE Trans on Pattern Anal. and Mach. Intell. (PAMI)*, **29**, 840-853, 2007.

[7] Y. Rubner, C. Tomasi and L.J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, **40**, 99-121, 2000.

Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

E-mail address: {takano10, yamamoto}@sk.tsukuba.ac.jp