Chapter 4

The Earth Mover's Distance (EMD)

A very general distance measure with applications in content-based image retrieval is the Earth Mover's Distance (EMD) between distributions ([68]). The EMD has been successfully used in a common framework for measuring image similarity with respect to color ([69, 67, 65, 68]) and texture ([69, 68, 66]). In this framework, the summary or *signature* of an image is a finite collection of weighted points. For example, in [69] the color signature of an image is a collection of dominant image colors in the CIE-Lab color space ([88]), where each color is weighted by the fraction of image pixels classified as that color. In [69], the texture signature of a single texture image is a collection of spatial frequencies in log-polar coordinates, where each frequency is weighted by the amount of energy present at that frequency. To complete the uniform framework, a distance measure on weight distributions is needed to measure similarity between image signatures.

The *Earth Mover's Distance* (EMD) between two distributions is proportional to the minimum amount of *work* required to change one distribution into the other. Here one unit of work is defined as the amount of work necessary to move one unit of weight by one unit of distance. The distance measure between weight locations is known as the *ground distance*. The morphing process between equal-weight distributions can be visualized as weight flowing from one distribution to the other until the distributions are identical. Figures 4.1(a)-(c) and 4.2(a)-(c) illustrate the minimum work morphing for three different pairs of equal-weight distributions.

In chapter 2, we used *mass* instead of *weight* in our EMD description because the EMD optimization problem was originally called the *mass transfer problem*. We consider the two terms interchangeable, although our notation given in the next section corresponds better

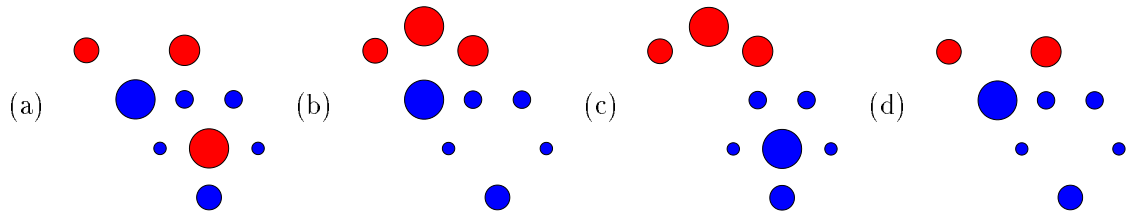


Figure 4.1: Example Distributions in 2D. Each of the examples (a), (b), (c), and (d) shows two distributions, one whose points are centered at the red discs and one whose points are centered at the blue discs. The area of a disc is equal to the weight at its center point in the distribution. The pairs (a), (b), and (c) are equal-weight pairs. In (d), the red distribution is lighter than the blue distribution. This example is the same as (b) with one of the red weights removed.

with *weight*, and in physics the units of weight $\times$ distance are the same as the units for work. On the other hand, *mass* corresponds better with the term *Earth Mover's Distance*. This name was suggested by Jorge Stolfi ([76]) who got the idea from some CAD programs for road design which have a function that computes the optimal earth displacement from roadcuts to roadfills.

An important property of the EMD is that it allows *partial matching*. When the total weights of the distributions are unequal, the EMD requires all the weight in the lighter distribution to be matched to weight in the heavier distribution. Some weight in the heavier distribution, however, will not be matched to weight in the lighter distribution. The matching process between unequal-weight distributions can be visualized as a flow in two different ways: (i) weight flows from the lighter distribution to the heavier distribution until the lighter distribution becomes a sub-distribution of the heavier one, or (ii) weight flows from the heavier distribution to the lighter distribution until all the weight in the lighter distribution has been covered. A type (i) flow visualization for the unequal-weight distributions in Figure 4.1(d) is shown in Figure 4.2(d).

The EMD matching process can also be visualized as filling holes with piles of dirt. The holes are located at the points in the lighter distribution, and the dirt piles are located at the points in the heavier distribution. The volume of a hole or dirt pile is given by the weight value of its position. In the equal-weight case, either distribution can be used to define the dirt piles or the holes, and all the dirt is needed to fill the holes. In the unequal-weight case, there is dirt leftover once all the holes are filled.

In the next section, we give the formal definition of the Earth Mover's Distance and discuss some of its properties. The work minimization problem which defines the EMD is a

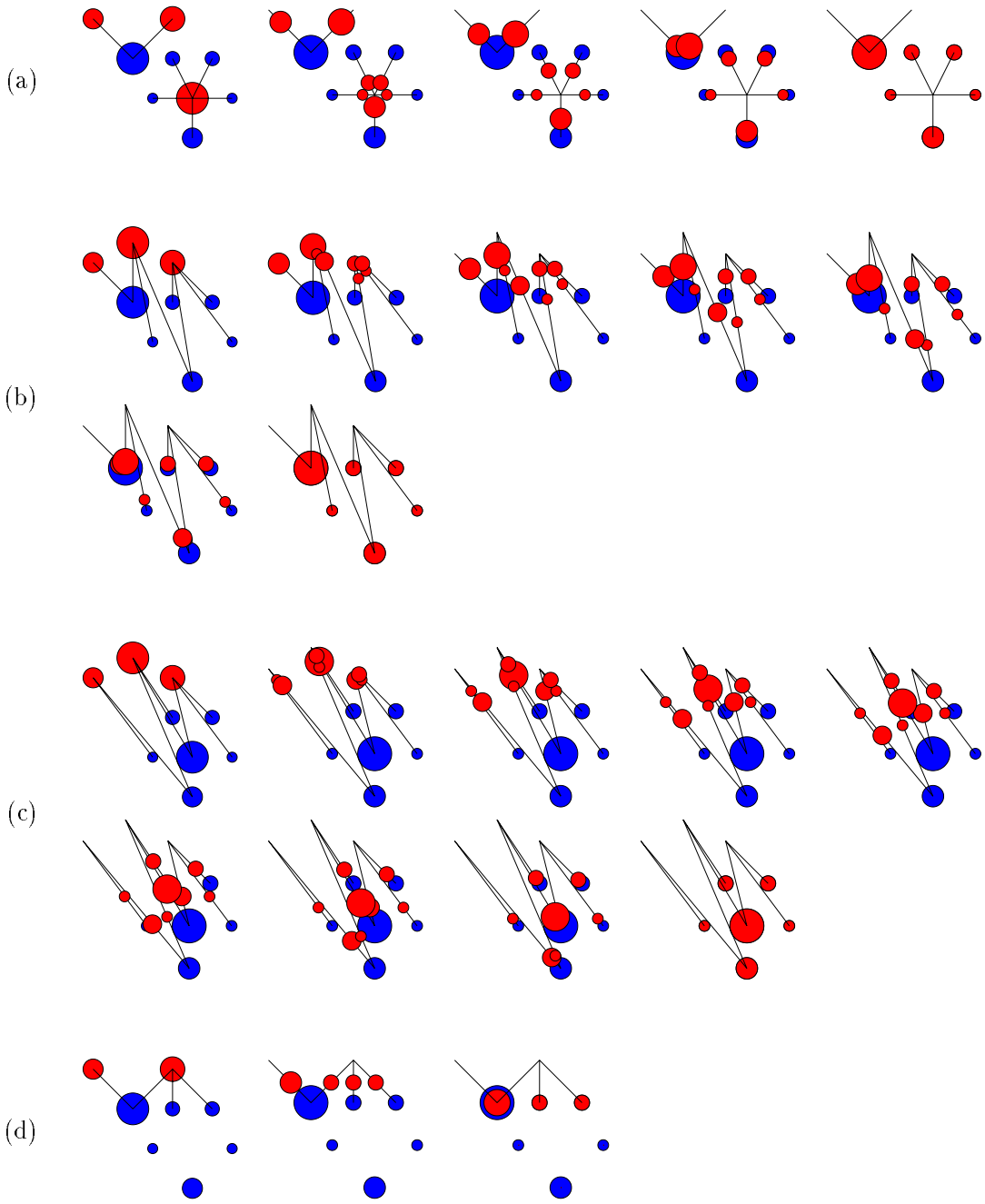


Figure 4.2: The EMD Morphing Process. Here we show the least work morphing for the equal-weight examples (a), (b), (c), and the unequal-weight example (d) in Figure 4.1. Weight flows from the red distributions to matching weight in the blue distributions. The amount of work done between frames is the same for every pair of adjacent frames shown (except possibly between the last two frames in each sequence). The EMD is smaller between pair (a) than pair (b), and smaller between pair (b) than pair (c). In (d), some of the blue weight is not matched to any red weight.

type of linear program known as the transportation problem. We discuss the transportation problem and its connection to the EMD in section 4.2. In section 4.3, we consider some special cases of matching (i) distributions which define ordinary point sets (section 4.3.1), and (ii) equal-weight distributions on the real line (section 4.3.2). In section 4.4, we give a couple of modifications to the EMD which make it more amenable to partial matching. In section 4.4.1, we present the partial EMD which forces only some fraction of the weight of the lighter distribution to be matched. In section 4.4.2, we discuss the τ-EMD which measures the amount of weight that *cannot* be matched if we only allow weight to flow over ground distances that do not exceed τ. Finally, in section 4.5 we use the EMD to estimate the size at which a color pattern may appear within an image. Please refer back to section 2.3 for a comparison of the EMD and bin-to-bin histogram distance measures.

4.1 Basic Definitions and Notation

We denote a discrete, finite *distribution* $\mathbf{x}$ as

$$\mathbf{x} = \{ (x_1, w_1), \ldots, (x_m, w_m) \} \equiv (X, w) \in \mathbf{D}^{K,m}$$

where $X = [\, x_1 \ \cdots \ x_m \,] \in \mathbf{R}^{K \times m}$ and $w_i \geq 0$, for all $i = 1, \ldots, m$. Here K is the dimension of ambient space of the points $x_i \in \mathbf{R}^K$, and m is the number of points. The (total) *weight* of the distribution $\mathbf{x}$ is $w_\Sigma = \sum_{i=1}^{m} w_i$. Given two distributions $\mathbf{x} = (X, w) \in \mathbf{D}^{K,m}$ and $\mathbf{y} = (Y, u) \in \mathbf{D}^{K,n}$, a *flow* between $\mathbf{x}$ and $\mathbf{y}$ is any matrix $F = (f_{ij}) \in \mathbf{R}^{m \times n}$. Intuitively, f_{ij} represents the amount of weight at x_i which is matched to weight at y_j. The term *flow* is meant to evoke the image of weight flowing from the points in the heavier distribution to the points in the lighter distribution until all the weight in the lighter distribution has been covered. If one distribution is known to be heavier than the other, then we shall write that a flow is *from* the heavier distribution *to* the lighter distribution. The flow F is a *feasible flow* between $\mathbf{x}$ and $\mathbf{y}$ iff

$$f_{ij} \geq 0 \qquad i = 1, \ldots, m, \ j = 1, \ldots, n, \tag{4.1}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_i \qquad i = 1, \ldots, m, \tag{4.2}$$

$$\sum_{i=1}^{m} f_{ij} \leq u_j \qquad j = 1, \ldots, n, \quad \text{and} \tag{4.3}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(w_\Sigma, u_\Sigma). \tag{4.4}$$

Constraint (4.1) requires the amount of x_i matched to y_j to be nonnegative. Constraint (4.2) ensures that the weight in $\mathbf{y}$ matched to x_i does not exceed w_i. Similarly, (4.3) ensures that the weight in $\mathbf{x}$ matched to y_j does not exceed u_j. Finally, constraint (4.4) forces the total amount of weight matched to be equal to the weight of the lighter distribution.

Let $\mathcal{F}(\mathbf{x}, \mathbf{y})$ denote the set of all feasible flows between $\mathbf{x}$ and $\mathbf{y}$. The work done by a feasible flow $F \in \mathcal{F}(\mathbf{x}, \mathbf{y})$ in matching $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\text{WORK}(F, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij},$$

where $d_{ij} = d(x_i, y_j)$ is the distance between x_i and y_j. An example ground distance is the Euclidean distance $d(x_i, y_j) = ||x_i \Leftrightarrow y_j||_2$. The *Earth Mover's Distance* $\text{EMD}(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ is the minimum amount of work to match $\mathbf{x}$ and $\mathbf{y}$, normalized by the weight of the lighter distribution:

$$\text{EMD}(\mathbf{x}, \mathbf{y}) = \frac{\min_{F=(f_{ij}) \in \mathcal{F}(\mathbf{x}, \mathbf{y})} \text{WORK}(F, \mathbf{x}, \mathbf{y})}{\min(w_\Sigma, u_\Sigma)}. \tag{4.5}$$

In the next section, we connect the work minimization problem in the numerator of (4.5) to a special type of linear program called the transportation problem ([32]). The normalization by the minimum weight makes the EMD equal to the average distance travelled by weight during an optimal (i.e. work minimizing) flow, and ensures that the EMD does not change if all the weights in both distributions are scaled by the same factor. Examples of feasible non-optimal and optimal flows between equal-weight distributions are shown in Figure 4.3, and between unequal-weight distributions are shown in Figure 4.4. In the unequal-weight case, some of the weight in the heavier distribution is unmatched by a feasible flow (more precisely, $w_\Sigma \Leftrightarrow u_\Sigma$ $\mathbf{x}$-weight is unmatched if $\mathbf{x}$ is heavier than $\mathbf{y}$).

The EMD is a metric when the total weights of the distributions are equal and the ground distance between weights is a metric ([68]). The only difficult part of the proof is showing the triangle inquality $\text{EMD}(\mathbf{x}, \mathbf{z}) \leq \text{EMD}(\mathbf{x}, \mathbf{y}) + \text{EMD}(\mathbf{y}, \mathbf{z})$. One way to morph $\mathbf{x}$ into $\mathbf{z}$ is to morph $\mathbf{x}$ into $\mathbf{y}$ and then $\mathbf{y}$ into $\mathbf{z}$. If dirt travels from x_i to y_j to z_k, then the metric assumption for the ground distance yields $d(x_i, z_k) \leq d(x_i, y_j) + d(y_j, z_k)$; i.e., it is cheaper just to transport the dirt directly from x_i to z_k. If we use an optimal matching $F^* = (f_{ij}^*)$ to change $\mathbf{x}$ into $\mathbf{y}$ and an optimal matching $G^* = (g_{jk}^*)$ to change $\mathbf{y}$ into $\mathbf{z}$, then the composite morphing $H = (h_{ik})$ to change $\mathbf{x}$ into $\mathbf{z}$ cannot cost more than $\text{EMD}(\mathbf{x}, \mathbf{y}) + \text{EMD}(\mathbf{y}, \mathbf{z})$. The EMD triangle inequality then follows from the fact that $\text{EMD}(\mathbf{x}, \mathbf{z})$ is the minimum cost of any morphing from $\mathbf{x}$ to $\mathbf{z}$. The composite flow H is

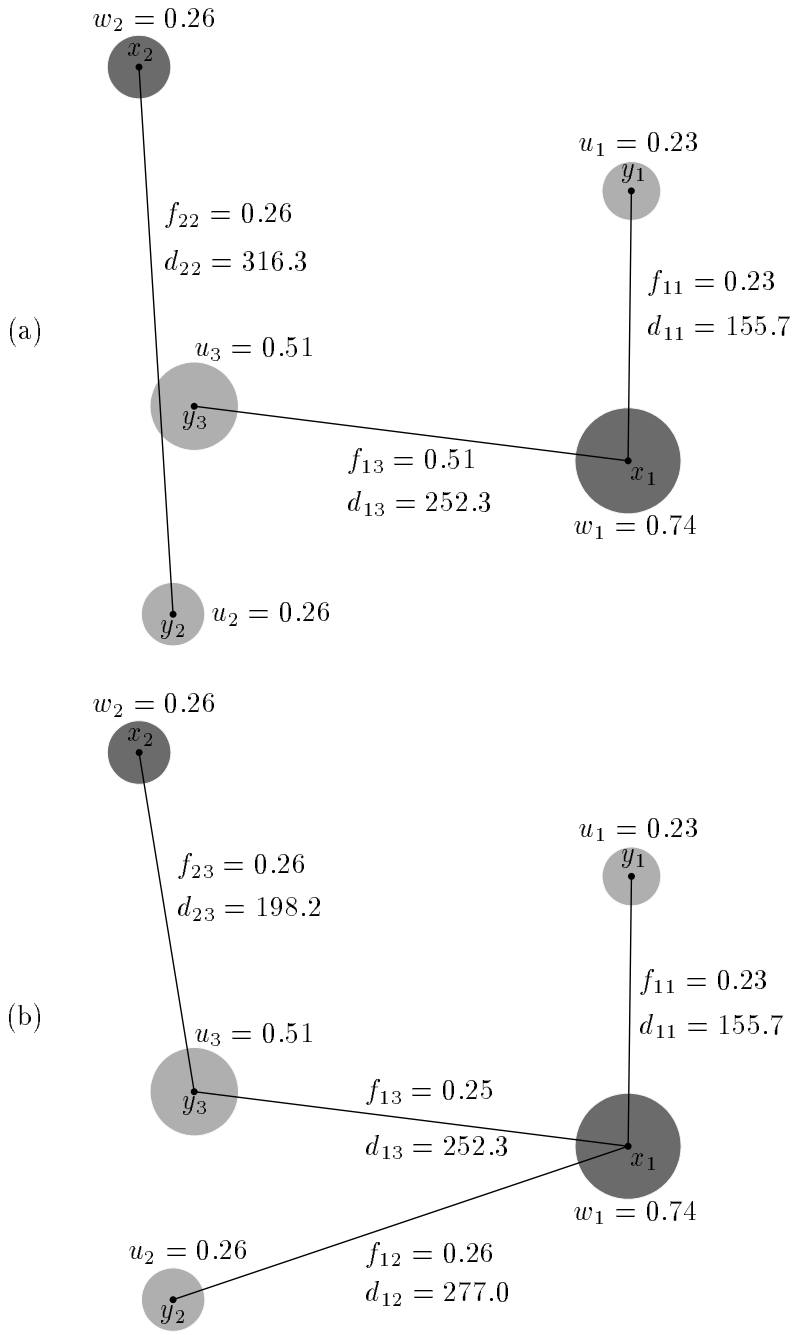


Figure 4.3: A Non-Optimal and an Optimal Flow between Equal-Weight Distributions. The area of the disc around a weight location is equal to the amount of weight at that location. (a) The amount of work done to match $\mathbf{x}$ and $\mathbf{y}$ by this feasible flow is $0.23 \times 155.7 + 0.51 \times 252.3 + 0.26 \times 316.3 = 246.7$. This flow is not optimal. (b) This flow is a work minimizing flow. The total amount of work done is $0.23 \times 155.7 + 0.26 \times 277.0 + 0.25 \times 252.3 + 0.26 \times 198.2 = 222.4$. Since the total weight of both $\mathbf{x}$ and $\mathbf{y}$ is one, the EMD is equal to the minimum amount of work: $\mathrm{EMD}(\mathbf{x}, \mathbf{y}) = 222.4$.

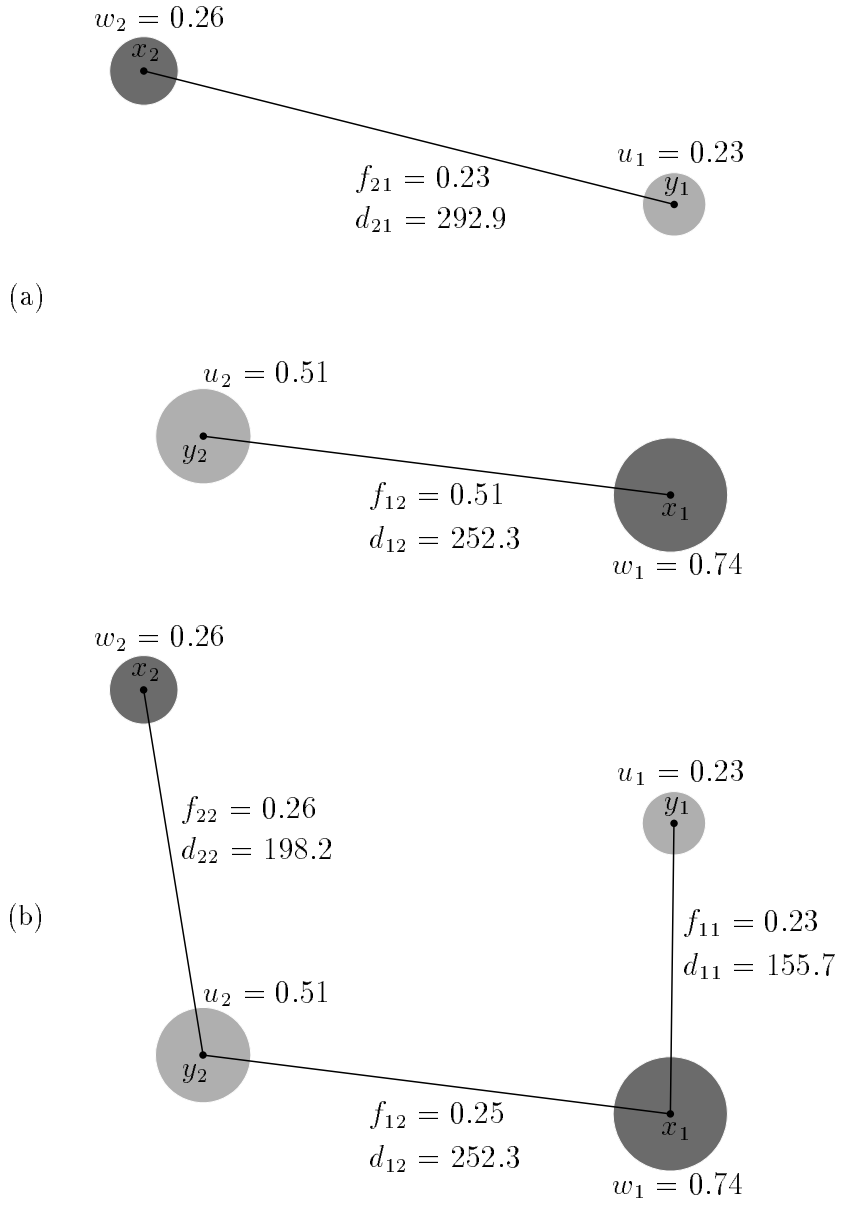


Figure 4.4: A Non-Optimal and an Optimal Flow between Unequal-Weight Distributions. Here $\mathbf{x}$ is heavier than $\mathbf{y}$. (a) The amount of work done to match $\mathbf{x}$ and $\mathbf{y}$ by this feasible flow is $0.51 \times 252.3 + 0.23 \times 292.9 = 196.0$. For this flow, 0.23 of the weight at x_1 and 0.03 of the weight at x_2 are not used in the matching. This flow is not optimal. (b) This flow is a work minimizing flow. The total amount of work for this flow to cover $\mathbf{y}$ is $0.23 \times 155.7 + 0.25 \times 252.3 + 0.26 \times 198.2 = 150.4$. For this flow, 0.26 of the weight at x_1 is not used in the matching. Since the total weight of the lighter distribution is 0.74, $\text{EMD}(\mathbf{x}, \mathbf{y}) = 150.4/0.74 = 203.3$.

derived from the flows F^* and G^* as the sum of interval intersections

$$h_{ik} = \sum_{j=1}^{n} \left| \left[\sum_{\widehat{i}=1}^{i-1} f^*_{\widehat{ij}}, \sum_{\widehat{i}=1}^{i} f^*_{\widehat{ij}} \right] \bigcap \left[\sum_{\widehat{k}=1}^{k-1} g^*_{j\widehat{k}}, \sum_{\widehat{k}=1}^{k} g^*_{j\widehat{k}} \right] \right| .$$

See [68] for the intuition for this formula. Since the L_p ($p \geq 1$) distance functions are metrics, the EMD is a metric between equal-weight distributions whenever the ground distance is an L_p distance.

Another commonly used distance function is $d = L_2^2$, the square of the ordinary L_2 distance. The L_2^2 distance function does not obey the triangle inequality, but it is a weak metric between points since

$$||p \Leftrightarrow q||_2^2 \leq 2(||p \Leftrightarrow r||_2^2 + ||q \Leftrightarrow r||_2^2) \qquad \forall p, q, r.$$

Thus, the morphing H from $\mathbf{x}$ to $\mathbf{z}$ costs no more than $2(\mathrm{EMD}(\mathbf{x}, \mathbf{y}) + \mathrm{EMD}(\mathbf{y}, \mathbf{z}))$ when the gorund distance is L_2^2. It follows that

$$\mathrm{EMD}^{L_2^2}(\mathbf{x}, \mathbf{z}) \leq 2(\mathrm{EMD}^{L_2^2}(\mathbf{x}, \mathbf{y}) + \mathrm{EMD}^{L_2^2}(\mathbf{y}, \mathbf{z})). \tag{4.6}$$

Thus, the EMD is a weak metric between equal-weight distributions when $d = L_2^2$.

When the distributions are not necessarily equal-weight, the EMD is no longer a metric. If $\mathbf{x}$ is lighter than $\mathbf{y}$, then a feasible flow matches all the weight in $\mathbf{x}$ to part of the weight in $\mathbf{y}$. If $\mathbf{x}$ and $\mathbf{z}$ are both lighter than $\mathbf{y}$, then it can happen that $\mathrm{EMD}(\mathbf{x}, \mathbf{y})$ and $\mathrm{EMD}(\mathbf{y}, \mathbf{z})$ are small, but $\mathrm{EMD}(\mathbf{x}, \mathbf{z})$ is large. This is because $\mathbf{x}$ and $\mathbf{z}$ might match well two parts of $\mathbf{y}$ that have little or no weight in common. There is no reason that two such parts of $\mathbf{y}$ must be similar under the EMD.

In the examples and discussion given thus far, the EMD measures the distance between two collections of weighted points based on a ground distance between points. This does not, however, expose the full generality of the EMD. The coordinates of distribution points are not used directly in the EMD formulation; only the ground distances between points are needed. Therefore, there is no need to work in a point feature space; the only requirement is that ground distances between features can be computed. In general, the EMD is a distance measure between two sets of weighted objects which is built upon a distance between individual objects. In this thesis, however, we focus mainly on the case of distributions of weight in some point feature space.

4.2 Connection to the Transportation Problem

The transportation problem (TP) is a special type of linear program (LP) which seeks to find the minimum cost way to transport goods from a set of sources or suppliers $i = 1, \ldots, m$ to a set of destinations or demanders $j = 1, \ldots, n$. Supplier i has a supply of s_i units, and demander j has a demand of d_j units. The cost per unit transported from supplier i to demander j is denoted by c_{ij}, and the number of units transported is denoted by x_{ij}. Assuming that the total supply $s_\Sigma = \sum_{i=1}^{m} s_i$ is equal to the total demand $d_\Sigma = \sum_{j=1}^{n} d_j$, the transportation problem is to compute

$$\min_{(x_{ij})} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

subject to

$$\begin{aligned} x_{ij} &\geq 0 \qquad i = 1, \ldots, m,\ j = 1, \ldots, n, \\ \sum_{j=1}^{n} x_{ij} &= s_i \qquad i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} &= d_j \qquad j = 1, \ldots, n. \end{aligned}$$

If the total supply and demand are not equal, then it is impossible to satisfy the given constraints. The equality constraints can be written as

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & & & & & & & & & & & & \\ & & & & 1 & 1 & \cdots & 1 & & & & & & & & \\ & & & & & & & & \ddots & & & & & & & \\ & & & & & & & & & & & 1 & 1 & \cdots & 1 \\ 1 & & & & 1 & & & & & & & 1 & & & & \\ & 1 & & & & 1 & & & \cdots & & 1 & & & & & \\ & & \ddots & & & & \ddots & & & & & & \ddots & & & \\ & & & 1 & & & & 1 & & & & & & & 1 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \\ \vdots \\ x_{m1} \\ x_{m2} \\ \vdots \\ x_{mn} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}.$$

The structure of this coefficient matrix can be exploited to improve both the time and space required by the simplex algorithm on a transportation problem. A detailed description of the *transportation simplex method* can be found in [32]. A C-code implementation of transportation simplex algorithm is currently available at `http://robotics.stanford.edu/~rubner/research.html`.

The transportation simplex algorithm can still be applied when the total supply s_Σ is greater than the total demand d_Σ. The goal is still to find the minimum cost way to satisfy all the demand. In this case, however, there will be some supply left over after the demand has been satisfied. The LP for the unbalanced case is

$$\min_{(x_{ij})} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

subject to

$$\begin{aligned} x_{ij} &\geq 0 \qquad i=1,\ldots,m,\ j=1,\ldots,n,\\ \sum_{j=1}^{n} x_{ij} &\leq s_i \qquad i=1,\ldots,m,\\ \sum_{i=1}^{m} x_{ij} &= d_j \qquad j=1,\ldots,n. \end{aligned}$$

In order to apply the transportation simplex method, we convert the unbalanced TP to an equivalent balanced TP. This is done by adding a dummy demander $n+1$ with demand $d_{n+1} = s_\Sigma \Leftrightarrow d_\Sigma$, and for which $c_{i,n+1} = 0$ for $i = 1,\ldots,m$. The total demand in the modified problem is equal to the total supply, and the minimum cost is the same for the balanced and unbalanced problems. The dummy demander gives the suppliers a place to dump their leftover supply at no cost.

Let us now connect the work minimization LP to the unbalanced transportation problem. If, for example, $u_\Sigma \leq w_\Sigma$, then the work minimization LP can be rewritten as

$$\min_{(f_{ij})} \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}d_{ij}$$

subject to

$$\begin{aligned} f_{ij} &\geq 0 \qquad i=1,\ldots,m,\ j=1,\ldots,n,\\ \sum_{j=1}^{n} f_{ij} &\leq w_i \qquad i=1,\ldots,m, \end{aligned}$$

$$\sum_{i=1}^{m} f_{ij} = u_j \qquad j = 1, \ldots, n.$$

This LP is an unbalanced transportation problem, where the supplies are $w_1, \ldots, w_m$, the demands are $u_1, \ldots, u_n$, and the costs are $d_{ij} = d(x_i, y_j)$. Similarly, if $w_\Sigma \leq u_\Sigma$, then the suppliers are from distribution $\mathbf{y} = (Y, u)$ and the demanders are from distribution $\mathbf{x} = (X, w)$. In the case of equal-weight distributions, $w_\Sigma = u_\Sigma$, the work LP reduces to

$$\min_{(f_{ij})} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij}$$

subject to

$$\begin{aligned} f_{ij} &\geq 0 \qquad i = 1, \ldots, m,\ j = 1, \ldots, n, \\ \sum_{j=1}^{n} f_{ij} &= w_i \qquad i = 1, \ldots, m, \\ \sum_{i=1}^{m} f_{ij} &= u_j \qquad j = 1, \ldots, n, \end{aligned}$$

which is a balanced transportation problem. Thus, the work minimization problem in the numerator of equation (4.5) is a transportation problem, and it can be solved efficiently by applying the transportation simplex method.

4.3 Special Cases

In this section, we examine two special cases of the EMD when the input distributions are restricted in some way. In section 4.3.1, we show that the EMD reduces to an optimal one-to-one matching of points when all point weights in the two distributions are equal to one. In section 4.3.2, we consider the case of equal-weight distributions on the real line. In this case, we give a very efficient algorithm to compute the EMD in one pass over the points.

4.3.1 Point Set Matching using the EMD

A point set is a special case of a distribution in which all weights are equal to one. In the language of the transportation problem, all the supplies and demands are equal to one unit. A slightly more general restricted input to the transportation problem is one in which all supplies and demands are equal to integers. Here we can assume that the transportation problem is balanced, for a dummy demander that absorbs any excess supply will also have

an integer demand. The integer input restriction adds structure to the transportation problem in what is usually known as the *integer solutions property*. This property states that when all the supplies and demands are integers, all feasible flows located at vertices of the feasible convex polytope $\mathcal{F}$ have integer values ([32]). Hence, all optimal feasible vertex flows consist only of integer values when all supplies and demands are integers. The transportation simplex method returns an optimal vertex flow.

Now let us return to the specific case when all supplies and demands are equal to one. From constraints (4.1), (4.2), and (4.3), it follows that $0 \leq f_{ij} \leq w_i$ and $0 \leq f_{ij} \leq u_j$ in every feasible flow $F = (f_{ij})$. This means that $0 \leq f_{ij} \leq 1$ in every feasible flow between point sets. Combining this fact with the integer solutions property, there exists an optimal feasible flow $F^* = (f^*_{ij})$ at a vertex of $\mathcal{F}$ with $f^*_{ij} \in \{0, 1\}\ \forall i, j$. As previously noted, the transportation simplex method will return such a solution. The flow values involving a dummy demander needed to create a balanced transportation problem will be integers, but need not be binary. This is irrelevant for our purposes since such flow variables are not really part of the solution. The bottom line is that for distributions $\mathbf{x} \in \mathbf{D}^{K,m}$ and $\mathbf{y} \in \mathbf{D}^{K,n}$ which are point sets in $\mathbf{R}^K$ with $m \geq n$,

$$\mathrm{EMD}(\mathbf{x}, \mathbf{y}) = \frac{\min_{F=(f_{ij}) \in \mathcal{F}(\mathbf{x},\mathbf{y})} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d(x_i, y_j)}{\min(w_\Sigma, u_\Sigma)} = \frac{\min_{\phi \in \Phi} \sum_{j=1}^{n} d(x_{\phi(j)}, y_j)}{n},$$

where Φ is the set of one-to-one correspondences

$$\Phi = \{\ \phi : \{\ 1, \ldots, n\ \} \rightarrow \{\ 1, \ldots, m\ \} \mid \phi(j_1) = \phi(j_2) \Leftrightarrow j_1 = j_2\ \}.$$

Thus, the EMD between point sets measures the average distance between corresponding points in an optimal one-to-one matching.

It is worthwhile to note that although the transportation simplex method will find an optimal one-to-one matching between point sets, it does not take advantage of the fact that the supplies and demands are all equal to one. A transportation problem with such supplies and demands is known as an *assignment problem*, and there are specialized algorithms to solve assignment problems ([75]). One word of caution is needed before applying an assignment problem algorithm instead of a transportation problem algorithm to match point sets. Creating a balanced assignment problem (i.e. one in which the two point sets have the same number of points) by adding dummy points to the smaller set will result in a large increase in the number of problem variables if the two sets have very different sizes. A transportation problem can be balanced with the addition of only one dummy demander which creates far fewer dummy variables than in the assignment case.

Thus for very unbalanced point set matching problems, it may be more efficient to apply the transportation simplex method than an assignment problem algorithm which operates only on balanced problems.

4.3.2 The EMD in One Dimension

Let $x = (X, w) \in \mathbf{D}^{1,m}$ and $y = (Y, u) \in \mathbf{D}^{1,n}$ be distributions on the real line. Assume the points in $\mathbf{x}$ and $\mathbf{y}$ are sorted by position:

$$x_1 < x_2 < \cdots < x_m \qquad \text{and} \qquad y_1 < y_2 < \cdots < y_n.$$

We also assume in this section that the ground distance between points is the absolute value ($d = L_1$).

Define the *cumulative distribution function* (CDF) of $\mathbf{x}$ as

$$W(t) = \begin{cases} 0 & \text{if } t \in (\Leftrightarrow\infty, x_1) \\ \sum_{i=1}^{k} w_i & \text{if } t \in [x_k, x_{k+1}),\ 1 \le k \le m \Leftrightarrow 1 \\ w_\Sigma = \sum_{i=1}^{m} w_i & \text{if } t \in [x_m, \infty). \end{cases}$$

Similarly, the CDF of $\mathbf{y}$ is

$$U(t) = \begin{cases} 0 & \text{if } t \in (\Leftrightarrow\infty, y_1) \\ \sum_{j=1}^{l} u_j & \text{if } t \in [y_l, y_{l+1}),\ 1 \le l \le n \Leftrightarrow 1 \\ u_\Sigma = \sum_{j=1}^{n} u_j & \text{if } t \in [y_n, \infty). \end{cases}$$

If $\mathbf{x}$ and $\mathbf{y}$ are equal weight, then the minimum work to transform one distribution into the other is the area between the graphs of the CDFs of $\mathbf{x}$ and $\mathbf{y}$. We shall prove this fact later in this section in Theorem 5. An example is shown in Figure 4.5.

The flow naturally defined by the CDFs is called the *CDF flow*, and is denoted $F^{\text{CDF}} = (f_{ij}^{\text{CDF}})$. Once again, see Figure 4.5. If we let

$$\begin{array}{rcccl} W_k & = & W(x_k) & = & \sum_{i=1}^{k} w_i \qquad \text{and} \\ U_l & = & U(y_l) & = & \sum_{j=1}^{l} u_j, \end{array}$$

then the CDF flow is given by

$$f_{ij}^{\text{CDF}} = |[W_{i-1}, W_i] \cap [U_{j-1}, U_j]|.$$

Theorem 4 *The flow F^{CDF} is a feasible flow between equal-weight distributions $\mathbf{x}$ and $\mathbf{y}$;*

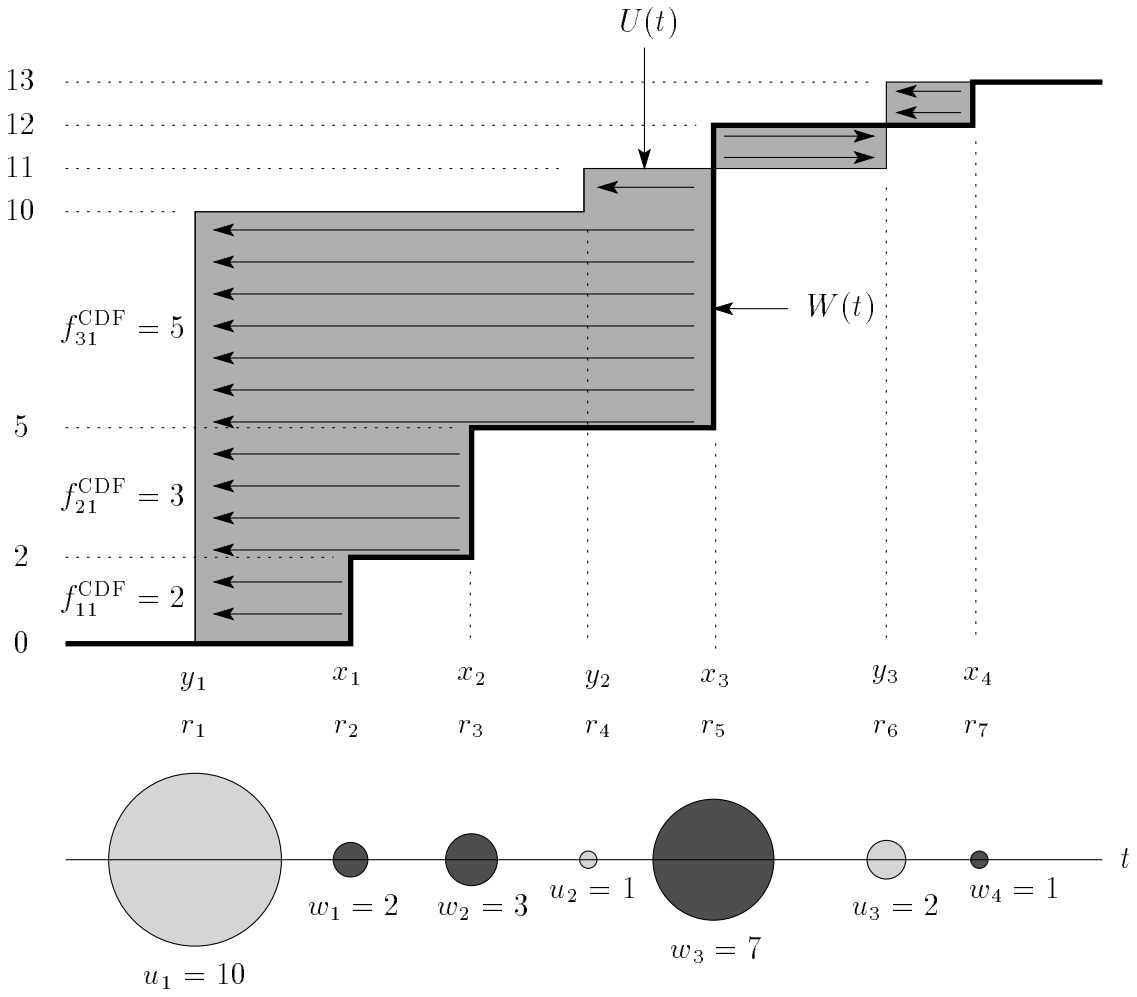


Figure 4.5: The EMD between Equal-Weight Distributions on the Real Line. The cumulative distribution functions (CDFs) for the equal-weight line distributions $\mathbf{x}$ and $\mathbf{y}$ are $W(t)$ and $U(t)$, repsectively. The minimum work to transform $\mathbf{x}$ into $\mathbf{y}$ is equal to the area between the two CDFs. An optimal transforming flow $F^{\text{CDF}} = (f_{ij}^{\text{CDF}})$, called the *CDF flow*, is shown with directed lines from $\mathbf{x}$-weight to matching $\mathbf{y}$-weight. The CDF flow is $f_{11}^{\text{CDF}} = 2$, $f_{21}^{\text{CDF}} = 3$, $f_{31}^{\text{CDF}} = 5$, $f_{32}^{\text{CDF}} = 1$, $f_{33}^{\text{CDF}} = 1$, $f_{43}^{\text{CDF}} = 1$, and $f_{ij}^{\text{CDF}} = 0$ for all other pairs (i, j). The EMD between $\mathbf{x}$ and $\mathbf{y}$ is obtained by dividing the minimum work by the total weight of the distributions ($w_\Sigma = u_\Sigma = 13$ in this example).

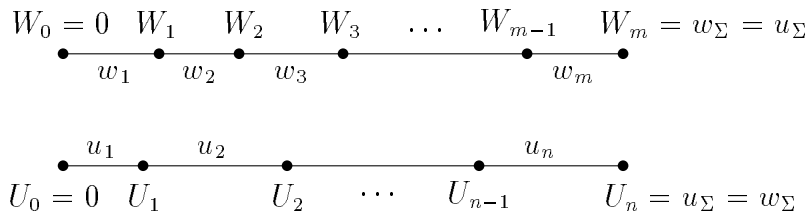


Figure 4.6: Feasibility of the CDF flow. The length of the **x**-interval $[W_{i-1}, W_i]$ is the **x**-weight w_i, and the length of the **y**-interval $[U_{j-1}, U_j]$ is the **y**-weight u_j. It should be clear from this figure that $\sum_{j=1}^{n} f_{ij}^{\mathrm{CDF}} = w_i$ and $\sum_{i=1}^{m} f_{ij}^{\mathrm{CDF}} = u_j$, where $f_{ij}^{\mathrm{CDF}} = |[W_{i-1}, W_i] \cap [U_{j-1}, U_j]|$.

i.e., $F^{\mathrm{CDF}} \in \mathcal{F}(\mathbf{x}, \mathbf{y})$.

Proof. Obviously $f_{ij}^{\mathrm{CDF}} \geq 0$. It remains to show that

$$\sum_{j=1}^{n} f_{ij}^{\mathrm{CDF}} = w_i \qquad \text{and} \qquad \sum_{i=1}^{m} f_{ij}^{\mathrm{CDF}} = u_j.$$

Note that the disjoint (except at interval endpoints) unions

$$\bigcup_{i=1}^{m} [W_{i-1}, W_i] = [0, w_\Sigma] \qquad \text{and} \qquad \bigcup_{j=1}^{n} [U_{j-1}, U_j] = [0, u_\Sigma]$$

cover exactly the same interval $[0, w_\Sigma] = [0, u_\Sigma]$. See Figure 4.6. It follows that

$$\begin{array}{lcll}
\sum_{j=1}^{n} f_{ij}^{\mathrm{CDF}} & = & \sum_{j=1}^{n} |[W_{i-1}, W_i] \cap [U_{j-1}, U_j]| & \\
& = & \left| [W_{i-1}, W_i] \cap \left(\bigcup_{j=1}^{n} [U_{j-1}, U_j] \right) \right| & \text{(interior disjointness of } [U_{j-1}, U_j]) \\
& = & |[W_{i-1}, W_i] \cap [0, u_\Sigma]| & \\
& = & |[W_{i-1}, W_i] \cap [0, w_\Sigma]| & \\
& = & |[W_{i-1}, W_i]| & ([W_{i-1}, W_i] \subseteq [0, w_\Sigma]) \\
\sum_{j=1}^{n} f_{ij}^{\mathrm{CDF}} & = & w_i. &
\end{array}$$

Similar reasoning proves that $\sum_{i=1}^{m} f_{ij}^{\mathrm{CDF}} = u_j$. ■

Now denote the sorted list of breakpoints $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n$ as

$$r_1 \leq r_2 \leq \cdots \leq r_{m+n}.$$

See Figure 4.5. In order to prove the optimality of F^{CDF}, we need the following lemma.

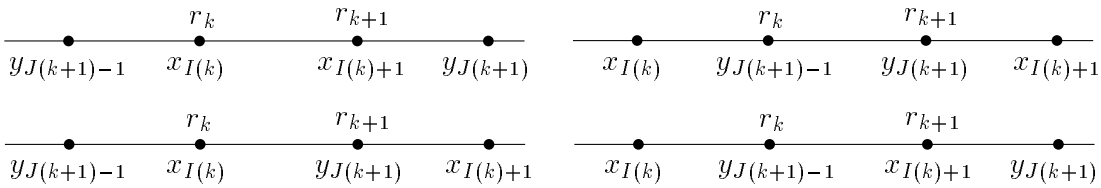


Figure 4.7: Breakpoint Notation Used in Lemma 2. $I(k)$ is the largest i such that $x_i \leq r_k$, and $J(k+1)$ is the smallest j such that $y_j \geq r_{k+1}$. The leftmost and rightmost labelled points are not necessarily r_{k-1} and r_{k+2}.

Lemma 2 *The feasible flow F^{CDF} between equal-weight distributions $\mathbf{x}$ and $\mathbf{y}$ moves exactly $|W(r_k) \Leftrightarrow U(r_k)|$ weight from $\mathbf{x}$ to $\mathbf{y}$ over the interval (r_k, r_{k+1}). More precisely, it moves $W(r_k) \Leftrightarrow U(r_k)$ $\mathbf{x}$-weight from r_k to r_{k+1} if $W(r_k) \geq U(r_k)$ and $U(r_k) \Leftrightarrow W(r_k)$ from r_{k+1} to r_k if $U(r_k) > W(r_k)$.*

Proof. Let $I(k)$ be the largest i such that $x_i \leq r_k$, and let $J(k+1)$ be the smallest j such that $y_j \geq r_{k+1}$. The four possible configurations are shown in Figure 4.7. Note also that $I(k)+1$ is the smallest i such that $x_i \geq r_{k+1}$ and $J(k+1) \Leftrightarrow 1$ is the largest j such that $y_j \leq r_k$. The key observations here are that

$$W_{I(k)} = W(r_k) \qquad \text{and} \qquad U_{J(k+1)-1} = U(r_k) \tag{4.7}$$

for all four possible configurations.

The amount of $\mathbf{x}$-weight $\alpha_{k \to k+1}$ that flows from r_k to r_{k+1} during the feasible flow F^{CDF} is

$$\begin{aligned}
\alpha_{k \to k+1} &= \sum_{i=1}^{I(k)} \sum_{j=J(k+1)}^{n} f_{ij}^{\text{CDF}} \\
&= \sum_{i=1}^{I(k)} \sum_{j=J(k+1)}^{n} |[W_{i-1}, W_i] \cap [U_{j-1}, U_j]| \\
&= \left| \left(\cup_{i=1}^{I(k)} [W_{i-1}, W_i] \right) \cap \left(\cup_{j=J(k+1)}^{n} [U_{j-1}, U_j] \right) \right| \qquad (4.8) \\
&= |[0, W_{I(k)}] \cap [U_{J(k+1)-1}, U_n]| \\
&= |[0, W(r_k)] \cap [U(r_k), u_\Sigma]| \qquad \text{(by (4.7))} \\
\alpha_{k \to k+1} &= \begin{cases} W(r_k) \Leftrightarrow U(r_k) & \text{if } W(r_k) \geq U(r_k) \\ 0 & \text{otherwise} \end{cases} . \qquad (4.9)
\end{aligned}$$

The line (4.8) follows from the previous line by the interior disjointness of the intervals $[W_{i-1}, W_i]$ and the interior disjointness of the intervals $[U_{j-1}, U_j]$. Similarly, the amount of

$\mathbf{x}$-weight $\alpha_{k+1\to k}$ that flows from r_{k+1} to r_k is

$$\begin{aligned}
\alpha_{k+1\to k} &= \sum_{i=I(k)+1}^{m} \sum_{j=1}^{J(k+1)-1} f_{ij}^{\mathrm{CDF}} \\
&= \sum_{i=I(k)+1}^{m} \sum_{j=1}^{J(k+1)-1} |[W_{i-1}, W_i] \cap [U_{j-1}, U_j]| \\
&= \left| \left(\cup_{i=I(k)+1}^{m} [W_{i-1}, W_i] \right) \cap \left(\cup_{j=1}^{J(k+1)-1} [U_{j-1}, U_j] \right) \right| \\
&= |[W_{I(k)}, W_m] \cap [0, U_{J(k+1)-1}]| \\
&= |[W(r_k), w_\Sigma] \cap [0, U(r_k)]| \qquad \text{(by (4.7))} \\
\alpha_{k+1\to k} &= \begin{cases} U(r_k) \Leftrightarrow W(r_k) & \text{if } U(r_k) > W(r_k) \\ 0 & \text{otherwise} \end{cases} . \qquad (4.10)
\end{aligned}$$

The desired result follows from (4.9) and (4.10). ■

We are now ready to prove the main result of this section.

Theorem 5 *If* $\mathbf{x} = (X, w) \in \mathbf{D}^{1,m}$ *and* $\mathbf{y} = (Y, u) \in \mathbf{D}^{1,n}$ *have equal weight* $w_\Sigma = u_\Sigma$, *then*

$$\mathrm{EMD}(\mathbf{x}, \mathbf{y}) = \frac{\int_{-\infty}^{\infty} |W(t) \Leftrightarrow U(t)| \, dt}{w_\Sigma}.$$

Furthermore, F^{CDF} *is an optimal feasible flow between* $\mathbf{x}$ *and* $\mathbf{y}$.

Proof. Note that $W(t)$ and $U(t)$ are constant over the interval $t \in [r_k, r_{k+1})$ for $k = 1, \ldots, m+n \Leftrightarrow 1$, $W(t) = U(t) \equiv 0$ for $t \in (\Leftrightarrow\infty, r_1)$, and $W(t) = U(t) \equiv w_\Sigma = u_\Sigma$ for $t \in [r_{m+n}, \infty)$. Therefore the integral of the absolute difference of the CDFs may be written as the finite summation

$$\int_{-\infty}^{\infty} |W(t) \Leftrightarrow U(t)| \, dt = \sum_{k=1}^{m+n-1} E_k \ , \qquad (4.11)$$

where

$$E_k = (r_{k+1} \Leftrightarrow r_k) \, |W(r_k) \Leftrightarrow U(r_k)|.$$

Consider the interval (r_k, r_{k+1}). At any position t in this interval, the absolute difference $|W(t) \Leftrightarrow U(t)|$ is equal to $|W(r_k) \Leftrightarrow U(r_k)|$. Suppose that $W(r_k) > U(r_k)$. Then in any feasible flow from $\mathbf{x}$ to $\mathbf{y}$, the net flow from r_k to r_{k+1} must be exactly $W(r_k) \Leftrightarrow U(r_k)$. If the net flow is less than this amount, then there will be less $\mathbf{x}$-weight than $\mathbf{y}$-weight in $[r_{k+1}, \infty)$ after the flow is complete. If the net flow is more than this amount, then there will be more $\mathbf{x}$-weight than $\mathbf{y}$-weight in $[r_{k+1}, \infty)$ after the flow is complete. See Figure 4.8(a).

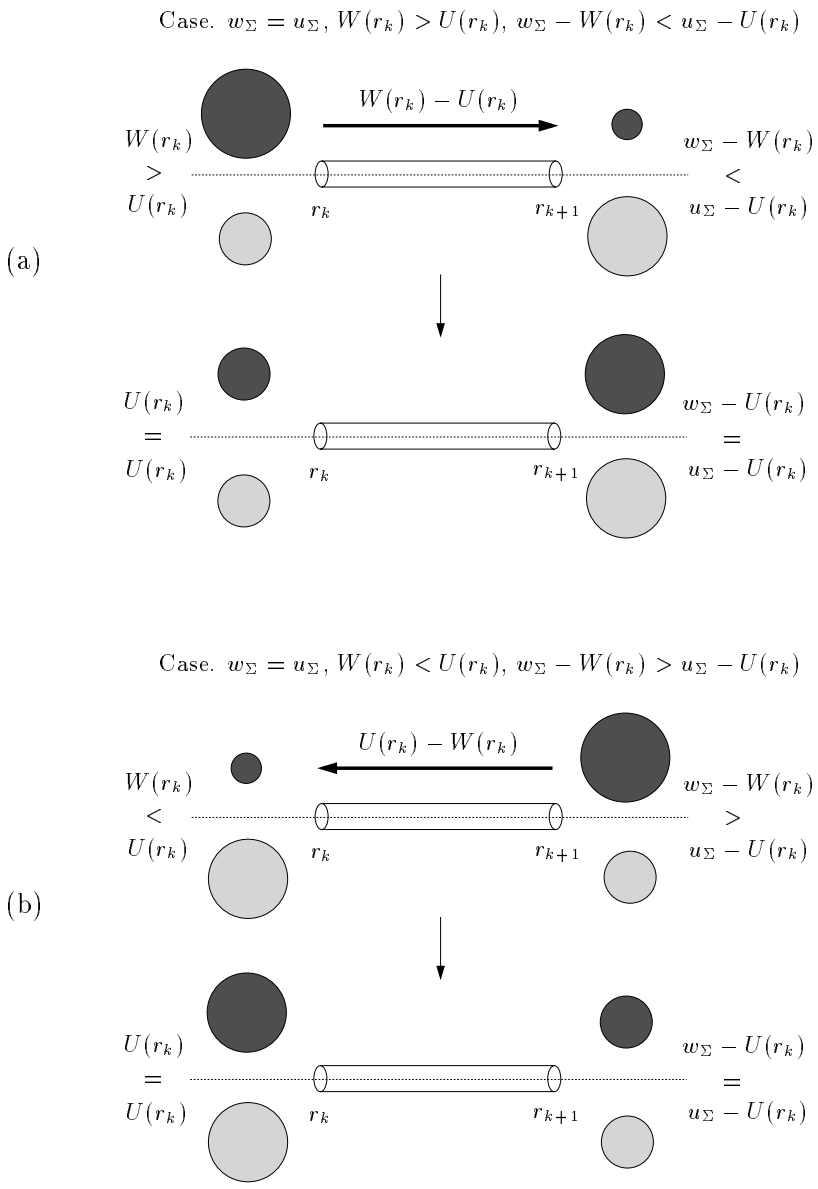


Figure 4.8: Flow Feasibility for Equal-Weight Distributions on the Real Line. $\mathbf{x} = (X, w)$ and $\mathbf{y} = (Y, u)$ are distributions in 1D. Here $r_1 \leq \cdots \leq r_{m+n}$ is the position-sorted list of points in $\mathbf{x}$ and $\mathbf{y}$, and $W(t)$ and $U(t)$ are the CDFs for $\mathbf{x}$ and $\mathbf{y}$, respectively. (a) $W(r_k) > U(r_k)$, $w_\Sigma \Leftrightarrow W(r_k) < u_\Sigma \Leftrightarrow U(r_k)$. In this case, a flow from $\mathbf{x}$ to $\mathbf{y}$ is feasible only if the net flow of $\mathbf{x}$-weight from r_k to r_{k+1} is exactly $W(r_k) \Leftrightarrow U(r_k)$. (b) $W(r_k) < U(r_k)$, $w_\Sigma \Leftrightarrow W(r_k) > u_\Sigma \Leftrightarrow U(r_k)$. In this case, a flow from $\mathbf{x}$ to $\mathbf{y}$ is feasible only if the net flow of $\mathbf{x}$-weight from r_{k+1} to r_k is exactly $U(r_k) \Leftrightarrow W(r_k)$.

Similar logic shows that if $U(r_k) > W(r_k)$, then the net flow of $\mathbf{x}$-weight from r_{k+1} to r_k must be exactly $U(r_k) \Leftrightarrow W(r_k)$. This case is illustrated in Figure 4.8(b). In either case, the

amount of work E_k done in moving weight from $\mathbf{x}$ over the interval (r_k, r_{k+1}) is at least E_k, and

$$\min_{F \in \mathcal{F}(\mathbf{x},\mathbf{y})} \mathrm{WORK}(F, \mathbf{x}, \mathbf{y}) \geq \sum_{k=1}^{m+n-1} E_k. \tag{4.12}$$

To complete the proof, note that Lemma 2 says that F^{CDF} is a feasible flow[1] which requires work $\sum_{k=1}^{m+n-1} E_k$ to match $\mathbf{x}$ and $\mathbf{y}$. It follows that

$$\min_{F \in \mathcal{F}(\mathbf{x},\mathbf{y})} \mathrm{WORK}(F, \mathbf{x}, \mathbf{y}) \leq \mathrm{WORK}(F^{\mathrm{CDF}}, \mathbf{x}, \mathbf{y}) = \sum_{k=1}^{m+n-1} E_k. \tag{4.13}$$

Combining (4.12), (4.13), and (4.11) gives the desired result after normalizing by $w_\Sigma = u_\Sigma$. ■

Pseudocode to compute the EMD between equal-weight distributions in one dimension (with ground distance equal to the L_1 distance) is given below. This code is a direct translation of (4.11) and computes the integral by a sweep over the position axis, summing areas of rectangles with bases $r_{k+1} \Leftrightarrow r_k$ and heights $|W(r_k) \Leftrightarrow U(r_k)|$. Again, see Figure 4.5.

```
function emd = EMD_1(x, y)
/* assumes K = 1, w_Σ = u_Σ, ground distance is L_1 */
/* assumes x_1 ≤ x_2 ≤ ··· ≤ x_m, y_1 ≤ y_2 ≤ ··· ≤ y_n */
    work = wsum = usum = r = 0
    /* first increment of work will be 0, regardless of r */
    i = j = 1
    xnext = x_1
    ynext = y_1
    while ((i ≤ m) or (j ≤ n))
        if (xnext ≤ ynext)
            work += |wsum-usum|*(xnext-r)
            wsum += w_i
            r = xnext
            i += 1
            xnext = (i ≤ m) ? x_i : ∞
        else
            work += |wsum-usum|*(ynext-r)
            usum += u_j
            r = ynext
            j += 1
```

[1] In [11], it is incorrectly stated that there is a unique feasible flow between equal-weight distributions in 1D. In fact, there may even be more than one optimal feasible flow. For example, suppose X = [0 1], w = [1 1], Y = [1 2], and u = [1 1]. Then F^{CDF} is given by $f_{11}^{\mathrm{CDF}} = f_{22}^{\mathrm{CDF}} = 1$, $f_{12}^{\mathrm{CDF}} = f_{21}^{\mathrm{CDF}} = 0$. The feasible flow F^* given by $f_{11}^* = f_{22}^* = 0$, $f_{12}^* = f_{21}^* = 1$ is also an optimal feasible flow between $\mathbf{x} = (X, w)$ and $\mathbf{y} = (Y, u)$.

```
            ynext = (j ≤ n) ? y_j : ∞
        end if
    end while
    return (work / usum)
end function
```

Assuming that the points in $x \in \mathbf{D}^{1,m}$ and $y \in \mathbf{D}^{1,n}$ are in sorted order, the routine EMD_1 runs in linear time $\Theta(m+n)$. The combined sorted list $r_1, \ldots, r_{m+n}$ of points in $\mathbf{x}$ and $\mathbf{y}$ is discovered by walking along the two sorted lists of points. At any time during the algorithm, there is a pointer to the next $\mathbf{x}$ and next $\mathbf{y}$ value to be considered. The value r_{k+1} then follows in constant time from the value of r_k.

The function EMD_1 does not compute the optimal CDF flow $F^{\mathrm{CDF}} = (f_{ij}^{\mathrm{CDF}})$. We can rewrite the EMD_1 routine as shown below so that it also returns the optimal CDF flow with the single EMD value. This code is a direct translation of the fact that $\int_{-\infty}^{\infty} |W(t) \Leftrightarrow U(t)|\ dt = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}^{\mathrm{CDF}} |x_i \Leftrightarrow y_j|$ (which follows from Theorem 5) and computes the integral by a sweep over the cumulative-weight axis, summing areas of rectangles with bases f_{ij}^{CDF} and heights $|x_i \Leftrightarrow y_j|$. See Figure 4.9.

```
 function [emd,CDFflow] = EMD_1(x, y)
/* assumes K = 1, w_Σ = u_Σ, ground distance is L_1 */
/* assumes x_1 ≤ x_2 ≤ ··· ≤ x_m, y_1 ≤ y_2 ≤ ··· ≤ y_n */
    work = prev = CDFflow.nFlow = 0
    i = j = 1
    wsum = w_1 /* holds W_i */
    usum = u_1 /* holds U_j */
    while ((i ≤ m) and (j ≤ n))
        CDFflow[CDFflow.nFlow].from = i
        CDFflow[CDFflow.nFlow].to = j
        if (usum ≤ wsum) /* check (U_j ≤ W_i) */
            f_ij^CDF = usum-prev
            work += f_ij^CDF × |x_i ⇔ y_j|
            prev = usum
            usum += u_j
            j += 1
        else
            f_ij^CDF = wsum-prev
            work += f_ij^CDF × |x_i ⇔ y_j|
            prev = wsum
            wsum += w_i
            i += 1
        end if
```

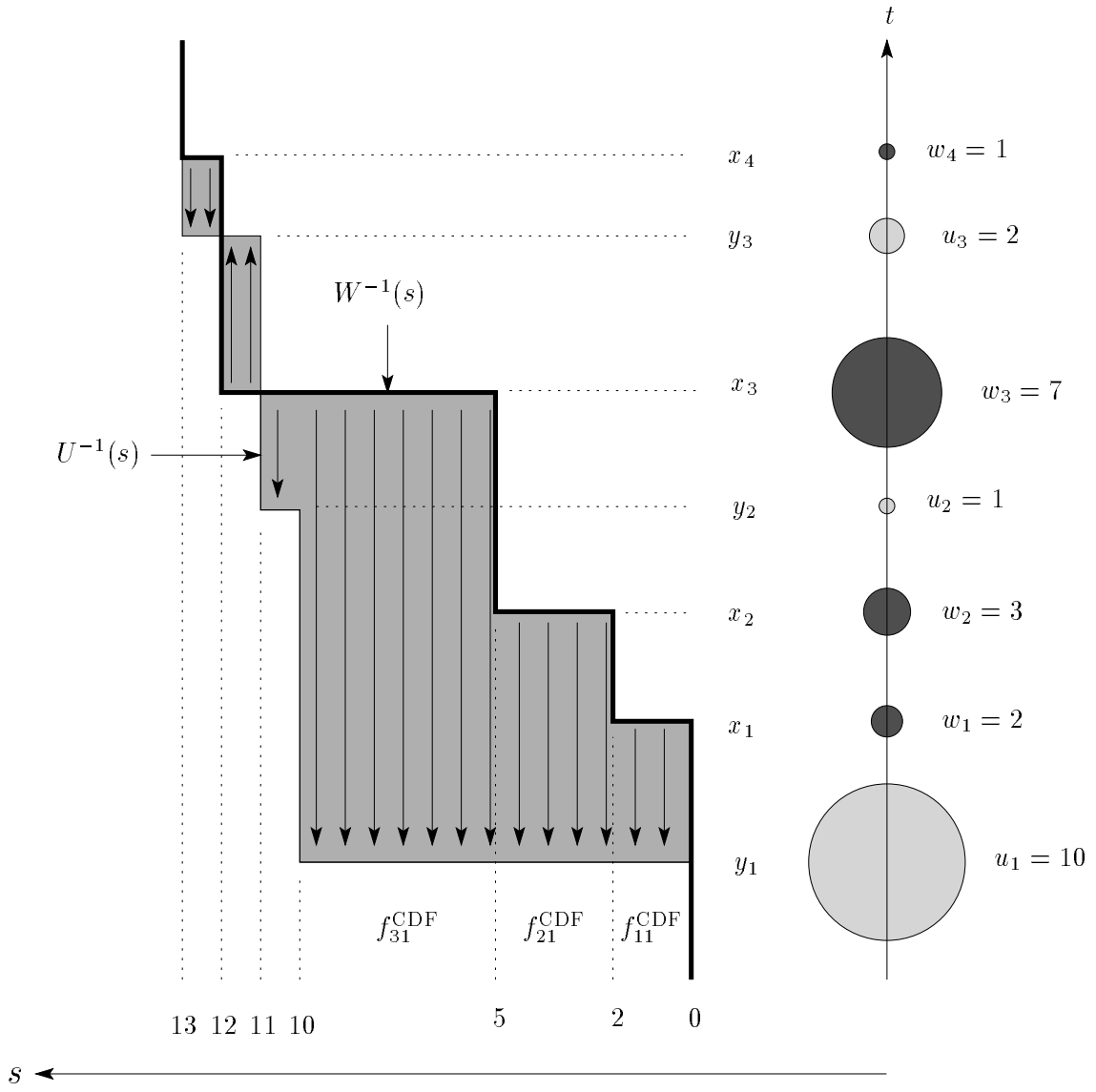


Figure 4.9: The Inverse CDFs. The area between the inverse CDFs $W^{-1}(s)$ and $U^{-1}(s)$ over $s \in [0, w_\Sigma] = [0, u_\Sigma]$ is clearly the same as the area between the CDFs $W(t)$ and $U(t)$ (see Figure 4.5) over $t \in (\Leftrightarrow\infty, \infty)$.

```
        CDFflow[CDFflow.nFlow].amount = f_ij^CDF
        CDFflow.nFlow += 1
    end while
    emd = work / usum
    return (emd,CDFflow)
end function
```

This version of EMD_1 also runs in $\Theta(m+n)$ time since there is a constant amount of computation done at each of the $m+n$ breakpoints $W_1, \ldots, W_m, U_1, \ldots, U_n$. All flow variables f_{ij}^{CDF} not explicitly contained in the variable CDFflow are equal to zero.

4.4 Modifications

We now discuss some useful modifications to the EMD. As initially stated, the EMD computation forces all the weight in the lighter distribution to match weight in the heavier distribution. In section 4.4.1, we extend the EMD to take another parameter $0 < \gamma \leq 1$ which specifies the fraction of the lighter distribution to be matched. The *partial Earth Mover's Distance* computation automatically selects the best weight from the lighter distribution to match.[2] The ability to compute the best partial match is important for robustness in the presence of outliers and/or missing data.[3] The γ parameter is an attempt to avoid penalizing the non-matching parts of two distributions which have a lot in common. Remember that the goal is to measure visual similarity by matching summary distributions, and visual similarity may follow from only a partial match. An alternative similarity measure which accounts for this fact asks "How much weight can be matched when flow distances are limited to at most some given ground distance τ?". This *restricted Earth Mover's Distance* is the subject of section 4.4.2.

4.4.1 The Partial Earth Mover's Distance

The *partial Earth Mover's Distance* EMD^{γ} matches only a given fraction $0 < \gamma \leq 1$ of the weight of the lighter distribution or some absolute amount of weight $0 < \gamma \leq \min(w_\Sigma, u_\Sigma)$. The former case in which γ is a relative quantity is called the *relative* partial EMD, and the latter case in which γ is an absolute quantity is called the *absolute* partial EMD. In a

[2] This name may be slightly misleading since the EMD already does partial matching. When one distribution is heavier than the other, all the weight in the lighter distribution is matched, but some of the weight in the heavier distribution is unmatched. With the partial EMD, some of the weight of the lighter distribution is unmatched.

[3] The EMD is robust to a small amount of outlier mass since the large ground distances needed to match the outlier mass are weighted by small fractions of mass moved.

relative partial EMD problem, the amount of weight matched is $M(\gamma) = \gamma \min(w_\Sigma, u_\Sigma)$; in an absolute partial EMD problem, the amount of weight matched is $M(\gamma) = \gamma$. In either case, the conditions for a feasible flow are

$$\begin{aligned} f_{ij} &\geq 0 \qquad i = 1, \ldots, m,\ j = 1, \ldots, n, \\ \sum_{j=1}^{n} f_{ij} &\leq w_i \qquad i = 1, \ldots, m, \\ \sum_{i=1}^{m} f_{ij} &\leq u_j \qquad j = 1, \ldots, n, \quad \text{and} \\ \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} &= M(\gamma). \end{aligned} \tag{4.14}$$

The only difference in the feasibility conditions for the partial EMD and the ordinary EMD are in the final conditions (4.14) and (4.4) which indicate the total amount of weight to match. If we denote the set of feasible flows between $\mathbf{x}$ and $\mathbf{y}$ for partial match parameter γ as $\mathcal{F}^\gamma(\mathbf{x}, \mathbf{y})$, then we define the partial EMD as

$$\text{EMD}^\gamma(\mathbf{x}, \mathbf{y}) = \frac{\min_{F=(f_{ij}) \in \mathcal{F}^\gamma(\mathbf{x},\mathbf{y})} \text{WORK}(F, \mathbf{x}, \mathbf{y})}{M(\gamma)}. \tag{4.15}$$

Since γ is given, the denominator of (4.15) is fixed for an EMD$^\gamma$ computation. An example partial EMD computation is shown in Figure 4.10. In section 4.2, we showed that the work minimization problem for the ordinary EMD computation can be solved as a transportation problem. We now show that the same is true for the (relative or absolute) partial EMD computation. Therefore, the same transportation problem code that is used to compute the EMD can also be used to compute the partial EMD.

Suppose that $\mathbf{x}$ is at least as heavy as $\mathbf{y}$; i.e., $w_\Sigma \geq u_\Sigma$. Then the work minimization problem in the numerator of (4.15) is the balanced transportation problem

$$\begin{aligned} s_i &= w_i \quad i = 1, \ldots, m \\ s_{m+1} &= u_\Sigma \Leftrightarrow M(\gamma) \\ d_j &= u_j, \quad j = 1, \ldots, n \\ d_{n+1} &= w_\Sigma \Leftrightarrow M(\gamma) \\ c_{ij} &= d(x_i, y_j) \quad i = 1, \ldots, m,\ j = 1, \ldots, n \\ c_{m+1,j} &= 0 \quad j = 1, \ldots, n \\ c_{i,n+1} &= 0 \quad i = 1, \ldots, m \\ c_{m+1,n+1} &= \infty. \end{aligned}$$

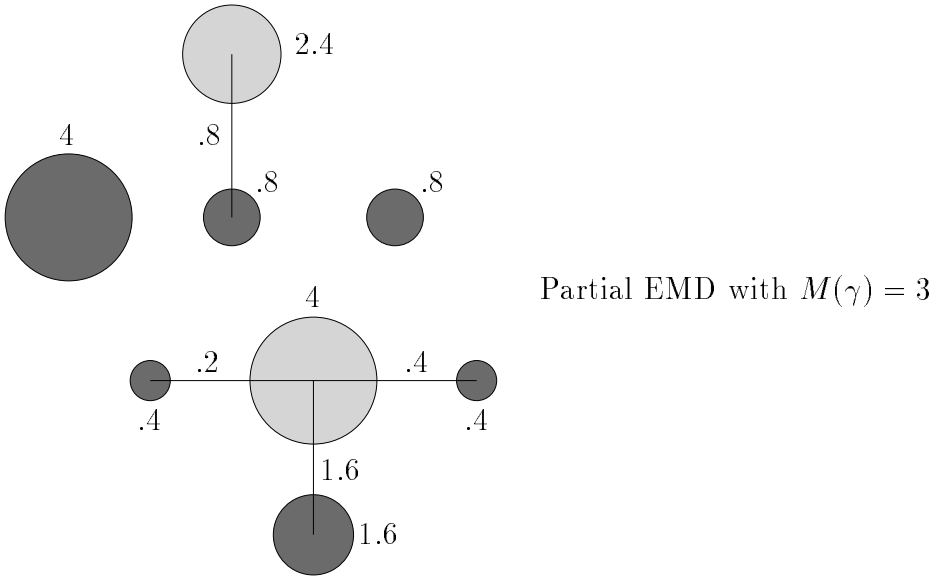


Figure 4.10: Partial EMD Example. The dark gray distribution has total weight 8, while the light gray distribution has total weight 6.4. An optimal flow for the partial EMD when $M(\gamma) = 3$ units of weight must be matched is shown by the labelled edges. All ground distances used in this flow are equal, and less than all ground distances not used.

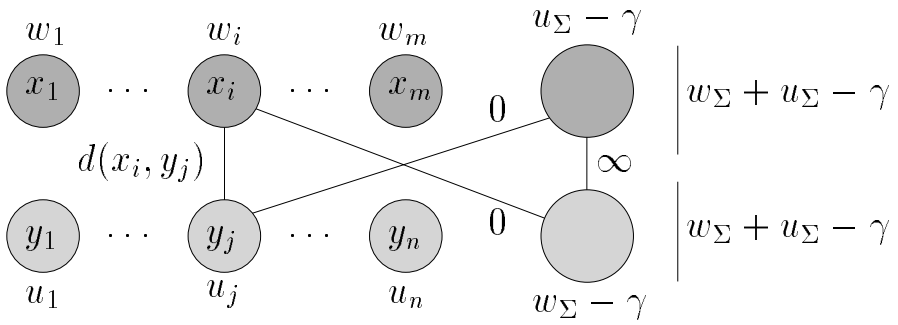


Figure 4.11: The Partial EMD as a Balanced Transportation Problem. See the text for an explanation.

A graphical representation is shown in Figure 4.11. The total supply s_Σ and total demand d_Σ are both equal to $w_\Sigma + u_\Sigma \Leftrightarrow M(\gamma)$. The dummy supplier $m+1$ is given a supply which equal to the unmatched weight of $\mathbf{y}$, while the dummy demander $n+1$ is given a demand which is equal to the unmatched weight of $\mathbf{x}$. The weight of the dummy supplier is prevented from matching the weight of the dummy demander with the requirement $c_{m+1,n+1} = \infty$, so all of the weight of the dummy supplier will be matched at no cost to demanders $j = 1, \ldots, n$. Of

the remaining supply w_Σ possessed by suppliers $i = 1, \ldots, m$, $w_\Sigma \Leftrightarrow M(\gamma)$ will be matched at no cost to the dummy demander. Therefore only $M(\gamma)$ weight will be matched at possibly nonzero cost. An algorithm to solve the transportation problem will find the optimal way to transport this weight from suppliers $i = 1, \ldots, m$ to demanders $j = 1, \ldots, n$. If $\mathbf{x}$ is lighter than $\mathbf{y}$, then the above formulation with the roles of $\mathbf{x}$ and $\mathbf{y}$ interchanged allows the partial EMD work minimization problem to be formulated as a balanced transportation problem.

In section 4.3.1, we discussed the special case in which the two distributions compared by the EMD are point sets. The EMD yields an optimal one-to-one matching in which each point in the smaller set is matched to a point in the larger set. Using the partial EMD instead of the EMD, we can find an optimal matching which matches only some subset of the points in the smaller set. The total number of points to be matched using EMD$^\gamma$ is $M(\gamma)$. As long as γ is selected so that $M(\gamma)$ is an integer, all the supplies and demands in the corresponding transportation problem will be integers (see the above formulations), with all the non-dummy supplies and demands equal to one. Applying the transportation simplex algorithm will yield an optimal one-to-one matching between size $M(\gamma)$ subsets of the smaller and larger sets. If the distributions $\mathbf{x} \in \mathbf{D}^{K,m}$ and $\mathbf{y} \in \mathbf{D}^{K,n}$ are point sets in $\mathbf{R}^K$ with $m \geq n$, then

$$\begin{aligned} \mathrm{EMD}^\gamma(\mathbf{x}, \mathbf{y}) &= \frac{\min_{F=(f_{ij}) \in \mathcal{F}^\gamma(\mathbf{x},\mathbf{y})} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d(x_i, y_j)}{M(\gamma)} \\ &= \frac{\min_{\phi \in \Phi^\gamma} \sum_{j \in \mathrm{domain}(\phi)} d(x_{\phi(j)}, y_j)}{M(\gamma)}, \end{aligned}$$

where Φ^γ is the set of one-to-one partial correspondences

$$\Phi^\gamma = \{\, \phi : S \to [1..m] \;\mid\; S \subseteq [1..n],\, |S| = M(\gamma),\, \phi(j_1) = \phi(j_2) \Leftrightarrow j_1 = j_2 \;\; \forall j_1, j_2 \in S \,\}.$$

It is important to note that only the number of points to be matched is given; the partial EMD figures out the best subsets to match.

4.4.2 The Restricted Earth Mover's Distance

The *restricted Earth Mover's Distance* τ-EMD is a measure of how much weight can be matched when ground distances for transportation are limited to a threshold τ. When comparing two distributions $\mathbf{x} = (X, w)$ and $\mathbf{y} = (Y, u)$, the maximum amount of weight that can be matched if transportation distances $d_{ij} = d(x_i, y_j)$ are unrestricted is $M = \min(w_\Sigma, u_\Sigma)$. Let M_τ denote the maximum amount of weight that can be matched using

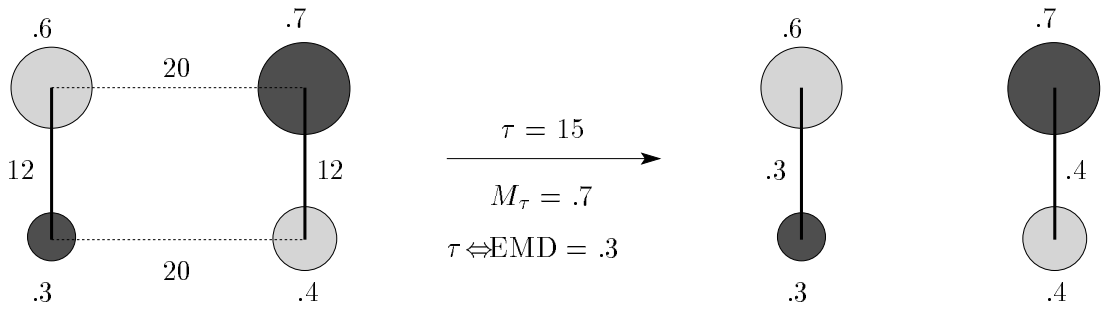


Figure 4.12: τ-EMD Example. (left) Equal-weight distributions with distances between weight locations. (right) If $\tau = 15$, then the maximum amount of weight that can be matched is $M_\tau = 0.7$. The maximal matching is indicated by the labelled edges. Weight cannot be matched between the locations which are $20 > \tau$ units apart. Since the total weight of both distributions is one, the fraction of weight that cannot be matched is τ-EMD $= 0.3$.

only distances $d_{ij} \leq \tau$. Then we define the restricted EMD as

$$\tau\text{-EMD}(\mathbf{x}, \mathbf{y}) = 1 \Leftrightarrow \frac{M_\tau(\mathbf{x}, \mathbf{y})}{\min(w_\Sigma, u_\Sigma)}. \tag{4.16}$$

Note that the τ-EMD actually equals the fraction of weight that *cannot* be matched, with $0 \leq \tau\text{-EMD}(\mathbf{x}, \mathbf{y}) \leq 1$, so that τ-EMD is a dissimilarity measure rather than a similarity measure. The extreme values are zero when the maximum amount of weight can be matched, and one when none of the weight can be matched. An example is shown in Figure 4.12. We now show that the computation of $M_\tau(\mathbf{x}, \mathbf{y})$ is a transportation problem.

The $M_\tau(\mathbf{x}, \mathbf{y})$ computation is the linear program

$$M_\tau(\mathbf{x}, \mathbf{y}) = \max_{F \in \mathcal{F}(\mathbf{x}, \mathbf{y})} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}[d(x_i, y_j) \leq \tau],$$

where

$$d_\tau(x_i, y_j) = [d(x_i, y_j) \leq \tau] = \begin{cases} 1 & \text{if } d(x_i, y_j) \leq \tau \\ 0 & \text{otherwise} \end{cases}.$$

The flow constraints are the same as for the original EMD computation, but now we sum up the matched weight f_{ij} whenever $d(x_i, y_j) \leq \tau$. The transportation problem here is

$$\Leftrightarrow M_\tau(\mathbf{x}, \mathbf{y}) = \min_{F \in \mathcal{F}(\mathbf{x}, \mathbf{y})} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}(\Leftrightarrow d_\tau(x_i, y_j))$$

In fact, this is an original EMD computation with ground distances $(\Leftrightarrow d_\tau(x_i, y_j))$. There is

no restriction in the transportation problem that costs be nonnegative, and the transportation simplex method makes no such assumption.

4.5 Use in Scale Estimation

In this section, we show how to use the EMD to estimate the scale at which a pattern appears within an image. Figure 4.13(a) shows an example of the color pattern problem. Scale estimation is a critical step in solving the pattern problem accurately and efficiently. A good scale estimate is important for accuracy because the scale determines how much information in the image is compared to the pattern; it is important for efficiency because trying many scales will be inefficient, especially if one is interested in finding very small occurrences of the pattern. Along with a scale estimate, our method also returns a measure indicating the distance between the pattern at the predicted scale and the image. If this distance is large, then the pattern probably does not occur within the image.

We regard an image as a distribution of color mass or curve orientation mass (for the shape pattern problem) in position space. Our scale estimation method uses the EMD to compare image summary distributions after marginalizing away position. Ignoring position information does throw away useful information, but it reduces the complexity of the summary distributions and, therefore, allows fast scale estimation. We will show that it is possible to get very good scale estimates without position information if the pattern has a single distinctive feature with respect to the image. In the color pattern problem, the marginalized distribution is a distribution in color space. In order to keep the distribution size small, the colors of an image are clustered into a small number of groups (approximately twenty). The weight of a color cluster in CIE-Lab space ([88]) is the fraction of the total image area classified as that color. Thus the total weight of a summary distribution is one.

Suppose that a pattern occurs in an image as a fraction $c^* \in (0, 1]$ of the total image area. An example is shown in Figure 4.13(a). Let $\mathbf{x}$ and $\mathbf{y} = (Y, u)$ denote unit-weight color signatures of the image and pattern, respectively. See Figure 4.13(b),(d). Since (Y, c^*u) is lighter than $\mathbf{x}$, the EMD finds the optimal matching between c^* of the image color weight and the color weight in (Y, c^*u). Consider the ideal case of an exact pattern occurrence, with the same color clusters used in $\mathbf{x}$ and $\mathbf{y}$ for the pattern colors. Then the c^* of $\mathbf{x}$'s color weight contributed by the pattern occurrence will match exactly the color weight in (Y, c^*u), and $\mathrm{EMD}(\mathbf{x}, (Y, c^*u)) = 0$. Furthermore, $\mathrm{EMD}(\mathbf{x}, (Y, cu)) = 0$ for $c \in (0, c^*]$ since there is still enough image weight of each pattern color to match all the weight in (Y, cu). In general, we will prove that $\mathrm{EMD}(\mathbf{x}, (Y, cu))$ decreases as c decreases and eventually becomes constant for $c \in (0, c^0]$, as shown in Figure 4.13(e). If the graph levels off at a small EMD,

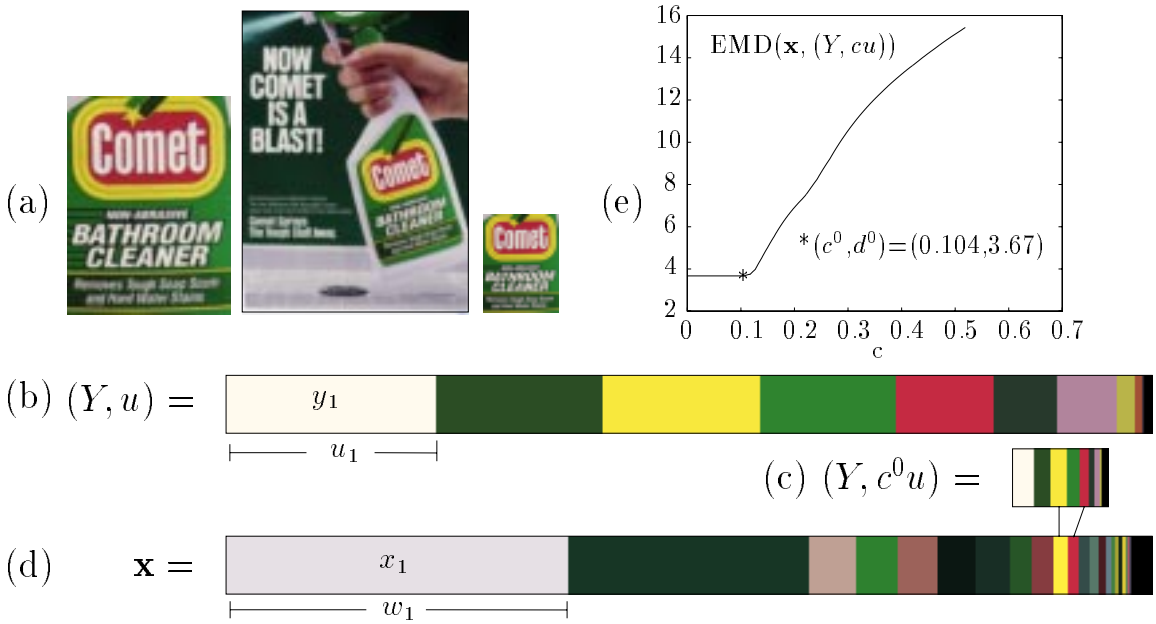


Figure 4.13: Scale Estimation – Main Idea. (a) pattern, image, and pattern scaled according to the scale estimate. (b) pattern signature. (c) pattern signature with weights scaled by the estimate. (d) image signature. (e) EMD($\mathbf{x}, (Y, cu)$) versus c.

then the pattern may occur in the image, and we take c^0 to be the scale estimate.

The main property of this scale estimation method is that in the ideal case it overestimates the scale by the *minimum* amount of background clutter over all pattern colors, where the amount of background clutter for a color is the amount of that color present in the image but not part of the pattern occurrence. Just one pattern color with a small amount of background clutter is enough to obtain an accurate scale estimate. Consider the example in Figure 4.13. The scale estimate c^0 is such that the amounts of red and yellow in the scaled pattern signature $(Y, c^0 u)$ are roughly equal to the amounts of red and yellow in the image, as shown in Figure 4.13(c). At scale c^0, there is still plenty of image weight to match the other pattern colors in $(Y, c^0 u)$. If there were a bit more red and yellow in the image, then the scale estimate c^0 would be a bit too high. In this example, red and yellow have zero background clutter since the only place that they occur in the image is within the pattern occurrence. Note that an accurate scale estimate is computed even in the presence of the dark green in the Comet label for which there is a lot of background clutter.

The preceeding discussion tacitly assumes that the pattern occurs only once in the image. Since our method does not use the positions of colors, it cannot tell the difference between two pattern occurrences at scales c_1 and c_2, and one larger occurrence at scale

$c_1 + c_2$. In this two pattern occurrence example, the computed scale estimate will be at least $c_1 + c_2$ if the same color clusters are used in $\mathbf{x}$ and $\mathbf{y}$ for the pattern colors.

We now study of the function $E(c) = \text{EMD}((X, w), (Y, cu))$, where $\mathbf{x} = (X, w)$ and $\mathbf{y} = (Y, u)$ are equal-weight distributions with total weight one, and $0 < c \leq 1$. The distribution (Y, cu) has total weight $c \leq 1$. The function $E(c)$ is thus given by

$$E(c) = \frac{\min_{(f_{ij}) \in \mathcal{F}(\mathbf{x}, (Y, cu))} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d(x_i, y_j)}{c},$$

where $(f_{ij}) \in \mathcal{F}(\mathbf{x}, (Y, cu))$ iff

$$\begin{aligned} f_{ij} &\geq 0 \qquad i = 1, \ldots, m,\ j = 1, \ldots, n, \\ \sum_{i=1}^{m} f_{ij} &= cu_j \quad j = 1, \ldots, n, \quad \text{and} \\ \sum_{j=1}^{n} f_{ij} &\leq w_i \quad i = 1, \ldots, m. \end{aligned}$$

Now set $h_{ij} = f_{ij}/c$. Then

$$E(c) = \min_{(h_{ij}) \in \mathcal{F}((X, \frac{1}{c}w), \mathbf{y})} \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} d(x_i, y_j),$$

where $(h_{ij}) \in \mathcal{F}((X, \frac{1}{c}w), \mathbf{y})$ iff

$$h_{ij} \geq 0 \qquad i = 1, \ldots, m,\ j = 1, \ldots, n, \tag{4.17}$$

$$\sum_{i=1}^{m} h_{ij} = u_j \qquad j = 1, \ldots, n, \quad \text{and} \tag{4.18}$$

$$\sum_{j=1}^{n} h_{ij} \leq \frac{1}{c} w_i \quad i = 1, \ldots, m. \tag{4.19}$$

Note that

$$\mathcal{F}((X, w/c_1), \mathbf{y}) \subseteq \mathcal{F}((X, w/c_2), \mathbf{y}) \Longleftrightarrow \frac{1}{c_1} \leq \frac{1}{c_2} \Longleftrightarrow c_2 \leq c_1. \tag{4.20}$$

Algebraically, the fact that the feasible region $\mathcal{F}((X, \frac{1}{c}w), \mathbf{y})$ increases as c decreases (and vice-versa) is because the final m constraints (4.19) involving the w_i's get weaker (stronger) as c decreases (increases). Logically, this fact make sense because the less mass that the EMD is asked to matched, the more ways there are to perform the matching.

Since $E(c)$ is a minimum over $\mathcal{F}((X, \frac{1}{c}w), \mathbf{y})$, it follows from (4.20) that $E(c)$ is a non-decreasing function of c:

$$E(c_1) \geq E(c_2) \quad \text{iff} \quad c_1 \geq c_2. \tag{4.21}$$

In fact, however, we can say something stronger than (4.21). Consider the convex polytope $Q \subseteq \mathbf{R}^{mn}$ defined by (4.17) and (4.18), and the convex polytope $P(c) \subseteq \mathbf{R}^{mn}$ defined by (4.19), so that

$$\mathcal{F}((X, w/c), \mathbf{y}) = Q \cap P(c). \tag{4.22}$$

Q is bounded since its constraints imply that $0 \leq h_{ij} \leq u_j$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$. The polytope $P(c)$ converges to $\mathbf{R}^{mn}$ as c decreases to zero since $\frac{1}{c}$ increases to ∞. Since Q is bounded, there is some c^0 for which $Q \subseteq P(c)$ $\forall c \leq c^0$. From this fact and (4.22), it follows that

$$\mathcal{F}((X, w/c), \mathbf{y}) = Q \quad \forall c \leq c^0,$$

and, hence,

$$E(c) = E(c^0) \quad \forall c \leq c^0. \tag{4.23}$$

Thus $E(c)$ decreases as c decreases, until some point c^0 at which the curve flattens out. Examples are shown in figures 4.13(e), 4.15, and 4.16.

To help with the intuition for (4.23), consider a simple color pattern problem example. Suppose the pattern contains 30% red, 40% white, and 30% blue, and the pattern is 20% of the image. In the ideal case of perfect color matches, the image has at least 6% red, 8% white, and 6% blue due to the presence of the pattern. If the pattern is scaled by less than $c = 0.20 = 20\%$, then its distribution will contain less than 6% red, 8% white, and 6% blue, and all of this color mass can be matched perfectly to the color masses of the image. The EMD will be zero for all values $0 < c \leq 0.20$.

We take as our scale estimate the largest c for which there is no real improvement in the EMD when c is decreased. What constitutes "no real improvement" in the value of the EMD is given as a parameter ε_d. There is also a parameter ε_c to specify the accuracy that is required for the scale estimate. Finally, the parameter $c_{\min}$ gives the smallest scale to be examined. The largest c for which there is no improvement in the EMD can be found via a binary search along the c-axis. See Figure 4.14. The pseudocode given below returns the scale estimate c^0, the EMD value $d^0 = \text{EMD}(\mathbf{x}, (Y, c^0 u))$, and an optimal flow flow^0 at the scale estimate.

function $[c^0, d^0, \text{flow}^0] = \text{ScaleEstimate}(\mathbf{x}, \mathbf{y}, c_{\min}, \varepsilon_c, \varepsilon_d)$
/* $\mathbf{x} = (X, w)$, $\mathbf{y} = (Y, u)$ */
/* *assumes* $w_\Sigma = u_\Sigma = 1$ */
    $c_{\max} = 1$
    $[d_{\min}, \text{flow}_{\min}] = \text{EMD}((X, w), (Y, c_{\min} u))$
    /* *loop invariant:* $c_{\min} \leq c^0 \leq c_{\max}$ */

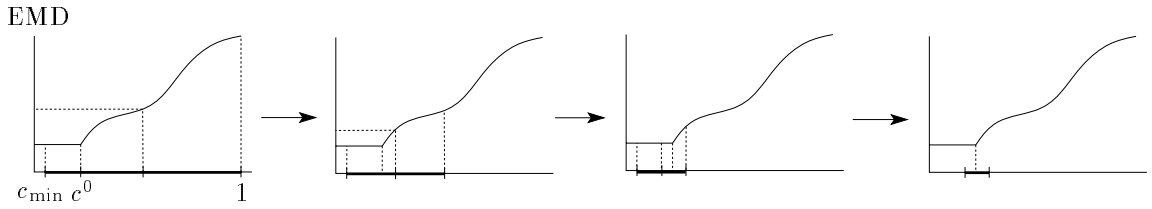


Figure 4.14: Scale Estimation Algorithm. Binary search narrows down the interval in which c^0 must occur.

while $(c_{\max} \Leftrightarrow c_{\min} > \varepsilon_c)$
  $c_{\text{mid}} = (c_{\min} + c_{\max})/2$
  $[d_{\text{mid}}, \text{flow}_{\text{mid}}] = \text{EMD}((X, w), (Y, c_{\text{mid}}u))$
  if $(|d_{\text{mid}} \Leftrightarrow d_{\min}| \leq \varepsilon_d)$
    $d_{\min} = d_{\text{mid}}$
    $c_{\min} = c_{\text{mid}}$
    $\text{flow}_{\min} = \text{flow}_{\text{mid}}$
  else
    $c_{\max} = c_{\text{mid}}$
  end if
end while
return ($c_{\min}$,$d_{\min}$,$\text{flow}_{\min}$)
end function

Here $c_{\min}$ is the smallest scale pattern that the user wants to find. If the returned distance d^0 is greater than a user supplied threshold τ, then we report that the pattern does not occur within the image. Otherwise, we take c^0 as an estimate of the pattern scale within the image.

The ScaleEstimate routine requires at most $\lceil \log_2(1/\varepsilon_c) \rceil + 1$ EMD computations since the length of the initial interval $[c_{\min}, c_{\max}] = [c_{\min}, 1]$ is at most one, and this interval is cut in half after each EMD call within the while loop (the "+1" is from the initial EMD call at $c = c_{\min}$ outside the loop). If, for example, $\varepsilon_c = 0.001$, then at most 11 EMD calls are made before $|[c_{\min}, c_{\max}]| \leq \varepsilon_c$. Note that the point sets of the distributions remain constant (X and Y) throughout the execution of ScaleEstimate. Thus the cost matrix (c_{ij}), $c_{ij} = d_{ij} = d(x_i, y_j)$, for the EMD transportation problems can be computed once at the beginning of ScaleEstimate and used for all subsequent EMD computations.

If we pass the threshold τ to ScaleEstimate, then the routine can exit after the first call $\text{EMD}((X, w), (Y, c_{\min}u))$ if this quantity is greater than τ. This can happen, for example, if a large part of the pattern is red, but there is no color similar to red in the image. No matter how small the value of c, the distances that image color mass must flow in color space to

cover the red pattern mass will be large in this case. Recall that the EMD is equal to the average ground distance that mass travels during an optimal flow. Since a large fraction of the total mass moved must travel large distances, the average distance moved, and hence the EMD, will be large. If $\mathrm{EMD}((X,w),(Y,c_{\min}u)) > \tau$, then ScaleEstimate performs only one EMD computation before concluding that the pattern does not occur in the image.

ScaleEstimate may also benefit in efficiency from the use of efficient, effective lower bounds on the EMD. Only the result of comparing $\mathrm{EMD}((X,w),(Y,c_{\mathrm{mid}}u))$ with the current $d_{\min}$ is needed to determine in which half of $[c_{\min},c_{\max}]$ the scale estimate c^0 occurs. The actual value of the EMD is not needed if it can be proven by other means that $\mathrm{EMD}((X,w),(Y,c_{\mathrm{mid}}u)) > d_{\min}$ (as in the first and second frames in Figure 4.14). If so, ScaleEstimate can update $c_{\max} = c_{\mathrm{mid}}$ without performing an EMD computation. Also, if a lower bound on $\mathrm{EMD}((X,w),(Y,c_{\min}u))$ is greater than τ, then ScaleEstimate can exit without performing a single EMD computation. Whether or not ScaleEstimate runs faster using lower bounds depends on how long the lower bounds take to compute and how often they succeed in pruning an EMD computation. Lower bounds on the EMD are discussed in Chapter 5. We now shift from our discussion of efficiency issues to a more general discussion of the ScaleEstimate algorithm.

Consider the ideal case of perfectly matching features in the pattern and image. As previously mentioned, our scale estimation method overestimates the scale by the *minimum* amount of background clutter over all pattern colors in the ideal case. Suppose, for example, the pattern distribution is 50% red, 20% white, and 30% blue, and that the pattern represents 20% of the total image area. Then the image has at least 10% red, 4% white, and 6% blue from the pattern. Suppose the exact distribution of the image is 40% red, 5% white, 25% blue, 15% green, and 15% yellow. The white mass from the image will not be covered completely until the pattern distribution is scaled by $c = 0.25 = 25\%$. After this point, there will be no gain in EMD with further decreases in scale. If the image had 4% white instead of 5% white, our scale estimate would have been exactly correct at $c^0 = 0.20 = 20\%$.

In general, the reasoning is not so clear cut because corresponding parts of the pattern and image will not have exactly the same color (and even if the color matches were perfect in the original images, clustering in color space to produce the small distributions will likely destroy that perfection), and optimal matching strategies can match color mass from one color in the pattern to several colors in the image. In practice, we have observed scale estimates which are a little smaller than predicted by an ideal case analysis. This is true in the Comet example shown in Figure 4.13, where there is zero background clutter for yellow and red but the scale is slightly underestimated.

4.5.1 Experiments with the Color Pattern Problem

Figures 4.15–4.19 illustrate the performance of our scale estimation algorithm for the color pattern problem, where the patterns are product logos and the images are product advertisements.

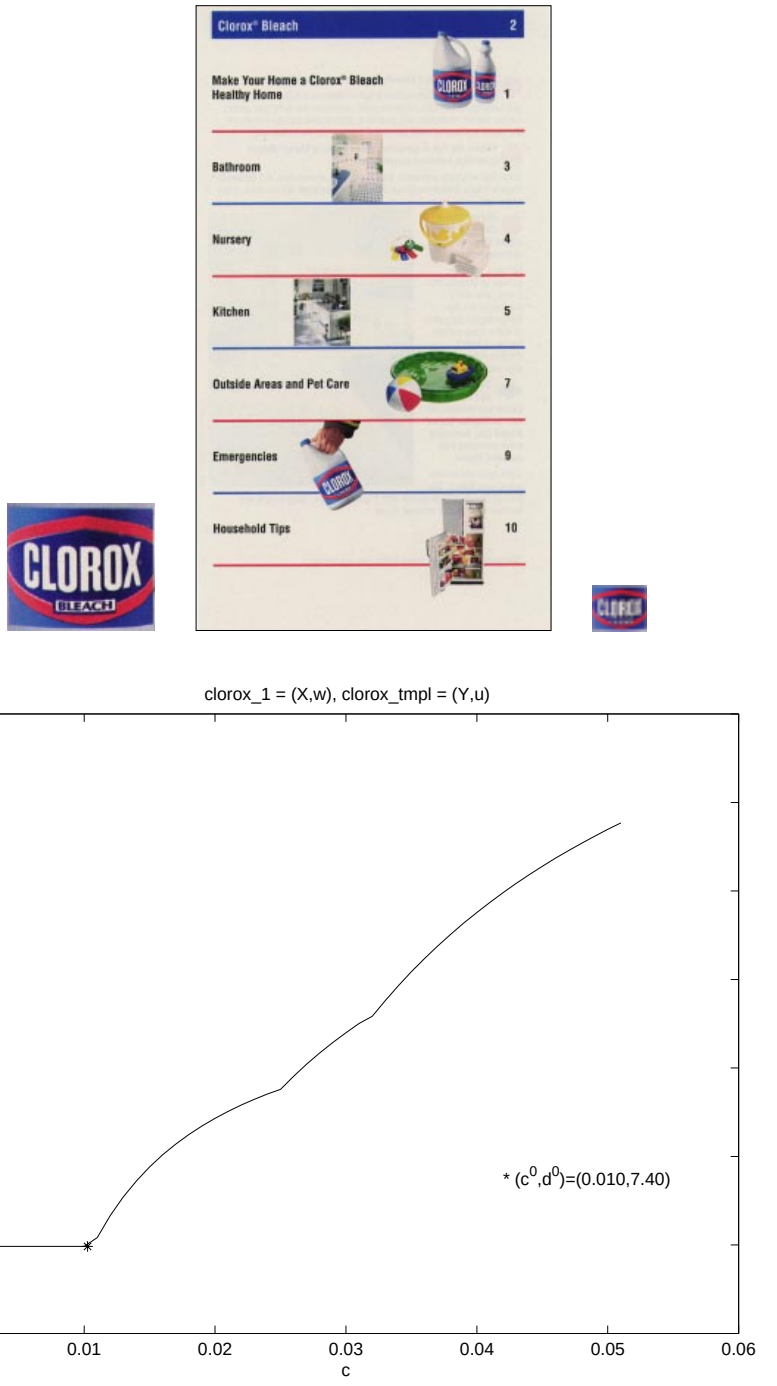


Figure 4.15: Scale Estimation – Clorox Example. From left to right in the first row, we see the Clorox logo (pattern), a Clorox advertisement, and the Clorox logo scaled according to the scale estimate given by the graph in the second row. The graph predicts that the pattern occurs at scale $c^0 = 1.0\%$ of the image. The top left, top right, and bottom Clorox logos occupy approximately 0.5%, 0.2%, and 0.5%, respectively, of the Clorox advertisement area.

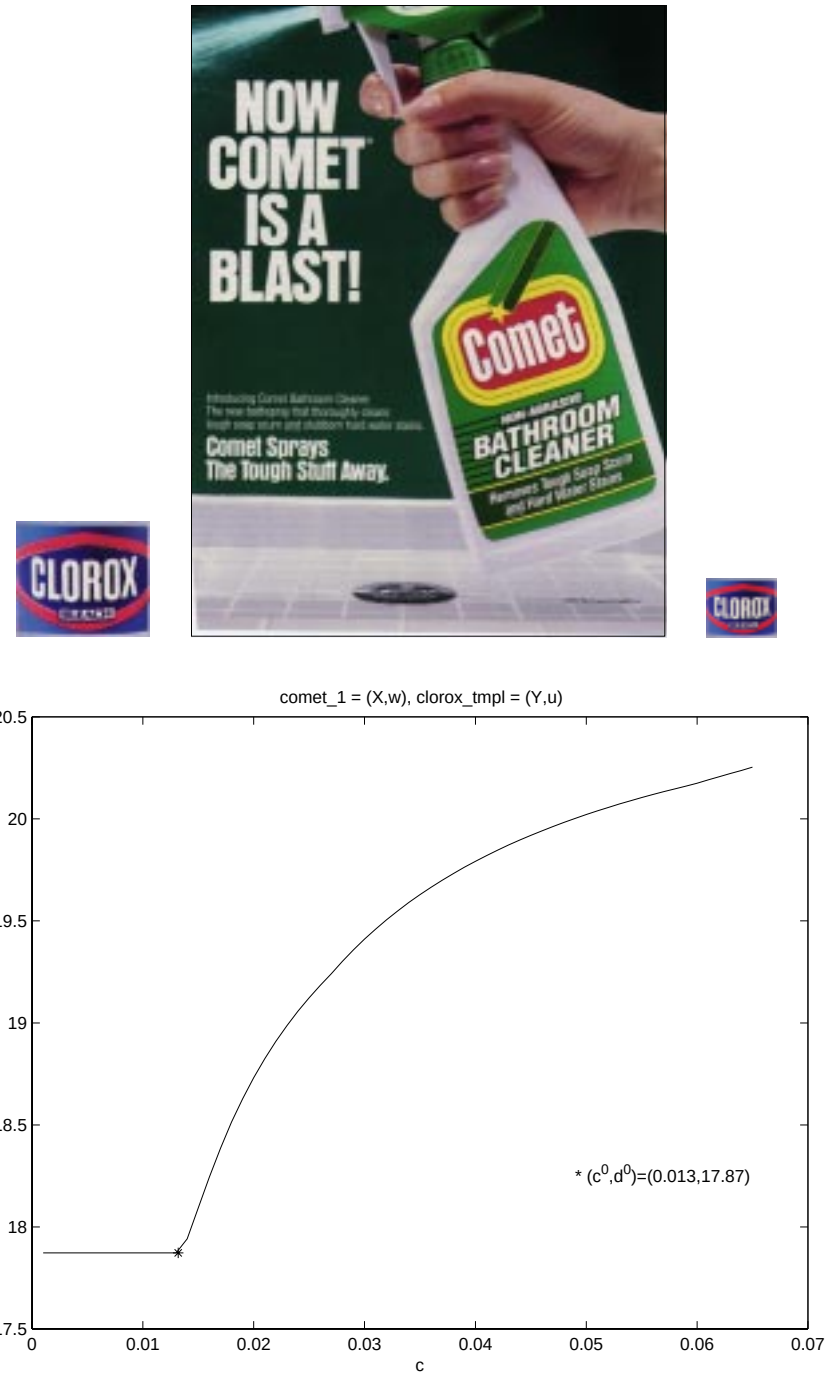


Figure 4.16: Scale Estimation – Pattern Not in the Image. From left to right in the first row, we see the Clorox logo (pattern), a Comet advertisement, and the Clorox logo scaled according to the scale estimate given by the graph in the second row. The graph predicts that the pattern occurs as $c^0 = 1.3\%$ of the image. However, the EMD at scale c^0 is $d^0 = 17.87$ units in CIE-Lab space. This large EMD value indicates that the pattern probably does *not* occur within the image.

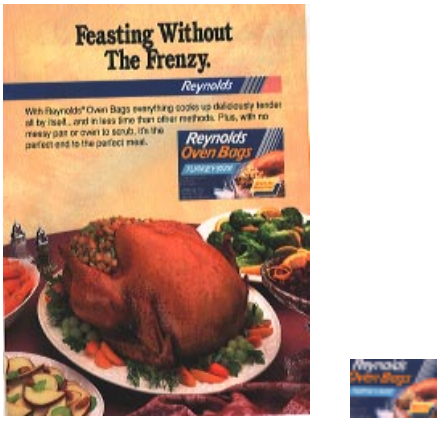
(a) (b) (c) (d)

Figure 4.17: Scale Estimation Results – Example Set 1. In each of the examples, the advertisement is shown on the left and the scaled (according to our prediction) logo is shown on the right. Let c^0 denote the predicted scale and c denote the (approximate) measured scale (in terms of fraction of advertisement area). (a) $c^0 = 2.2\%$, $c = 3.8\%$. (b) $c^0 = 4.2\%$, $c = 6.3\%$. (c) $c^0 = 2.6\%$, $c = 4.0\%$. (d) $c^0 = 6.3\%$, $c = 7.9\%$.

(a)

(b)

(c)



(d)

Figure 4.18: Scale Estimation Results – Example Set 2. In each of the examples, the advertisement is shown on the left and the scaled (according to our prediction) logo is shown on the right. Let c^0 denote the predicted scale and c denote the (approximate) measured scale (in terms of fraction of advertisement area). (a) $c^0 = 5.1\%$, $c = 3.9\%$. (b) $c^0 = 5.0\%$, $c = 6.0\%$. (c) $c^0 = 3.6\%$, $c = 4.7\%$. (d) $c^0 = 3.8\%$, $c = 5.7\%$.

Figure 4.19: Scale Estimation Results – Example Set 3. In each of the examples, the advertisement is shown on the left and the scaled (according to our prediction) logo is shown on the right. Let c^0 denote the predicted scale and c denote the (approximate) measured scale (in terms of fraction of advertisement area). (a) $c^0 = 3.8\%$, $c = 4.3\%$. (b) One of the cigarette boxes is $c = 4.0\%$ of the image. Our scale estimate $c^0 = 8.0\%$ is too large because the pattern occurs twice in the image. (c) $c^0 = 3.5\%$, $c = 2.4\%$. (d) The box of Tide occupies $c = 1.4\%$. Our scale estimate $c^0 = 3.5\%$ is too large because the pattern occurs twice in the image.