that it is decreasing rightward; while the constraints on the dual function for the M_r' problem with the Manhattan metric are that the function has vertical slope between -1 and 1, has rightward slope between $-\infty$ and 1, and is 0 on the boundary of the square. Again, we show how to construct a dual function for one problem by using a dual function for the other on a fraction of the square, and adjusting it to obtain a dual function for the entire square. This procedure does not change the asymptotic behavior of the dual function, so it proves that $D_r = \Theta(D_r')$.

We first prove the constraints on the dual function for M_r. Recall in the problem M_r we are allowed to match points to the top and bottom edges, and that a $-$ point may only be matched to a $+$ point on its right. We minimize the sum of the vertical lengths of the edges. In the dual problem, we may consider a $-$ point and a $+$ point matched to the boundary to be matched to each other through the boundary. Thus, a $-$ point P may be matched to a $+$ point Q either directly or through the boundary. If the points matched directly, then the $+$ point must be to its right, and the weight of the edge is $|y_Q - y_P|$, where y_Q and y_P are the y-coordinates of Q and P. If the points are matched through the boundary, the weight of the edge is $d(P, \mathbf{B}) + d(Q, \mathbf{B})$, where $d(X, \mathbf{B})$ is the minimum distance from the point X to the top or bottom boundary of the square. Thus, for a $-$ point P and a $+$ point Q, $w(Q) - w(P) \le f(P, Q)$, where

$$f(P,Q) = \begin{cases} \min(d(P,\mathbf{B}) + d(Q,\mathbf{B}), |y_Q - y_P|), & P \text{ left of } Q \\ d(P,\mathbf{B}) + d(Q,\mathbf{B}), & P \text{ right of } Q. \end{cases}$$

This function satisfies the triangle inequality because it is the "length" of the shortest path from P to Q, if the length of a path is measured only by vertical distance, paths can only go to the right, and can go from any point on the top or bottom boundaries to any other point on these boundaries. Thus, by Lemma

1.5.1 we can find a function w' mapping the unit square into $\mathbf{R}$ satisfying

$$w'(X_2) - w'(X_1) \leq f(X_1, X_2).$$

We claim that a function w' satisfies this equation if and only if it is constant on the top and bottom boundaries, decreasing rightward, and has vertical slope with absolute value at most 1.

Suppose a function w' satisfies

$$w'(X_2) - w'(X_1) \leq f(X_1, X_2),$$

where $f(X_1, X_2)$ is the edge weight for the problem M_r given above. We will how that it satisfies the conditions listed above. If X_1 and X_2 have the same y-coordinates and X_1 is to the left of X_2, $f(X_1, X_2) = 0$, so $w'(X_1) \geq w'(X_2)$. Thus w' is decreasing rightward. If X_1 and X_2 have the same x-coordinates, then $w'(X_2) - w'(X_1) \leq |y_{X_2} - y_{X_1}| \leq d(X_1, X_2)$, so w' has slope at most 1 vertically. If X_1 and X_2 are on the top or bottom boundary of the square, $|w'(X_1) - w'(X_2)| \leq d(X_1, \mathbf{B}) + d(X_2, \mathbf{B}) = 0$, so $w'(X_1) = w'(X_2)$.

Now, suppose w' is decreasing rightward, is 0 on the top and bottom boundaries, and has vertical slope at most 1. We will show that $f(X_1, X_2) \geq w(X_2) - w(X_1)$. Let X_1 and X_2 be two points in the unit square. Since the vertical slope of w' is at most 1 and $w' = 0$ on the top and bottom boundaries, $d(X, \mathbf{B}) \geq |w'(X)|$ for any point X, so $d(X_1, \mathbf{B}) + d(X_2, \mathbf{B}) \geq w'(X_2) - w'(X_1)$. Now, suppose X_1 and X_2 are two points with X_2 to the right of X_1. Let X_3 be the point at the x-coordinate of P and the y-coordinate of Q. Since w' is decreasing rightward, $w'(X_1) \geq w'(X_3)$. Since w' has vertical slope at most 1, $|w'(X_2) - w'(X_3)| \leq |y_{X_2} - y_{X_3}|$. Thus,

$$\begin{aligned} w'(X_2) - w'(X_1) &\leq w'(X_2) - w'(X_3) \\ &\leq |y_{X_2} - y_{X_1}|. \end{aligned}$$

This shows that $w'(X_2) - w'(X_1) \leq f(X_1, X_2)$.

We now look at the rightward matching problem M_r'. We will use the Manhattan metric in analyzing this problem. This makes the analysis easier, and at worst, changes the edge lengths by a factor of 2. In M_r' problem, $-$'s can be matched to the top, bottom, and right sides of the square, while $+$'s can be matched to the top, bottom, and left sides. We will let $\mathbf{B}_r$ stand for the top, bottom and right sides of the square and $\mathbf{B}_l$ stand for the top, bottom and left sides. For this problem, the weight function on edges is

$$f(P,Q) = \begin{cases} d(P,\mathbf{B}_r) + d(Q,\mathbf{B}_l) & Q \text{ left of } P \\ \min(d(P,\mathbf{B}_r) + d(Q,\mathbf{B}_l), d_M(P,Q)) & \text{otherwise,} \end{cases}$$

where $d_M(P,Q) = |y_P - y_Q| + |x_P - x_Q|$ is the ditance in the Manhattan metric between P and Q. This function satisfies the triangle inequality since it is the "length" of the shortest path between P and Q, if the Manhattan metric is used to measure length and paths are allowed to leave the square on the top, bottom or right boundaries and re-enter in the top, bottom, or left boundaries for no cost. Thus, by Lemma 1.5.1 we can assume that the dual function w is defined on the whole unit square, and satisfies $w(X_2) - w(X_1) \leq f(X_1, X_2)$. We claim that the set of functions w satisfying $w(X_2) - w(X_1) \leq f(X_1, X_2)$ are exactly those w such that w has vertical slope at most 1 in absolute value, has horizontal slope less than 1 (i.e., between $-\infty$ and 1), and such that w is constant on the boundary of the square.

Suppose that w is a function satisfying $w(X_2) - w(X_1) \leq f(X_1, X_2)$, where f is the edge weight for the problem M_r' given above. Then w has slope 1 vertically since if X_1 and X_2, have the same x-coordinate,

$$w(X_2) - w(X_1) \leq d(X_1, X_2).$$

The slope of w rightward is at most 1, since if X_1 and X_2 have the same y-coordinate and X_2 is to the right of X_1, then $w(X_2) - w(X_1) \leq d(X_1, X_2)$. If

X_1 and X_2 are points on the top and bottom boundaries, then $w(X_1) = w(X_2)$. We can assume without loss of generality that this value on the top and bottom boundaries is 0. If X_1 is on the right boundary, and X_2 is on the top or bottom boundary, $w(X_2) - w(X_1) \leq 0$, so $w(X_1) \leq w(X_2) = 0$. Thus, points on the right boundary have negative values. Similarly, w is positive on the left boundary of the square. However, since the function w can decrease arbitrarily fast rightwards, we can change w so that the left and right boundaries have value 0, and still have w satisfy all the conditions.

We now assume that w is 0 on the boundaries of the square, has rightward slope at most 1, and vertical slope between -1 and 1. If point X_2 is to the right of X_1, then by the slope conditions, $w(X_2) - w(X_1) \leq d_M(X_2, X_1)$. Furthermore, for any points X_1 and X_2, $w(X_2) \leq d(X_2, B_l)$ and $-w(X_1) \leq d(X_1, B_r)$, so $w(X_2) - (X_1) \leq d(X_2, B_l) + d(X_1, B_r)$. This shows that the function w satisfies $w(X_2) - w(X_1) \leq f(X_1, X_2)$.

Theorem 1.5.3: $D_r = \Theta(D_r')$.

Proof: We show that, given a dual function for M_r on the middle ninth of the square, we can find a dual function for M_r' on the entire square, which has the same expected value. This shows that $D_r(n) = \Omega(D_r'(\frac{n}{9}))$, which implies $D_r = \Omega(D_r')$. Similarly, we show that given a dual function for M_r' on the middle ninth of the square, we can find a dual function for M_r on the entire square, implying $D_r = O(D_r')$.

Recall the requirements for a dual function for the problems M_r and M_r'. A dual function for M_r' must have slope rightward at most 1, vertical slope between ± 1, and be 0 on the boundary of the square. The dual function for M_r must be decreasing rightward, must have vertical slope between ± 1, and must be 0 on the top and bottom boundaries of the square.

Suppose we are given a dual function for M_r on the middle ninth of the square. It has slope at most 1 on the boundary of the middle ninth. We need to extend it to the entire square and make the boundary on the entire square 0. We can multiply w by $\frac{1}{2}$ so that its slope on the boundary of the middle ninth never exceeds $\frac{1}{2}$, and then we use the same argument as in Theorem 1.5.2.

Suppose we are given a dual function w' for M_r' on the middle ninth of the square. We need to make w' be decreasing rightwards. To do this, all we need to do is add the function $\frac{1}{2} - x$. Unfortunately, this function does not satisfy the requirement that it be 0 on the top and bottom boundaries of the square. However, we can let

$$w = \begin{cases} y(\frac{1}{2} - x), & y \leq \frac{1}{3} \\ \frac{1}{3}(\frac{1}{2} - x), & \frac{1}{3} \leq y \leq \frac{2}{3} \\ (1 - y)(\frac{1}{2} - x), & \frac{2}{3} \leq y. \end{cases}$$

This function will have all the desired properties, so will be a dual function for M_r. ∎

1.6. Variations

In all the problems presented so far, we have assumed that there were n $+$ points and n $-$ points and that they were distributed independently and uniformly in the unit square. For some applications we need different models for the distribution of the points. In this section we will give three possible variations in the way the points are chosen. We will show that in most cases, these variations make no difference in the asymptotic behavior.

In the first variation the difference is that instead of matching between two kinds of points each distributed at random (we will call this *two color matching*) we match from one kind of point distributed randomly to a fixed grid of n points in the unit square (we call this *grid matching*). Grid matching is more difficult than two color matching. Suppose that we can do grid matching for some problem. To do two color matching for this problem all we do is match each color to a grid, and then match points matched to the same grid point to each other. The expected sum of the lengths of the edges (or number of unmatched points, or maximum edge length) in two color matching is thus less than twice the expected sum for grid matching. We will prove the upper bound for maximum edge length and for average edge length matching in the case of grid matching, and the lower bound for average edge length and for up-right matching in the case of two color matching. Since grid matching is harder than two color matching, this proves the bounds both for two color and for grid matching.

In the next variation we will discuss there are no longer an equal number of $+$ and $-$ points. Instead of there being n $+$ points and n $-$ points, there are $2n$ points, each having an equal probability of being a $+$ or $-$ point. Since there are no longer an equal number of each kind of point, we cannot expect a perfect matching. We must permit points to be matched to the sides of the square. Thus, there will be an average of $\Theta(\sqrt{n})$ difference between the number of $+$ and of $-$ points.

We show that this variation does not change the asymptotic behavior of the expected value. In a matching, if k points are deleted, the sum of the edge lengths changes by at most k, since unmatched points can be matched to an edge with cost at most 1. Suppose we are given a matching with $(n+k)$ $+$ points and $(n-k)$ $-$ points. We choose k of the $+$ points at random and turn them into $-$ points. We now have a matching with n $+$ and n $-$ points. Furthermore, the distribution of points is the same as if they had been chosen independently and uniformly in the unit square. Since we changed k points, the sum of the edge lengths in the matching can change by at most k. The expected value of

$|k|$ is $O(\sqrt{n})$, where $n+k$ is the number of $+$ points. Thus, the expected sum of the edge lengths for our variant model the other is different by at most $O(\sqrt{n})$. Since this quantity is smaller than the expected sum of the edge lengths for all our problems, it does not change the expectation.

For high probability analysis, we have that with probability $1-\frac{1}{n^\alpha}$, the difference of the number of points is $O(\sqrt{n \log n})$ and with probability $1-n^{\sqrt{\log n}}$, the difference is $O(\sqrt{n}\log^{3/4} n)$. These probabilities are small enough so that the high probability results do not change in up-right matching or in the upper bound for average edge length matching. The proof of the lower bound for average edge length matching can be seen to apply to both models.

For maximum edge length the technique above does not apply. However, you can match both $+$ and $-$ of points to two grids with different numbers of points using edges of $O(n^{-1/2}\log^{3/4})$. We then match the two different kinds of grid points to each other, matching extra ones to the side of the square, with edges of length $O(n^{-1/2}\log^{3/4} n)$, assuming that the number of points in the two grids differs by $O(n^{1/2}\log^{3/4} n)$. Thus, this variation does not make any difference in the results for maximum edge length matching

The third variation is that the points are evenly spaced vertically, or horizontally, or both instead of being randomly distributed in the unit square. We can change one model into the other here by moving all the points a small distance. Suppose we have points randomly distributed in a unit square. We move all the points up or down until they are evenly spaced vertically. The average distance a point moves is on the average $\Theta(n^{-1/2})$. With probability $1-\frac{1}{n^\alpha}$ no point moves farther than $O(n^{-1/2}\log^{1/2} n)$ and with probability $1-n^{-\sqrt{\log n}}$ no point moves farther than $O(n^{-1/2}\log^{3/4} n)$. These probabilities give the results.

Chapter 2. Average Edge Length Matching

2.1. The Lower Bound

In this section, we will prove the lower bound for the average edge length in a matching. This proof is essentially the same as in [AKT], but it is somewhat simplified, so we do not have to appeal to the strong theorems about probability that they use. We also prove that the lower bound of $\Theta(\sqrt{n \log n})$ holds with probability $1 - 2^{n^\epsilon}$ for any $\epsilon < 1$, while they prove that the expected value is $\Theta(\sqrt{n \log n})$.

To obtain this bound, we need two lemmas. The first is essentially a lemma on martingales. It is stated as a theorem on binary trees, but it can be thought of as saying that if a martingale has a variance of at most 1 for each step, then after n steps, the variance is at most n. Thus, the probability of a value above $\alpha\sqrt{n}$ is at most $1/\alpha^2$. The second lemma says that if you flip k coins, then with constant probability, there is a $\Omega(\sqrt{n})$ excess of either heads or tails.

The first lemma can be thought of in terms of martingales by considering paths through the tree. Each path will have equal probability. At each vertex, the children will have weights a more and a less than the current vertex. Thus, at every step, there is an equal probability of adding or subtracting some value a, where a depends on the previous path you have taken through the tree.

Lemma 2.1.1: Suppose that for a binary tree T of depth d, every node v has some weight $w(v)$. Suppose that the root has weight 0 and that every non-leaf node v with weight $w(v)$ has two children with weights $w(v) + a$ and $w(v) - a$, where $0 \leq a \leq 1$. Then the weights of the leaves l_i satisfy

$$\frac{1}{2^d} \sum_{i=1}^{2^d} w(l_i)^2 \leq d.$$

Proof: We prove this lemma by induction on the height of the tree. Take a node v with weight $w(v)$. Then its children have weights $w(v) + \alpha$ and $w(v) - \alpha$ for some $\alpha \leq 1$. The sum the squares of these values is the contribution of these nodes to the sum. We have

$$(w(v) + \alpha)^2 + (w(v) - \alpha)^2 = 2(w(v)^2 + \alpha^2).$$

Let v_{ki} be the ith node on level k of the tree. Then, summing the above equation, we get

$$\frac{1}{2^{k+1}} \sum_{i=1}^{2^{k+1}} w(v_{k+1,i})^2 = \frac{1}{2^{k+1}} \sum_{i=1}^{2^k} (w(v_{ki}) + \alpha_{ki})^2 + (w(v_{ki}) - \alpha_{ki})^2 = \frac{1}{2^k} \sum_{i=1}^{2^k} (w_{ki}^2 + \alpha_{ki}^2).$$

From $\alpha_{ki} \leq 1$, we obtain $\frac{1}{2^k} \sum_i \alpha_{ki}^2 \leq 1$. Thus,

$$\frac{1}{2^{k+1}} \sum_{i=1}^{2^{k+1}} w(v_{k+1,i})^2 \leq 1 + \frac{1}{2^k} \sum_{i=1}^{2^k} w(v_{ki})^2.$$

Since the value at the root is 0, this proves the lemma by induction. ∎

We now prove a lemma saying that if you flip n coins, then with probability $\frac{3}{4}$, the difference between the number of heads and the expected number of heads, $n/2$, is $\Omega(\sqrt{n})$.

Lemma 2.1.2: If there are n independent events, each of which has probability $\frac{1}{2}$ of occurring, and X is the number which occur, then with probability $\frac{3}{4}$, $|X - n/2| \geq \sqrt{n}/8$.

Proof: The probability that $X = k$ for any given k is small. It is maximized for $k = \lceil \frac{n}{2} \rceil$, in which case it is

$$2^{-n} \binom{n}{\frac{n}{2}} \approx 2^{-n} \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\pi n \left(\frac{n}{2e}\right)^n} = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{n}}.$$

Thus, the probability that $|X - \frac{n}{2}| \geq \sqrt{n}/8$ is at most

$$\frac{\sqrt{n}}{4}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}} \leq \frac{1}{4},$$

since there are $\sqrt{n}/4$ numbers between $\frac{n}{2}$ and $\frac{\sqrt{n}}{8}$, and X has at most a $\sqrt{\frac{2}{\pi}}/\sqrt{n}$ chance of being each of these.

We now prove the lower bound on average edge length matching.

Theorem 2.1.3: Suppose n points are distributed independently and uniformly in the unit square, and each of these has an equal probability of being a $-$ point or a $+$ point. Then, with probability at least $1 - 2^{-n^\epsilon}$ for any $\epsilon < 1$, the minimum edge length matching between $-$ and $+$ points has sum of edge lengths $\Omega(\sqrt{n \log n})$, where points are also allowed to be matched to an edge of the square.

To prove the lower bound, we will produce a dual function w that lower bounds the weight of any matching. The function will have the following properties.

1. With high probability, $\sum w(P^+) - \sum w(P_-) = \Theta(\sqrt{n \log n})$.
2. The slope is less than 1. We accomplish this by making $\left|\frac{\partial f}{\partial x}\right| \leq \frac{1}{\sqrt{2}}$ and $\left|\frac{\partial f}{\partial y}\right| \leq \frac{1}{\sqrt{2}}$.
3. The function w is 0 on the boundary of the square.

To obtain this function w, we will produce the function in stages, adding a component at each stage. Each stage adds $\Theta(\sqrt{n}/\sqrt{\log n})$ to the sum $\sum w(P_+) - \sum w(P_-)$. At stage i, we divide the unit square into a grid of $2^{2i} \times 2^{2i}$ smaller squares S_{ij}, $1 \leq j \leq 2^{4i}$. The function we add will be defined on one of these squares by $s \cdot d(x, S_{ij})$, where $d(x, S_{ij})$ is the distance from the boundary of the square S_{ij} to the point x if $x \in S_{ij}$ and 0 if $x \notin S_{ij}$, and $s = 1/\sqrt{\log n}$.

In each stage, we divide every square S_{ij} into 16 smaller square $S_{i+1,j'}$, where $16j - 15 \leq j' \leq 16j$. These 16 smaller squares will be paired, to obtain 8 pairs

A	E	E	B
H	A	B	F
H	D	C	F
D	G	G	C

Figure 2.1.1: Dividing S_{ij} into 16 smaller squares $S_{i+1,j'}$.

of squares. We pair the two middle squares on each side and the two adjacent squares along each diagonal In Figure 2.1.1, squares given the same letter are paired. We call squares from pairs A, B, C or D ***diagonal squares*** and square from E, F, G or H ***edge squares***. Whatever we do to one square in a pair, we do the opposite to the other. That is, if for one square S_{ij} we add $s \cdot d(x, S_{ij})$, to the other square in its pair, $S_{ij'}$, we add $-s \cdot d(x, S_{ij'})$. If we add 0 to one square of a pair, we add 0 to both of them.

We pair the squares in this way because in both the square in any pair, the function is the same up to translation and adding a constant. Specifically, the slope of the function is the same in two squares in any pair. In an edge square, the slope has only one value. In a diagonal square, the two regions on each side of the diagonal may have different slopes. (See Figure 2.1.2.) We are thus producing a tree of slopes as in Lemma 2.1.1. Each of the two squares in a pair has the same slopes, and opposite slopes are added to each of them. By Lemma 2.1.1 this gives that on at most $\frac{1}{4}$ of the squares, the slope has ever exceeded 1 after $\Theta(\log n)$ stages. This is what we want, because this means that we can stop refining the function on any square where the slope would exceed

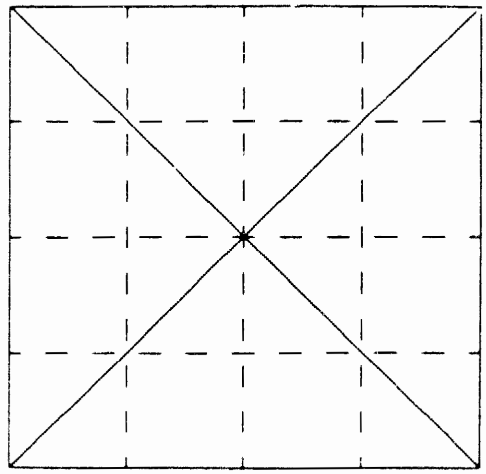

Figure 2.1.2: The slope is constant in each of the four regions.

1, and still be construction the function in $\frac{3}{4}$ of the small squares at every step. The expected number of points in a small square at the ith stage is $n/16^i$. Each of these points has an equal probability of being chosen a + or − point. Thus, by Lemma 2.1.2, each of these 16^i small squares has a probability of $\frac{3}{4}$ of contributing $\Omega(s4^{-i} \cdot \sqrt{n/16^i}) = \Omega(s16^{-i})$ to the value of the dual function. Since there are $m = 16^i$ of these small squares, the probability that more than $\frac{1}{2}$ of these small squares contribute less than this to the dual function is less than

$$\frac{m}{2}\left(\frac{3}{4}\right)^{m/2}\left(\frac{1}{4}\right)^{m/2}\binom{m}{\frac{1}{2}m} \leq m\left(\frac{3}{4}\right)^{m/2}.$$

Thus, the ith level adds, with probability $1-\alpha^{16^i}$, $\Omega(s\sqrt{n})$ to the dual function. Since there are $\Theta(\log n)$ levels, and $s = \Theta(1/\sqrt{\log n})$, this gives a lower bound of $\Omega(\sqrt{n}\sqrt{\log n})$. By starting on the ith level, where $16^i = n^\epsilon$ for $\epsilon < 1$, we obtain the bound of $\Omega(\sqrt{n \log n})$ with probability $1 - c^{-n^\epsilon}$ for some constant c. ∎

2.2. The Upper Bound

To show the upper bound for the average edge length problem, what we do is construct a matching with expected average edge length $\Theta(\sqrt{n \log n})$. We will

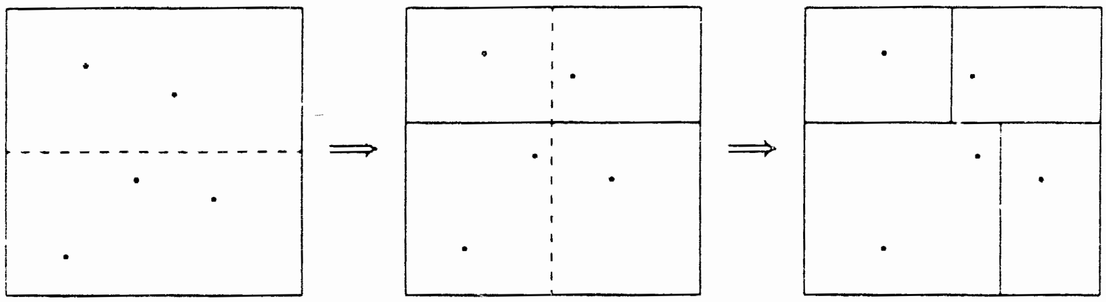

Figure 2.2.1: Dividing the square into rectangles.

do this by recursively subdividing the square. First, we divide the square in half horizontally and linearly transform each half so that it has area proportional to the number of points in it. (See Figure 2.2.1). We then subdivide each of these halves vertically and transform each to have area proportional to the number of points they contain (See Figure 2.2.1). We apply the same procedure recursively to each section, alternating horizontal and vertical division. After doing this $\Theta(\log n)$ times, each point is contained in a rectangle with area $1/n$. We then divide the square into n grid squares and use Hall's Theorem to match each point to a grid square which overlaps its rectangle.

There are two questions to be answered. First, how far did we move each point while constructing the rectangles. Second, when we match the rectangles to the squares, what is the diameter of the rectangles. If these two quantities both average $O(\sqrt{\log n}/\sqrt{n})$, then we are done, since we moved each point $O(\sqrt{\log n}/\sqrt{n})$ in the first stage, and in the second stage, every point is within diam(grid square) + diam(rectangle) of the corresponding grid point.

The diameter of the rectangles is determined entirely by their aspect ratios, since all the rectangles have the same area of $1/n$. We will show that with high probability all the rectangles have a bounded aspect ratio. In the early steps, with high probability, we change the aspect ration by very little. We must determine how much the aspect ratio changes in the last few steps.

We will show that the average distance a point is moved is $\Theta(\sqrt{\log n}/\sqrt{n})$. At each step, a point is moved an average of $1/\sqrt{n}$ in a random direction. This is a random walk, so the average distance moved after $\Theta(\log n)$ steps is $\Theta(\sqrt{\log n})$ times the average distance moved at each step, or $\Theta(\sqrt{\log n}/\sqrt{n})$, as desired.

There are several observations which will make the proof easier. The first observation is that we can stop with $\Theta(\log n)$ points in each rectangle, instead of continuing until there is exactly one point in each rectangle. The second observation is that at the ith stage, the regions are subsquares of the original square with a $2^{i/2} \times 2^{i/2}$ grid. The third observation is that if the aspect ratios of the rectangles are constant, we only need to keep track of the movement of the middle of each square.

The reason that we can stop when the rectangles contain $\Theta(\log n)$ points is that a rectangle containing $\log n$ points has diameter $\Theta(\sqrt{\log n}/\sqrt{n})$ (the aspect ratio is bounded by a constant) and is thus small enough that we can move the points anywhere in it without moving them farther than $O(\sqrt{\log n}/\sqrt{n})$, the distance we wish to show the average point is moved.

The second observation is easy: at each stage we divide every region into two equal-sized regions, alternating between horizontal and vertical division. If we do not apply any transformations, this gives a $2^i \times 2^i$ grid after $2i$ divisions.

The reason that we only need to keep track of the middle of each square is that the middle of the square determines how much the rest of the square moves. Given a rectangle and a square, there is only one affine transformation taking the rectangle to the square. If they have the same area and the rectangle has an aspect ratio of r, then no point moves farther than diam(rectangle) + diam(square). (See figure 2.2.2). Since the rectangle has aspect ratio r, if the square has side length s, the diameter of the rectangle is $s\sqrt{r^{1/2} + r^{-1/2}} = O(sr^{1/4})$. If r is bounded, this is $O(s)$, which is sufficiently small.