Optimal Matching and Deterministic Sampling

A Thesis

Submitted to the Faculty

of

Drexel University

by

Jeff Abrahamson

in partial fulfillment of the

requirements for the degree

of

PhD in Computer Science

2 November 2007

© Copyright 2 November 2007
Jeff Abrahamson.

This work is licensed under the terms of the Creative Commons Attribution-ShareAlike license. The license is available at http://creativecommons.org/licenses/by-sa/2.0/.

Acknowledgements

First and foremost I would like to thank my advisor Ali Shokoufandeh for his patient help and advice during this ongoing journey. Béla Csaba also deserves special thanks for the many long hours we spent working together developing some of the probabilistic matching theory that has made its way into this work and become so much of its substance. Without those collaborations this dissertation would be thin indeed. I also extend my thanks to my committee members Ali Shokoufandeh, Béla Csaba, Jeremy Johnson, Pawel Hitczenko, Dario Salvucci, and Eric Schmutz for their help and patience. In addition I would like to thank M. Fatih Demirci, Trip Denton, John Novatnack, Dave Richardson, Adam O'Donnell, and Craig Schroeder for their many helpful insights and suggestions. Yelena Kushleyeva and Walt Mankowski provided helpful encouragement and chocolate, without which this might have been a work of suffering. John Parejko and his father Ken Parejko were of great assistance in unearthing a bit of useful 19th century mathematics. Stéphane Birklé merits acknowledgement on multifarious grounds.

More formally, this work was partially supported by NSF grant IIS-0456001 and by ONR grant ONR-N000140410363.

Dedications

Many characters have influenced the course and quality of my life and of this most recent period that produced this work. Two stand out. Misha was present at the beginning. Stéphane has been present for the middle and the end and, more than anyone else, has motivated its completion. Remarkably, and against all expectation, an emotional mapping has developed in my mind between these two. Most wondrous, I have discovered this mapping to be injective. That such an expression of humanity should accompany such a theoretical work deserves mention here, for it validates more than any theoretical result the undertaking and completion of this dissertation.

Table of Contents

List of Figures vii

Abstract viii

1. Introduction 1

    1.1 Related Work 2

    1.2 Contribution 2

    1.3 Outline 4

2. Optimal Matching 5

    2.1 Mathematical Preliminaries 5

    2.2 Combinatorics 8

    2.3 The Earth Mover's Distance 9

    2.4 The Monge-Kantorovich Problem 11

    2.5 Defining the Problem 12

    2.6 Hierarchically Separated Trees 13

    2.7 Approaching a Solution 14

3. Optimal Unweighted Matching 16

    3.1 One Point is Known 16

    3.2 Discrepancy 23

    3.3 The Upper Bound Matching Theorem 32

    3.4 Optimal Unweighted Matching on $[0, 1]$ 37

    3.5 Optimal Matching on $[0, 1]^2$ 40

    3.6 Optimal Matching on $[0, 1]^d$, $d \geq 3$ 42

    3.7 Conclusions 44

4. Optimal Weighted Matching 45

4.1 A Special Case: Matching n points against m points, Uniform Mass 46

4.2 A Naïve Approach to Optimal Weighted Matching in $[0,1]^d$ 47

4.3 Optimal Weighted Matching 49

4.4 Conclusions 52

5. Matching on Non-Rectangular Sets and Future Work 53

5.1 Optimal Matching on Non-Rectangular Sets 53

5.2 Optimal Matching on Sets of Non-Positive Measure 55

5.3 Future Work on Optimal Matching 60

5.4 Conclusions 62

6. Deterministic Sampling 63

6.1 Overview and Related Work 63

6.2 Defining the Problem 65

6.3 Approaching a Solution 65

6.4 Preserving the Centroid 66

6.4.1 Martingales Prove the Obvious 66

6.4.2 Variance and Antipodal Point Algorithms 68

6.4.3 An Application of the Variance 68

6.4.4 Choosing Points One at a Time: Lots of Points is Easy 70

6.4.5 Choosing Points One at a Time: Fewer Points is Harder 72

6.5 Future Work: Preserving the Earth Mover's Distance 79

6.5.1 Integer and Linear Programming 79

6.5.2 Greedy Nearest Neighbor Search 84

6.6 Conclusions 85

7. Conclusions 86

A. GNU Free Documentation License 87

A.1 Applicability and Definitions 88

A.2 VERBATIM COPYING 89

A.3 COPYING IN QUANTITY 90

A.4 Modifications 91

A.5 Combining Documents 93

A.6 Collections of Documents 93

A.7 Aggregation with Independent Works 94

A.8 Translation 94

A.9 Termination 95

A.10 Future Revisions of this License 95

A.11 How to use this License for your Documents 95

Bibliography 97

List of Figures

2.1 A balanced 3-HST with branching number $b = 2$ and $\Delta = 26$. 14

3.1 The events H_0, H_1, and H_2. 18

3.2 A Gaussian and its two tails. 24

3.3 The top two levels of an HST, showing the partition of red and blue points. 32

3.4 The solid lines represent the optimal matching of the expanded set of red and blue points in the right set (lower blue points, on $[1/3, 1]$). The dotted lines represent the optimal matching of the original right set (upper blue points). Remember that the vertical is purely illustrative: the matching weight is the horizontal distance only. 38

3.5 A balanced 2-HST with branching factor 2 whose metric dominates the Euclidean metric on $[0, 1]$. 39

3.6 A balanced 2-HST with branching factor 4 whose metric dominates the Euclidean metric on $[0, 1]^2$. The coordinates of three points are indicated to show how the tree embeds in the unit square without too cluttering the diagram. The tree in the drawing is slightly distorted so as not to overtrace lines. 41

5.1 The recursive decomposition of a triangle into four similar triangles. 54

5.2 Construction of the (ternary) Cantor set: each successive horizontal section is an iteration of the recursive removing of middle thirds. The Cantor set has measure zero (and so contains no segment), is uncountable, and is closed. 55

5.3 The Sierpinski triangle. We find an upper bound using a balanced 2-HST with branching factor 3. 56

5.4 The first four iterations of a Menger sponge. A 20-HST provides a dominating metric (cf. Theorem 5.6). 57

5.5 A Sierpinski carpet [199], constructed by recursively dividing the square into nine equal subsquares and removing the middle subsquare. 57

5.6 The first four steps in Process 5.8: removing upper-left and lower-right squares. 59

5.7 The first four steps in Process 5.10: removing upper-left and lower-right squares alternating with removing upper-right and lower-left squares. 60

5.8 Non-uniform distributions may be approximated by 2-HST's with non-uniform branching factors. On the left, we modify the branching factor to change the distribution, leaving the probability the same at each node (50% chance of distributing to either side). On the right, we modify the probabilities at the interior nodes, leaving the branching factor constant. 61

Abstract

Optimal Matching and Deterministic Sampling

Jeff Abrahamson

Advisor: Ali Shokoufandeh, Ph.D.

Randomness is a fundamental problem in theoretical computer science. This research considers two questions concerning randomness. First, it examines some extremal point matching problems, exploring the dependence of matching weight with partition cardinality in vertex-weighted bipartite graphs. Second, it considers the problem of subset selection, providing several deterministic algorithms for point selection that are as good as or better than random subset selection according to various criteria.

1. Introduction

Mathematics studies mappings between objects. Computer science concerns itself with projections of mathematical objects onto certain sorts of machines in the physical world. Theoretical computer science occupies a nether realm between the two, its practitioners neither the swift beasts of the earth nor the ethereal creatures of the air, but some odd chimera that believes itself blessed with a privileged and unique view. This work is of that ilk.

A fundamental question (and quest) in theoretical computer science concerns randomness. We know randomized algorithms that outperform their deterministic peers [45]. A classic example concerns computing minimum spanning trees. While the textbook algorithms attributed to Prim and to Kruskal [53] run in $O(m \log n)$ time (and Borůvka's more complicated algorithm only slightly better [30, 85, 133]), far more efficient algorithms exist [48, 76, 78, 198, 43], even achieving linear time under rather special circumstances [77]. Karger, however, showed an algorithm with expected linear running time [108]. It is an open question whether a deterministic algorithm can compute the general MST as fast. (Remarkably, sublinear approximations exist as well: cf. Chazelle [46, 47].)

As Saks [153] notes in his 1996 survey, the questions "Does randomness ever provide more than a polynomial advantage in time over determinism?" and "Does randomness provide more than a constant factor advantage in space over determinism?" remain unsolved.

One problem of interest in this domain is extremal random matchings. Presented with two random point sets, how do we expect matching weight to vary with data set size? The seminal work on this subject is Ajtai et al's 1984 paper [4], a very deep work that bears no risk of accusation of being overly detailed in its presentation. The question I address here, in somewhat more general context than Ajtai *et al.* considered, is what the asymptotic behavior of optimal matchings of weighted random point sets is as the size of the sets increases.

A second problem of interest concerns subset selection [122]: given a large data set, we wish to approximate it (according to some criterion that we specify) by a subset of its members. For many criteria, random subselection is a good engineering choice with a very high probability of successful outcome. But it begs the question: must we resort to the brain-dead idiocy of blindly picking data points when we believe ourselves to be so much more clever? We suppose for the sake of argument

that we are very patient as well as very clever and so do not hope to beat the neatly speedy solution of random selection.

1.1 Related Work

Humans have the remarkable ability to see the world around them and to make visual sense of it. This is an enormously difficult computational problem. Indeed, a fundamental unsolved problem of computer vision asks how to match features of objects from the visual world in order to determine whether those objects are similar. Features in computer vision are abstractions of the underlying shapes: silhouettes, edges, corners, skeletons, etc. Thus, matching objects by features corresponds (especially in the context of this dissertation) to point set matching. A related and equally important facet of human vision is the ability to simplify point sets (features) so as to preserve the essence of the underlying object in some meaningful way. This dissertation addresses, in a rather abstract way, each of these problems.

More precisely, this dissertation addresses two related but separate questions. The first concerns the asymptotic weight of optimal matchings. The second concerns algorithms for pattern simplification. Detailed literature reviews of work related to each of those questions appears with the discussion of those two problems, in Sections 2.2, 2.3, and 2.4 for optimal matching and in Section 6.1 for pattern simplification.

1.2 Contribution

Our main contribution in the context of optimal matching will be the development of a new technique using hierarchically separated trees (HST's) to prove upper bounds on optimal matching problems. (We will define HST's in Definition 2.16. For now, it is reasonable to picture a balanced tree each of whose non-leaf nodes has the same number of children.) The general approach will be to approximate matching problems in Euclidean space with matching problems in HST's, and then to bound the tree approximation. This will result in several theorems (3.26, 3.28, 3.31, 4.11, 4.12, 4.14), which we can summarize as follows:

> **Theorem:** Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$ and let M_n be the expected weight of an optimal matching

of B against R. Then

$$M_n = \Theta(\sqrt{n}) \qquad \text{if } d = 1$$

$$M_n = O(\sqrt{n}\log n) \qquad \text{if } d = 2$$

$$M_n = \Theta(n^{(d-1)/d}) \qquad \text{if } d \geq 3.$$

In the process of proving this, we will prove a supporting result that is important in its own right:

Upper Bound Matching Theorem: Let T be a k-HST with m leaves, uniform branching factor b, and diameter Δ, with $b \neq k^2$, and let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points respectively distributed uniformly and at random on the leaves of T. Then the expected weight of an optimal matching of R and B in T is

$$\left(\frac{k\Delta\sqrt{n}}{m^{\log_b k}}\right)\left(\frac{\sqrt{m} - m^{\log_b k}}{\sqrt{b} - k}\right).$$

Under the same assumptions, if $b = k^2$, then the expected weight of an optimal matching of R and B in T is

$$\Delta\sqrt{n}\log_b m.$$

We will use the Upper Bound Matching Theorem to prove upper bounds on several sets for which showing reasonable matching results would previously have been intractable: triangles, tetrahedrons, the Cantor set, and various fractals. In the process of exploring these more exotic sets, we will prove a further upper bound theorem, a significant insight despite its simplicity once stated:

Generalized Upper Bound Theorem: Let $X \subset \mathbb{R}^d$ be a bounded set, and let $Y \subseteq X$. If the Upper Bound Matching Theorem provides an upper bound of $f(n)$ for optimal matching on X, then $f(n)$ is also an upper bound for optimal matching on Y.

As an avenue for future work, we propose an algorithm using these results that holds promise to permit comparisons of shape matching results even when different comparisons were made using differently sampled shapes.

Finally, we will explore one family of deterministic algorithms for subset selection. The result will be a non-obvious but not overly deep result: that it is possible to choose points deterministically from a set, with the goal of choosing a subset whose centroid is close to the centroid of the original set, and

yet to expect to do at least as well as choosing them randomly. We'll present three variations on an algorithm that does this (and prove that it is correct). That algorithm will lead us to a more useful algorithm for subset selection that preserves the Earth Mover's Distance. Proving that algorithm correct remains an open problem for future work, although experimental results suggest that it is correct.

1.3 Outline

Let us end this introduction with a brief roadmap of where we will go. Chapter 2 introduces some notation and definitions, as well as some probability theory and other mathematics necessary for this thesis. This chapter also introduces the problem of optimal matching and discusses the new approach we will follow. Chapter 3 solves the matching problems from Chapter 2 in the context of unweighted point sets. Chapter 4 parallels Chapter 3 in structure, but extending the results to weighted matching. Chapter 5 discusses matching outside the context of rectangular regions and outlines potential applications of this work for future study. Chapter 6 describes some deterministic subset selection problems and their solutions, with notes on unsolved problems.

2. Optimal Matching

2.1 Mathematical Preliminaries

While this dissertation does not depend on any too terribly esoteric mathematics, it does make use of a smattering of facts from probability theory and related disciplines. We review in this section some mostly well-known definitions and theorems for later use. The impatient reader may safely skim this section.

Variance and expectation are related:

Lemma 2.1. *Let X be a random variable with finite expectation and variance. Then* $\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}X)^2$.

Proof.

$$\begin{aligned}
\mathbf{Var}(X) &= \mathbf{E}(X - \mathbf{E}X)^2 \\
&= \mathbf{E}\left(X^2 - 2X\mathbf{E}X + (\mathbf{E}X)^2\right) \\
&= \mathbf{E}\left(X^2\right) - 2\mathbf{E}(X\mathbf{E}X) + \mathbf{E}\left((\mathbf{E}X)^2\right) \\
&= \mathbf{E}\left(X^2\right) - 2(\mathbf{E}X)^2 + (\mathbf{E}X)^2 \\
&= \mathbf{E}\left(X^2\right) - (\mathbf{E}X)^2
\end{aligned}$$

□

Stirling's formula will often be of use:

Theorem 2.2 (Stirling)**.**

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1.$$

This is often written

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

A common proof, such as that of Feller [74, II.9], uses the Riemann integral to approximate $\ln n$,

and so $\ln n!$, and then shows convergence of the sequence of differences between $\ln n!$ and the approximation. Aficionados of proofs of Stirling's formula may also delight in articles by Feller [73] and Robbins [150] in the *American Mathematical Monthly.*

Theorem 2.3 (Markov's Inequality for non-negative random variables)**.** *Let X be a non-negative random variable. Then*

$$\forall c > 0,\ \Pr(X \geq c) \leq \frac{\mathbf{E}X}{c}.$$

Proof. Let

$$I\{X \geq c\} = \begin{cases} 1 & \text{if} X \geq c \\ 0 & \text{otherwise} \end{cases}$$

Since $X \geq 0$, we have that

$$X \geq cI\{X \geq c\}.$$

The result follows by taking expectations. □

In fact, Markov's Inequality is an instance of the following more general statement [87]:

Theorem 2.4. *Let $h : \mathbb{R} \to [0, \infty)$ be a non-negative real function. Then*

$$\forall a > 0,\ \Pr(h(X) \geq a) \leq \frac{\mathbf{E}h(X)}{a}.$$

Proof. Consider the event $A = \{h(X) \geq a\}$. Then $h(X) \geq aI_A$. Taking expectations, we have $\mathbf{E}h(X) \geq \mathbf{E}(aI_A) = a\Pr(A)$. Dividing by a yields the result. □

Markov's Inequality follows as a corollary:

Corollary 2.5 (Markov Inequality)**.** *Let X be a random variable. Then for*

$$\forall c > 0,\ P(|X| \geq c) \leq \frac{\mathbf{E}|X|}{c}.$$

Alternate Proof of Markov's Inequality. Let $h(x) = |X|$ in Theorem 2.4. □

Corollary 2.6 (Boole's Inequality)**.** *Let A_i, $i \in [n]$ be a set of events and $I\{A_i\}$ be their indicator functions. Then*

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i).$$

Proof. Let

$$N = \sum_{i=1}^{n} I\{A_i\}$$

denote the number of events A_i that occur. Clearly $\Pr\left(\cup_{i=1}^{n} A_i\right) = \mathbf{E}N$. The Markov inequality states that

$$\Pr(N \geq 1) \leq \mathbf{E}N.$$

If $N \geq 1$, at least one of the events must occur, so

$$\mathbf{E}N = \sum_{i=1}^{n} \mathbf{E}I\{A_j\} = \sum_{i=1}^{n} \Pr(A_i).$$

The result follows. □

Theorem 2.7 (Chernoff Bound)**.** *Let X have moment generating function $\phi(t) = \mathbf{E}e^{tX}$. Then for any $c > 0$,*

$$\begin{aligned} P(X \geq c) &\leq e^{-tc}\phi(t), \qquad \text{if } t > 0 \\ P(X \leq c) &\leq e^{-tc}\phi(t), \qquad \text{if } t < 0. \end{aligned}$$

Proof. For $t > 0$,

$$\Pr(X \geq c) = \Pr\left(e^{tX} \geq e^{tc}\right) \leq \mathbf{E}\left[e^{tX}\right] e^{-tc}$$

by Markov's inequality in the final step. The case that $t < 0$ follows similarly. □

Theorem 2.8 (Kolmogorov's Inequality, as stated in Grimmett and Stirzaker [87])**.** *Let $X_1, X_2, \ldots$ be independent random variables with finite means and variances, and let $S_n = X_1 + \cdots + X_n$. Then*

$$\forall \epsilon > 0,\ \Pr\left(\max_{k \in [n]} |S_k - \mathbf{E}S_k| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\mathbf{Var}(S_n)$$

Theorem 2.9 (Hölder's Inequality)**.** *Let $p, q \in \mathbb{R}^+$ with*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then

$$\int f(x)g(x)\,dx \leq \left(\int [f(x)]^p\,dx\right)^{1/p} \left(\int [g(x)]^q\,dx\right)^{1/q}$$

Note that the Cauchy-Schwarz inequality is the special case that $p = q = 2$.

Definition 2.10. Given a probability space $(\Omega, \mathcal{F})$ that we also call the *sample space*, a *stochastic process* is a collection of random variables $\mathbf{X} = \{X_t \mid 0 \leq t < \infty\}$ on $(\Omega, \mathcal{F})$ taking values in a measure space $(S, \mathcal{S})$, which we call the *state space*.

For our meager purposes, Ω is either $\mathbb{R}^d$ or a discretization of $\mathbb{R}^d$, $S = \mathbb{R}^d$, and $\mathcal{S} = \mathcal{B}(S)$.

Definition 2.11. Let $\{X_i\}$ be a sequence of random variables. A *martingale with respect to the sequence* $\{X_i\}$ is a stochastic process $\{Z_i\}$ with the following properties:

1. Z_n is a function of $X_0, \ldots, X_n$,
2. $\mathbf{E}Z_n < \infty$, and
3. $\mathbf{E}[Z_{n+1} \mid X_0, \ldots, X_n] = Z_n$

When the difference between successive values of a martingale may be bounded, we can derive a bound on the tail of the martingale:

Theorem 2.12 (Hoeffding-Azuma Inequality)**.** *Let $(Y, \mathcal{F})$ be a martingale and $X_1, X_2, \ldots,$ a sequence of real numbers such that $\Pr(|Y_n - Y_{n-1}| \leq X_n) = 1$ for all n. Then for $x > 0$,*

$$\Pr(|Y_n - Y_0| \geq x) \leq 2 \exp\left\{ \frac{-\frac{1}{2}x^2}{\sum_{i=1}^n X_i^2} \right\}$$

Thus, the Hoeffding-Azuma inequality holds whenever $Z_i - Z_{i-1}$ is constrained to lie in a random interval $I(X_1, \ldots, x_{i-1})$ whose length is bounded by the constant c_i.

2.2 Combinatorics

A number of combinatorial results seem relevant to a discussion of extremal matching problems. If a graph has hk vertices and minimum degree at least $(h-1)k$, then it contains k vertex-disjoint copies of K_h [92]. Viewed differently, this result, asserts that an equitable k-coloring exists; that is, a coloring in which no two color classes differ in cardinality by more than one. The theorem is often stated inversely: that if a graph on n vertices has maximum degree no more than r, then an equitable $(r+1)$-coloring exists.

Given n points in the plane, the number of distinct interpoint distances, call it here $f(n)$, is bounded above and below by the following result of Erdös [67]:

$$\sqrt{n - \frac{3}{4}} - \frac{1}{2} \leq f(n) \leq \frac{cn}{\sqrt{\log n}}$$

Given, moreover, t lines in the plane with $\sqrt{n} \leq t \leq \binom{n}{2}$, there exist constants, independent of n and t, such that the following hold [175, 174]:

- The number of incidences between the points and the lines is less than $c_1 n^{2/3} t^{2/3}$.
- If, moreover, $k \leq \sqrt{n}$, then the number of lines containing at least k points is less than $c_2 n^2/k^3$.
- If the points are not all colinear, then some point must lie on $c_3 n$ of the lines determined by the n points.
- There exist fewer than $\exp(c_4\sqrt{n})$ sequences of the form $2 \leq y_1 \leq y_2 \leq y_3 \leq \cdots \leq y_t$ for which there exist n points and t lines $l_1, l_2, \ldots, l_t$ such that line l_j contains y_j points.

The first two of these properties, at least, extend to the complex plane and are tight in both real and complex planes [182].

Indyk discusses combinatorial bounds on pattern matching [105]. The discrepancy result we present in Lemma 3.11 is reflected obscurely in [160].

2.3 The Earth Mover's Distance

Cohen and Guibas [52, 51] and Rubner *et al.* [151] introduced a variant of the transshipment, or transportation, problem [49] in the context of computer vision and shape recognition [186]. Archer *et al.* [16] applied the EMD to metric labeling. It is equivalent to the Mallows distance on probability distributions [120]. The EMD has also found use in recognizing musical fragments [185, 191].

We define the Earth Mover's Distance as a transportation problem. Let (M, ρ) be a metric space, and consider two weighted point sets, $A = \{a_i\}_{i=1}^n$ and $B = \{b_i\}_{i=1}^m$ with weight functions w_A and w_B. Let $W_A = \sum_{i=1}^n w_A(a_i)$ and $W_B = \sum_{i=1}^m w_B(b_i)$ be the total weight of each point set. Then the *Earth Mover's Distance* is

$$\mathrm{EMD}(A, B) = \min_{F \in \mathcal{F}} \frac{\sum_{i=1}^n \sum_{j=1}^m f_{i,j} \rho(a_i, b_j)}{\min(W_A, W_B)}$$

where $\mathcal{F}$ is the set of feasible flows from A to B and $F \in \mathcal{F}$ is written $F = (f_{i,j})$.

The EMD may be formulated as a linear programming problem, since it is a transportation problem:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i,j}\rho(a_i, b_j) \\ \text{subject to} \quad & \\ f_{i,j} &\geq 0 \\ \sum_{j=1}^{n} f_{i,j} &\leq w_A(a_i) \\ \sum_{i=1}^{m} f_{i,j} &\leq w_B(b_j) \\ \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i,j} &\leq \min(W_A, W_B) \end{aligned}$$

If $W_A = W_B$, the last equation drops out and the inequalities become equality.

In practice, it is generally most efficient to compute the EMD using the simplex method than the LP formulation above, although Cabello *et al.* [37] provide $(1+\epsilon)$- and $(2+\epsilon)$-approximations for computing minimum EMD under translation and rigid motion respectively. Further approximations come from work of Assent *et al.* [17]. Theoretically, however, we find the most useful complexity bound by considering EMD as a network flow problem. In this case an algorithm of Orlin ([138], [9] in [116]) provides an $O((nm)^2 \log(n+m))$ worst case complexity.

If the masses of the two point sets are not equal, one may either create a phantom mass on the lighter side to make up the difference, with zero cost links to the heavier side, or else scale the masses to make the two sides of equal mass. This later technique is sometimes referred to as the *Proportional Transportation Distance*: the minimum cost flow that exhausts capacity of the source. In particular, as above, given a metric space (M, ρ), consider two weighted point sets, $A = \{a_i\}_{i=1}^{n}$ and $B = \{b_i\}_{i=1}^{m}$ with weight functions w_A and w_B. Let $W_A = \sum_{i=1}^{n} w_A(a_i)$ and $W_B = \sum_{i=1}^{m} w_B(b_i)$ be the total weight of each point set. The *Proportional Transportation Distance*

$$\mathrm{PTD}(A, B) = \min_{F \in \mathcal{F}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i,j}\rho(a_i, b_j)}{W_A},$$

where $\mathcal{F}$ is the set of feasible flows from A to B and $F \in \mathcal{F}$, written $F = (f_{i,j})$, has the following

properties:

1. $f_{i,j} \geq 0,\ \forall i \in [n], \forall j \in [m]$
2. $\sum_{j=1}^{m} f_{i,j} = w_A(a_i),\ \forall i \in [n]$
3. $\sum_{i=1}^{n} f_{i,j} = (W_A/W_B)b_j,\ \forall j \in [m]$
4. $\sum_{i=1}^{n} \sum_{j=1}^{m} f_{i,j} = W_A$.

2.4 The Monge-Kantorovich Problem

The Earth Mover's Distance is eminently a discrete distance. The continuous case is known as the Monge-Kantorovich Problem (MKP). (Both the discrete and the continuous cases receive the moniker "mass transportation problem.") Markowich [124] presents an excellent measure-theoretic introduction to the topic, a small portion of which I summarize in simplified form below. The problem apparently has its origins in work by Monge [131] on city planning.

A proper definition of the Monge-Kantorovich Problem requires measure theory. In order to avoid several pages of mathematical background that we would promptly abandon, we define the MKP somewhat informally. Let f and g be bounded non-negative Radon measures on $\mathbb{R}^d$. (It suffices to think of ordinary probability densities.) For convenience, let us assume that f and g have compact support X and Y respectively, with $f(X)$, $g(Y)$ both finite and equal. Let $\mathcal{T} = \{T : \mathbb{R}^d \to \mathbb{R}^d \mid f(T^{-1}(A)) = g(A), \forall A \in \mathcal{B}(\mathbb{R}^d)\}$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel algebra on $\mathbb{R}^d$. That is, $\mathcal{T}$ is the set of all one-to-one measurable mappings that push f into g, mappings that we will call transportation plans.

If we look at some $x \in \mathbb{R}^d$ and some $T \in \mathcal{T}$, then T is a transportation function and $T(x)$ is the location of x after transportation. Denote by $r : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ a non-negative measure function, which in the simplest case is Euclidean distance, and which we call work, since $r(x, T(x))$ is the work of moving point x to $T(x)$. The total work (transportation cost) of moving X to Y given the work is then

$$C_r(f, g; T) = \int_X r(x, T(x)) df.$$

If we take the infimum over all possible transportation plans $T \in \mathcal{T}$, we have

$$O_r(f,g) = \inf_{T \in \mathcal{T}} C_r(f,g;T).$$

Clearly the optimization desired is computationally intractable in all but the most trivial instances. Kantorovich [106, 107], taking advantage of mathematical techniques developed since the time of Monge, offered a modern formulation and a slight relaxation that is analytically (but not necessarily computationally) more tractable.

The Monge-Kantorovich problem (related to the Wasserstein metrics [193]) is the subject of ongoing research in pure mathematics [86, 35, 132, 28, 31, 112, 24, 25, 111, 156, 13, 194]. It has applications to PDE's [68, 55, 130], dynamical systems [140], physical motion in the presence of obstacles [72], image processing [15], image morphing [172, 203, 204, 94] (and so also video), image registration [93], shape recognition [1], and mechanics [193, 195]. Gangbo's survey article [80] and Villani's book [187] offer comprehensive overviews of the MKP, both far beyond our needs here.

2.5 Defining the Problem

Let $\{X_i\}$ and $\{Y_i\}$ be two sets of points chosen uniformly at random in $[0,1]^d$, with $|X_i| = |Y_i| = i$. We wish to find (asymptotic) bounds on the sequence $\{d_i\}$, where $d_i = \mathbf{E}\textsc{emd}(X_i, Y_i)$ is the expected matching weight. In fact, for unweighted point sets the problem was solved by Ajtai et al. [4] in 1984 in dimension $d = 2$. Shor [164] and Leighton and Shor [118] addressed the problem differently shortly after Ajtai *et al.*. Rhee and Talagrand [148] have explored upward matching (in $[0,1]^2$), where points from X must be matched to points of Y that have greater x- and y-coordinates. They explored a similar problem in the cube [149]. Talagrand [179] provides a bound for $[0,1]^{\geq 3}$ with arbitrary normed metric. With Yukich [178] he addresses square exponential transportation cost. Talagrand [176, 180] also explores the problem with majorizing measures. He derives more theoretical results as well [177].

We first provide results and proofs for other dimensions. Ajtai *et al.* claim that dimensions other than 2 are trivial, but proofs for $d \neq 2$ apparently have escaped the literature (and are apparently non-trivial to lesser minds than theirs). We continue by expanding their result to weighted point sets.

It is worth noting, as an aside, that the division technique that Ajtai *et al.* use is echoed in work on bin packing [50, 109]. Ajtai *et al.*'s optimal matching results see application in measure theory [200, 42] as well.

2.6 Hierarchically Separated Trees

In order to address this optimal matching problem and, it will turn out, to extend the domain of solutions, we introduce some new machinery. This will turn out, in Chapter 5, to have rather far-reaching and surprising consequences.

Definition 2.13. A metric (M, d) *dominates* a metric $(M, \tilde{d})$ if $\forall x, y \in M$, $d(x, y) \geq \tilde{d}(x, y)$.

Definition 2.14. A metric space (M, d) α-approximates a metric space $(M, \tilde{d})$ if $\forall x, y \in M$, $\tilde{d}(x, y) \leq d(x, y) \leq \alpha \cdot \tilde{d}(x, y)$.

That is, (M, d) α-approximates $(M, \tilde{d})$ if it dominates it but not by too much.

Definition 2.15. Let (M, d) be a metric space and $S = \{(M, d_i)\}$ be a family of metric spaces over the same set. We say that S *α-probabilistically approximates* (M, d) if there exists a probability distribution over S such that $\forall x, y \in M$, $\mathbf{E} d_i(x, y) \leq \alpha \cdot d(x, y)$.

Definition 2.16. A *k-hierarchically well-separated tree* (k-HST) is a rooted weighted tree with two additional properties: (1) edge weights between nodes and their children are the same for any given parent node, and (2) the ratio of edge weights along a root-leaf path is at least k (so edges get lighter by a factor of at least k as one approaches leaves).

Bartal [21] (extended in [22]) proved the following important result:

Theorem 2.17 (Bartal [21], Theorem 8)**.** *Every weighted connected graph G can be α-probabilistically approximated by the set of k-HST's of diameter* $\mathrm{diam}(T) = O(\mathrm{diam}(G))$*, where* $\alpha = O(k \log n \log[k](\min(n, \Delta)))$*. If G is a tree, then* $\alpha = O(k \log[k] \min(n, \Delta))$*, where n is the number of nodes in the tree and Δ is the weighted diameter of the tree.*

The last part of Theorem 2.17 concerning approximating metrics by families of trees was improved to $O(\log n)$ by Fakcharoenphol [70], and, indeed, it is their tree construction that will motivate our approach. This will allow us to approximate matching problems with trees and then to bound that

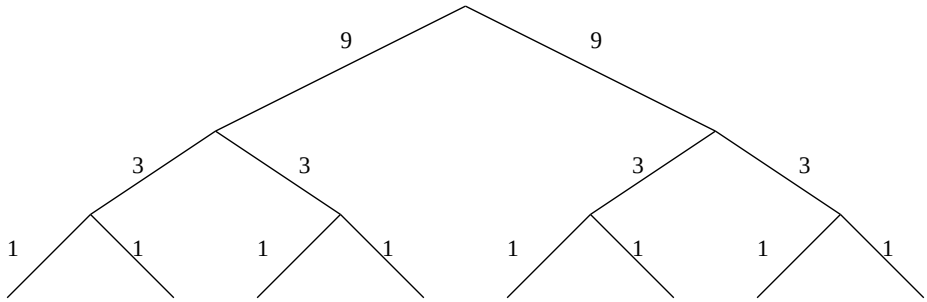

Figure 2.1: A balanced 3-HST with branching number $b = 2$ and $\Delta = 26$.

approximation.[1] Guénoche *et al.* [89] showed how to extend partial metrics to tree metrics. Linial *et al.*[121] discuss embeddings into Euclidean spaces, while Gupta [90] shows an $\tilde{O}(n^{1/(d-1)})$ distortion bound on embedding.

One final definition will be useful concerning HST's:

Definition 2.18. If points x and y match by a path in T whose node closest to the root is p, then we'll say that the match of x and y *transits* p, and we'll refer to their match as a transit of p.

2.7 Approaching a Solution

We will begin by approximating point sets in $[0,1]^d$ by an appropriately dense lattice, allowing us to impose a tree metric with small approximation (discretization) error. For example, if we use m leaves for n points, with $m \gg n$, then in one dimension each point is within $1/(2m)$ of a leaf point. Motivated by the spirit of Theorem 2.17, we will weight the edges of the tree such that the metric it imposes dominates the Euclidean metric on $[0,1]^d$. Note that we are motivated by, but do not directly use, work of Bartal [21, 22] that an appropriate family of trees can approximate any metric with a distortion factor of $O(\log n \log\log n)$, since improved to $O(\log n)$. We then consider the matching problem in the tree metrics and show bounds on the matching weight.

The strategy is to consider how many points must be matched across any given node of the tree. For example (Lemma 3.16), only $\Theta(\sqrt{n})$ points (the standard deviation of the distribution) must be matched across the root in the tree for optimal matching in $[0,1]^d$, the rest of the points matching within subtrees of the root. The matching weight thus computed is an upper bound of the actual

[1] It turns out, however, that we will not need their $O(\log n)$ bound, since we will prove upper bounds with HST's and then be able to prove corresponding and tight lower bounds by other means.

matching weight, since the tree metric dominates the Euclidean metric. It will turn out that lower bounds usually are straight-forward to find using combinatorial means. Indeed, it is the upper bounds that are of most interest, and we compute the lower bounds principally to show that we have done a good job of finding an upper bound.

Our results using trees to provide upper bounds for unweighted matching problems agree with the results of Ajtai et al. [4] except in dimension 2, where our bound is slightly looser.

3. Optimal Unweighted Matching

Having framed the problem in general terms and developed the outline of our approach, we now consider the specific problem of finding the expected weight of matching unweighted points in $[0,1]^d$. We'll begin in Section 3.1 with an overly simple case, the hydrogen atom of the subject, since it is completely solvable. (That is, we can find a differential equation for the distribution.) We'll then go on to explore the general cases of $[0,1]^d$.

3.1 One Point is Known

The simplest case of random matching concerns a single red point against a single blue point on the unit interval $[0,1]$. The field of integral geometry has answered that question in full. The technique involved dates from two papers of William Crofton [56, 58], although it also appears in his 1885 Encyclopedia Britannica article [57].

Klain and Rota [114] and Santaló [155, 12] offer introductory books on integral geometry. Nisan's survey of extracting randomness [134] is steeped in this theory. An important problem in integral geometry has been finding the expected area of a triangle formed by three points dropped at random within another triangle, first solved by Alagar [5] in 1977 using Crofton's technique. Several variations of this problem have since appeared in the literature [183, 184, 40, 117, 159, 158, 64, 98], as well as related problems concerning convex hulls of random points [32, 63, 125, 115] and [34]. Buchta [33] looked at random polygons in quadrilaterals.

A related problem is Sylvester's four point problem [196, 197, 26, 139] (and, apparently, [173], which I have been unable to find). Sylvester's problem asks for the probability that four points dropped at random form a triangle (i.e., that one is inside the convex hull of the other three).

Integral geometry has found application more widely. Carmichael [38] computed the mean length of line segments divided by randomly dropped points. Eisenberg [65] showed an equivalence under certain circumstances between Crofton's differential equation and computing expectations by conditioning. Bagchi [20] applies it to finding a natural real-valued function continuous only at rationals, a result more commonly seen in the context of analysis. The field has also seen success in constructing Buffon curves [60], computing the average distance between points in shapes [61], computing

constant-width curves [39], characterizing Hamiltonian circuits [137], computing Stoke's Theorem in $\mathbb{R}^2$ [66], characterizing pseudorandom number generators [81], finding distributions of random secants in convex bodies [113], studying random walks on a simplex[145, 88], differential forms [154], the central limit theorem [201], and boundaries of convex regions [202]. For completeness, I also cite Finch's chapter on geometric probability constants [75] and several Sloane sequences related to distributions of points in convex figures[166, 167, 168, 169].

Let X be a finite set of points in $D \subset \mathbb{R}^d$, $\mathcal{P}$ be some property, P be the fraction of subsets of X where $\mathcal{P}$ holds, P' be the enlargement of P, and V measure of D. In particular, let $\mathcal{P}$ be some property dependent only on the relationship of the points of X to each other that is invariant under isometric transformation.

Then Crofton's formula is

$$\frac{dP}{dV} = n\left(\frac{P' - P}{V}\right) = n\left(\frac{P(\mathcal{P}_{F'}) - P(\mathcal{P}_F)}{\gamma(D)}\right). \tag{3.1}$$

Theorem 3.1. *If two points are dropped uniformly and at random on* $[0, 1]$*, then the probability distribution function of the distance between them is given by*

$$\Pr(d < t \mid L = \alpha) = \frac{2t\alpha - t^2}{\alpha^2}.$$

Proof. Consider two points x_1 and x_2 dropped at random on the interval $[0, \alpha]$. Without our knowledge and just before the random placement, someone extends the interval to be of length $\alpha' = \alpha + \Delta\alpha$. Consider the following events (Cf. Figure 3.1):

$$\begin{aligned} H_0 &= \text{event that both } x_i < \alpha \\ H_1 &= \text{event that precisely one } x_i < \alpha \\ H_2 &= \text{event that neither } x_i < \alpha \end{aligned}$$

Denote by d the distance (a random variable) between x_1 and x_2 and by L the length of the line

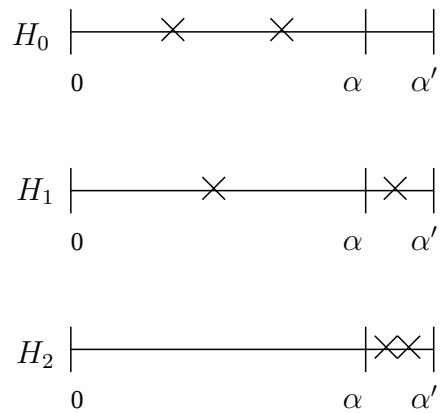

Figure 3.1: The events H_0, H_1, and H_2.

segment (α or α' for us). Then we have

$$\begin{aligned}
\Pr(d < t \mid L = \alpha') &= \sum_{i=0}^{2} \Pr(d < t \mid H_i, L = \alpha') \Pr(H_i \mid L = \alpha') \\
&= \sum_{i=0}^{2} \Pr(d < t \mid H_i, L = \alpha') \cdot \binom{2}{i} \frac{\alpha^{2-i}(\Delta\alpha)^i}{(\alpha')^2}
\end{aligned}$$

To make the notation somewhat simpler, let $f_\alpha(t) = \Pr(d < t \mid L = \alpha)$ and denote the same conditioned on H_i by $f^i_\alpha(t) = \Pr(d < t \mid L = \alpha, H_i)$. We then have

$$\begin{aligned}
\Pr(d < t \mid L = \alpha') &= f_{\alpha'}(t) \\
&= f^0_{\alpha'}(t) \cdot \binom{2}{0} \frac{\alpha^2(\Delta\alpha)^0}{(\alpha')^2} + f^1_{\alpha'}(t) \cdot \binom{2}{1} \frac{\alpha^1(\Delta\alpha)^1}{(\alpha')^2} + f^2_{\alpha'}(t) \cdot \binom{2}{2} \frac{\alpha^0(\Delta\alpha)^2}{(\alpha')^2} \\
&= f^0_{\alpha'}(t) \cdot \frac{\alpha^2}{(\alpha')^2} + f^1_{\alpha'}(t) \cdot 2\frac{\alpha(\Delta\alpha)}{(\alpha')^2} + f^2_{\alpha'}(t) \cdot \frac{(\Delta\alpha)^2}{(\alpha')^2}
\end{aligned}$$

Subtract $f_\alpha(t) = \Pr(d < t \mid L = \alpha)$ from both sides, divide by $\Delta\alpha$, and take the limit as $\Delta\alpha \to 0$ (almost equivalently, as $\alpha' \to \alpha$):

$$f_{\alpha'}(t) - f_\alpha(t) = f^0_{\alpha'}(t) \cdot \frac{\alpha^2}{(\alpha')^2} + f^1_{\alpha'}(t) \cdot 2\frac{\alpha(\Delta\alpha)}{(\alpha')^2} + f^2_{\alpha'}(t) \cdot \frac{(\Delta\alpha)^2}{(\alpha')^2} - f_\alpha(t)$$

$$\frac{f_{\alpha'}(t) - f_\alpha(t)}{\Delta\alpha} = \left[\frac{f^0_{\alpha'}(t)}{\Delta\alpha} \cdot \frac{\alpha^2}{(\alpha')^2}\right] + \left[\frac{f^1_{\alpha'}(t)}{\Delta\alpha} \cdot 2\frac{\alpha(\Delta\alpha)}{(\alpha')^2}\right] + \left[\frac{f^2_{\alpha'}(t)}{\Delta\alpha} \cdot \frac{(\Delta\alpha)^2}{(\alpha')^2}\right] - \left[\frac{f_\alpha(t)}{\Delta\alpha}\right]$$

When we take the limit, we note that $\alpha' \to \alpha$ and $f_\alpha^0 = f_\alpha$ almost surely. The first and fourth terms are thus identical with opposite sign, while the third term goes to zero, leaving only the second term. The conditions H_i, in the limit, tell us how many points are *on* the right endpoint of the interval.

$$
\begin{aligned}
\lim_{\Delta\alpha\to 0}\frac{f_{\alpha'}(t)-f_{\alpha}(t)}{\Delta\alpha} &= \lim_{\Delta\alpha\to 0}\left\{\left[\frac{f^0_{\alpha'}(t)}{\Delta\alpha}\cdot\frac{\alpha^2}{(\alpha')^2}\right]+\left[\frac{f^1_{\alpha'}(t)}{\Delta\alpha}\cdot 2\frac{\alpha(\Delta\alpha)}{(\alpha')^2}\right]\right. \\
&\qquad \left.+\left[\frac{f^2_{\alpha'}(t)}{\Delta\alpha}\cdot\frac{(\Delta\alpha)^2}{(\alpha')^2}\right]-\left[\frac{f_{\alpha}(t)}{\Delta\alpha}\right]\right\} \\
&= \lim_{\Delta\alpha\to 0}\left\{\left[\frac{f^0_{\alpha'}(t)}{\Delta\alpha}\cdot\frac{\alpha^2}{(\alpha')^2}\right]-\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right]+\left[\frac{f^2_{\alpha'}(t)(\Delta\alpha)^1}{(\alpha')^2}\right]+\left[\frac{f_{\alpha}(t)}{\Delta\alpha}\right]\right\} \\
&= \lim_{\Delta\alpha\to 0}\left\{\left[\frac{f^0_{\alpha'}(t)}{\Delta\alpha}\cdot\frac{\alpha^2}{(\alpha')^2}\right]-\left[\frac{f_{\alpha}(t)}{\Delta\alpha}\right]\right\} \\
&\qquad +\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right]+\lim_{\Delta\alpha\to 0}\left[\frac{f^2_{\alpha'}(t)(\Delta\alpha)^1}{(\alpha')^2}\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{\left[\frac{f^0_{\alpha'}(t)}{\Delta\alpha}\cdot\frac{\alpha^2}{(\alpha')^2}\right]-\left[\frac{f_{\alpha}(t)}{\Delta\alpha}\right]\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right]+\left[0\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{\frac{f_{\alpha'}(t)}{\Delta\alpha}\cdot\left(\frac{\alpha^2}{(\alpha')^2}-1\right)\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{\frac{f_{\alpha'}(t)}{\Delta\alpha}\cdot\left(\frac{\alpha^2-(\alpha')^2}{(\alpha')^2}\right)\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{\frac{f_{\alpha'}(t)}{\Delta\alpha}\cdot\left(\frac{\alpha^2-\alpha^2-2(\Delta\alpha)\alpha-(\Delta\alpha)^2}{(\alpha')^2}\right)\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{\frac{f_{\alpha'}(t)}{\Delta\alpha}\cdot\left(\frac{-2(\Delta\alpha)\alpha-(\Delta\alpha)^2}{(\alpha')^2}\right)\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right] \\
&= \lim_{\Delta\alpha\to 0}\left\{f_{\alpha'}(t)\cdot\left(\frac{-2\alpha-(\Delta\alpha)}{(\alpha')^2}\right)\right\}+\lim_{\Delta\alpha\to 0}\left[\frac{f^1_{\alpha'}(t)\,2\alpha}{(\alpha')^2}\right] \\
&= \left[\frac{-2f_{\alpha}(t)}{\alpha}\right]+\left[\frac{2f^1_{\alpha}(t)}{\alpha}\right]
\end{aligned}
$$

But

$$f_\alpha^1(t) = \Pr(d < t \mid \text{one point is on the right boundary of the interval}) = \frac{t}{\alpha}\,.$$

In brief, therefore, we have that

$$\frac{df_\alpha(t)}{d\alpha} = \frac{-2f_\alpha(t)}{\alpha} + \frac{2f_\alpha^1(t)}{\alpha} = \frac{-2f_\alpha(t)}{\alpha} + \frac{2t}{\alpha^2},$$

a differential equation whose solution is

$$\frac{2t\alpha + c(t)}{\alpha^2}.$$

To solve for the constant, we must satisfy the initial conditions $f_\alpha(0) = 0$ and $f_\alpha(\alpha) = 1$:

$$\frac{2 \cdot 0\alpha + c(0)}{\alpha^2} \quad = \quad 0$$

and

$$\frac{2\alpha\alpha + c(1)}{\alpha^2} \quad = \quad 1$$

from which we conclude that $c(0) = 0$ and $c(\alpha) = -\alpha^2$. Setting $c(t) = -t^2$, then, we write

$$f_\alpha(t) = \frac{2t\alpha - t^2}{\alpha^2},$$

or, reverting to the original probabilistic notation,

$$\Pr(d < t \mid L = \alpha) = \frac{2t\alpha - t^2}{\alpha^2}.$$

□

Kendall and Moran [110] derive the same result by applying Crofton's formula directly.

Fix the value of L. We can do some elementary arithmetic to compute the expectation, whose value perfectly follows intuition, and the variance.

Corollary 3.2. $\mathbf{E}d = L/3$.

Proof.

$$\begin{aligned}
\mathbf{E}d &= \int_0^L t\left[\frac{d}{dt}\Pr(d<t)\right]\,dt \\
&= \frac{1}{L^2}\int_0^L t(2L-2t)\,dt \\
&= \frac{1}{L^2}\int_0^L 2tL-2t^2\,dt \\
&= \frac{1}{L^2}\left[t^2L-\frac{2t^3}{3}\right]_0^L \\
&= \frac{1}{L^2}\left(L^3-\frac{2}{3}L^3\right) \\
&= \frac{L}{3}.
\end{aligned}$$

□

Corollary 3.3. $\mathbf{Var}(d) = L^2/18$.

Proof.

$$\begin{aligned}
\mathbf{E}d^2 &= \int_0^L t^2\left[\frac{d}{dt}\Pr(d<t)\right]\,dt \\
&= \frac{1}{L^2}\int_0^L t^2(2L-2t)\,dt \\
&= \frac{1}{L^2}\left[\frac{2Lt^2}{3}-\frac{2t^4}{4}\right]_0^L \\
&= \frac{2L^2}{3}-\frac{2L^2}{4} \\
&= \frac{L^2}{6}
\end{aligned}$$

Noting Lemma 2.1, then, the result follows:

$$\mathbf{Var}(d) = \mathbf{E}d^2 - (\mathbf{E}d)^2 = \frac{L^2}{6} - \frac{L^2}{9} = \frac{L^2}{18}$$

□

Thus, for two points on the unit interval, we can compute the complete probability distribution

function (Theorem 3.1), from which we can easily find the mean (Corollary 3.2) and variance (Corollary 3.3).

3.2 Discrepancy

Littlewood and Offord[1] proved the following result in 1943:

Theorem 3.4. *Let $a_1, \ldots, a_n$ be complex numbers with $|a_i| \geq 1 \, \forall i \in [n]$ and consider the 2^n linear combinations $\sum_{i=1}^n \epsilon_i a_i$ with $\epsilon_i \in \{-1, 1\}$. Then the number of these sums that lie in the interior of any circle of radius 1 is no greater than $c2^n(\log n)/\sqrt{n}$ for some positive constant c.*

A related and useful result is this:

Theorem 3.5 (Sperner's Theorem on Sets [170])**.** *Given a set system $(\mathcal{S}, S)$, where $\mathcal{S} \subseteq 2^S$ is an antichain in the inclusion lattice over 2^S, then*

$$|S| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$$

where $n = |S|$.

The assertion that $\mathcal{S}$ is an antichain in the inclusion lattice merely states that if $A, B \in \mathcal{S}$ then either $A = B$ or else both $A - B$ and $B - A$ are non empty (neither is included in the other).

Theorem 3.6 (Kleitman)**.** *Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be vectors in $\mathbb{R}^d$, each of length at least 1, and let $R_1, \ldots, R_k$ be k open regions of $\mathbb{R}^d$, each of diameter less than 2. Then the number of linear combinations $\sum_{i=1}^n \epsilon_i \mathbf{a}_i$, with $\epsilon_i \in \{-1, 1\}$, that lie in the union $\cup_i R_i$ is at most the sum of the k largest binomial coefficients $\binom{n}{j}$.*

Corollary 3.7. *If $k = 1$ in Theorem 3.6, then the number of linear combinations is bounded by $\binom{n}{\lfloor n/2 \rfloor}$.*

Consider now a discrete random walk on the one-dimensional integer lattice. We may well ask what is the expected displacement from the origin after n steps. We'll answer this with Lemma 3.11. In the mean time, Corollary 3.7 assures us that the probability of returning to $\{-1, 0, 1\}$ is no more than $\binom{n}{\lfloor n/2 \rfloor}/2^n$, which Theorem 3.9 assures us is asymptotically equal to $1/\sqrt{n}$:

[1] The theorems of Littlewood and Offord, Sperner, and Kleitman I cite from Aigner and Ziegler [3], who offer (necessarily) elegant proofs.

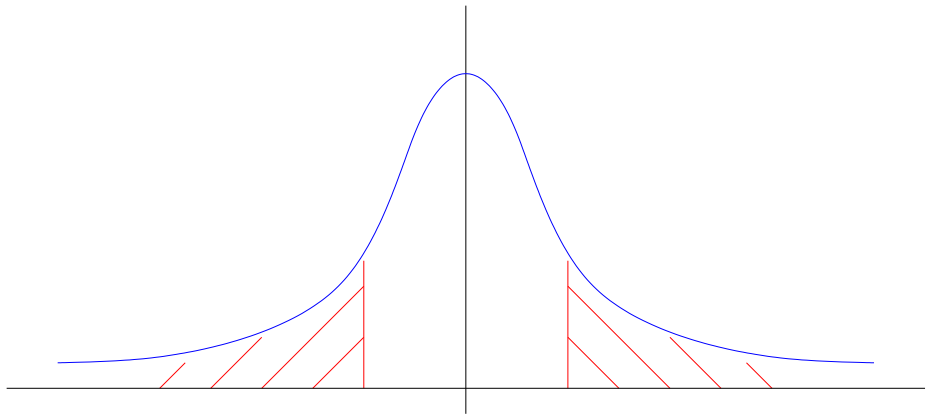

Figure 3.2: A Gaussian and its two tails.

Theorem 3.8. *The probability that a one-dimensional discrete random walk returns to* $\{-1, 0, 1\}$ *after* n *steps is at most* $1/\sqrt{n}$.

Claim 3.9.

$$\binom{n}{\frac{n}{2}} \sim c \cdot \frac{2^n}{\sqrt{n}}.$$

Proof.

$$\begin{aligned}
\binom{n}{\frac{n}{2}} &= \frac{n!}{\left(\frac{n}{2}\right)!^2} \\
&\sim \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\left[\sqrt{2\pi\left(\frac{n}{2}\right)}\left(\frac{n}{2e}\right)^{\frac{n}{2}}\right]^2} \\
&= \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{2\pi n\left(\frac{1}{2}\right)\left(\frac{n}{2e}\right)^n} \\
&= \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\left(\frac{1}{2}\right)2\pi n\left(\frac{n}{e}\right)^n\left(\frac{1}{2}\right)^n} \\
&= c \cdot \frac{2^n}{\sqrt{n}}.
\end{aligned}$$

□

Claim 3.10.

$$\frac{\binom{n}{\frac{n}{2}+c\sqrt{n}}}{\binom{n}{\frac{n}{2}}} \sim c'$$

Proof.

$$\begin{aligned}
\frac{\binom{n}{\frac{n}{2}+c\sqrt{n}}}{\binom{n}{\frac{n}{2}}} &= \frac{\left[\frac{n!}{\left(\frac{n}{2}+c\sqrt{n}\right)!\left(\frac{n}{2}-c\sqrt{n}\right)!}\right]}{\left[\frac{n!}{\left(\frac{n}{2}\right)!^2}\right]} = \frac{\left(\frac{n}{2}\right)!^2}{\left(\frac{n}{2}+c\sqrt{n}\right)!\left(\frac{n}{2}-c\sqrt{n}\right)!} \\
&\sim \frac{\left(\sqrt{2\pi\frac{n}{2}}\left(\frac{n}{2e}\right)^{\frac{n}{2}}\right)^2}{\sqrt{2\pi\left(\frac{n}{2}+c\sqrt{n}\right)}\left(\frac{n+2c\sqrt{n}}{2e}\right)^{\frac{n}{2}+c\sqrt{n}}\sqrt{2\pi\left(\frac{n}{2}-c\sqrt{n}\right)}\left(\frac{n-2c\sqrt{n}}{2e}\right)^{\frac{n}{2}-c\sqrt{n}}} \\
&= \frac{\pi n\left(\frac{n}{2e}\right)^n}{\sqrt{4\pi^2\left(\frac{n^2}{4}-c^2n\right)}\left(\frac{n+2c\sqrt{n}}{2e}\right)^{\frac{n}{2}}\left(\frac{n-2c\sqrt{n}}{2e}\right)^{\frac{n}{2}}} \\
&= \frac{\pi n\left(\frac{n}{2e}\right)^n}{\sqrt{4\pi^2\left(\frac{n^2}{4}-c^2n\right)}\left(\frac{n^2-4c^2n}{4e^2}\right)^{\frac{n}{2}}} \\
&= \frac{\pi n\left(\frac{n}{2e}\right)^n}{\sqrt{4\pi^2\left(\frac{n^2}{4}-c^2n\right)}\left(\sqrt{\frac{n^2-4c^2n}{4e^2}}\right)^n} \\
&\sim \frac{\alpha_1 n}{n}\cdot\frac{\left(\frac{n}{2e}\right)^n}{\left(\sqrt{\frac{n^2-4c^2n}{4e^2}}\right)^n} \sim \alpha_3\cdot\frac{n^n}{\left(\sqrt{n^2-4c^2n}\right)^n} \sim \alpha_3\cdot\alpha_4 = c'
\end{aligned}$$

where the various α_i are constants independent of n. □

Claims 3.9 and 3.10 permit us to prove

Lemma 3.11. *Consider flipping n coins, and let X be a random variable measuring the number of heads. Let $Y = |X - \frac{n}{2}|$. Then $\mathbf{E}Y \sim c\sqrt{n}$.*

An alternate phrasing, motivated by our promise in posing Theorem 3.8, is to say that the expected deviation from the origin of a random walk after n steps is $\Theta(\sqrt{n})$. Since this dissertation mostly considers discrete processes, we offer first a combinatorial proof, although a proof using calculus is more straightforward.

Proof. Asymptotically we have

$$\begin{aligned}
\Pr(Y \le c\sqrt{n}) &= \Pr(\frac{n}{2} - c\sqrt{n} \le X \le \frac{n}{2} + c\sqrt{n}) \\
&= \sum_{i=\frac{n}{2}-c\sqrt{n}}^{\frac{n}{2}+c\sqrt{n}} \binom{n}{i} 2^{-n} \\
&\sim (2c\sqrt{n})\binom{n}{n/2} 2^{-n} \qquad \textit{Claim 3.10} \\
&\sim (2c\sqrt{n})\left(c' 2^n/\sqrt{n}\right) 2^{-n} \qquad \textit{Claim 3.9} \\
&\sim c_2 > 0,
\end{aligned}$$

But then we also have that $\Pr(Y \ge c\sqrt{n}) \sim c_3 = 1 - c_2$. Since c_2 and c_3 are non-zero, $\mathbf{E}Y \sim c\sqrt{n}$. Too see this, note that Markov's Inequality (Theorem 2.3 and Corollary 2.5) gives $\mathbf{E}Y \ge c\sqrt{n}$. But if $\mathbf{E}Y < c\sqrt{n}$, then $\mathbf{E}Y/(c\sqrt{n}) = o(1)$, which would be a contradiction. So $\mathbf{E}Y = \Theta(\sqrt{n})$. □

As hinted, a simpler proof uses calculus. Here it is natural to think of a slightly revised formulation, so we restate the lemma:

Lemma 3.12 (The same statement as Lemma 3.11)**.** *Let $X_1, X_2, X_3, \ldots, X_n$ be random variables taking values -1 and 1 with equal probability and let $Y_k = \sum_{i=1}^{k} X_i$ for $k = 1, \ldots, n$. Then $\mathbf{E}|Y| = \Theta(\sqrt{n})$.*

Proof. Since the binomial process provides a normal distribution in the limit, we may write

$$\begin{aligned}
\mathbf{E}|Y| &= \frac{2}{\sqrt{2\pi}\sigma} \int_0^\infty x e^{-x^2/2\sigma^2}\, dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty e^{-y/2\sigma^2}\, dy \qquad \textit{(substituting } y = x^2,\ dy = 2x\, dx\textit{)} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \left[-2\sigma^2 e^{-y/2\sigma^2}\right]_0^\infty \\
&= \frac{2\sigma^2}{\sqrt{2\pi}\sigma} \\
&= \sqrt{\frac{2}{\pi}} \cdot \sigma = \sqrt{\frac{2}{\pi}} \cdot \sqrt{n} = \Theta(\sqrt{n}).
\end{aligned}$$

□

Lemma 3.11 has the following corollary:

Corollary 3.13. *Consider placing n blue points and n red points on the unit line segment $[0,1]$. Then the expected difference in number of red and blue points in $[0,\frac{1}{3}]$, the first third of the segment, is $\Theta(\sqrt{n})$.*

Proof. Divide the unit interval $[0,1] = [0,\frac{1}{3}) \cup [\frac{1}{3},\frac{2}{3}) \cup [\frac{2}{3},1]$ and refer to the three components as $A = [0,\frac{1}{3})$, $B = [\frac{1}{3},\frac{2}{3})$, and $C = [\frac{2}{3},1]$. Denote by the random variables A_r, B_r, and C_r the number of red points (of n total) that fall in each of A, B, and C respectively, and denote by A_b, B_b, and C_b the number of blue points in each.

Suppose without loss of generality that $A_b > A_r$. By Lemma 3.11, suppose without loss of generality that $A_b > A_r \sim c\sqrt{n}$ with probability $\epsilon > 0$, for some constant c. Suppose moreover that $|B_b - B_r| \sim c\sqrt{n}$.

We will match the red points of A to the blue points of A by matching the left-most red to the left most blue, and so forth. Then we will match the red points of B to the blue points of B in the same fashion. Finally, in a third step, we match the remaining points from A and B as well as the points of C. We'd like to know the value of $\Pr(|A_b - A_r| \geq c\sqrt{n})$.

Consider the process of placing the red and the blue points. Number the points $1,2,3,\ldots,n$ as they are placed (i.e., in time order rather than in spatial order). We can write indicator variables for placement in segment A:

$$A_r^i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ red point is in } A \\ 0 & \text{otherwise} \end{cases}$$

$$A_b^i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ blue point is in } A \\ 0 & \text{otherwise} \end{cases}$$

Then clearly $A_r = \sum A_r^i$ and $A_b = \sum A_b^i$. Consider $A_b^1 - A_r^1$. Since the A_r^i and A_b^i are independent, we have

$$\Pr\left(A_b^1 - A_r^1 = k\right) = \begin{cases} \frac{2}{3} & \text{if } k = 1 \\ \frac{5}{9} = \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 & \text{if } k = 0 \\ \frac{2}{3} & \text{if } k = -1 \end{cases}$$

Let $X_i = A_b^i - A_r^i$ and $X = A_b - A_r = \sum \left(A_b^i - A_r^i\right)$. We will show that

$$\mathbf{E}|X| \geq \left(\frac{4}{9}\right)^{1/2} \sqrt{n}.$$

Let $f = g$ in Hölder's inequality (Theorem 2.9), with $\frac{1}{p} = \frac{2}{3}$ and $\frac{1}{q} = \frac{1}{3}$ and so $|f|^2 = |f|^{2/3}|f|^{4/3}$.

$$\begin{aligned}
\int |f|^2 \, dx &\leq \left(\int f(x)^p \, dx\right)^{1/p} \left(\int f(x)^q \, dx\right)^{1/q} \\
&= \left(\int \left[|f|^{2/3}\right]^{3/2}\right)^{2/3} \left(\int \left[|f|^{4/3}\right]^3\right)^{1/3} \\
&= \left(\int |f|\right)^{2/3} \left(\int |f|^4\right)^{1/3}
\end{aligned}$$

Rearranging terms,

$$\left(\int |f|\right)^{2/3} \geq \frac{\int |f|^2}{\left(\int |f|^4\right)^{1/3}}$$

and so

$$\left(\int |f|\right) \geq \frac{\left(\int |f|^2\right)^{3/2}}{\left(\int |f|^4\right)^{1/3 \cdot 3/2}} = \frac{\left(\int |f|^2\right)^{3/2}}{\left(\int |f|^4\right)^{1/2}}$$

Changing integrals to expectations, we can write

$$\mathbf{E}|X| \geq \frac{\left(\mathbf{E}(X^2)\right)^{3/2}}{\left(\mathbf{E}(X^4)\right)^{1/2}}. \tag{3.2}$$

Returning to the proof at hand, Equation 3.2 yields

$$\mathbf{E}|X| \geq \frac{\left(\mathbf{E}(X^2)\right)^{3/2}}{\left(\mathbf{E}(X^4)\right)^{1/2}} = \frac{\left(\frac{4}{9}\right)^{3/2} n^{3/2}}{\sqrt{\frac{16}{81}n^2 + \frac{4n}{9}}} \sim \left(\frac{4}{9}\right)^{1/2} \sqrt{n}.$$

To see the numerator, note that

$$\begin{aligned}
\mathbf{E}\left(X^2\right) &= \mathbf{E}\left[\left(\sum X_i\right)^2\right] \\
&= \mathbf{E}\left(\sum_i X_i^2 + \sum_{i\neq j} X_i X_j\right) \\
&= \sum_i \mathbf{E}\left(X_i^2\right) + \sum_{i\neq j} (\mathbf{E}X_i)(\mathbf{E}X_j) \\
&= \sum_i \frac{4}{9} + \sum_{i\neq j} 0\cdot 0 \qquad \textit{Since } \mathbf{E}X_i = 0,\ \forall i \in [n] \\
&= \frac{4n}{9}. \qquad \textit{Since summation is over } i \in [n]
\end{aligned}$$

To see the denominator, consider

$$\begin{aligned}
\mathbf{E}\left(X^4\right) &= \mathbf{E}\left[\left(\sum_i X_i\right)^2 \left(\sum_i X_i\right)^2\right] \\
&= \mathbf{E}\left[\left(\sum_i X_i^2 + \sum_{i\neq j} X_i X_j\right)\left(\sum_k X_k^2 + \sum_{k\neq l} X_k X_l\right)\right] \\
&= \mathbf{E}\left[\left(\sum_i X_i^2\right)\left(\sum_k X_k^2\right)\right] \qquad \textit{Other terms are zero, since } (i,j) \textit{ cancels } (j,i) \\
&= \sum_i \mathbf{E}\left(X_i^4\right) + \sum_{i\neq k} \mathbf{E}\left(X_i^2\right)\mathbf{E}\left(X_k^2\right) \\
&= \frac{4n}{9} + \left(\frac{4}{9}\right)^2 n^2 \\
&= \frac{16}{81}n^2 + \frac{4n}{9}
\end{aligned}$$

□

Although $\mathbf{E}X = 0$, the absolute value has positive expectation:

Claim 3.14. $\mathbf{Var}(X) = \frac{4}{9}n$.

Proof.

$$\mathbf{Var}(X) = \mathbf{E}\left(X^2\right) - (\mathbf{E}X)^2 = \mathbf{E}\left(X^2\right) = \frac{4}{9}n,$$

since $\mathbf{E}X = 0$ and, by Corollary 3.13, $\mathbf{E}X^2 = \frac{4}{9}n$. □

Corollary 3.15. $\mathbf{D}X = \frac{2}{3}\sqrt{n}$.

Proof.

$$\mathbf{D}X = \sqrt{\mathbf{Var}(X)}.$$

□

Finally, let us apply some of the above directly to HST's, where it will prove particularly fruitful in the following sections.

Lemma 3.16. *The expected number of points that transit a node v in the HST T is $\Theta(\sqrt{l})$, where l is the number of leaf nodes below v.*

Proof. Let us at first suppose that the tree's branching factor is 2^d. We can think of the division from one into 2^d subnodes as flipping d coins, and so we imagine, in place of the single node v with its 2^d children, a fictional balanced binary tree of height d with 2^d leaves. We want to count the number of points that would transit *any* node of this binary tree, assuming that m points are scattered at the leaves. The sum of those transit numbers is the number of points that would transit the node v with its 2^d children.

Somewhat informally, Claim 3.17 leads us to expect $c\sqrt{m}$ points to transit the top node, $c\sqrt{m/2}$ points to transit the nodes one level below, and so forth, such that the desired sum is

$$c\sqrt{m} + c\sqrt{m/2} + c\sqrt{m/4} + \cdots + c\sqrt{m/2^{d-1}} \approx c'\sqrt{m}. \tag{3.3}$$

The derivation of Equation (3.3) is straightforward, since $\sqrt{m}$ distributes out, leaving a geometric

series:

$$\begin{aligned}
\text{number of points transiting} &= c\sqrt{m} + c\sqrt{m/2} + c\sqrt{m/4} + \cdots + c\sqrt{m/2^{d-1}} \\
&= c\sqrt{m}\left(\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{4}} + \cdots + \frac{1}{\sqrt{2^{d-1}}}\right) \\
&= c\sqrt{m}\left(\sqrt{2}^{-0} + \sqrt{2}^{-1} + \sqrt{2}^{-2} + \cdots + \sqrt{2}^{-(d-1)}\right) \\
&= c\sqrt{m}\left(\frac{\sqrt{2}^{-n} - 1}{\sqrt{2}^{-1} - 1}\right) \\
&= c\sqrt{m}\left(\frac{1 - \sqrt{2}^{n}}{\sqrt{2}^{m-1} - \sqrt{2}^{n}}\right) \\
&= c\sqrt{m}\left(\frac{1 - 2^{m/2}}{2^{m/2}\left(\frac{1-\sqrt{2}}{2}\right)}\right) \\
&= c\sqrt{m}\left(\frac{2^{m/2} - 1}{2^{m/2}}\right)\left(\frac{2}{\sqrt{2} - 1}\right) \\
&\approx c'\sqrt{m}.
\end{aligned}$$

More formally, we can inductively prove a bound on the discrepancy. Let $a_0 = n$, the number of red points at the root (zero) level. Recursively define $a_{i+1} = a_i/2 - c\sqrt{a_i}$. Then $a_i \geq n/2^i - 2c\sqrt{m/2^i}$. To see this, note that it is trivially true for $i = 0$ and suppose that it is true for all positive integers less than or equal to i. Then

$$a_{i+1} = \frac{a_i}{2} - c\sqrt{a_i} \geq \frac{m}{2^{i+1}} - c\sqrt{\frac{m}{2^i}} - c\sqrt{\frac{m}{2^i} - 2c\sqrt{\frac{m}{2^i}}} \geq \frac{m}{2^{i+1}} - 2c\sqrt{\frac{m}{2^{i+1}}}\,.$$

Noting that the discrepancy is $\sqrt{l}$ for every branching factor that is an integral power of 2, and noting moreover that the discrepancy is non-increasing with branching factor, the result follows. □

It is worth noting, above, that the discrepancy at a node about which we speak is just the number of points that transit the node.

In the previous lemma, we used this result:

Claim 3.17. *Let R_L and B_L represent the number of red and blue points respectively at the first node below the root on the left and R_R and B_R the number of points on the right; cf. Figure 3.3.*

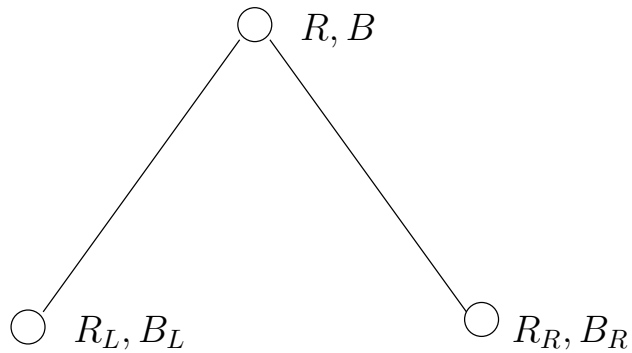

Figure 3.3: The top two levels of an HST, showing the partition of red and blue points.

Clearly $R = R_L + R_R$ *and* $B = B_L + B_R$*. We claim moreover the following:* $\mathbf{E}|R_L - B_L| = \Theta(\sqrt{n})$.

Proof. This is merely a restatement of Theorem 3.11. □

3.3 The Upper Bound Matching Theorem

In the remainder of this chapter, our modus operandi will be to prove lower bounds for various matching problems using combinatorial means, then to prove upper bounds using HST's. To avoid working through the calculus of a new HST for each separate problem we encounter, we would like to develop here the general theory, which we can then hope to apply to each problem as needed to produce an upper bound. As long as we can produce an appropriate HST for the problem, showing that it's metric dominates the natural metric of the problem at hand, we will be able to produce an upper bound without need to resort to extensive computation.

Let us first provide a definition to restrict somewhat the variety of HST that we will consider:

Definition 3.18. We call a k-HST *balanced* if

1. every edge parented to the root has the same weight,
2. every edge not parented to the root has weight equal to k times the weight of the edge immediately closer to the root,
3. every non-leaf node has the same number of children (which we will call the *branching number or branching factor*), and

4. every leaf has the same depth.

Lemma 3.19. *Let T be a balanced k-HST, $k > 1$, with depth i and whose edges below the root have weight α. Then the weighted diameter of T is*

$$d_i = 2\alpha\left(\frac{1-k^{1-i}}{1-k^{-1}}\right) < 2\alpha\frac{k}{k-1}.$$

Proof. Since T is a balanced k-HST, the sequence of edge weights as we walk from the root to a leaf is α, α/k, α/k^2, etc. The diameter is therefore

$$\begin{aligned} d_i &= 2\left(\alpha + \frac{\alpha}{k} + \frac{\alpha}{k^2} + \frac{\alpha}{k^3} + \cdots + \frac{\alpha}{k^{i-1}}\right) \\ &= 2\alpha\left(1 + \frac{1}{k} + \frac{1}{k^2} + \frac{1}{k^3} + \cdots + \frac{\alpha}{k^{i-1}}\right) \\ &= 2\alpha\left(\frac{\left(\frac{1}{k}\right)^i - 1}{\frac{1}{k} - 1}\right) \\ &= 2\alpha\left(\frac{1-k^{1-i}}{1-k^{-1}}\right) \end{aligned}$$

Since we are interested in the upper bound, we may derive a simpler but slightly greater expression by considering a family of such trees $\{T_i\}$ of depth j, each with diameter d_j, and compute instead the weighted diameter of the limit tree, whose weighted diameter, if it exists, is surely greater than that of T, which is one of the T_j. (The sequence of weights is monotonically increasing.) Since each T_j is a balanced k-HST, the sequence of edge weights as we walk from the root to a leaf is α, α/k, α/k^2, etc. The limit diameter is therefore

$$\begin{aligned} \lim_{j\to\infty} d_j &= 2\left(\alpha + \frac{\alpha}{k} + \frac{\alpha}{k^2} + \frac{\alpha}{k^3} + \cdots\right) \\ &= 2\alpha\left(1 + \frac{1}{k} + \frac{1}{k^2} + \frac{1}{k^3} + \cdots\right) \\ &= \frac{2\alpha}{1-\frac{1}{k}} \\ &= \frac{2\alpha k}{k-1}. \end{aligned}$$

□

Note that we needed the assertion that $k > 1$ in order to have the weight sequence converge. Henceforth, we will assume that $k > 1$ for all k-HST's without explicitly requiring it in the exposition.

Corollary 3.20. *Let T be a balanced k-HST of depth i with weighted diameter d_i. Then the edges immediately below the root have weight*

$$\alpha = \left(\frac{d_i}{2}\right)\left(\frac{1-k^{-1}}{1-k^{1-i}}\right).$$

We can now state the main result of this section:

Theorem 3.21 (Upper Bound Matching Theorem)**.** *Let T be a balanced k-HST with m leaves, uniform branching factor b, and diameter Δ, with $b \neq k^2$, and let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points respectively distributed uniformly and at random on the leaves of T. Then the expected weight of an optimal matching of R and B in T is*

$$\left(\frac{k\Delta\sqrt{n}}{m^{\log_b k}}\right)\left(\frac{\sqrt{m} - m^{\log_b k}}{\sqrt{b} - k}\right).$$

Under the same assumptions, if $b = k^2$, then the expected weight of an optimal matching of R and B in T is

$$\Delta\sqrt{n}\log_b m.$$

An important observation is that the diameter of the tree—the actual weights on the edges—plays no role in the asymptotic behavior (in the big-O sense) of the matching weight. Thus, once we prove an asymptotic upper bound for, say, the unit interval (Section 3.4, Lemma 3.25), it applies just as well to any line segment, regardless of length. Said differently, we have a much more subtle and profound result: if two matching scenarios are dominated by trees with the same weight factor (k) and branching number (b), then we conclude the same upper bounds, as a function of n, for both scenarios. We will use this extensively in Chapter 5, where we will formalize this notion in Theorem 5.7.

Proof. We first assume that $b \neq k^2$. The expected weight of an optimal matching of R and B in T is equal to the sum over the nodes of T of the product of the number of points to transit the node and the leaf to node to leaf path that those transiting matches take. Any nodes that match at the leaves contribute no weight to the matching.

We approach the tree level by level. At the root, we expect $\sqrt{n}$ points to transit (Lemmas 3.11 and 3.16). The weight of the leaf to node to leaf path is Δ. The children of the root (there are b of them) expect $\sqrt{n/b}$ transits each, and each transit incurs a cost of Δ/k. The series terminates at the leaves: level $\lambda = \log_b m - 1$. The sum is thus

$$\begin{aligned}
W &= (1)(\Delta)(\sqrt{n}) + (b)\left(\sqrt{\frac{n}{b}}\right)\left(\frac{\Delta}{k}\right) + (b^2)\left(\sqrt{\frac{n}{b^2}}\right)\left(\frac{\Delta}{k^2}\right) + \cdots + (b^\lambda)\left(\sqrt{\frac{n}{b^\lambda}}\right)\left(\frac{\Delta}{k^\lambda}\right) \\
&= \Delta\sqrt{n}\left(1 + \frac{\sqrt{b}}{k} + \left(\frac{\sqrt{b}}{k}\right)^2 + \cdots + \left(\frac{\sqrt{b}}{k}\right)^\lambda\right) \\
&= \Delta\sqrt{n}\left(\frac{\left(\frac{\sqrt{b}}{k}\right)^{\lambda+1} - 1}{\frac{\sqrt{b}}{k} - 1}\right) \qquad \text{(sum of geometric series, since } b \neq k^2\text{)} \\
&= \Delta\sqrt{n}\left(\frac{\frac{\sqrt{m}}{k^{\lambda+1}} - 1}{\frac{\sqrt{b}}{k} - 1}\right) \qquad \text{(since } b^{\lambda+1} = m\text{)} \\
&= \Delta\sqrt{n}\left(\frac{\sqrt{m} - k^{\lambda+1}}{\sqrt{b}k^\lambda - k^{\lambda+1}}\right) \\
&= \Delta\sqrt{n}\left(\frac{\sqrt{m} - m^{\log_b k}}{\frac{1}{k}\sqrt{b}m^{\log_b k} - m^{\log_b k}}\right) \qquad \text{(since } k^{\lambda+1} = k^{\log_b m} = m^{\log_b k}\text{)} \\
&= \left(\frac{\Delta\sqrt{n}}{m^{\log_b k}}\right)\left(\frac{\sqrt{m} - m^{\log_b k}}{\sqrt{b}/k - 1}\right) \\
&= \left(\frac{k\Delta\sqrt{n}}{m^{\log_b k}}\right)\left(\frac{\sqrt{m} - m^{\log_b k}}{\sqrt{b} - k}\right).
\end{aligned}$$

Let us now suppose that $b = k^2$. Then the series is no longer geometric, and we have instead

$$\begin{aligned}
W &= (1)(\Delta)(\sqrt{n}) + (b)\left(\sqrt{\frac{n}{b}}\right)\left(\frac{\Delta}{k}\right) + (b^2)\left(\sqrt{\frac{n}{b^2}}\right)\left(\frac{\Delta}{k^2}\right) + \cdots + (b^\lambda)\left(\sqrt{\frac{n}{b^\lambda}}\right)\left(\frac{\Delta}{k^\lambda}\right) \\
&= \Delta\sqrt{n}\left(1 + \frac{\sqrt{b}}{k} + \left(\frac{\sqrt{b}}{k}\right)^2 + \cdots + \left(\frac{\sqrt{b}}{k}\right)^\lambda\right) \\
&= \Delta\sqrt{n}(\lambda + 1) \\
&= \Delta\sqrt{n}\log_b m.
\end{aligned}$$

□

Note 3.22. It is tempting in the above to simplify the sum in the $b \neq k^2$ case by permitting the tree to be arbitrarily deep, taking comfort in the ever smaller weights of the edges and the knowledge that we only need an upper bound. Of course, the problem will be that Lemma 3.16 no longer applies in the limit. Hope springs eternal, however, and if we pass to the limit and sum the infinite series, we do, indeed, get an incorrect result:

$$\begin{aligned} W &= (1)(\Delta)(\sqrt{n}) + (b)\left(\sqrt{\frac{n}{b}}\right)\left(\frac{\Delta}{b}\right) + (b^2)\left(\sqrt{\frac{n}{b^2}}\right)\left(\frac{\Delta}{k^2}\right) + \cdots \\ &= \Delta\sqrt{n}\left(1 + \frac{\sqrt{b}}{k} + \left(\frac{\sqrt{b}}{k}\right)^2 + \cdots\right) \\ &= \Delta\sqrt{n}\left(\frac{1}{1 - \frac{\sqrt{b}}{k}}\right) \\ &= \Delta\sqrt{n}\left(\frac{k}{k - \sqrt{b}}\right). \end{aligned}$$

While we have not yet seen application of the Upper Bound Matching Theorem, let us present here a short counterexample to show that the above simplified computation does not help us with computing upper bounds. In particular, while this formulation happens to provide a correct result for matching on $[0, 1]$, it does not for matching on $[0, 1]^d$.

Example 3.23. On $[0, 1]$ (see Section 3.4), we have $k = 2$, $b = 2$, and $\Delta = 1$. So the above (incorrect) formulation yields

$$W = \frac{1\sqrt{n}2}{2 - \sqrt{2}} = O(\sqrt{n}).$$

This happens to be the correct result.

On $[0, 1]^3$ (see Section 3.6), however, we have $k = 2$, $b = 8$, $\Delta = \sqrt{3}$. In this case, we get

$$\sqrt{3}\sqrt{n}\left(\frac{2}{2 - 2\sqrt{2}}\right) = O(\sqrt{n}).$$

The correct answer on $[0, 1]^3$, however, is $n^{2/3}$ (cf. Lemma 3.30).

We will use balanced 2-HST's exclusively, since they correspond nicely with Euclidean distance. In particular, they correspond to the divide-and-conquer approach of considering successive refinements of the shape whose Euclidean metric we want to approximate with a dominating metric. Note, however, that our results extend directly to all Minkowski metrics (L_p).

Using balanced 2-HST's permits us to look at the problem from two angles. In one formulation, we can think of the points as given. The tree summarizes where the points are and so shows one way, optimal in the tree metric, of matching those points. Formulating differently, however, we can think of the tree as distributing the points, which are all in some sense provided at the root and distributed from node to node on direct paths to the leaves. That these two models are equivalent and that they do provide for uniform distribution is the basis for urn models in probability theory.

A final comment deserves attention before we move on to application of the Upper Bound Matching Theorem. Using the Upper Bound Matching Theorem to find upper bounds for matching in Euclidean space introduces two sources of distortion (error): the distortion from the dominating metric imposed by the tree and the discretization distortion resulting from approximating the points to a lattice. These distortions are both determined by the value m, the number of leaves in the tree, and so the number of lattice points. Fortunately, as the distortions tend to cancel each other to some extent, but not for arbitrary value of m! It is a straight-forward matter of arithmetic for the settings we will examine to find a value of m that makes everything work out properly. Precisely because it is so straight-forward, it is easy to forget and to see m as a magic number. Nothing could be further from the case! It is merely a number chosen with malice of foresight so that everything works out correctly. We will see that for $d = 1$, $m = n^2$ works quite well, and for $d \geq 3$, $m = n$ suits our needs. No magic: just arithmetic.

3.4 Optimal Unweighted Matching on $[0, 1]$

Ajtai *et al.* [4] note without proof that computing optimal unweighted matchings is easy on $[0, 1]$. Perhaps it is easy, but the proof is surely not obvious. With the theory we developed in Sections 3.2 and 3.3, however, it is not too difficult. We prove separate lower and upper bounds (Lemmas 3.24 and 3.25 respectively) before summarizing them in Theorem 3.26.

Lemma 3.24. *Given n red points and n blue points distributed uniformly on $[0, 1]$, the expected weight of an optimal matching is $\Omega(\sqrt{n})$.*

Proof. Without loss of generality, suppose that the subinterval $[0, 1/3]$ has $n/3 + \Theta(\sqrt{n})$ red points. Let $\alpha = \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$. Then we expect $\sqrt{n}$ points to fall into any interval of length α, so we expect to have about $\sqrt{n}$ points in the interval $I = [1/3, 1/3 + \alpha]$. Let $a = \frac{1}{3} + \alpha$.

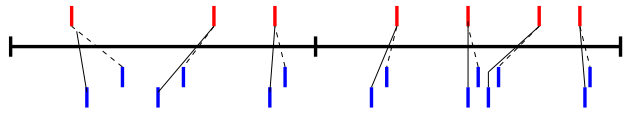

Figure 3.4: The solid lines represent the optimal matching of the expanded set of red and blue points in the right set (lower blue points, on $[1/3, 1]$). The dotted lines represent the optimal matching of the original right set (upper blue points). Remember that the vertical is purely illustrative: the matching weight is the horizontal distance only.

Now distribute n blue points uniformly and at random on $[0, 1]$. Clearly the blue points on $[a, 1]$ are also a uniformly distributed set of random points. The probability that at most $\sqrt{n}$ blue points falls into I is bounded from below by a positive constant. So with some positive probability we will have at least as many blue points on $[a, 1]$ as red points on $[1/3, 1]$. Denote by n_b the number of blue points in $[a, 1]$ and by n_r the number of red points in $[1/3, 1]$. We expect that $n_b > n_r$ (i.e., it is more probably than not), so set $m = n_b - n_r$.

If we now want to match the red points of $[1/3, 1]$ to the blue points of $[a, 1]$, we have m points missing, and so we can't form a perfect matching. So let us pick m additional points (uniformly, at random) from the set of all red and blue points on $[1/3, 1]$. Adding this set temporarily to the red points on $[1/3, 1]$, we can form a perfect matching of n_b red points against n_b blue points.

Consider for the moment only those blue points in $[a, 1]$ and those red points (including the temporary additions) in $[1/3, 1]$. In the following, we'll call these points the *right set* of points. Consider an expansion of only the blue points in the right set so that they cover $[1/3, 1]$:

$$x \mapsto \left(\frac{2\sqrt{n}}{2\sqrt{n} + 3c}\right)(x - 1) + 1.$$

Let us now compare the weight of optimally matching the right set versus optimally matching the expanded right set (Figure 3.4). On average each blue point moves $\Theta(1/(2\sqrt{n}))$ to the left. Let us call the line segment joining a red point to a blue point a *right edge* if the blue point is to the right of (greater than) the red point.

In Figure 3.4, the dotted line segments represent the optimal matching of the right set before expansion. The solid line segments represent the optimal matching after expansion. Since the blue points move only to the right in the expansion, clearly a solid right edge came from a dotted right edge. By symmetry, we expect half of the solid line segments to be right edges, since they are

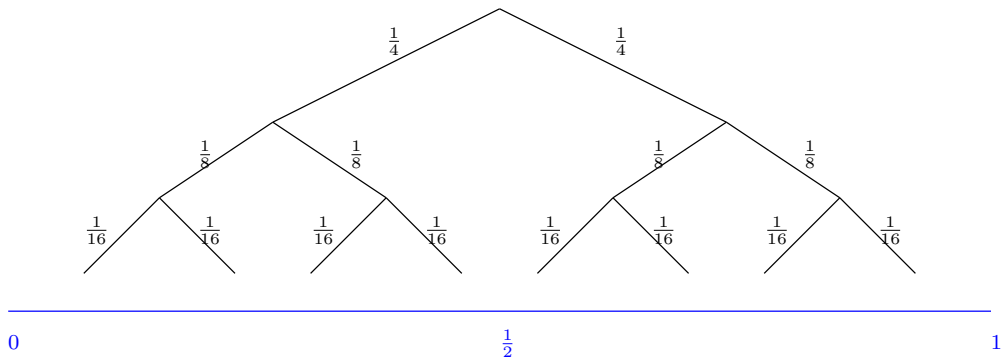

Figure 3.5: A balanced 2-HST with branching factor 2 whose metric dominates the Euclidean metric on $[0, 1]$.

uniformly distributed independent random points on $[1/3, 1]$.

Since the blue points in the right set move on average $1/(2\sqrt{n})$ to the left, the expected weight of the dotted edges in $[1/3, 1]$ is greater than the product of the number of right edges, the number of points, and the expected shift, or

$$\Theta\left(\left(\frac{1}{2}\right)\left(\frac{2}{3}\cdot n\right)\right)\Theta\left(\left(\frac{1}{2\sqrt{n}}\right)\right) = \Theta(c\sqrt{n}).$$

Now we must explain about the temporary points. Let us delete the temporary red points as well as the left-most blue points in the right set (which, post-transformation, are on $[1/3, 1]$). If there is a right edge between an old red point (not one of the temporary reds) and a blue point, then after deletion it will remain a right edge. If we redo the transformation and delete these points, we will find that a right edge will become a long right edge if the two points are far enough to the left. Let us use $[1/3, 5/6]$ for "enough." Long here means $\Omega(1/\sqrt{n})$. Then we have $(5/6-1/3)n/2-m = n/4-m$ long right edges. But the number of long right edges in the optimal matching between the original two random sets is at least $n/4-m$ with positive constant probability, and all long edges have length $O(1/\sqrt{n})$. Thus the optimal matching will have length at least their product, $O(\sqrt{n})$, with positive constant probability. The expected optimal matching weight is thus $O(\sqrt{n})$. □

Lemma 3.25. *Given n red points and n blue points distributed uniformly on $[0, 1]$, the expected weight of an optimal matching is $O(\sqrt{n})$.*

Proof. Consider n red points and n blue points distributed uniformly on $[0, 1]$. Construct a balanced

2-HST($k = 2$) with $m = n^2$ leaves, branching factor $b = 2$ and diameter $\Delta = 1$. The tree dominates the Euclidean metric on the unit interval, since, in the best case the distances are equal. (Some distances are much longer.) See Figure 3.5. The discretization overhead associated with approximating the red and blue points with the leaves is no more than the cost of moving each point to the nearest leaf: $(2n)(1/2n) = 1$.

We are now prepared to apply the Upper Bound Matching Theorem (Theorem 3.21), since $b = 2 \neq 4 = k^2$. Noting that $\log_b k = \log_2 2 = 1$, we have

$$W = \left(\frac{2 \cdot 1 \cdot \sqrt{n}}{m}\right)\left(\frac{\sqrt{m} - m}{\sqrt{2} - 2}\right) = \frac{2\sqrt{n}}{n^2} \cdot \frac{n^2 - n}{2 - \sqrt{2}} \sim \sqrt{n}.$$

Since the tree metric dominates the Euclidean metric (by the triangle inequality), and since the expected matching weight in the balanced 2-HST is asymptotically $\sqrt{n}$, we have proved an upper bound for the Euclidean unit interval. □

In the above proof, we set $m = n^2$ without any explanation. In general, the value of m is set with malice of foresight. That is, we need enough points that the discretization approximation does not add too much weight to the matching, but not so many that the tree adds too much distortion.

With Lemmas 3.24 and 3.25 we have proved the following:

Theorem 3.26. *Given n red points and n blue points distributed uniformly on $[0, 1]$, the expected weight of an optimal matching is $\Theta(\sqrt{n})$.*

3.5 Optimal Matching on $[0, 1]^2$

In the plane our HST technique offers loose results, but it is worth pursuing nonetheless. Ajtai *et al.* [4] showed that $\mathbf{E}M_n = \Omega(\sqrt{n \log n})$, a result on which we can neither improve nor offer new techniques. In Lemma 3.27, we use the Upper Bound Matching Theorem (Theorem 3.21) to show that $\mathbf{E}M_n = O(\sqrt{n} \log n)$. For good measure, we summarize what we know about matching on $[0, 1]^2$ in Theorem 3.28.

Lemma 3.27. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0, 1]^2$ and let M_n be the expected weight of an optimal matching of B against R. Then $\mathbf{E}M_n = O(\sqrt{n} \log n)$.*

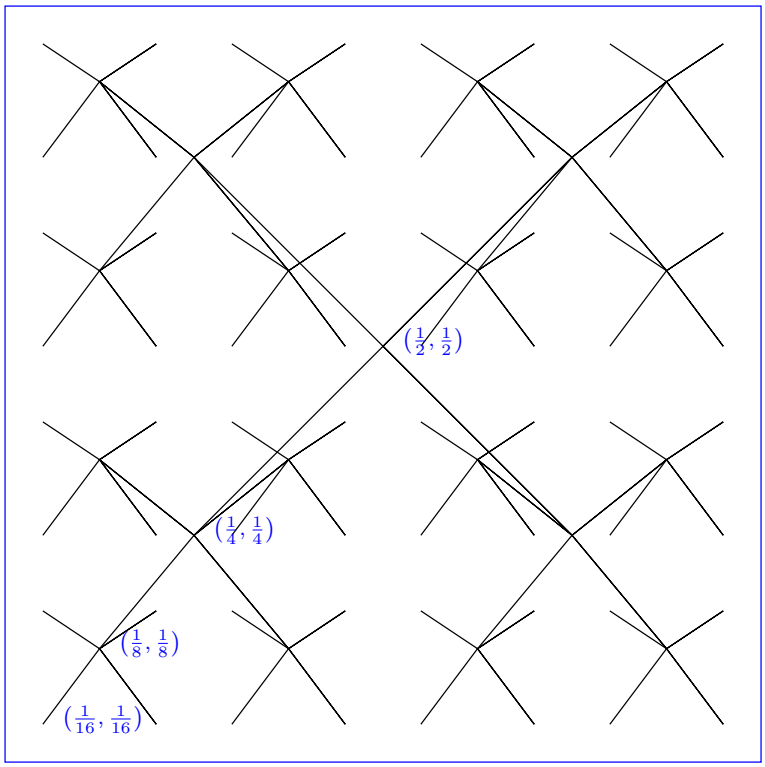

Figure 3.6: A balanced 2-HST with branching factor 4 whose metric dominates the Euclidean metric on $[0, 1]^2$. The coordinates of three points are indicated to show how the tree embeds in the unit square without too cluttering the diagram. The tree in the drawing is slightly distorted so as not to overtrace lines.

Proof. Consider n red points and n blue points distributed uniformly on $[0,1]^2$. Construct a balanced 2-HST ($k = 2$) with $m = n^2$ leaves, branching factor $b = 4$ and diameter $\Delta = \sqrt{2}$. The tree dominates the Euclidean metric on the unit interval, since, in the best case the distances are equal. (Some distances are much longer.) See Figure 3.6. The discretization overhead associated with approximating the red and blue points with the leaves is no more than the cost of moving each point to the nearest leaf: $(2n)(1/(\sqrt{2}n)) = \sqrt{2}$.

We are now prepared to apply the Upper Bound Matching Theorem (Theorem 3.21), since $b = 4 = k^2$. We have

$$W = \Delta\sqrt{n}\log_b m \sim \sqrt{n}\log n.$$

Since the tree metric dominates the Euclidean metric, and since the expected matching weight in the tree is $\sqrt{n}\log n$, we have proved an upper bound for $[0,1]^2$. □

In order to achieve the upper bound of Ajtai *et al.*, we could set $m = \exp(\frac{1}{2}\sqrt{\log n})$. Unfortunately, the discretization error then becomes too large, overwhelming the result from matching in the tree.

Using Ajtai *et al.*'s lower bound and our upper bound from Lemma 3.27, we immediately have

Theorem 3.28. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^2$ and let M_n be the expected weight of an optimal matching of B against R. Then $\mathbf{E}M_n \in \Omega(\sqrt{n\log n}) \cap O(\sqrt{n}\log n)$.*

As noted, Ajtai *et al.*'s result is in fact $\Theta(\sqrt{n\log n})$.

3.6 Optimal Matching on $[0,1]^d$, $d \geq 3$

We now jump to real dimension 3 and above, showing (Lemma 3.29) that $\mathbf{E}M_n = \Omega(n^{(d-1)/d})$, then (Lemma 3.30) that $\mathbf{E}M_n = O(n^{(d-1)/d})$, which we summarize with Theorem 3.31.

Lemma 3.29. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$, $d \geq 3$, and let M_n be the optimal matching of B against R. Then $\mathbf{E}M_n = \Omega(n^{(d-1)/d})$.*

Proof. For each red point r_i, consider a ball s_i of volume $1/n$ around it. The probability that blue point b_j falls in the ball s_i for some $j \in [n]$ is $1/n$, and so the probability that b_j does not fall in s_i

is $1 - 1/n$. The probability that no blue point at all falls in s_i is thus $(1 - 1/n)^n \sim 1/e$.

Let R_i be the random variable taking value 1 if some blue point is within s_i and 0 if not. Then

$$\mathbf{E} \sum_{i \in [n]} R_i = \sum_{i \in [n]} \mathbf{E} R_i = n \mathbf{E} R_1 \sim \frac{n}{e}$$

is the expected number of red points that match to a blue point with distance no more than the radius ρ of a $1/n$-volume d-ball. Since $V_d(\rho) = \Theta(\rho^d)$,

$$\frac{1}{n} = \Theta\left(\rho^d\right) \qquad \Rightarrow \qquad \rho = \Theta\left(n^{-1/d}\right).$$

The expected weight of a matching in $\mathbb{R}^d$ can be no smaller, then, than

$$\left(\frac{n}{e}\right)\left(c' n^{-1/d}\right) = c n^{(d-1)/d}.$$

We thus have that $\mathbf{E} M_n = \Omega(n^{(d-1)/d})$. □

Lemma 3.30. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$, $d \geq 3$, and let M_n be the optimal matching of B against R. Then $\mathbf{E} M_n = O(n^{(d-1)/d})$.*

One approach to proving the upper bound is by means of a combinatorial sieve. One divides the unit d-cube into n identical sub-cubes and places a marker at the center of each. One then considers the expected number of markers that match to a red point within their own sub-cubes. Then one considers the expected number of marker points that don't match a red point in their own sub-cubes, but do match a red point in their own 2-cube (a sub-cube of side length twice the original sub-cubes, formed by joining together 2^d sub-cubes). One continues in this fashion with 4-cubes, 8-cubes, and so on. The computation becomes terribly messy, and the brave soul who arrives at the end of such a proof finds the victory Pyrrhic for its obfuscating arithmetic and lack of elegance.

An alternate and clearer approach uses HST's (and, of course, the Upper Bound Matching Theorem (Theorem 3.21)):

Proof of Lemma 3.30. Consider n red points and n blue points distributed uniformly on $[0,1]^d$. Construct a balanced 2-HST ($k = 2$) with $m = n$ leaves, branching factor $b = 2^d$, and diameter

$\Delta = \sqrt{d}$. The tree dominates the Euclidean metric on the unit interval, since, in the best case the distances are equal. (Some distances are much longer.) The discretization overhead associated with approximating the red and blue points with the leaves is no more than the cost of moving each point to the nearest leaf: $(2n)(1/((\sqrt{d}/2)n^{-1/d})) = \sqrt{d}n^{(d-1)/d}$.

We are now prepared to apply the Upper Bound Matching Theorem (Theorem 3.21), since $b = 2^d \neq 4 = k^2$. Noting that $\log_b k = \log_{2^d} 2 = \frac{1}{d}$, we have

$$W = \left(\frac{2\sqrt{d}\sqrt{n}}{n^{1/d}}\right)\left(\frac{\sqrt{n} - n^{1/d}}{2^{d/2} - 2}\right) = \Theta(n^{1/2}n^{-1/d}n^{1/2}) = \Theta(n^{(d-1)/d}).$$

Since the tree metric dominates the Euclidean metric, and since the expected matching weight in the tree is $n^{(d-1)/d}$, we have proved an upper bound for $[0,1]^d$. □

From Lemmas 3.30 and 3.30, we immediately have

Theorem 3.31. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$, $d \geq 3$, and let M_n be the optimal matching of B against R. Then $\mathbf{E}M_n = \Theta(n^{(d-1)/d})$.*

3.7 Conclusions

In this section we proved a general result, the Upper Bound Matching Theorem (Theorem 3.21), that permits us easily to provide upper bounds on unweighted matching weights using HST's. We used this to bound the cost of unweighted matching in the unit cube for all finite dimensions, bounds that we know to be tight in all cases except in $[0,1]^2$. Using combinatorial arguments, we showed lower bounds for $[0,1]$ and for $[0,1]^d$ for $d \geq 3$.

4. Optimal Weighted Matching

Thus far we have developed new machinery that permits us to approach optimal unweighted matching problems more easily than has previously been possible. We have, moreover, used that machinery to duplicate the results heretofore available in the literature (with a slightly loose result in the single case of dimension 2). In this chapter, we use our new machinery to extend our optimal matching results to weighted point sets. We will find that giving weights to the points does not affect the expected matching weight by more than a constant factor. This and the next chapter, where we extend our matching results to non-rectangular sets, are the most interesting and most important applications of the newly developed machinery, and are important contributions of this dissertation, since they have apparently not been treated at all in the literature.

First, though, let us consider what we mean by weighted matching. We begin by distinguishing mass functions from weight functions, a semantic distinction that is useful here but not in common usage.

Definition 4.1. Given a graph $G = (V, E)$, a *vertex mass function* is a function $m_v : V \to \mathbb{R}^+$. We say, moreover, that $m(G) = m(V) = \sum_{u \in V} m_v(u)$.

Definition 4.2. Given a graph $G = (V, E)$ with vertex mass function m_v, an *edge mass function* is a function $m_e : E \to \mathbb{R}^+$ that obeys the following condition:

$$\forall v \in V,\ m_v(v) = \sum_{e \in E(v)} m_e(e), \tag{4.1}$$

where $E(v)$ is the subset of edges with one endpoint at v.

Definition 4.3. Given a graph $G = (V, E)$ with vertex and edge mass functions, an *edge weight function* is a function $w_e : E \to \mathbb{R}^+$. We say, moreover, that $w(G) = \sum_{e \in E} w_e(e) m_e(e)$.

So mass concerns vertices and weight concerns edges. In what follows we will always assume that our graphs are embedded in a metric space, such that w_e corresponds to distance in that space.

Definition 4.4. Given point sets R and B with vertex mass functions m_v such that $m(R) = m(B)$, a bipartite graph $G = (R, B, E)$ whose edge set E satisfies Equation 4.1 is called a *weighted matching* of R and B. The weight of the matching is $w(G)$. If $m(R) \neq m(B)$, we extend the definition by

adding a point of mass $|m(R) - m(B)|$ to the lighter side and at distance zero from all other points. The matching is the induced subgraph that results from removing the added point at distance zero.

Definition 4.5. Given point sets R and B with vertex mass function m_v such that $m(R) = m(B)$, an *optimal weighted matching* is a minimum weight matching over all weighted graphs $G = (R, B, E)$:

$$\min_{E \in 2^{R \times B}} w((R, B, E)).$$

Armed with these definitions, we now address the following problem:

Problem 4.6. Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of points distributed uniformly and at random on the unit d-cube $I = [0, 1]^d$. Let $m_R : R \to [0, 1]$ and $m_B : B \to [0, 1]$ be vertex mass functions, assigned uniformly and at random. What is the expected weight of the optimal weighted matching?

In what follows, it will be convenient to have at our disposal the following constant, the expected minimum of two numbers chosen uniformly at random from $[0, 1]$:

Definition 4.7. Let x and y be two points chosen uniformly at random from the unit interval $[0, 1]$. Define $\eta = \mathbf{E} \min d(x, y)$.

This quantity could surely be computed, but for our purposes it suffices that it exists and is a constant (or, more pedantically, a function of the number 2, since we could define the family η_i for collections of i points: we will not follow this distraction further). In fact, Corollary 3.2 and an appeal to symmetry immediately gives $\eta = \frac{1}{3}$.

4.1 A Special Case: Matching n points against m points, Uniform Mass

While the general problem of optimal weighted matching presents significant challenges, we may nonetheless discuss at least one special case worthy of note in simpler terms.

In the unit cube $[0, 1]^d$, consider the set $R = \{r_1, \ldots, r_n\}$ with points of unit weight and the set $B = \{b_1, \ldots, b_m\}$ $(m > n)$ whose points each have weight $\frac{n}{m}$, where $m = kn$ for some integer $k > 1$. We can consider sets $B_1, \ldots, B_k$ each randomly sampled from B, with $\cup B_i = B$ and $B_i \cap B_j = \emptyset$ for $i \neq j$. Then the minimum transportation distance between R and B_i we know from Chapter 3, and so the minimum transportation distance between R and $B = \cup B_i$ is no more than k times that

weight. We have proved

Theorem 4.8. *Given the set $R = \{r_1, \ldots, r_n\}$ with points of unit weight and the weighted set $B = \{b_1, \ldots, b_m\}$ whose points each have weight $\frac{n}{m}$, where $m = kn$ for some integer $k > 1$, with R, $B \subset [0,1]^d$, the optimal weighted matching between R and B is*

$$\begin{array}{ll} O(k\sqrt{n}) & \textit{if } d = 1 \textit{ (cf. Lemma 3.25)} \\ O(kn \log n) & \textit{if } d = 2 \textit{ (cf. Lemma 3.27)} \\ O(kn^{(d-1)/d}) & \textit{if } d \geq 3 \textit{ (cf. Lemma 3.30).} \end{array}$$

4.2 A Naïve Approach to Optimal Weighted Matching in $[0,1]^d$

Section 4.1 suggests an obvious approach to considering weighted matching. In fact, the motivation leads to false first steps, but is instructive nonetheless. Indeed, the following discussion suggests that Theorem 4.8 is probably a very rough approximation at best.

As a first step to addressing the general weighted matching problem, let us discretize the vertex mass functions. Let $M_k = \{\frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1\}$. Then we have $m_R : R \to M_k$ and $m_B : B \to M_k$. The idea will be to divide R and B into horizontal (mass) slices $R = R^{(1)} \cup R^{(2)} \cup \cdots \cup R^{(k)}$ and $B = B^{(1)} \cup B^{(2)} \cup \cdots \cup B^{(k)}$ and then to consider unweighted matching on each slice. To simplify theoretical concerns, however, we construct the sets $R^{(i)}$ and $B^{(i)}$ by the following process.

Let $R_1, R_2, \ldots, R_k$ initially be empty sets. Choose $R = \{r_i\}_{i=1}^n \subset I$ uniformly and at random. For each point r_i, choose $m \in M_k$ uniformly and at random. Assign r_i to R_m. Clearly each R_m is a uniformly distributed random point set on I. Since the R_m are mutually independent, we can set

$$\begin{aligned} R^{(1)} &= R_1 \\ R^{(2)} &= R_1 \cup R_2 \\ R^{(3)} &= R_1 \cup R_2 \cup R_3 \\ \vdots \quad & \quad \vdots \\ R^{(k)} &= \bigcup_{m=1}^{k} R_m \end{aligned}$$

and the $R^{(i)}$'s are uniformly randomly distributed, since they are unions of independent uniformly

distributed random sets. We thus have a vertex mass function on R that is "well-behaved" from the standpoint of analyzing different mass classes.

If $f(n)$ is the expected weight of an optimal unweighted matching on two sets of cardinality n, then the weighted matching (up to discrepancy in the size of the weight partitions) is at least upper bounded by

$$f\left(\frac{n}{k}\right) + f\left(\frac{2n}{k}\right) + f\left(\frac{3n}{k}\right) + \cdots + f\left(n\right). \tag{4.2}$$

It suffices to compute this upper bound for the weighted matching on the unit interval to see that this approach will be inadequate. On the unit interval, an unweighted matching has weight $c\sqrt{n}$. The weighted matching, then, is upper bounded by

$$W = \sum_{i \in [k]} c\sqrt{i \cdot \frac{n}{k}}$$

We bound W by the following approximation:

$$\begin{aligned}
\Theta(\sqrt{n}k) = c\sqrt{\frac{n}{k}}\left(\frac{2}{3}\right)k^{3/2} \quad &= \\
c\sqrt{\frac{n}{k}}\left(\frac{2}{3}\right)k^{3/2} \quad &= \\
c\sqrt{\frac{n}{k}}\left(\frac{2}{3}\right)x^{3/2}\Big|_0^k \quad &= \\
c\sqrt{\frac{n}{k}}\int_0^k \sqrt{x}\,dx \quad &= \\
\int_0^k c\sqrt{\frac{xn}{k}}\,dx \quad < W < \quad &\int_1^{k+1} c\sqrt{\frac{xn}{k}}\,dx \\
= \quad &c\sqrt{\frac{n}{k}}\left(\frac{2}{3}\right)\left[(k+1)^{3/2} - 1\right] = \Theta(\sqrt{n}k)
\end{aligned}$$

The problem with this scheme becomes more apparent when we note that k copies of the same set of n/k points has optimal matching score $k\sqrt{n/k} = \sqrt{nk}$. If the k copies were independently distributed, the optimal matching would have weight $\sqrt{n}$.

4.3 Optimal Weighted Matching

We now prove that weighted matching is no heavier than unweighted matching.

Theorem 4.9. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$ with masses assigned uniformly and at random from $[0,1]$. Let $\mathbf{E}M_n$ be the expected weight of the optimal unweighted matching of B against R and let $\mathbf{E}W_n$ be the expected weight of the optimal weighted matching. Then $\mathbf{E}W_n \leq \mathbf{E}M_n$.*

Proof. Find the optimal unweighted matching of R and B. For any point pair, the expected mass actually available to transfer is η, and so the mass transferred by this matching is $\eta\mathbf{E}M_n$. Consider now a second matching of R and B, but with masses reduced by the mass moved in the first matching. The expected unweighted matching weight is of course still $\mathbf{E}M_n$, and so the expected weighted matching weight is $\eta(1-\eta)\mathbf{E}M_n$. We continue this process, and in the limit get a weighted matching with weight

$$\sum_{i=0}^{\infty}(1-\eta)^i\eta\mathbf{E}M_n = \left(\frac{1}{\eta}\right)\eta\mathbf{E}M_n = \mathbf{E}M_n.$$

Since this process is no more efficient than the optimal matching, we have that $\mathbf{E}W_n \leq \mathbf{E}M_n$. □

It remains to prove lower bounds.

Lemma 4.10. *Given n red points and n blue points distributed uniformly on $[0,1]$ with weights distributed uniformly in $[0,1]$. Then the expected weight of an optimal matching is $\Omega(\sqrt{n})$.*

The proof closely parallels the proof of Lemma 3.24.

Proof. The expected total mass of all n red points is $n/2$. Without loss of generality, suppose that the subinterval $[0,1/3]$ has $n/6 + \Theta(\sqrt{n})$ red mass. Let $\alpha = \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$. Then we expect $\Theta(\sqrt{n})$ mass to fall into any interval of length α, so we expect to have $\Theta(\sqrt{n})$ mass in the interval $I = [1/3, 1/3+\alpha]$. Let $a = \frac{1}{3} + \alpha$.

Now distribute n blue points uniformly and at random on $[0,1]$. Clearly the blue points on $[a,1]$ are also a uniformly distributed set of random points. The probability that at most $\Theta(\sqrt{n})$ blue mass falls into I is bounded from below by a positive constant. So with some positive probability we will have at least as much blue mass on $[a,1]$ as red mass on $[1/3,1]$. Denote by n_b the mass of blue in

$[a, 1]$ and by n_r the mass of red in $[1/3, 1]$. We expect that $n_b > n_r$ (i.e., it is more probably than not), so set $m = n_b - n_r$.

If we now want to match the red mass of $[1/3, 1]$ to the blue mass of $[a, 1]$, we have m mass missing, and so we can't form a perfect matching. So let us pick m additional mass (uniformly, at random) from the set of all red and blue mass on $[1/3, 1]$. Adding this set temporarily to the red mass on $[1/3, 1]$, we can form a perfect matching of n_b red mass against n_b blue mass.

Consider for the moment only the blue mass in $[a, 1]$ and the red mass (including the temporary additions) in $[1/3, 1]$. In the following, we'll call this mass the *right set* of mass. Consider an expansion of only the blue mass in the right set so that they cover $[1/3, 1]$:

$$x \mapsto \left(\frac{2\sqrt{n}}{2\sqrt{n} + 3c}\right)(x - 1) + 1.$$

Let us now compare the weight of optimally matching the right set versus optimally matching the expanded right set. On average each blue mass moves $\Theta(1/(2\sqrt{n}))$ to the left. Let us call the line segment joining a red mass to a blue mass a *right edge* if the blue mass is to the right of (greater than) the red mass. Since the blue mass move only to the right in the expansion, clearly a solid right edge came from a dotted right edge. By symmetry, we expect half of the solid line segments to be right edges, since they are uniformly distributed independent random mass on $[1/3, 1]$.

Since the blue mass in the right set moves on average $1/(2\sqrt{n})$ to the left, the expected weight of the dotted edges in $[1/3, 1]$ is greater than the product of the number of right edges, the number of mass, and the expected shift, or

$$\Theta\left(\left(\frac{1}{2}\right)\left(\frac{2}{3} \cdot n\right)\right)\Theta\left(\left(\frac{1}{2\sqrt{n}}\right)\right) = \Theta(c\sqrt{n}).$$

Now we must explain about the temporary mass. Let us delete the temporary red mass as well as the left-most blue mass in the right set (which, post-transformation, are on $[1/3, 1]$). If there is a right edge between an old red point (not one of the temporary reds) and a blue point, then after deletion it will remain a right edge. If we redo the transformation and delete this mass, we will find that a right edge will become a long right edge if the two points are far enough to the left. Let us use $[1/3, 5/6]$ for "enough." Long here means $\Omega(1/\sqrt{n})$. Then we have $(5/6 - 1/3)n/2 - m = n/4 - m$ long right edges. But the number of long right edges in the optimal matching between the original

two random sets is at least $n/4-m$ with positive constant probability, and all long edges have length $O(1/\sqrt{n})$. Thus the optimal matching will have length at least their product, $O(\sqrt{n})$, with positive constant probability. The expected optimal matching weight is thus $O(\sqrt{n})$. □

With Lemmas 4.10 and Theorem 4.9 we have proved the following:

Theorem 4.11. *Given n red points and n blue points distributed uniformly on $[0,1]$, the expected weight of an optimal matching is $\Theta(\sqrt{n})$.*

In the two dimensional case, the best we can do is less than we'd like. Lemma 4.13 provides a lower bound of $\Omega(\sqrt{n})$, while Theorem 4.9 provides an upper bound of $O(\sqrt{n\log n})$ (using Ajtai et al.'s result for unweighted matching). We can therefore state only the following:

Theorem 4.12. *Let $B=\{b_i\}_{i=1}^n$ and $R=\{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^2$ and let W_n be the expected weight of an optimal weighted matching of B against R. Then $\mathbf{E}W_n \in \Omega(\sqrt{n}) \cap O(\sqrt{n\log n})$.*

Lemma 4.13. *Let $B=\{b_i\}_{i=1}^n$ and $R=\{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$, $d\geq 3$, with weights distributed uniformly in $[0,1]$, and let W_n be the optimal matching of B against R. Then $\mathbf{E}W_n = \Omega(n^{(d-1)/d})$.*

Proof. For each red point r_i, consider a ball s_i of volume $1/n$ around it. The probability that blue point b_j falls in the ball s_i for some $j\in[n]$ is $1/n$, and so the probability that b_j does not fall in s_i is $1-1/n$. The probability that no blue point at all falls in s_i is thus $(1-1/n)^n \sim 1/e$.

Let R_i be the random variable taking value 1 if some blue point is within s_i and 0 if not. Then

$$\mathbf{E}\sum_{i\in[n]} R_i = \sum_{i\in[n]} \mathbf{E}R_i = n\mathbf{E}R_1 \sim \frac{n}{e}$$

is the expected number of red points that match to a blue point with distance no more than the radius ρ of a $1/n$-volume d-ball. Since $V_d(\rho)=\Theta(\rho^d)$,

$$\frac{1}{n} = \Theta\left(\rho^d\right) \qquad \Rightarrow \qquad \rho = \Theta\left(n^{-1/d}\right).$$

Since the expected mass exchanged (when any is) is η, the expected weight of a matching in $\mathbb{R}^d$ can

be no smaller, then, than

$$\eta \frac{1}{2} \left(\frac{n}{e}\right) \left(c' n^{-1/d}\right) = cn^{(d-1)/d}.$$

We thus have that $\mathbf{E}W_n = \Omega(n^{(d-1)/d})$. □

As foreshadowed, Lemma 4.13 and Theorem 4.9 prove the following:

Theorem 4.14. *Let $B = \{b_i\}_{i=1}^n$ and $R = \{r_i\}_{i=1}^n$ be sets of blue and red points distributed uniformly at random in $[0,1]^d$, $d \geq 3$, with weights distributed uniformly in $[0,1]$, and let M_n be the optimal matching of B against R. Then $\mathbf{E}M_n = \Theta(n^{(d-1)/d})$.*

4.4 Conclusions

In this section we discussed some special cases of weighted matching before defining a general version of the problem. We showed that the general optimal weighted matching is bounded above by optimal unweighted matching.

5. Matching on Non-Rectangular Sets and Future Work

The HST machinery we have developed allows us to extend our optimal matching results in directions that seem unlikely using the machinery previously described in the literature. We are able to establish upper bounds on expected optimal matching weight for non-rectangular sets—indeed, for any set that submits to a regular tiling. Moreover, we can establish reasonable upper bounds for expected optimal matching on certain sets of measure zero or whose measure is not obvious to determine. By Bartal's results on approximating arbitrary metrics with tree metrics [21], we will know our upper bounds are no worse than $\log n$ greater than optimal, although lower bound computations are not as clear here. Perhaps surprisingly, these discussions will lead us to further insight into matchings on rectangular sets. At the close of this chapter, we will propose a practical algorithm and avenues for future work.

5.1 Optimal Matching on Non-Rectangular Sets

We now turn our attention to computing the expected weight of optimally matching n red points and n blue points distributed uniformly and at random in the interior of a triangle. A triangle is easily decomposed into four similar triangles by connecting the midpoints of the three sides (Figure 5.1). Since this may be done recursively, this affords a decomposition amenable to using a balanced 2-HST with branching factor 4. If at each level we give each edge the weight required for the longest edge, we have a tree that is identical, up to a constant in edge length, to the tree we used for optimal matching in $[0,1]^2$. It follows, then, from the Upper Bound Matching Theorem (Theorem 3.21) that the same result must hold: the expected weight of an optimal matching is bounded above by $O(\sqrt{n}\log n)$.

Theorem 5.1. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random in the interior of a triangle contained in $[0,1]^2$. Then the expected weight of an optimal matching between R and B is $O(\sqrt{n}\log n)$.*

We can continue this technique in higher dimensions. Consider the tetrahedron, which we consider as embedded in $[0,1]^3$, since we must give it some bounding box in order to speak meaningfully of expectation. As in the case of the triangle, we can similarly divide the tetrahedron recursively

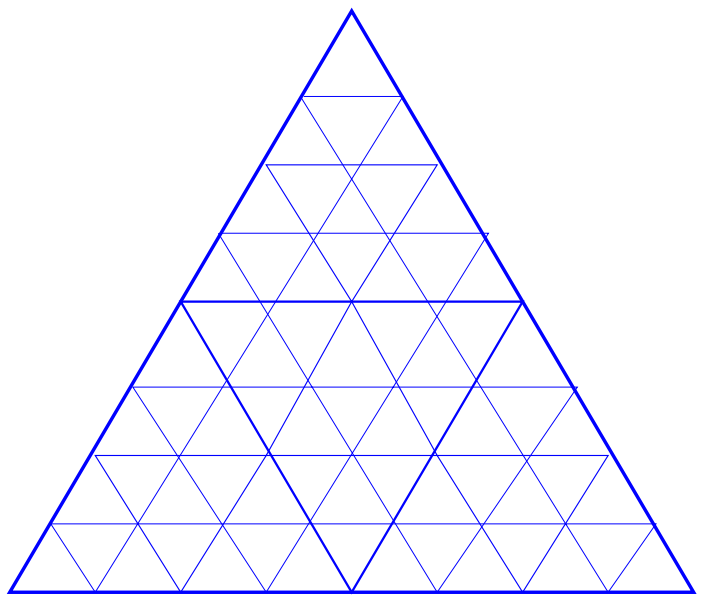

Figure 5.1: The recursive decomposition of a triangle into four similar triangles.

into sets of five similar sub-tetrahedrons. Then a balanced 2-HST ($k = 2$) with branching factor 5 ($b = 5$) upper bounds the Euclidean metric on the tetrahedron, and we have

Theorem 5.2. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random in the interior of a tetrahedron embedded in $[0, 1]^3$. Then the expected weight of an optimal matching between R and B is $O(n^{1-\log_5 2})$.*

Proof. Since $b = 5 \neq 4 = k^2$, and calling $L = \log_5 2 \approx .431$, the Upper Bound Matching Theorem (Theorem 3.21) tells us that

$$W = \left(\frac{2\Delta\sqrt{n}}{m^L}\right)\left(\frac{\sqrt{m} - m^L}{\sqrt{5} - 2}\right) = \Theta(n^{1/2}m^{-L}m^{1/2}) = \Theta(n^{1/2}m^{-L+\frac{1}{2}}).$$

We know from the proof of Lemma 3.30 that $m = n$ adds only constant discretization overhead in $[0, 1]^3$, so the same is true for a tetrahedron embedded in $[0, 1]^3$. The result follows by substituting $m = n$ above. □

In fact, we will see with Theorem 5.7 that we can do better: $O(n^{2/3})$.

Figure 5.2: Construction of the (ternary) Cantor set: each successive horizontal section is an iteration of the recursive removing of middle thirds. The Cantor set has measure zero (and so contains no segment), is uncountable, and is closed.

5.2 Optimal Matching on Sets of Non-Positive Measure

This section necessarily begins with a disclaimer: this text is a dissertation in computer science and not in pure mathematics. We have no particular interest aside from curiosity in claiming any particular matching bounds for sets that a computer scientist is unlikely ever to attempt to construct. It is visually and mentally stimulating in the following to discuss optimal unweighted matching on, for example, the Cantor set, but it would be a distraction here to prove that the appropriate results remain true in the limits of the processes that generate the fractals and other exotic sets of which we will speak. The truly interesting result is that our machinery is general enough to consider these things and to give us useful results. The reader, therefore, is implored to remember that we always mean to consider matching on the $\log n^{\text{th}}$ iteration of these processes, not the limit. We will be careful in the statement of our theorems, but it would be a shame to lose visual panache by prohibiting reference altogether to the limits of these processes.

With that overly long disclaimer, let us begin by noting that the (ternary) Cantor set is formed by repeatedly removing the open middle third of each line segment in a set of line segments, starting with $[0, 1]$, see Figure 5.2. The Cantor set arises naturally in topology and, perhaps more commonly, in elementary measure theory, where it serves as a simple example of an uncountable set of measure zero.

Consider sets $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ of red and blue points respectively, distributed uniformly at random on the Cantor set. We are interested in the expected weight of an optimal matching. Using balanced 2-HST's, we can at least establish an upper bound. Consider a balanced 2-HST with branching factor $b = 2$ and $m = n^2$ leaves over the unit interval, in which we think of the Cantor set as being embedded. The discretization overhead associated with approximating the red and blue points with the leaves is no more than the cost of moving each point to the nearest

Figure 5.3: The Sierpinski triangle. We find an upper bound using a balanced 2-HST with branching factor 3.

leaf: $(2n)(1/2n) = 1$. (In fact, the distance to the nearest leaf is certainly less than $1/(2n)$, but $1/(2n)$ suffices and the real distance presents distracting complexity.) By the Upper Bound Matching Theorem (Theorem 3.21), then, since $b = 2 \neq 4 = k^2$, and noting that $\log_b k = \log_2 2 = 1$, we have an upper bound of

$$W = \left(\frac{2 \cdot 1 \cdot \sqrt{n}}{m}\right)\left(\frac{\sqrt{m} - m}{\sqrt{2} - 2}\right) = \frac{2\sqrt{n}}{n^2} \cdot \frac{n^2 - n}{2 - \sqrt{2}} = \Theta(\sqrt{n}).$$

Since the tree metric dominates the Euclidean metric, and since the expected matching weight in the balanced 2-HST is asymptotically $\sqrt{n}$, we have proved the unexpected result that the expected weight of an optional matching of n points distributed on the Cantor set is no heavier than the same points distributed on $[0, 1]$ itself. We have proved

Theorem 5.3. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random in the set remaining after $\log n$ iterations of the Cantor set process. Then the expected weight of an optimal matching between R and B is $O(\sqrt{n})$.*

It is interesting to note that the Smith-Volterra-Cantor set (in which one removes successively smaller intervals such that the set has positive measure but contains no interval) is dominated by the same tree, and so we have for it the same upper bound, even though it has positive measure.

Consider now the Sierpinski triangle (see Figure 5.3). Here a balanced 2-HST with branching factor $b = 3$ and m leaves dominates the Euclidean metric, offering a nice approximation. Calling $L = \log_3 2 \approx .631$, we have

$$W = \left(\frac{2 \cdot \Delta \cdot \sqrt{n}}{m^L}\right)\left(\frac{\sqrt{m} - m^L}{\sqrt{3} - 2}\right) = \frac{c\sqrt{n}}{m^L} \cdot \left(m^L - \sqrt{m}\right) = \Theta(\sqrt{n}).$$

We have thus proved

Theorem 5.4. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed*

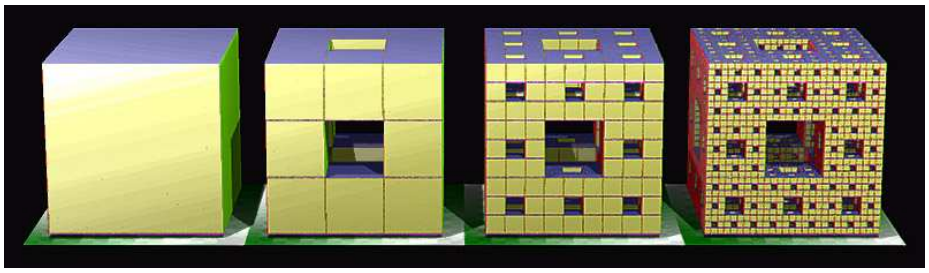

Figure 5.4: The first four iterations of a Menger sponge. A 20-HST provides a dominating metric (cf. Theorem 5.6).

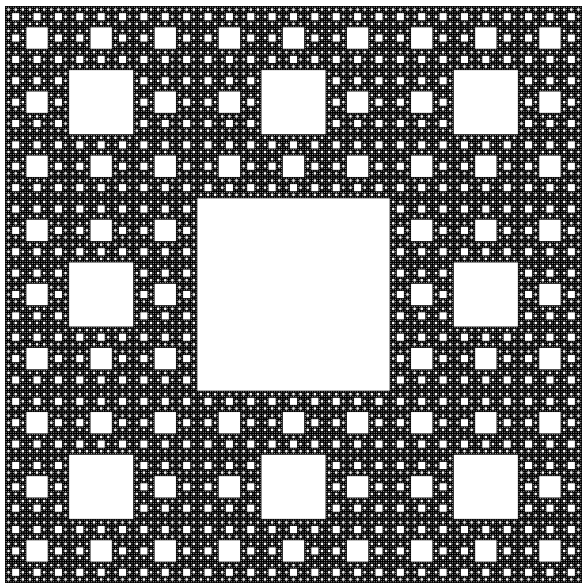

Figure 5.5: A Sierpinski carpet [199], constructed by recursively dividing the square into nine equal subsquares and removing the middle subsquare.

uniformly and at random in the interior of the $\log_3 n^{th}$ *iteration of a Sierpinski triangle. Then the expected weight of an optimal matching between* R *and* B *is* $O(\sqrt{n})$.

Note that the value of m here is not important as long as it is a function of n and $m > n$.

We can find upper bounds for optimal matching on other fractals. A Sierpinski carpet (cf. Figure 5.5) results from removing sub-squares from a square, but let us consider instead the Menger sponge, after which we will be well placed to generalize what we have done so far in this chapter.

Definition 5.5. A *Menger sponge* results from recursively dividing the unit cube into $3^3 = 27$ subcubes, removing the middle cube on each face and the cube in the center, then recursing on each

subcube (cf. Figure 5.4).

Theorem 5.6. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random in the interior of the $\log_{20} n^{th}$ iteration of a Menger sponge. For simplicity of notation, let $L = \log_{20} 2 \approx .231$. Then the expected weight of an optimal matching between R and B is $O(n^{1-\log_{20} 2})$.*

Proof. To find an upper bound on the expected weight of matchings on the Menger sponge, consider a balanced 2-HST with $m = n$ leaves and branching factor $b = 20$ (the number of cubes left behind at each iteration). Then an upper bound on the weight of an expected matching is

$$W = \left(\frac{2 \cdot \Delta \cdot \sqrt{n}}{m^L}\right)\left(\frac{\sqrt{m} - m^L}{\sqrt{2}0 - 2}\right) = c_1\sqrt{n}n^{-L}n^{1/2} = n^{1-L} = n^{1-\log_{20} 2} \approx n^{.769}.$$

While each level of the tree places its children in the 20 cubes left by the division of the nodes cube, the maximum distance to a leaf is certainly no greater than if we simply distributed the leaves uniformly in the cube (on a lattice). If that were the case, we would have a maximum distance from each point to the nearest leaf of $\sqrt{3}/(2\sqrt[3]{n})$, and so we would be adding total weight $(2n\sqrt{3})/(2\sqrt[3]{n}) \sim n^{2/3} \approx n^{.667}$ to the matching. Since the matching within the tree has greater expected weight, the result follows. □

The astute reader has perhaps noticed a pattern of similarity between optimal matching on non-rectangular sets and optimal matching on rectangular sets. Let us now state this generalization more formally with the following important theorem:

Theorem 5.7 (Generalized Upper Bound Theorem)**.** *Let $X \subset \mathbb{R}^d$ be a bounded set, and let $Y \subseteq X$. If the Upper Bound Matching Theorem provides an upper bound of $f(n)$ for optimal matching on X, then $f(n)$ is also an upper bound for optimal matching on Y.*

In other words, removing points from a set can not increase the upper bound on the expected optimal matching weight, at least as far as we can tell using HST's and the Upper Bound Matching Theorem. Note, however, that nothing prevents Y from having a lesser upper bound. (Consider $[0, 1] \subset [0, 1]^2$.) But we can not have a greater upper bound for Y using the Upper Bound Matching Theorem.

Proof. By hypothesis, we can construct an HST T on X such that the metric of T dominates the

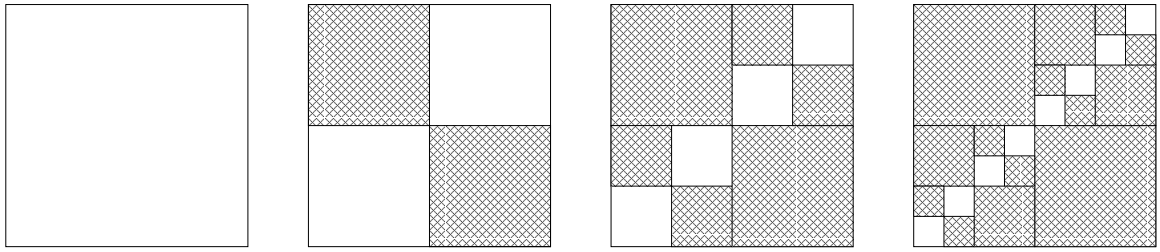

Figure 5.6: The first four steps in Process 5.8: removing upper-left and lower-right squares.

metric of X and, taking into account discretization error, the Upper Bound Matching Theorem gives us an upper bound of $f(n)$ for optimally matching two sets of n points each, distributed uniformly at random on X. Suppose that the expected weight of an optimal matching on Y is $g(n) > f(n)$. Then we could use the same tree T on Y to find that f is also an upper bound on Y, which contradicts our assumption that $g > f$. $\square$

As immediate corollaries to the Generalized Upper Bound Theorem we see that an upper bound for matching on triangles is $O(\sqrt{n}\log n)$ (Theorem 5.1), for tetrahedrons is $O(n^{2/3})$ (Theorem 5.2 in the generalization we promised at the time of its proof), the Cantor set is $O(\sqrt{n})$ (Theorem 5.3), as is the Smith-Volterra-Cantor set, and the Menger sponge has upper bound $O(n^{2/3})$, a slight improvement on Theorem 5.6.

Let us now consider the following process, see Figure 5.6.

Process 5.8. Begin with the unit square, removing the top left and bottom right subsquares. Repeat this with the remaining two squares, then the remaining four, and so forth. If x is a point on the boundary of a square we remove, then we do not remove x if every neighborhood of x contains some point that has neither been removed nor is itself on the boundary of the square we are currently removing.

Clearly the limit of this process is the diagonal, and so the expected weight of an optimal matching of two sets of n points is $\Theta(\sqrt{n})$.

If we compute an upper bound on the expected weight of an optimal matching, we should surely expect to find the answer to be no greater than $O(\sqrt{n}\log n)$. Indeed, we might start with a 4-branching balanced 2-HST, but then note that two of each four branches will be pruned. But the result for a 2-branching balanced 2-HST is, up to a constant, the same tree as we used in

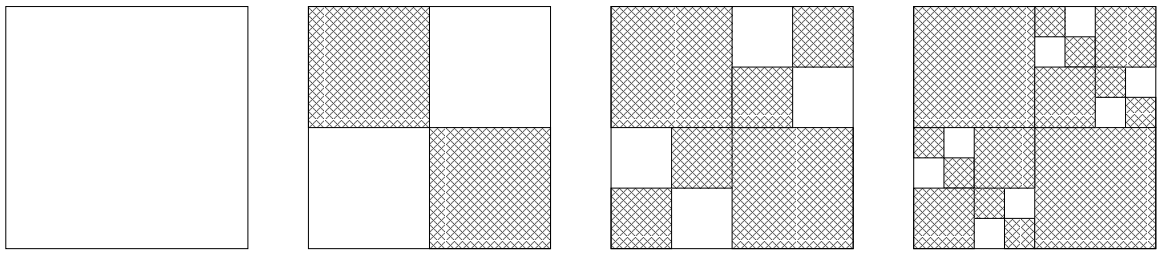

Figure 5.7: The first four steps in Process 5.10: removing upper-left and lower-right squares alternating with removing upper-right and lower-left squares.

Theorem 3.25. Or we might appeal directly to the Upper Bound Matching Theorem (Theorem 3.21).

Theorem 5.9. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random on the interior of the area remaining after $\log n$ iterations of Process 5.8. Then the expected weight of an optimal matching between R and B is $\Theta(\sqrt{n})$.*

Consider now the same process as in Process 5.8, except that in alternate steps we remove instead the upper-right and lower left squares, see Figure 5.7:

Process 5.10. Begin with the unit square. In even steps, remove the top-left and bottom-right quarter of each square. In odd steps, remove the bottom-left and top-right quarters. We treat boundary points as in Process 5.8.

Clearly the limit is no longer a line segment. Without worrying about whether the limit of Process 5.10 contains any points at all, let us consider the question of optimal matchings among red and blue points distributed uniformly and at random in the interior of the $\log n^{\text{th}}$ iteration of this process, since we are considering finite trees in any case. Clearly the HST in this event is the same as the tree in the case Process 5.8. We therefore achieve the same upper bound of $O(\sqrt{n})$.

Theorem 5.11. *Let $R = \{r_1, \ldots, r_n\}$ and $B = \{b_1, \ldots, b_n\}$ be sets of red and blue points distributed uniformly and at random on the interior of the area remaining after $\log n$ iterations of Process 5.10. Then the expected weight of an optimal matching between R and B is $O(\sqrt{n})$.*

5.3 Future Work on Optimal Matching

An obvious extension to this work that we did not have time to pursue considers the question of non-uniform distributions. Our tree method should be fruitful even in this context by considering

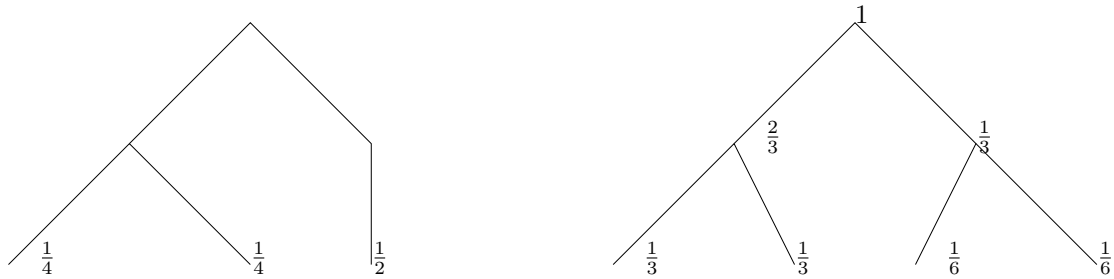

Figure 5.8: Non-uniform distributions may be approximated by 2-HST's with non-uniform branching factors. On the left, we modify the branching factor to change the distribution, leaving the probability the same at each node (50% chance of distributing to either side). On the right, we modify the probabilities at the interior nodes, leaving the branching factor constant.

non-uniform branching factors. Figure 5.8 shows an illustrative example. The interest in non-uniform distributions, beyond the pure theoretical pleasure that comes from being able to make claims about distributions other than the uniform, lies in the domain of computer vision. The point distributions that come from randomly sampling the shapes and objects we see around us in the real world are certainly examples of non-uniform probability distributions.

Suppose we have the sort of point set we often find in computer vision applications: points uniformly sampled from a shape (a silhouette, a surface, a compact manifold). We are interested in the behavior of the transportation distance between two such (equal cardinality) random samplings. We may approximately tile the shape with cubes, but not perfectly. Our results, whether weighted or unweighted, apply to each cube, regardless of its size. The question, then, is whether these various cubes are additive in the limit.

Intuition suggests that they should be. We have seen in the proofs that most points match in the lower levels (smaller cubes or lower down in the HST's, which is roughly the same thing). This suggests experimental work to validate the utility of the following procedure for comparing comparisons of compact manifolds ("shapes") at different resolution.

Proposition 5.12. *Let M be a real, compact, well-behaved d-manifold. Let $R = \{r_1, \ldots, r_N\}$ and $B = \{b_1, \ldots, b_N\}$ be independent, uniformly sampled random point sets on M. Then*

$$\mathrm{EMD}(R, B) \sim \begin{cases} V\sqrt{n} & \textit{if } d = 1, \\ V\sqrt{n \log n} & \textit{if } d = 2, \textit{and} \\ V n^{(d-1)/d} & \textit{if } d \geq 3, \end{cases}$$

where V is the measure of M and $n = N/V$, the sampling density.

This presents a sampling-density-independent measure of object similarity. Given two compact manifolds M_1 and M_2 each sampled at n_1 and n_2 points per unit volume, comparisons of the manifolds are comparable by scaling by the ratios indicated in Proposition 5.12. Said another way, we may define a normalized EMD measure:

NORMALIZED-EMD(S_1, S_2)

$\triangleright$ S_1 and S_2 are point sets representing uniformly sampled manifolds.

1 Scale objects to have the same volume (based on triangulation, for example)

2 $N \leftarrow n_1 + n_2$, the sum of the numbers of points in the two objects

3 **return** EMD (possibly under transformation) / the factor from Proposition 5.12.

It will be interesting to see if this similarity function indeed returns about 1 for similar object comparisons regardless of sampling density (as long as the objects are not under-sampled). Algorithm NORMALIZED-EMD promises to permit object comparisons despite different sampling densities, facilitating the use of different point set databases in the same set of experiments.

5.4 Conclusions

In this section we used the Upper Bound Matching Theorem (Theorem 3.21) to extend our matching results to various tileable sets: triangles, tetrahedrons, and so forth. The Upper Bound Matching Theorem served as well to extend our matching results to more exotic sets, such as the Cantor set and various fractals. These examples let us to prove the Generalized Upper Bound Theorem, Theorem 5.7. We finished by proposing application of the current work in shape matching in the context of computer vision.

6. Deterministic Sampling

6.1 Overview and Related Work

Data compression[23, 29, 36, 100, 119, 71, 163, 192, 205, 206], like subset selection, seeks to represent points as space efficiently as possible. Claude Shannon [163, 162, 161] formalized the measure of information content given the probability distribution of the point set. That work provided a theoretical framework for determining the maximum capacity of a communications channel. Closely related is the (incomputable) Kolmogorov complexity [165, p. 236].

Rate distortion theory [54] represents a set of points by an optimal reduced point set. It constructs new points as needed rather than selecting a subset of the original set. The information bottleneck method [181] permits a trade-off between compressed size and fully detailed quantization.

In the clustering problem[123] a point set is partitioned based on some given similarity measure, with goal that points in each partition are more similar among themselves than to points in other partitions. Indyk [102, 103, 62, 102] discusses these problems at length, especially from a geometric perspective, noting the complexity that increasing dimension presents and offering several approaches to the problem. Clustering is closely related to the problem of nearest neighbor and k-nearest neighbor searching [104, 103].

Gupta *et al.* [91] compiled a review with extensive bibliography of many techniques used in the design of randomized algorithms. Amir and Dar [14] demonstrate deterministic polynomial algorithms for uniform random sampling. Even *et al.* [69] describe efficient deterministic constructions of small probability spaces. Such constructions have applications in constructing small low discrepancy sets in high dimensions. Chari *et al.* [41] present NC algorithms[1] for approximating probability distributions. Chazelle and Friedman [45, 44] have shown that randomized algorithms combined with divide-and-conquer techniques can be derandomized with only polynomial overhead. Reingold *et al.* [146, 147] discuss using "small" amounts of randomness to achieve almost true randomness. Saks [153] provides a related survey.

The problem we address here was initially motivated by the study of reference functions in shape

[1] The class NC is the set of decision problems for which there exists a family of circuits of polynomial size and polylogarithmic depth. The following formulation is equivalent: the set of decision problems decidable in polylogarithmic time on a parallel computer with a polynomial number of random access machines [188, p. 107ff].

recognition. A reference function [10, 2, 9, 11] is a Lipschitz continuous selector function (q.v., below: Definition 6.2) that is equivariant with respect to a given family of transformations. More formally,

Definition 6.1. Given a class $\mathcal{C}$ of shapes in $\mathbb{R}^d$ and a set $\mathcal{T}$ of transformations of $\mathbb{R}^d$, a *reference function* is a function $f : \mathcal{C} \to \mathbb{R}^d$ with the following two properties:

- f is equivariant with respect to $\mathcal{T}$: $\forall X \in \mathcal{C}$, $\forall T \in \mathcal{T}, f(T(X)) = T(f(X))$
- f is Lipschitz continuous: $\exists L > 0$ such that $\forall X, Y \in \mathcal{C}$, $\forall T \in \mathcal{T}, \| f(X) - f(Y) \| \leq L \cdot d(X, Y)$.

The Lipschitz constant L is called the *quality of the reference function.* The image of a shape X under f is called the *reference point* of f.

Typically $\mathcal{C}$ is the set of compact sets in $\mathbb{R}^d$, sometimes with some additional constraints such as connectivity, simple connectivity, or not too spiny. The set $\mathcal{T}$ is typically "translations," "rotations," "isometries," or "isometries plus scaling". Some authors conflate the mapping with its image, thus identifying reference functions and reference points. The goal of reference function research is to aid in shape matching by reducing the number of degrees of freedom between two candidate shapes to be matched. If the reference points of the shapes are aligned, then the theory hopes to show that the error in distance between the shapes is bounded. Reference functions arise in a mathematics as selector functions [127, 128, 141, 152, 7, 8, 83, 157, 79, 144, 18, 19, 96, 101, 171, 6, 59, 99, 126, 27, 136, 84, 135, 97, 129, 142, 143].

Definition 6.2. Given topological spaces X and Y and a mapping $\phi : X \to \mathcal{F}(Y)$, a *selection for* ϕ [129] is a mapping $f : X \to Y$ such that $\forall x \in X$, $f(x) \in \phi(x)$.

Definition 6.3. A *selector function* $\varphi : \mathcal{C}(T) \to T$ maps the set of compact subsets of a topological space T to points in T. In this paper, when we talk of selector functions, we will require in addition and without repeating it each time that the function φ be continuous and that, moreover, it be the identity function on points: $\forall p \in T$, $\varphi(\{p\}) = p$.

One may (less usefully) define selector functions in a more general context by replacing $\mathcal{C}(T)$ with $\mathcal{P}(T)$, the power set of T.

Neither the mathematics of selector functions nor the work on reference functions interests us greatly at this point. We note merely that it has influenced this work in its early stages.

6.2 Defining the Problem

Arguably the simplest technique for reducing the complexity of a point set while nearly preserving some given intrinsic structural property $\mathcal{P}$ is random selection. While no algorithm can hope to exceed the time bounds of random selection (assuming a free source of randomness), one might hope to develop deterministic techniques with improved performance as measured by closeness of preservation of $\mathcal{P}$.

We address the following problem:

Problem 6.4. Let $\gamma : C \to \mathbb{R}^d$ be the centroid function on compact sets. Given a finite set $S \subset \mathbb{R}^d$ of points and $m \in \mathbb{Z}^+$, $m \ll |S|$, find a set $T \subset S$ with $|T| = m$ that minimizes $|\gamma(S) - \gamma(T)|$.

Global minimization, of course, is combinatorially expensive. In reality, we look at the closely related problem in which the goal is to find a T such that $|\gamma(S) - \gamma(T)| < \epsilon$, where we choose the ϵ that we would expect from uniformly choosing the points of T at random.

We also begin to address the following problem:

Problem 6.5. Given a finite set $S \subset \mathbb{R}^d$ of points and $m \in \mathbb{Z}^+$, $m \ll |S|$, find a set $T \subset S$ with $|T| = m$ that minimizes $\textsc{emd}(S, T)$.

Again, in reality we look at the related problem of deterministically selecting T such that $\textsc{emd}(S, T) < \epsilon$, as above.

Giné [82] offers a survey of work on how well sample means approach expected value, as does Wellner [189, 190].

6.3 Approaching a Solution

The difficulty of the problem lies purely in its combinatorial complexity. The path to solution, then, lies in computing a combinatorially smaller number of things that permit a greedy solution. In the case of approximating the centroid, those things are pairs of points. By computing point pairs, of which there are only $O(n^2)$, we can consider greedy algorithms on point pairs rather than more complicated optimization on individual points. In the case of the EMD version we consider greedy algorithms on k-nearest neighbor clusters.

6.4 Preserving the Centroid

6.4.1 Martingales Prove the Obvious

It is often useful to note that what we think is obvious is probably true. In that spirit, what follows is an application of the Hoeffding-Azuma Inequality that shows that random sampling stands a very good chance of preserving the centroid.

Let $S = \{s_i\}_{i=1}^n$ be a set of points in $[0,1]$ and $S' = \{s'_i\}_{i=1}^m$ be a subset of S with $m \ll n$. (We can consider this problem on $[0,1]^d$ for $d > 1$, but the problem is clearly decomposable as far as this section is concerned, so we restrict our attention to the notationally simpler case of $[0,1]$.) Let $\gamma : 2^S \to \mathbb{R}$ be the centroid function.

We'll think of S' as randomly drawing m points from S without replacement. Let $\{X_i\}_{i=1}^m$ be a sequence of random variables, where the value of X_i is the member of S that we pick on the ith draw. (Without loss of generality, X_i is the value of s'_i.) Let us formalize this as a martingale. Let $Z_0 = \mathbf{E}c(S')$, and let $\{Z_i\}_{i=1}^m$ be a sequence of random variables defined by

$$Z_i = \mathbf{E}\left[c(S') \mid X_1, \ldots, X_i\right].$$

Our goal is to show that $|c(S) - c(S')|$ is small. In particular, we'll show in Claim 6.6 that $\{Z_i\}_{i=1}^m$ is a martingale. The bound will then follow by the Hoefding-Azuma Inequality in Theorem 6.7.

Claim 6.6. *$\{Z_i\}_{i=1}^m$ is a martingale with respect to $\{X_i\}_{i=1}^m$.*

Proof. To be a martingale, we must satisfy the following three conditions:

- Z_i is function of $X_1, \ldots, X_i$,
- $\mathbf{E}Z_i < \infty$, and
- $\mathbf{E}[Z_{i+1} \mid X_1, \ldots, X_i] = Z_i$ for all $i \in [m-1]$

The first two conditions are obvious, the second following from the fact that the image of c is

bounded. The third requires some work.

$$\begin{aligned}
\mathbf{E}[Z_{i+1} \mid X_1, \ldots, X_i] &= \mathbf{E}\left[\mathbf{E}\left[c(S') \mid X_1, \ldots, X_i, X_{i+1}\right] \mid X_1, \ldots, X_i\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[c(S') \mid X_1, \ldots, X_i\right] \mid X_1, \ldots, X_i\right] && (6.1) \\
&= \mathbf{E}\left[c(S') \mid X_1, \ldots, X_i\right] && (6.2) \\
&= Z_i
\end{aligned}$$

where Equation (6.2) follows because $\mathbf{E}X \mid X = X$ for any random variable X, from which, in particular, $\mathbf{E}(\mathbf{E}X \mid Y) \mid Y = \mathbf{E}X \mid Y$. Equation (6.1) is derived by noting that

$$\begin{aligned}
\mathbf{E}\left[c(S') \mid X_1, \ldots, X_i, X_{i+1}\right] &= \sum_{x_{i+1}} \mathbf{E}\left[c(S') \mid X_1, \ldots, X_i, X_{i+1} = x_{i+1}\right] \Pr(X_{i+1} = x_{i+1}) \\
&= \sum_{x_{i+1}} \mathbf{E}\left[c(S') \mid X_1, \ldots, X_i, X_{i+1} = x_{i+1}\right] \frac{1}{n-i} \\
&= \mathbf{E}\left[c(S') \mid X_1, \ldots, X_i\right]
\end{aligned}$$

□

We would like to show that $\Pr(|c(S) - c(S')| > t) < \epsilon$ for some suitable definition of t, and ϵ.

The Hoeffding-Azuma Inequality provides that

Theorem 6.7. *Let $Z_0 = \gamma$ and suppose there exist constants c_i, such that*

$$|Z_i - Z_{i-1}| \leq c_i.$$

Then for $i \geq 0$ and $\epsilon > 0$,

$$\Pr(Z_i - Z_0 \geq \epsilon) \leq \Pr(\max_{1 \leq k \leq n} Z_k \geq \epsilon) \leq \exp\left\{\frac{-\epsilon^2}{2\sum_{k=1}^{i} c_k^2}\right\}.$$

Proof. Setting $c_i = -2d/i$, where d is the diameter of S, is surely adequate to fulfill the requirements above, since the centroid can not change by more than the maximum interpoint distance weighted by the points $X_1, \ldots, X_{i-1}$. □

6.4.2 Variance and Antipodal Point Algorithms

Let us try to solve the following problems:

Problem 6.8. Given $\epsilon > 0$, find the smallest set Y we can with the following properties:

- $Y \subseteq X$,
- $|Y| = m$, and
- $|\gamma(Y) - \gamma(X)| < \epsilon$.

Here the function $\gamma : 2^X \to \mathbb{R}$ is the centroid function.

A variation of Problem 6.8 is the following:

Problem 6.9. Given a positive integer $m < n$, find the set Y with the above three properties so as to achieve the smallest ϵ that we can.

6.4.3 An Application of the Variance

Hypothesis 6.10. *Let $X_1, X_2, \ldots X_{n+k}$ be $n+k$ random variables (points selected independently from a uniform distribution) on the square $[0,1]^d$. For each m, let $\gamma_m = \frac{1}{m}\sum_{k=1}^{m} X_i$ be the centroid of the first m points, and let $\widetilde{D}_{n,k} = \gamma_{n+k} - \gamma_n$.*

We thus have that $\widetilde{D}_{n,k}$ is the difference between the $n+k$'th centroid and the n'th centroid, so that the variance $\mathbf{Var}(\widetilde{D}_{n,k})$ is the mean square distance between the two centroids.

Claim 6.11 (Eric Schmutz)**.** *Let $X = \{X_1, \ldots, X_{n+k}\}$ and $\widetilde{D}_{n,k}$ be as in Hypothesis 6.10, with $d = 1$. Then $\mathbf{Var}(\widetilde{D}_{n,k}) = \frac{k}{12n(n+k)}$.*

Proof. Note that

$$\widetilde{D}_{n,k} = \gamma_{n+k} - \gamma_n = \left(\frac{1}{n+k} - \frac{1}{n}\right)\sum_{i=1}^{n} X_i + \frac{1}{n+k}\sum_{i=n+1}^{n+k} X_i.$$

The variance of a sum of independent random variables is the sum of their variances, therefore

$$\begin{aligned}\mathbf{Var}(\widetilde{D}_{n,k}) &= \left(\frac{1}{n+k}-\frac{1}{n}\right)^2 n\,\mathbf{Var}(X_1)+\left(\frac{1}{(n+k)^2}\right)k\,\mathbf{Var}X_1\\ &= \left(\frac{k}{n(n+k)}\right)\mathbf{Var}(X_1).\end{aligned}$$

But $\mathbf{Var}(X_i)=\frac{1}{12}$, and the result follows. □

Corollary 6.12. *Let X and $\widetilde{D}_{n,k}$ be as in Hypothesis 6.10, for arbitrary $d\in\mathbb{Z}^+$. Then $\mathbf{Var}(X)=\frac{d}{12}$.*

Proof.

$$\begin{aligned}\mathbf{E}X &= \int_0^1\cdots\int_0^1 X\,dx_1\cdots dx_d\\ &= \int_0^1\cdots\int_0^1 (X_1,\ldots,X_d)\,dx_1\cdots dx_d\\ &= \left(\frac{1}{2},\ldots,\frac{1}{2}\right),\end{aligned}$$

$$(\mathbf{E}X)^2 = \mathbf{E}X\cdot\mathbf{E}X=\frac{d}{4},$$

and

$$\begin{aligned}\mathbf{E}(X^2) &= \int_0^1\cdots\int_0^1 X^2\,dx_1\cdots dx_d\\ &= \int_0^1\cdots\int_0^1 X_1^2+\cdots+X_d^2\,dx_1\cdots dx_d\\ &= \frac{d}{3}\end{aligned}$$

From Lemma 2.1, then,

$$\mathbf{Var}(X)=\mathbf{E}(X^2)-(\mathbf{E}X)^2=\frac{d}{3}-\frac{d}{4}=\frac{d}{12}.$$

□

Corollary 6.13. *Let X and $\widetilde{D}_{n,k}$ be as in Hypothesis 6.10, for arbitrary $d \in \mathbb{Z}^+$. Then* $\mathbf{Var}(\widetilde{D}_{n,k}) = \frac{kd}{12n(n+k)}$.

Proof. The proof is identical to that of Claim 6.11 except in the last step where we instead use Corollary 6.12. □

Definition 6.14. Consider integers m and n such that $0 < m < n$. Let $X = \{x_1, \ldots, x_n\}$ be a random set of points, and denote γ_k the centroid of the first k points of X, $\{x_1, \ldots, x_k\}$. We define $D_{n,m} = \gamma_n - \gamma_m$.

Corollary 6.15.

$$\mathbf{Var}(D_{n,m}) = \frac{(n-m)d}{12mn}.$$

Proof. $D_{n,m} = \widetilde{D}_{m,n-m}$. □

Note that if $m \ll n$, then $\mathbf{Var}(D_{n,m}) \sim \frac{d}{12m}$.

6.4.4 Choosing Points One at a Time: Lots of Points is Easy

Definition 6.16. Let V be a vector space. Denote the *centroid function*

$$\begin{aligned} \gamma : 2^V &\rightarrow V \\ \{\xi_1, \ldots, \xi_j\} &\mapsto \frac{1}{j}\sum_{i=1}^{j} \xi_i. \end{aligned}$$

Theorem 6.17. *Let $S = \{s_1, \ldots, s_n\}$ be a set of points in $\mathbb{R}^d$ viewed as a normed vector space. Consider the k-set $X = \{x_1, \ldots, x_k\} \subset S$. Then the set $Y = \{y_1, \ldots, y_{n-k}\} \subset S - X$ has the property that $\gamma(Y) = (n/(n-k))\gamma(S) - (k/(n-k))\gamma(X)$.*

Proof.

$$\begin{aligned}\gamma(S) &= \frac{1}{n}\sum_{i=1}^{n} \\ &= \frac{1}{n}\left(\sum_{i=1}^{k} x_i + \sum_{i=1}^{n-k} y_i\right) \\ &= \left(\frac{k}{n}\right)\gamma(X) + \left(\frac{n-k}{n}\right)\gamma(Y).\end{aligned}$$

Rearranging terms, we have

$$\left(\frac{n-k}{n}\right)\gamma(Y) = \gamma(S) - \left(\frac{k}{n}\right)\gamma(X),$$

and so

$$\gamma(Y) = \left(\frac{n}{n-k}\right)\gamma(S) - \left(\frac{n}{n-k}\right)\left(\frac{k}{n}\right)\gamma(X) = \left(\frac{n}{n-k}\right)\gamma(S) - \left(\frac{k}{n-k}\right)\gamma(X).$$

□

It follows that if $n = 2k$ and we identify a set of points X, then an individual point of the set $Y = S - X$ may be anywhere in $\mathbb{R}^d$, but with each point of Y that we identify, the remaining points are further and further constrained. If the point set S is sufficiently large relative to X (that is, $n \gg k$), and if its points are sufficiently distributed according to some sufficiently nice probability distribution, then we can make the following statement:

Not a Theorem. *If the hypotheses of Theorem 6.17 are satisfied, and, moreover, if S is distributed "sufficiently randomly," then there exists a set $Y = \{y_1, \ldots, y_k\} \subset S - X$ and an $\epsilon > 0$ such that $|\frac{1}{2}[\gamma(Y) + \gamma(X)] - \gamma(S)| < \epsilon$.*

Note that the above non-theorem is tautologically true: ϵ may be as large as desired. We may correct the problem with

Theorem 6.18. *Let f be a probability distribution function on $\mathbb{R}^d$. Let $S_n = \{s_1, \ldots, s_n\}$ ($n \in \mathbb{Z}^+$) be a family of sets of points in $\mathbb{R}^d$ viewed as a normed vector space such that each S_n is drawn randomly according to f. Let $X = \{x_1, \ldots, x_k\} \subset S_n$ be a k-set in S_k. Then $\forall \epsilon > 0$, $\forall \delta \in (0, 1)$, $\exists N > 0$ such that with probability $\geq \delta$, $\forall n > N$, $\exists Y = \{y_1, \ldots, y_k\} \subset S_n - X$ such that $|\frac{1}{2}[\gamma(Y) + \gamma(X)] - \gamma(S_n)| < \epsilon$.*

Proof. I present two proof sketches.

1. (Gambler's Ruin) Define a region R of radius no more than ϵ around $\gamma(S_n) - \gamma(X)$. The probability of fewer than k points falling within R approaches zero as n increases without bound.

2. (Random graph theory) Consider the points of S_n as vertices of a graph. Consider the family $G_{n,p}(S_n)$ of random graphs. If the selector function (cf. Definition 6.3, below) is first order (centroid is not), then by the 0-1 Law the set Y almost surely exists.

□

6.4.5 Choosing Points One at a Time: Fewer Points is Harder

We start with a definition (recall Definitions 6.2 and 6.3):

Definition 6.19. Given a set $S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$, a (continuous) selector function φ on $\mathbb{R}^d$, and a point $p \in S$, then the point *antipodal to p in S relative to φ* is a point $q \in S$ that minimizes $|\varphi(\{p, q\}) - \varphi(S)|$.

Clearly if $\varphi = \gamma$, the centroid function, then the point antipodal to $p \in S$ relative to γ is just the point closest to $2\gamma(S) - p$.

Consider the following algorithm for finding antipodal points relative to a continuous selector function:

```
FIND-ANTIPODAL-POINT(S, φ, p)
    ▷ p ∈ S = {s_1, ..., s_n} ⊂ R^d.
1   min-dist ← ∞
2   for q ∈ S\{p}
3         do
4             dist ← |φ({p, q}) − φ(S)|
5             if dist < min-dist:
6                then
7                       min-dist ← dist
8                       min-q ← q
9   return min-q
```

We will see FIND-ANTIPODAL-POINT again in procedure FIND-ANTIPODALS on page 76.

Note that if $\varphi = \gamma$, the procedure appears to simplify to the following:

```
FIND-ANTIPODAL-POINT(S, γ, p)
    ▷ p ∈ S = {s_1, ..., s_n} ⊂ R^d.
1   Find the point q ∈ S\{p} closest to 2γ(S) − p
2   return q
```

Nearest neighbor search, however, is a difficult problem to solve in sublinear time, although some sublinear approximation algorithms are known [103, 95]. Given knowledge of the selector, in any event, the search space can sometimes be reduced, as in the case of centroid. Once we can compute antipodal points, we can find the image under φ of each point and its antipodal point:

```
FIND-ANTIPODAL-PAIR-IMAGE(S, φ, p)
    ▷ p ∈ S = {s_1, ..., s_n} ⊂ R^d.
1   q ← FIND-ANTIPODAL-POINT(S, φ, p)
2   return γ({p, q})
```

Lemma 6.20. *Let $S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$. Select a point p at random from S. Then* FIND-ANTIPODAL-PAIR-IMAGE(S, γ, p) *returns a point r such that $\mathbf{E}|r - \gamma(S)| < n^{-1/d}$. Moreover,*

if x and y are two points chosen at random from S, then

$$\mathbf{E}|\gamma(\{x,y\}) - \gamma(S)| > \mathbf{E}|r - \gamma(S)|.$$

Proof. Given the first point p, we expect to find a point q in the d-cube of edge length $n^{-1/d}$ nearest $2\gamma(S) - p$. On the other hand, by Corollary 6.15, we expect

$$|\gamma(\{x,y\}) - \gamma(S)| = D_{n,2} = \frac{(n-2)d}{24n} \sim \frac{d}{24}.$$

□

We can extend the idea of "a point antipodal to a point" to the idea of "a point antipodal to a set of points."

Definition 6.21. Given a set $S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$, a subset $T \subset S$ with $|T| \ll |S|$, and a continuous selector function φ on $\mathbb{R}^d$, then the *point antipodal to T in S relative to φ* is a point $q \in S$ that minimizes $|\varphi(T \cup \{q\}) - \varphi(S)|$.

Consider the following algorithm for finding antipodal points of sets:

```
SET-FIND-ANTIPODAL-POINT(S, φ, T)
  ▷ T ⊂ S = {s_1, ..., s_n} ⊂ R^d, with |T| ≪ |S|.
  min-dist ← ∞
  for q ∈ S\T
      do
          dist ← |φ(T ∪ {q}) − φ(S)|
          if dist < min-dist:
             then
                  min-dist ← dist
                  min-q ← q
  return min-q
```

Note that for $\varphi = \gamma$, this simplifies to the following:

SET-FIND-ANTIPODAL-POINT(S, γ, T)

$\triangleright$ $T \subset S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$, with $|T| \ll |S|$.

1 Find the point $q \in S$ closest to $(|T|+1)\gamma(S) - |T|\varphi(T)$ subject to $q \notin T$.

2 **return** q

Claim 6.22. *The centroid version of* SET-FIND-ANTIPODAL-POINT *is correct.*

Proof. Note that we want a q (ideally) such that $\varphi(S) = \varphi(T \cup \{q\})$. Writing $m = |T|$ for convenience, we can then compute

$$\begin{aligned} \varphi(S) &= \varphi(T \cup \{q\}) \\ &= \frac{\varphi(\{t_1\}) + \cdots + \varphi(\{t_m\}) + \varphi(\{q\})}{m+1} \\ &= \frac{m\varphi(T) + \varphi(\{q\})}{m+1} \end{aligned}$$

Rearranging terms, then, we have

$$\varphi(\{q\}) = (m+1)\varphi(S) - m\varphi(T).$$

□

Theorem 6.23. *Let $S = \{s_1, \ldots, s_n\} \subset \mathbb{R}^d$. Select a point, say p_1, at random from S and set $X_1 = \{p_1\}$. For each of $i = 1, \ldots, m-1$, set $X_{i+1} = X_i \cup \{$*SET-FIND-ANTIPODAL-POINT$(S, X_i)\}$. Let T be an m-element subset of S chosen at random. Then*

$$\mathbf{E}\,|\gamma(X_m) - \gamma(S)| < \mathbf{E}\,|\gamma(S) - \gamma(T)|\,.$$

Proof. Let T_i be a family of randomly chosen subsets of S such that $|T_i| = i$. The T_i may have non-empty intersection.

The expected deviation of X_2 from the centroid is $n^{-1/d}$, since, given the first point p_1, we expect to find a point p_2 in the d-cube of edge length $n^{-1/d}$ nearest $\gamma(S) - p_1$. On the other hand, we expect

$$|\gamma(S) - \gamma(T_2)| = D_{n,2} = \frac{(n-2)d}{24n} \sim \frac{d}{24}.$$

The assertion is therefore true for 2 points. We proceed by induction.

Suppose the assertion is true for $1, \ldots, k$. In the deterministic case, we pick the point closest to

$$\left(\frac{k}{n-k}\right)\gamma(X - X_i).$$

The expected distance from the optimal point is again $n^{-1/d}$, so the expected deviation of $\gamma(X_{i+1})$ from $\gamma(S)$ does not exceed $n^{-1/d}$, since the point chosen will be near the optimal direction. In the random case, we have

$$\mathbf{E}\,|\gamma(S) - \gamma(T_{k+1})| = \frac{(n-m)d}{2mn}.$$

□

We now define Find-Antipodals based on Find-Antipodal-Point from page 73. In Find-Antipodals we find a set of n points near the centroid of a set $S = \{s_1, \ldots, s_n\}$, each such point being the image under φ of two distinct points in S.

```
Find-Antipodals(S, φ)
    ▷ S = {s_1, ..., s_n} ⊂ R^d.
1   T ← ∅
2   for s ∈ S
3         do
4             q ← Find-Antipodal-Point(S, φ, s)
5             c ← φ({p, q})
6             T ← T ∪ {(c, first(p, q), second(p, q))}
7   return T
```

Here the functions first and second simply assure that (c, p, q) and (c, q, p) are not both represented in T, should they both arise. The order (first and second) may be determined, for example, by using the coordinates inherited from the real vector space structure.

We now use Find-Antipodals to make a selection T of $m \ll n$ points from S with the desired property that $|\varphi(T) - \varphi(S)|$ is small.

```
SUBSELECT-ANTIPODAL-FILL-BY-SET(S, φ, m)
    ▷ S = {s_1, ..., s_n} ⊂ ℝ^d, m ≪ n
1   ref-point ← φ(S)
2   T ← FIND-ANTIPODALS(S, φ)
3   Sort T in ascending order on the c values of each triple.
4   U ← T[1..m/2]
    ▷ Now set V to be the union of the p and q values of U
5   V ← {x | ∃u ∈ U, x = u[p] ∨ x = u[q]}
6   while |V| < m
7         do
8               V ← SET-FIND-ANTIPODAL-POINT(S, φ, V) ∪ V
9   return V
```

The while loop in SUBSELECT-ANTIPODAL-FILL-BY-SET compensates for the possibility that V after line 5 contains duplicate p and q components. Theorem 6.25 justifies this while loop.

Lemma 6.24. *Let $U_c = \{x \mid \exists u \in U, x = u[c]\}$, where U is the set in line 4 of* SUBSELECT-ANTIPODAL-FILL-BY-SET. *Then $\mathbf{E}|\varphi(U_c) - \varphi(S)| < n^{-1/d}$.*

Proof. By Lemma 6.20 the assertion is true for each point $c \in U_c$. Since the image under φ of a finite point set lies within the convex hull of the set of images of the individual points **(why?)**, the result follows. □

Theorem 6.25. *Procedure* SUBSELECT-ANTIPODAL-FILL-BY-SET *returns a set V with the property that $\mathbf{E}|\varphi(V) - \varphi(S)| < 2n^{-1/d}$.*

Proof. Since S is random and uniform, and since $m \ll n$, the probability that the points added by the while loop do not lie within $2n^{-1/d}$ of their desired location approaches zero. By the same reasoning, the probability of all the missing points lying in the same half-plane approaches zero, so the changes almost surely cancel each other to within $n^{-1/d}$. □

It is interesting to note that a slightly modified version of the algorithm using multisets might be slightly more efficient (although asymptotically the same):

```
SUBSELECT-ANTIPODAL-FILL-BY-PAIR(S, φ, m)
    ▷ S = {s_1, ..., s_n} ⊂ ℝ^d, m ≪ n
    ref-point ← φ(S)
    T ← FIND-ANTIPODALS(S, φ)
    Sort T on the c values of each triple.
    U ← T[1..m/2]
    ▷ Now set V to be the union of the p and q values of U
    V ← {x | ∃u ∈ U, x = u[p]} ∪_multi {x | ∃u ∈ U, x = u[q]}
    for v ∈ V duplicated
        do
            ▷ Fix each duplicate
            Remove one copy of v from V
            u ← the point nearest v ∈ S that is not already in V
            V ← V ∪ {u}
    return V
```

Theorem 6.26. *Procedure* SUBSELECT-ANTIPODAL-FILL-BY-PAIR *returns a set V with the property that* $\mathbf{E}|\varphi(V) - \varphi(S)| < 2n^{-1/d}$.

Proof. For each point that must be fixed, the probability that an unchosen point does not lie within $2n^{-1/d}$ approaches zero. □

It is interesting to note that a slightly modified version of the algorithm using multisets might be slightly more efficient (although asymptotically the same):

```
SUBSELECT-ANTIPODAL-FILL-FRESH-PAIRS(S, φ, m)
    ▷ S = {s_1, ..., s_n} ⊂ ℝ^d, m ≪ n
ref-point ← φ(S)
T ← FIND-ANTIPODALS(S, φ)
Sort T on the c values of each triple.
V ← ∅
t ← T[0]
while |V| < m
      do
          ▷ Add pairs of points if both elements of the pair are new to V.
          if t[0] ∉ V and t[1] ∉ V
             then
                  V ← V ∪ t
          t ← t.next        ▷ As long as m < n/3, we won't run out of pairs.
                                If m ≪ n/3, we'll moreover do well.
return V
```

Theorem 6.27. *Procedure* SUBSELECT-ANTIPODAL-FILL-FRESH-PAIRS *returns a set V with the property that* $\mathbf{E}|\varphi(V) - \varphi(S)| < 2n^{-1/d}$.

Proof. For each point that must be fixed, the probability that an unchosen point does not lie within $2n^{-1/d}$ approaches zero. □

6.5 Future Work: Preserving the Earth Mover's Distance

In this section we consider the related (and perhaps more practical) problem of preserving the Earth Mover's Distance between a point set and a sampled copy of itself. Experimental work suggests the algorithms of Section 6.5.2 will bear fruit.

6.5.1 Integer and Linear Programming

We first define the Earth Mover's Distance and note that it is easily solved by linear programming. That it is so easily solved by linear programming suggests an integer programming algorithm for EMD-preserving subset selection. Sadly, that algorithm does not work, but provides a starting

point for an algorithm, EMD-CLUSTER, that might (Conjecture 6.28) work, and which leads nicely to some open questions.

Consider two point sets $P = \{p_i\}_{i=1}^n$ and $Q = \{q_i\}_{i=1}^m$ together with a mass function $m : P \cup Q \to \mathbb{R}$. From these points we form the directed bipartite graph $G = (V, E)$, where $V = P \cup Q$ and $E = \{(p_i, q_j)\}$. We will denote the adjacency matrix by $A = (a(p_i, q_j))$. Edge weights equal the Euclidean distance between vertices, $d(p_i, q_j)$. We call P the source masses and Q the destination masses. If the sum of their masses is unequal, we may either scale the masses of one uniformly so that the sum on each side is the same or else add a single phantom mass on the lighter side to make up the difference, with zero weight edges from the phantom mass to all the masses on the other side.

The scaling technique is often called the *proportional transportation distance* [116], although for our purposes it is not useful to distinguish it from the original EMD. The phantom mass technique is generally called the Earth Mover's Distance with unequal masses. To simplify notation and without loss of generality, we assume that if a phantom point need be added, it already has been. We will be concerned later only with the proportional technique, but the exposition henceforth is unaffected by the difference between these two variations on mass transport optimization.

Let $C = (c(p_i, q_j))$ be the matrix of edge capacities, $W = (d(p_i, q_j))$ the edge weight matrix, and $F = (f(p_i, q_j))$ the flow matrix. (For our purposes here I am not interested in bounding capacity, so the elements of C may be taken as any numbers greater than the total mass of the points of P and Q.) Then

$$
\begin{aligned}
\{p_1, \ldots, p_n\} = P &\subset V \quad \text{(sources)} \\
\{q_1, \ldots, q_m\} = Q &\subset V \quad \text{(destinations)} \\
a(q_j, p_i) &= 0 \\
a(p_i, q_j) &= 1 \\
0 \le f(p_i, q_j) &\le c(p_i, q_j) \quad \forall\, p_i \in P,\, q_j \in Q \\
\sum_{i=1}^{n} m(p_i) &= \sum_{j=1}^{m} m(q_j)
\end{aligned}
$$

We can formulate this optimization problem in linear programming terms as shown in Table 6.1.

$$
\begin{array}{rcl}
\text{Minimize} & & \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} d(p_i,q_j)f(p_i,q_j) \\
\text{subject to} & & \\
0 \le f(p_i,q_j) & \le & c(p_i,q_j) \\
a(p_1,q_1)f(p_1,q_1)+\cdots+a(p_1,q_m)f(p_1,q_m) & = & w(p_1) \\
 & \vdots & \\
a(p_n,q_1)f(p_n,q_1)+\cdots+a(p_n,q_m)f(p_n,q_m) & = & w(p_n) \\
a(p_1,q_1)f(p_1,q_1)+\cdots+a(p_n,q_1)f(p_n,q_1) & = & w(q_1) \\
 & \vdots & \\
a(p_1,q_m)f(p_1,q_m)+\cdots+a(p_n,q_m)f(p_n,q_m) & = & w(q_m)
\end{array}
$$

Table 6.1: The Earth Mover's Distance formulation.

More succinctly,

$$
\text{Minimize } W \bullet F \text{ subject to } \begin{cases} F \le C \\ \operatorname{diag}(AF') = \mathbf{w}(P) \\ \operatorname{diag}(A'F) = \mathbf{w}(Q). \end{cases}
$$

In terms of our original sets S and S', we are interested in the situation where $P = S$ and $Q = S'$, which we view as a copy of S and whose members we will indicate $\underline{s}_i$ to remind ourselves that they are separate points that merely co-occupy the same space as the points of S. (In other words, the graph edge $(s_i, \underline{s}_i)$ is not a self-loop since we view $\underline{s}_i$ as a member of S'. Let $X = \{x_1, \ldots, x_n\}$ be the set of indicator variables for S. Indulging in a slight abuse of notation, since we don't know S', we may summarize this formulation as follows:

$$
\text{Minimize EMD}(S, S')
$$

$$
\text{subject to } \sum_{i=1}^{n} x_i \quad = \quad m.
$$

For the full glory of its algebraic minutia, however, see Table 6.2.

We relax the IP formulation to LP by letting $x_i \in [0, 1]$. Note, moreover, that $a(s_i, \underline{s}_j) = 1$ always, so we may drop the references to the adjacency matrix entirely. We may also drop the capacity constraint on flow. Finally, we note that if x_j only appears when multiplied by some $f(s_i, \underline{s}_j)$, then we may let the x_j's themselves be absorbed into the flow matrix. In order to drop the x_j, note

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} x_j d(s_i, \underline{s}_j) f(s_i, \underline{s}_j) \\
\text{subject to} \quad & \\
0 \leq f(s_i, \underline{s}_j) &\leq c(s_i, \underline{s}_j) \\
\sum_{i=1}^{n} x_i &= m \\
x_1 a(s_1, \underline{s}_1) f(s_1, \underline{s}_1) + \cdots + x_n a(s_1, \underline{s}_n) f(s_1, \underline{s}_n) &= 1 \\
&\vdots \\
x_1 a(s_n, \underline{s}_1) f(s_n, \underline{s}_1) + \cdots + x_n a(s_n, \underline{s}_n) f(s_n, \underline{s}_n) &= 1 \\
x_1 a(s_1, \underline{s}_1) f(s_1, \underline{s}_1) + \cdots + x_1 a(s_n, \underline{s}_1) f(s_n, \underline{s}_1) &= x_1 \cdot n/m \\
&\vdots \\
x_n a(s_1, \underline{s}_n) f(s_1, \underline{s}_n) + \cdots + x_n a(s_n, \underline{s}_n) f(s_n, \underline{s}_n) &= x_n \cdot n/m
\end{aligned}
$$

Table 6.2: Integer programming using the Earth Mover's Distance in the objective function.

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} d(s_i, \underline{s}_j) f(s_i, \underline{s}_j) \\
\text{subject to} \quad & \\
f(s_1, \underline{s}_1) + \cdots + f(s_1, \underline{s}_n) &= 1 \\
&\vdots \\
f(s_n, \underline{s}_1) + \cdots + f(s_n, \underline{s}_n) &= 1 \\
f(s_1, \underline{s}_1) + \cdots + f(s_n, \underline{s}_1) &\leq n/m \\
&\vdots \\
f(s_1, \underline{s}_n) + \cdots + f(s_n, \underline{s}_n) &\leq n/m
\end{aligned}
$$

Table 6.3: LP relaxation using the Earth Mover's Distance in the objective function.

that $x_j \leq 1$ in the equalities $\ldots = x_i \cdot n/m$, so we may drop those x_j and replace the equality with $\ldots \leq n/m$. We can almost recover the original x_j values by noting that

$$x_j f(s_1, \underline{s}_1) + \cdots + x_j f(s_n, \underline{s}_1) = x_j \left(f(s_1, \underline{s}_1) + \cdots + f(s_n, \underline{s}_1) \right) \leq n/m$$

and so

$$x_j \leq \frac{n/m}{f(s_1, \underline{s}_1) + \cdots + f(s_n, \underline{s}_1)} \, .$$

In fact, however, we will not need to recover the x_j at all, for the flow will provide the information we need. (More formally, we replace the x_j and the flow matrix $F = (f(s_i, \underline{s}_j))$ with a new flow matrix $\tilde{F} = \tilde{f}(s_i, \underline{s}_j)) = (x_j f(s_i, \underline{s}_j))$. For simplicity, we will not write the tilde's in the following.) Note, too, that the flow constraints relieve us of having to sum the x_j's, since the flow from S must match the flow to S'. When m of the inequalities have been changed to equality, all the flow will have been consumed at S' and the other flows must necessarily become 0. The LP relaxation is summarized in Table 6.3. The algorithm corresponding to Algorithm EMD-LP is the following.

EMD-LP(S)

▷ Using the LP relaxation of the IP formulation to find a good set S' when similarity is EMD.

$S' \leftarrow \emptyset$

Pick some s_i and put it in S'. (Can this be done better than randomly?)

Set the corresponding inequality to equality.

while $|S'| < m$

  do

    Solve LP problem from Table 6.3 with appropriate $\ldots \leq n/m$ replaced with $\ldots = n/m$.

    Select $\underline{s}_j$ that has maximum $\sum_{i=1}^{n} f(s_i, \underline{s}_j)$.

    Set the corresponding inequality to equality.

    Add corresponding s_i to S'.

Sadly, as noted at the beginning of this section, algorithm EMD-LP does not provide useful results in its current form. It suggests a modification, however, in the form of EMD-Cluster in Section 6.5.2.

6.5.2 Greedy Nearest Neighbor Search

Suppose we have a set S of N points distributed uniformly and independently at random in $[0,1]^d$ and that we select a subset $T \subset S$ of N/k of them at random for some fixed positive integer $k \ll N$. We know from Ajtai *et al.* [4] and our own extensions (cf. Chapter 3) that the optimal matching has expected weight $O(\sqrt{n})$ for $d = 1$, $O(\sqrt{n \log n})$ for $d = 2$, and $O(n^{(d-1)/d})$ for $d \geq 3$. Can we do at least as well deterministically?

Note that weighted EMD as a matching algorithm is an unassisted fractional clustering algorithm: each point in T matches to k points in S. Consider, then, the following algorithm:

```
EMD-Cluster(S)
    ▷ Pick a subset of N/k points from N so as to minimize the EMD.
T ← ∅
U ← S
while |U| ≥ k
      do
          Choose u ∈ U such that kNN-Weight(u, U, k) is minimal.
          T ← T ∪ NN-Set(u, U, k)
          U ← U − kNN-Set(u, U, k)
return T
```

EMD-Cluster uses the helper function kNN-Weight and kNN-Set, which find the k nearest neighbors to u in U and return respectively either their total edge weight (the sum of the distances) or else the set itself.

An open question is the following:

Conjecture 6.28. *Algorithm* EMD-cluster *has expected performance better than random subset selection.*

In the unit interval, the proof appears approachable by a sieve argument, although the details remain elusive. In higher dimensions, however, the conjecture appears to be related to optimal sphere packing (or, in L_1, optimal axis-aligned cube packing). As such, an approach motivated by bin packing may prove fruitful.

6.6 Conclusions

In this section we considered two subset selection problems. The first of these we are able to solve: choosing a subset of points whose centroid is close to the centroid of the original set. Our solution achieves polynomial time by reducing the problem to a combinatorially simpler problem, greedily choosing pairs of points. The second of the problems, choosing a subset whose Earth Mover's Distance from the original set is small, remains an open problem. We present an algorithm that experiment suggests is promising, but we can not offer a proof of its correctness. In prelude to this, we paused at the beginning to show that the seemingly obvious assertion that random selection has a good chance to provide a good subset is, in fact, justified.

7. Conclusions

In this dissertation we have explored two problems. First, we considered the question of the asymptotic weight of random matchings. In the unweighted case, we introduced new machinery that replicates previously known results, providing relatively straight-forward proofs for cases whose proofs have apparently escaped the literature. The new machinery, however, allowed us to extend our results to weighted matchings and to random matchings on a number of exotic sets. We proposed an algorithm applying these results to shape matching.

In the second problem we considered point set simplification. We showed a polynomial time simplification algorithm that preserved the centroid of the set. Tantalizingly, as future work, we offer an algorithm that appears to preserve the EMD.

Appendix A. GNU Free Documentation License

This appendix is present because at least one image is licensed under the GFDL, which requires that a copy of the license be included with this work.

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.

51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

A.1 Applicability and Definitions

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters.

A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

A.2 VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever

to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

A.3 COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

A.4 Modifications

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title,

year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously

added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

A.5 Combining Documents

You may combine the Document with other documents released under this License, under the terms defined in section A.4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

A.6 Collections of Documents

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this

License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

A.7 Aggregation with Independent Works

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A.3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

A.8 Translation

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A.4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section A.4) to Preserve its Title (section A.1) will typically require changing the actual title.

A.9 Termination

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

A.10 Future Revisions of this License

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

A.11 How to use this License for your Documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Bibliography

[1] Najma Ahmad. *The Geometry of Shape Recognition via the Monge-Kantorovich Optimal Transport Problem.* PhD thesis, Brown University, Providence, Rhode Island, May 2004.

[2] Oswin Aichholzer, Helmut Alt, and Günter Rote. Matching shapes with a reference point. *International Journal of Computational Geometry and Applications*, 1994.

[3] Martin Aigner and Günter M. Ziegler. *Proofs from THE BOOK.* Springer, 3 edition, 2004.

[4] Ajtai, Komlos, and Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.

[5] V. S. Alagar. On the distribution of a random triangle. *Journal of Applied Probability*, 14:284–297, 1977.

[6] H. Alt, O. Aichholzer, and G. Rote. Matching shapes with a reference point. *International Journal of Computational Geometry and its Applications*, 7:349–363, 1997.

[7] H. Alt, J. Blömer, M. Godau, and H. Wagener. Approximations of convex polygons. In *Proceedings 17th ICALP*, number 443 in LNCS, pages 703–716, 1990.

[8] H. Alt, K. Mehlhorn, H. Wagener, and E. Welzl. Congruence, similarity and symmetries of geometric objects. *Discrete and Computational Geometry*, 3:237–256, 1988.

[9] Helmut Alt, Oswin Aichholzer, and Günter Rote. Matching shapes with a reference point. In *SCG '94: Proceedings of the tenth annual symposium on Computational geometry*, pages 85–92, New York, NY, USA, 1994. ACM Press.

[10] Helmut Alt, Bernd Behrends, and Johannes Blömer. Approximate matching of polygonal shapes (extended abstract). In *SCG '91: Proceedings of the seventh annual symposium on Computational geometry*, pages 186–193, New York, NY, USA, 1991. ACM Press.

[11] Helmut Alt, Ulrich Fuchs, Günter Rote, and Gerald Weber. Matching cnovex shapes with respect to the symmetric difference. Technical Report 87(1c), Karl-Franzens-Universität Graz & Technische Universität Graz, Mar 1997.

[12] R. V. Ambartzumian, editor. *Stochastic and Integral Geometry.* Reidel, Dordrecht, Netherlands, 1987.

[13] Luigi Ambrosio. Optimal transport maps in Monge-Kantorovich problem. *BEIJING*, 3:131, 2002.

[14] Amihood Amir and Emanuel Dar. An improved deterministic algorithms for generalized random sampling. In *CIAC 1997: Proceedings of the Third Italian Conference on Algorithms and Complexity*, pages 159–170, London, UK, 1997. Springer-Verlag.

[15] Sigurd Angenent, Steven Haker, and Allen Tannenbaum. Minimzing flows for the monge-kantorovich problem. *Siam Journal of Mathematical Analysis*, 35(1):61–97, 2003. Society for Industrial and Applied Mathematics.

[16] Aaron Archer, Jittat Fakcharoenphol, Chris Harrelson, Robert Krauthgamer, Kunal Talwar, and Éva Tardos. Approximate classification via earthmover metrics. In *SODA 2004: Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete algorithms*, pages 1079–1087, Philadelphia, PA, USA, 2004. Society for Industrial and Applied Mathematics.

[17] Ira Assent, Andrea Wenning, and Thomas Seidl. Approximation techniques for indexing the earth mover's distance in multimedia databases. In *ICDE 2006: 22nd International Conference on Data Engineering*, page 11, 2006.

[18] M. J. Atallah. A matching problem in the plane. *Journal of Comput. Systems Sci.*, 31(1):63–70, Aug 1985.

[19] M. D. Atkinson. An optimal algorithm for geometric congruence. *Journal of Algorithms*, 8(2):159–172, 1987.

[20] Amitava Bagchi and Edward M. Reingold. A naturally occurring function continuous only at irrationals. *The American Mathematical Monthly*, 89(6):411–417, Jun–Jul 1982.

[21] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. *focs*, 00:184, 1996.

[22] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *STOC 1998: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 161–168, New York, NY, USA, 1998. ACM Press.

[23] Jon Louis Bentley, Daniel D. Sleator, Robert E. Tarjan, and Victor K. Wei. A locally adaptive data compression scheme. *Communications of the ACM*, 29(4):320–330, 1986.

[24] Patrick Bernard and Boris Buffoni. Weak KAM pairs and Monge-Kantorovich duality, 2005.

[25] Patrick Bernard and Boris Buffoni. Optimal mass transportation and mather theory, Jan 2007.

[26] W. Blaschke. Über affine Geometrie XI: Lösung des "Vvierpunkproblems" von Sylvester aus der Theorie der geometrischen Wwahrscheinlichkeiten. *Lepziger Berichte*, 69:436–453, 1917.

[27] N. Blum. A new approach to maximum matching in general graphs. In *Proceedings of the 17th ICALP*, volume 443 of *Lecture Notes in Computer Science*, pages 586–597. Springer Verlag, 1990.

[28] François Bolley and Cédric Villani. Weighted Csiszàr-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse Mathematiques*, 14(3):331–352, 2005.

[29] Abraham Bookstein, Shmuel T. Klein, and Timo Raita. Is huffman coding dead? (extended abstract). In *SIGIR 1993: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 80–87, New York, NY, USA, 1993. ACM Press.

[30] O. Borůvka. O jistém problému minimálním. *Práce Moravské Přirodovědecké Společnosti*, 3:37–58, 1926. (In Czech).

[31] Guy Bouchitté and Giuseppe Buttazzo. Characterization of optimal shapes and masses through Monge-Kantorovich equation. 2000.

[32] C. Buchta. Über die konvexe Hülle von Zufallspunkten in Eibereichen. *Elem. Math.*, 38:153–156, 1983.

[33] C. Buchta. Zufallspolygone in konvexen Vielecken. *Journal für die reine und angewandte Mathematik*, 347:212–220, 1984.

[34] Christian Buchta. An identiy relating moments of functionals of convex hulls. *Discrete and Computational Geometry*, 33:125–142, 2005.

[35] G. Buttazzo and E. Stepanov. Transport density in Monge-Kantorovich problems with Dirichlet conditions. *Discrete Contin. Dyn. Syst.*, 12(4):607–628, 2005.

[36] John W. Byers, Michael Luby, Michael Mitzenmacher, and Ashutosh Rege. A digital fountain approach to reliable distribution of bulk data. In *SIGCOMM 1998: Proceedings of the ACM SIGCOMM 1998 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 56–67, New York, NY, USA, 1998. ACM Press.

[37] S. Cabello, P. Giannopoulos, C. Knauer, and G. Rote. Matching point sets with respect to the earth mover's distance. In *Proceedings of the European Symposium on Algorithms*, 2005.

[38] R. D. Carmichael. Average and probability. *The American Mathematical Monthly*, 13(11):216–217, Nov 1906.

[39] G. D. Chakerian. A characterization of curves of constant width. *The American Mathematical Monthly*, 81(2):153–155, Feb 1974.

[40] Chern-Ching Chao, Hsien-Kuei Hwang, and Wen-Qi Liang. Expected measure of the union of random rectangles. *Journal of Applied Probability*, 35(2):495–500, Jun 1998.

[41] Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions (extended abstract). In *STOC 1994: Proceedings of the 26th annual ACM symposium on Theory of Computing*, pages 584–592, New York, NY, USA, 1994. ACM Press.

[42] Sourav Chatterjee, Ron Peled, Yuval Peres, and Dan Romik. Gravitational allocation to poisson points. *arXiv*, 2006.

[43] Bernard Chazelle. A minimum spanning tree algorithm with inverse-Ackermann type complexity. *Journal of the ACM*, 47(6):1028–1047, 2000.

[44] Bernard Chazelle and Joel Friedman. A deterministic view of random sampling and its uses in geometry. *Combinatorica*, 10(3):229–249, Sept 1990.

[45] Bernard Chazelle and Joel Friedman. A deterministic view of random sampling and its uses in geometry. In *FOCS 1988: 29th Symposiom on the Foundations of Computer Science*, pages 539–549, 1998.

[46] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. Approximating the minimum spanning tree weight in sublinear time. In F. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *ICALP 2001, LNCS 2076*, pages 190–200, Springer-Verlag Berlin, 2001.

[47] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. Approximating the minimum spanning tree weight in sublinear time. *SIAM Journal of Computing*, 34(6):1370–1379, 2005.

[48] D. Cheriton and R. E. Tarjan. Finding minimum spanning trees. *SIAM Journal of Computing*, 5:724–742, 1976.

[49] Vašek Chvàtal. *Linear Programming*. W. H. Freeman and Company, 1983.

[50] E. G. Coffman, George S. Lueker, Joel Spencer, and Peter M. Winkler. Packing random rectangles. (99-44), 2, 1999.

[51] Scott Cohen. *Finding Color and Shape Patterns in Images*. PhD thesis, 1999.

[52] Scott D. Cohen and Leonidas J. Guibas. The earth mover's distance under transformation sets. In *ICCV (2)*, pages 1076–1083, 1999.

[53] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, second edition, 2001.

[54] T. Cover and J. Thomas. *Elements of Information Theory: Rate Distortion Theory*. John Wiley & Sons, 1991.

[55] G. Crasta and A. Malusa. On a system of partial differential equations of Monge-Kantorovich type. *ArXiv Mathematics e-prints*, December 2006.

[56] M. W. Crofton. On the theory of probability, applied to random straight lines. *Proceedings of the Royal Society of London*, 16:266–269, 1867–1868.

[57] M. W. Crofton. *Encyclopedia Britannica*, chapter Probability, pages 768–788. , 9th edition edition, 1885.

[58] Morgan W. Crofton. On the theory of local probability, applied to straight lines drawn at random in a plane; the methods used being also extended to the proof of certain new theorems in the integral calculus. *Philosophical Transactions of the Royal Society of London*, 158:181–199, 1868.

[59] M. de Berg, O. Devilliers, M. van Kreveld, O. Schwarzkopf, and M. Teillaud. Computing the maximum overlap of two convex polygons under translations. In *Proceedings of the 7th International Symposium on Algorithms and Computation (ISAAC 1996)*, volume 1178 of *Lecture Notes in Computer Science*, pages 126–135. Springer Verlag, 1996.

[60] Duane W. DeTemple and Jack M. Robertson. Constructing buffon curves from their distributions. *The American Mathematical Monthly*, 87(10):779–784, Dec 1980.

[61] Steven R. Dunbar. The average distance between points in geometric figures. *The College Mathematics Journal*, 28(3):187–197, May 1997.

[62] Alon Efrat, Piotr Indyk, and Suresh Venkatasubramanian. Pattern matching for sets of segments. In *SODA 2001: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 295–304, Philadelphia, PA, USA, 2001. Society for Industrial and Applied Mathematics.

[63] B. Efron. The convex hull of a random set of points. *Biometrika*, 52:331–343, 1965.

[64] Bennett Eisenberg and Rosemary Sullivan. Random triangles in n dimensions. *The American Mathematical Monthly*, 103(4):308–318, Apr 1996.

[65] Bennett Eisenberg and Rosemary Sullivan. Crofton's differential equation. *The American Mathematical Monthly*, 107(2):129–139, Feb 2000.

[66] Bennett Eisenberg and Rosemary Sullivan. The fundamental theorem of calculus in two dimensions. *The American Mathematical Monthly*, 109(9):806–817, Nov 2002.

[67] Paul Erdös. On sets of distances of n points. *The American Mathematical Monthly*, 53(5):248–250, 1946.

[68] Lawrence Craig Evans. Partial differential equations and monge-kantorovich mass transer. An earlier version appeared in Current Developments in Mathematics, 1997, edited by S. T. Yau, Sept 2001.

[69] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Veličkovic. Approximations of general independent distributions. In *STOC 1992: Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 10–16, New York, NY, USA, 1992. ACM Press.

[70] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. 2003.

[71] Robert M. Fano. *Transmission of Information: A Statistical Theory of Communication*. MIT Press, 1961. (also seen cited as appearing in 1949).

[72] Mikhail Feldman. Growth of a sandpile around an obstacle. *Contemporary Mathematics*, 226:55–78, 1999. American Mathematical Society.

[73] William Feller. A direct proof of Stirling's formula. *The American Mathematical Monthly*, 74(10):1223–1225, Dec 1967.

[74] William Feller. *An Introduction to Probability Theory and Its Applications*, volume I, 3rd Edition. John Wiley & Sons, New York, 1968.

[75] S. R. Finch. *Mathematical Constants*, chapter 8.1: Geometric Probability Constants, pages 479–484. Cambridge University Press, Cambridge, England, 2003.

[76] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM*, 34:596–615, 1987.

[77] M. L. Fredman and D. E. Willard. Trans-dichotomous algorithms for minimum spanning trees and shortest paths. *Journal of Computer and System Sciences*, 48(3):533–551, 1994.

[78] H. N. Gabow, Z. Galil, T. Spencer, and R. E. Tarjan. Efficient algorithms for finding minimum spanning trees in undirected and directed graphs. *Combinatorica*, 6:109–122, 1986.

[79] Wilfrid Gangbo. The monge mass transfer problem and its applications. In *NSF-CBMS Conference on Contemporary Mathematics*, volume 226, Dec 1997.

[80] Wilfrid Gangbo. An introduction to the mass transportation theory and its application. Gangbo's notes from lectures given at the 2004 Summer Institute at Carnegie Mellon University and at IMA in March 2005. `gangbo@math.gatech.edu`, Mar 2005.

[81] Bernd Gärtner. Pitfalls in computing with pseudorandom determinants. In *SCG '00: Proceedings of the Sixteenth Annual Symposium on Computational Geometry*, pages 148–155, New York, NY, USA, 2000. ACM Press.

[82] Evarist Giné. Empirical processes and applications: An overview. *Bernoulli*, 2(1):1–28, Mar 1996.

[83] M. Godau. A natural metric for curves—computing the distance for polygonal chains and approximation algorithms. In *Proceedings 8th STACS*, volume 480 of *LNCS*, pages 127–136 (?), 1991.

[84] A. Goshtasby and G. C. Stockman. Point pattern matching using convex hull edges. *IEEE Transactions Syst. Man Cyber*, SMC-15(5):631–636, 1985.

[85] R. L. Graham and P. Hell. On the history of the minimum spanning tree problem. *Annals of the History of Computing*, 7:43–57, 1985.

[86] Luca Granieri. On action minimizing measures for the monge-kantorovich problem, 2005.

[87] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford Science Publications. Clarendon Press, Oxford, second edition, 1992.

[88] H. Groemer. On some mean values associated with a randomly selected simplex in a convex set. *Pacific Journal of Mathematics*, 45:525–533, 1973.

[89] Alain Guénoche, Bruno Leclerc, and Vladimir Makarenkov. On the extension of a partial metric to a tree metric. *Discrete Matehmatics*, 276(1–3):229–248, 2000.

[90] Anupam Gupta. Embedding tree metrics into low dimensional Euclidean spaces. pages 694–700, 1999.

[91] Rajiv Gupta, Scott A. Smolka, and Shaji Bhaskar. On randomization in sequential and distributed algorithms. *ACM Computing Surveys*, 26(1):7–86, 1994.

[92] A. Hajnal and E. Szemerédi. Combinatorial theory and its applications. In P. Erdös, A. Rényi, and V. T. Sós, editors, *Proof of a Conjecture of P. Erdös*, volume II, pages 601–623. North-Holland, London, 1970.

[93] Steven Haker, Allen Tannenbaum, and Ron Kikinis. Mass preserving mappings and image registration. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention: MICCAI 2001, Utrecht, The Netherlands*, volume 2208 of *Lecture Notes in Computer Science*, pages 120–127, Oct, 2001. Springer Berlin / Heidelberg.

[94] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[95] Sariel Har-Peled. A replacement for voronoi diagrams of near linear size. In *Proceedings of the 42nd Annual IEEE Symposium on the Foundations of Computer Science*, pages 94–103, 2001.

[96] S. Hart and M. Sharir. Nonlinearity of davenport-schinzel sequences and of generalized path compression schemes. *Combinatorica*, 6(2):151–177, 1986.

[97] Paul J. Heffernan and Stefan Schirra. Approximate decision algorithms for point set congruence. In *SCG '92: Proceedings of the eighth annual symposium on Computational geometry*, pages 93–101, New York, NY, USA, 1992. ACM Press.

[98] N. Henze. Random triangles in convex regions. *Journal of Applied Probability*, 20:111–125, 1983.

[99] J. E. Hopcroft and R. M. Karp. An $n^{5/2}$ algorithm for maximum matching in bipartite graphs. *SIAM Journal Comput*, 2:225–231, 1973.

[100] D. Huffman. A method for the construction of minimum redundancy codes. In *Proceedings of the IRE*, volume 40, pages 1098–1101, 1952.

[101] K. Imai, S. Sumino, and H. Imai. Minimax geometric fitting of two corresponding sets of points. *Proceedings of the ACM Symposium on Computational Geometry*, pages 266–275, 1989.

[102] Piotr Indyk. A sublinear time approximation scheme for clustering in metric spaces. In *IEEE Symposium on Foundations of Computer Science*, pages 154–159, 1999.

[103] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *STOC 1998: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 604–613, New York, NY, USA, 1998. ACM Press.

[104] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. pages 604–613, 1998.

[105] Piotr Indyk, Rajeev Motwani, and Suresh Venkatasubramanian. Geometric matching under noise: Combinatorial bounds and algorithms. In *SODA 1999: Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 457–465, Philadelphia, PA, USA, 1999. Society for Industrial and Applied Mathematics.

[106] L. Kantorovich. On the translocation of masses. *C. R. (Doklady) Academy of Sciences URSS (N.S.)*, 37:199–201, 1942. As cited in [80].

[107] L. Kantorovich. On a problem of monge. *Uspekhi Math. Nauk.*, 3:225–226, 1948. In Russian. As cited in [80].

[108] D. R. Karger, P. N. Klein, and R. E. Tarjan. A randomized linear time algorithm to find minimum spanning trees. *Journal of the ACM*, 42:321–328, 1995.

[109] Richard M. Karp, Michael Luby, and A. Marchetti-Spaccamela. A probabilistic analysis of multidimensional bin packing problems. In *STOC 1984: Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, pages 289–298, New York, NY, USA, 1984. ACM Press.

[110] M. G. Kendall and P. A. P. Moran. *Geometrical Probability*, volume 10 of *Griffin's Statistical Monographs & Courses*. Hafner Publishing Company, New York, 1963.

[111] David Kinderlehrer. Diffusion mediated transport and the brownian motor.

[112] David Kinderlehrer and Adrian Tudorascu. Transport via mass transportation. *Discrete AND Continuous Dynamical Systems Series B*, 6(2):311–338, 2006.

[113] J. F. C. Kingman. Random secants of a convex body. *Journal of Applied Probability*, 6:660–672, 1969.

[114] Daniel A. Klain and Gian-Carlo Rota. *Introduction to Geometric Probability*. Cambridge University Press, 1997.

[115] Victor Klee. What is the expected volume of a simplex whose vertices are chosen at random from a given convex body? *The American Mathematical Monthly*, 76(3):286–288, Mar 1969.

[116] Oliver Klein and Remco C. Veltkamp. Approximation algorithms for the earth mover's distance under transformations using reference points. Technical Report UU-CS-2005-003, Institute of Information and Computing Sciences, `www.cs.uu.nl`, 2005.

[117] Eric Langford. The probability that a random triangle is obtuse. *Biometrika*, 56(3):689–690, Dec 1969.

[118] Tom Leighton and Peter Shor. Tight bounds for minimax grid matching, with applications to the average case analysis of algorithms. In *STOC 1986: Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 91–103, New York, NY, USA, 1986. ACM Press.

[119] Debra A. Lelewer and Daniel S. Hirschberg. Data compression. *ACM Computing Surveys*, 19(3):261–296, 1987.

[120] Elizaveta Levina and Peter Bickel. The earth mover's distance is the mallows distance: Some insights from statistics.

[121] Nathan Linial, Avner Magen, and Michael E. Saks. Trees and Euclidean metrics. pages 169–175, 1998.

[122] H. Liu and H. Motoda. Feature transformation and subset selection. *IEEE Intelligent Systems*, 13(2):26–28, 1998.

[123] David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[124] Peter Markowich. Chapter 8: Optimal transportation and monge-ampere equations. `http://homepage.univie.ac.at/peter.markowich/download/monge.pdf`, Mar 2006.

[125] Bruno Massé. On the LLN for the number of vertices of a random convex hull. *Advances in Applied Probability*, 32:675–681, 2000.

[126] S. Micali and V. V. Vazirani. An $O(\sqrt{|V|} \cdot |E|$ algorithm for finding maximum matchings in general graphs. *21st FOCS (IEEE Symposium on Foundations of Computer Science)*, pages 12–27, 1980.

[127] E. Michael. Continuous selections i. *Annals of Mathematics*, 63(2):361–382, Mar 1956.

[128] E. Michael. Continuous selections ii. *Annals of Mathematics*, 64(3):562–580, Nov 1956.

[129] E. Michael and C. Pixley. A unified theorem on continuous selections. *Pacific Journal of Mathematics*, 1:187–188, 1980.

[130] Philippe Michel, Stéphane Mischler, and Benoît Perthame. General relative entropy inequality: an illustration on growth models. *Journal des Mathematiques Pures et Appliquéees*, 84:1235–1260, 2005.

[131] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Academie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.

[132] J. M. Morel and F. Santambrogio. Comparison of distances between measures. *Appl. Math. Lett.*, 2006.

[133] J. Nešetřil. A few remarks on the history of MST-problem. *Archivum Mathematicum (Brno)*, 33:15–22, 1997. (Preliminary version in KAM Series, Charles University, Prague, 97-338, 1997.).

[134] Noam Nisan. Extracting randomness: How and why—a survey. In *11th Annual IEEE Conference on Computational Complexity (CCC 1996)*, page 44, 1996.

[135] H. Ogawa. Labeled point pattern matching by fuzzy relaxation. *Pattern Recognition*, 17(5):569–574, 1984.

[136] H. Ogawa. Labeled point pattern matching by delaunay triangulation and maximal cliques. *Pattern Recognition*, 19(1):35–40, 1986.

[137] Oystein Ore. Note on hamilton circuits. *The American Mathematical Monthly*, 67(1):55, Jan 1960.

[138] J. B. Orlin. A faster strongly polynomial minimum cost flow algorithm. *Operations Research*, 41(2):338–350, mar–apr 1993.

[139] Richard E. Pfiefer. The historical development of J. J. Sylvester's four point problem. *Mathematics Magazine*, 62(5):309–317, Dec 1989.

[140] Leonid Prigozhin. Solutions to Monge-Kantorovich Equations as Stationary Points of a Dynamical System. *ArXiv Mathematics e-prints*, July 2005.

[141] K. Przeslawski. Linear and lipschitz-continuous selectors for the family of convex sets in euclidean vector spaces. *Bull. Pol. Acad. Sci.*, 33:31–33, 1985.

[142] K. Przeslawski and D. Yost. Continuity properties of selectors and michael's theorem. *Michigan Mathematics Journal*, 36(1):113–134, 1989.

[143] Krzysztof Przeslawski. Lipschitz continuous selectors, part i: Linear selectors. *Journal of Convex Analysis*, 5(2):249–267, 1998.

[144] S. T. Rachev. The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and its Applications*, 29(4):647–676, 1984.

[145] W. J. Reed. Random points in a simplex. *Pacific Journal of Mathematics*, 54(2):183–198, 1974.

[146] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. In *Proceddings of the 41st FOCS*, pages 22–31, 2000.

[147] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. *SIAM Journal on Computing*, 35(5):1185–1209, 2006.

[148] Wansoo T. Rhee and Michel Talagrand. Exact bounds for the stochastic upward matching problem. *Transactions of the American Mathematical Society*, 307(1):109–125, May 1988.

[149] WanSoo T. Rhee and Michel Talagrand. Matching random subsets of the cube with a tight control on one coordinate. *The Annals of Applied Probability*, 2(3):695–713, Aug 1992.

[150] Herbert Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, Jan 1955.

[151] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–122, 2000.

[152] D. Rutovitz. Some parameters associated with a finite dimensional banach space. *J. London Math. Soc.*, 2:241–255, 1965.

[153] Michael Saks. Randomization and derandomization in space-bounded computation. In *SCT: Annual Conference on Structure in Complexity Theory*, 1996.

[154] Hans Samelson. Differential forms, the early days; or the stories of deahna's theorem and of volterra's theorem. *The American Mathematical Monthly*, 108(6):522–530, Jun-Jul 2001.

[155] L. A. Santaló. *Integral Geometry and Geometric Probability.* Addison-Wesley, Reading, MA, 1976.

[156] Walter Schachermayer and Josef Teichmann. Solution of a problem in Villani's book. FAM-Seminar, TU Wien, Austria, 15 Dec 2005.

[157] S. Schirra. *Über die Bitkomplexität der ϵ-Kongruenz.* PhD thesis, Fachbereich Informatik, Universität des Saarlandes, 1988.

[158] Z. F. Seidov. Letters: Random triangle. *Mathematica Journal*, 7:414, 2000.

[159] Zakir F. Seidov. Random triangle problem: Geometrical approach, 2000.

[160] David F. Shanno and Roman L. Weil. 'linear' programming with abolute-value functionals. *Operations Research*, 19(1):120–124, Jan-Feb 1971.

[161] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, 1949.

[162] Claude E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656, July,October 1948.

[163] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, 1963.

[164] Peter W. Shor. The average-case analysis of some on-line algorithms for bin packing. *FOCS 1984: 25th Annual Symposium on Foundations of Computer Science*, pages 193–200, Oct 1986.

[165] Michael Sipser. *Introduction to the Theory of Computation*. Thomson, 2nd edition, 2006.

[166] N. J. A. Sloane. The on-line encyclopedia of integer sequences: A093072. `http://www.research.att.com/~njas/sequences/?q=a093072`.

[167] N. J. A. Sloane. The on-line encyclopedia of integer sequences: A093158. `http://www.research.att.com/~njas/sequences/?q=a093158`.

[168] N. J. A. Sloane. The on-line encyclopedia of integer sequences: A093159. `http://www.research.att.com/~njas/sequences/?q=a093159`.

[169] N. J. A. Sloane. The on-line encyclopedia of integer sequences: A103281. `http://www.research.att.com/~njas/sequences/?q=a103281`.

[170] `http://en.wikipedia.org/wiki/Sperner_family`.

[171] J. Sprinzak. title unknown. Master's thesis, Hebrew University, Department of Computer Science, 1990. Cited in arkin1991matching-points.

[172] Timo Stich and Marcus Magnor. Keyframe animation from video. In *ICP 2006: IEEE International Conference on Image Processing*, pages 2713–2716, 2006.

[173] J. J. Sylvester. . *The Educational Times*, Apr 1864.

[174] Endre Szemerédi and Jr. William T. Trotter. A combinatorial distinction between the euclidean and projective planes. *European Journal of Combinatorics*, 4:385–394, 1983.

[175] Endre Szemerédi and Jr. William T. Trotter. Extremal problems in discrete geometry. *Combinatorica*, 3(3–4):381–392, 1983.

[176] M. Talagrand. Matching theorems and empirical discrepancy computations using majorizing measures. *Journal of the American Mathematical Society*, 7(2):455–537, Apr 1994.

[177] M. Talagrand. The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *The Annals of Probability*, 22(2):919–959, Apr 1994.

[178] M. Talagrand and J. E. Yukich. The integrability of the square exponential transportation cost. *The Annals of Applied Probability*, 3(4):1100–1111, Nov 1993.

[179] Michel Talagrand. Matching random samples in many dimensions. *The Annals of Applied Probability*, 2(4):846–856, Nov 1992.

[180] Michel Talagrand. Majorizing measures: The generic chaining. *The Annals of Probability*, 24(3):1049–1103, Jul 1996.

[181] N. Tishby, F. Pereira, and W. Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377, 1999.

[182] Csaba D. Toth. The szemeredi-trotter theorem in the complex plane. *arXiv.org*, `http://arxiv.org/abs/math/0305283`, May 2003.

[183] M. Trott. The area of a random triangle. *Mathematica Journal*, 7(2):189–198, 1998. `://library.wolfram.com/infocenter/Articles/3413/`.

[184] M. Trott. *The Mathematica GuideBook for Symbolics*, chapter 1.10.1: Area of a Random Triangle in a Square, pages 298–311. Springer-Verlag, New York, 2006.

[185] Rainer Typke, Frans Wiering, and Remco C. Veltkamp. Evaluating the earth mover's distance for measuring symbolic melodic similarity. In *MIREX 2005*, 2005.

[186] R. C. Veltkamp. Shape matching: similarity measures and algorithms. pages 188–199, 2001.

[187] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Association, 2003.

[188] Heribert Vollmer. *Introduction to Circuit Complexity: A Uniform Approach*. Springer, Berlin, 1999.

[189] Jon A. Wellner. Empirical processes in action: A review. *International Statistical Review / Revue Internationale de Statistique*, 60(3):247–269, Dec 1992.

[190] Jon A. Wellner. [empirical processes and applications: An overview]: Discussion. *Bernoulli*, 2(1):29–37, 1996.

[191] Frans Wiering, Rainer Typke, and Remco C. Veltkamp. What transportation distances can do for music notation retrieval. *Computing in Musicology*, 13 (Music Query: Methods, Models, and User Studies):113–128, 2004.

[192] Ian H. Witten, Radford M. Neal, and John G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30(6):520–540, 1987.

[193] Gershon Wolansky. Mass transportation, the Monge-Kantorovich problem, Hamilton-Jacobi equation and metrics on measure-valued functions.

[194] Gershon Wolansky. Optimal transportation in the presence of a prescribed pressure field, 2003.

[195] Gershon Wolansky. Optimal transportation in the presence of a prescribed pressure field, 2003.

[196] W. S. B. Woolhouse. Question no. 2471. *Mathematical Questions and their Solutions from the Educational Times*, 8:100–105, 1867.

[197] W. S. B. Woolhouse. Some additional observations on the four-point problem. *Mathematical Questions and their Solutions from the Educational Times*, 7:81, 1867.

[198] A. Yao. An $O(|E| \log\log |V|)$ algorithm for finding minimum spanning trees. *Information Processing Letters*, 4:21–23, 1975.

[199] Damian Yerrick. Sierpinski carpet png. `http://en.wikipedia.org/wiki/Image:Sierpinski6.png`, 2002. Image is licensed under the GNU Free Documentation License, see Appendix A. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License." The specific use in the document modifies the image by scaling and changing file format.

[200] J. E. Yukich. Optimal matching and empirical measures. *Proceedings of the American Mathematical Society*, 107(4):1051–1059, Dec 1989.

[201] S. L. Zabell. Alan turing and the central limit theorem. *The American Mathematical Monthly*, 102(6):483–494, Jun-Jul 1995.

[202] V. A. Zalgaller and Sung Soo Kim. An application of crofton's theorem. *The American Mathematical Monthly*, 104(8):774, Oct 1997.

[203] Lei Zhu and Allen Tannenbaum. Image interpolation based on optimal mass preserving mappings. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2004*, volume 1, pages 21–24, 2004.

[204] Lei Zhu, yan Yang, Allen Tannenbaum, and Steven Haker. Image morphing based on mutual information and optimal mass transport. In *International Conference on Image Processing (ICIP 2004)*, volume 3, pages 1675–1678, Oct 2004.

[205] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.

[206] Jacob Ziv and Abraham Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.