

# Part I. Random Planar Matching

## Chapter 1. The Matching Problems

In this chapter, we will define the planar matching problems that we investigate, and prove certain easy relationships between them. There are four main problems, but each of these has several variations that we also discuss. For the most part, these variations do not affect the expected value of the optimum matching by more than a constant factor.

Many of these possible variations will be the same for all the problems. The variations are in the way the points are distributed in the square; once the distribution of the points has been chosen, the points can be matched in the manner of any of the four basic problems. To make the exposition clearer, we will first define the problems without giving all the variations, and introduce all the variations in a later section.

For the first four sections, we will assume that we have a set  $\mathcal{P}^+$  of  $n +$  points, and a set  $\mathcal{P}^-$  of  $n -$  points. These points are independently and uniformly distributed in a unit square. For each of the problems, we will wish to match the  $-$  points to the  $+$  points. In different problems, there may be different constraints on the matching, and we may be optimizing different functions of the matching.

## 1.1. Average Edge Length Matching

The average edge length problem is possibly the most natural of the matching problems we discuss in this thesis. We will call this problem  $M_a$ . The problem is: given  $n +$  points and  $n -$  points uniformly distributed in a unit square, match the  $-$  points to the  $+$  points so as to minimize the sum of the edge lengths. Let the expected sum of the lengths of the edges of an optimal matching be  $\mathcal{D}_a(n)$ . What is  $\mathcal{D}_a(n)$ ?

The problem was first investigated by Ajtai, Komlós and Tusnády. They show [AKT] that  $\mathcal{D}_a(n) = \Theta(\sqrt{n \log n})$ . In this thesis, we have simplified their proof of the lower bound. By using a more complex construction than theirs, we can avoid appealing to the difficult theorems on probability that they use. We also show the stronger result that with probability  $1 - 2^{-n^\epsilon}$  for  $\epsilon < 1$ , the average edge length of an optimal matching is  $\Omega(\sqrt{n \log n})$ , where the constant depends on  $\epsilon$ . It is easy to see that the upper bound does not hold with this high probability. With probability  $1 - \frac{1}{n^\alpha}$ , the sum of the  $x$ -coordinates of the  $-$  points and of the  $+$  points differ by  $\Omega(\sqrt{n \log n})$ . The horizontal components of the edge lengths must sum to at least this difference, so the sum of the edge lengths is  $\Omega(\sqrt{n \log n})$ .

We also discuss a slight variation of this problem. Suppose that points may be matched not only to points of the opposite sign, but also to the boundary of the square. Again, we wish to minimize the sum of the edge lengths. We will call this problem  $M'_a$ . Let the expected sum of the edge lengths of an optimal matching of this kind be  $\mathcal{D}'_a(n)$ . This gives a lower bound on the sum of the edge lengths inside the square, even if points are permitted to be matched to points outside the square. By going to the dual problem, we will show that  $\mathcal{D}'_a(n) = \Theta(\mathcal{D}_a(n))$ .

## 1.2. Rightward Matching

The rightward matching problem seems much less natural than the average edge length problem. This problem is interesting because it arises in the proof of a lower bound that applies to any on-line bin packing algorithm. It is also intermediate between two previously considered problems: average edge length matching and up-right matching. This is, to the best of our knowledge, the first time the problem has been considered.

The rightward matching problem  $M_r$  is: given  $n$   $+$  points and  $n$   $-$  points in the unit square, match every  $-$  point to a  $+$  point to its right or to the top or bottom edge of the square and match every  $+$  point to a  $-$  point to its left or to the top or bottom of the square. Find such a matching minimizing the sum of the vertical lengths of the edges. Let the expected sum of the vertical lengths of the edges of an optimal such matching be  $\mathcal{D}_r(n)$ . What is  $\mathcal{D}_r(n)$ ?

This problem has an equivalent variation which more closely resembles the average edge length problem. Match the  $-$  points to their right and the  $+$  points to their left, either to a point of the opposite kind or to any edge of the square. Minimize the sum of the edge lengths. We call this problem  $M'_r$ . Let the expected value of the sum of edge lengths in an optimal matching of this kind be  $\mathcal{D}'_r(n)$ . We can show, by going through the dual problem, that  $\mathcal{D}'_r(n) = \Theta(\mathcal{D}_r(n))$ .

It is clear that  $\mathcal{D}'_a(n) < \mathcal{D}'_r(n)$ . This is because the  $M'_a$  and  $M'_r$  problems only differ in the addition restriction on  $M'_r$  that edges go to the right. Any rightward matching is a matching, so the optimal average edge length matching has sum of edge lengths less than the optimal rightward matching. Thus,  $\mathcal{D}_a(n) = O(\mathcal{D}_r(n))$ .

## 1.3. Up-Right Matching

In the up-right matching problem  $M_{ur}$  a  $-$  point may only be matched with

a  $+$  point which is above and to the right of it. The goal is to minimize the number of unmatched points. This problem has previously arisen in at least two contexts. Karp, Luby and Marchetti arrived at it through an investigation of two-dimensional bin packing [KLM]. Dudley investigated the discrepancy of lower regions during research on lower regions [Du]. Both of these investigations resulted in bounds of  $O(\sqrt{n} \log n)$  and  $\Omega(\sqrt{n \log n})$ . We will prove that the correct bound is  $\Theta(\sqrt{n} \log^{3/4} n)$ .

As in average edge length and rightward matching, we can produce an equivalent problem  $M'_{ur}$  which has an answer of the same order and which resembles the other problems  $M'_a$  and  $M'_r$ . In this equivalent problem  $M'_{ur}$ ,  $-$  points must be matched up and to the right and  $+$  points down and to the left, but points may be matched off the square in the appropriate direction (i.e.,  $-$  points must be matched to the top or right edge, and  $+$  points to the bottom or left).

The problem  $M_{ur}$  of up-right matching is: given  $+$  and  $-$  points randomly distributed in a unit square, match the  $-$  points to the  $+$  points so that each  $-$  point is matched up and to the right to a  $+$  point. What is the expected number of points that are left unmatched by an optimum up-right matching? We call this quantity  $\mathcal{N}_{ur}$ .

The equivalent problem  $M'_{ur}$  is the following. Given  $n$   $+$  and  $n$   $-$  points randomly distributed in a unit square, match the  $-$  points up and right to a  $+$  point. Points can also be matched to an appropriate border. Minimize the sum of the edge lengths of an optimal matching. Let this sum be  $\mathcal{D}'_{ur}$ . This quantity has the same asymptotic behavior as  $\mathcal{N}_{ur}$ .

**Theorem 1.3.1:**  $\mathcal{D}'_{ur} = \Theta(\mathcal{N}_{ur})$ .

**Step 1:**  $\mathcal{N}_{ur} = O(\mathcal{D}'_{ur})$ .

**Proof:** Consider squares which have the same center as our unit square  $S$  but are smaller. Suppose we are given an up-right matching  $M$ . Let  $x$  be the size of a

side of the smaller square, and let  $N_M(x)$  be the number of edges of matching  $M$  crossing the border of the smaller square. We have for any matching  $M$ ,

$$\int_0^1 N_M(x) dx \leq D_M,$$

where  $D_M$  is the sum of the edge lengths of matching  $M$ . The integral counts the component of the edge length perpendicular to the border of the smaller square, which is less than the total edge length. The number of edges crossing the border of the smaller square is at least the number of unmatched points in the smaller square. Thus,  $E(N_M(x)) \geq \mathcal{N}_{ur}(nx^2) \geq \mathcal{N}_{ur}(n/4)$  for  $x \geq \frac{1}{2}$ . Integrating, we get

$$\frac{1}{2} \mathcal{N}_{ur}(n/4) \leq E\left(\int_0^1 N_M(x) dx\right) \leq D_M.$$

Thus,  $\mathcal{N}_{ur} = O(D'_{ur})$ . ■

**Step 2:**  $D'_{ur} = O(\mathcal{N}_{ur})$ .

Suppose we have a matching  $M$  with  $K$  unmatched points. Each of the unmatched points can be matched to the boundary of the square with edge length at most 1. Thus, these unmatched points contribute at most  $K$  to the sum of the edge lengths  $D_M$ . We now show the same bound for the edge lengths of the matched points.

We show the sum of the edge lengths of the matched points is  $O(K)$ . We bound the vertical and horizontal edge lengths separately. The total edge length is less than the sum of the horizontal and the vertical edge lengths because the Euclidean distance between two points is always less than the distance in the Manhattan metric. We bound the horizontal sum; the vertical bound is identical. Consider the quantity

$$\left| \sum_{Q \in \mathcal{P}^+} x_Q - \sum_{P \in \mathcal{P}^-} x_P \right|,$$

the difference of the sums of the  $x$ -coordinates of the  $+$  points and of the  $-$  points. This quantity averages  $\Theta(\sqrt{n})$  because the  $+$  and  $-$  points are distributed uniformly (and there are an equal number of each). With probability  $1 - \frac{1}{n^\alpha}$ , this is less than  $\Theta(\sqrt{n \log n})$ . Let  $M^+$  and  $M^-$  be the sets of  $+$  and  $-$  points adjacent to an edge in the matching  $M$ . Then

$$\sum_{Q \in P^+} x_Q - \sum_{Q \in M^+} x_Q = \sum_{Q \in P^+ - M^+} x_Q \leq K,$$

since the number of  $+$  points in  $P^+ - M^+$  is the number of unmatched points. Similarly,

$$\sum_{P \in P^-} x_P - \sum_{P \in M^-} x_P \leq K.$$

Thus,

$$E\left(\sum_{Q \in M^+} x_Q - \sum_{P \in M^-} x_P\right) = O(\log^{1/2} \sqrt{n} + K).$$

However,  $\sum_{Q \in M^+} x_Q - \sum_{P \in M^-} x_P$  is the sum of the horizontal edge lengths. If  $K > \sqrt{n}$  this bounds the expected sum of the edge lengths by  $O(K)$ , the bound we wished to show. ■

## 1.4. Maximum Edge Length Matching

The maximum edge length matching problem  $M_m$  is: given  $n$   $+$  and  $n$   $-$  points randomly distributed in the unit square, match the  $+$  points to the  $-$  points, minimizing the maximum edge length. This problem has arisen in the context of some VLSI problems [LL], and also seems a fairly natural question. We will show in Chapter 3 that the optimal maximum edge length is with high probability  $\Theta(\log^{3/4} n / \sqrt{n})$ . We now show that the maximum edge length is at least as large as the average edge length for an up-right matching.

Let  $\mathcal{D}_m$  be the expected maximum edge length of an optimal matching. We can produce an up-right matching with the same expected edge length  $\mathcal{D}_m$ , and thus  $O(n\mathcal{D}_m)$  unmatched points.

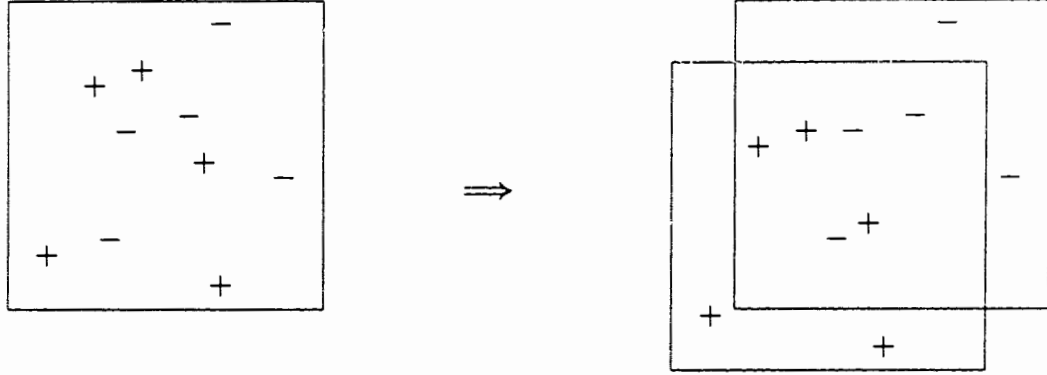


Figure 1.4.1: Shifting the  $+$  points down and left.

**Theorem 1.4.1:**  $\mathcal{N}_{ur} = O(n\mathcal{D}_m)$ .

**Proof:** Suppose that with high probability we can find a matching with maximum distance  $\alpha$ . We shift all the  $+$  points in the square down and left by  $\alpha$ . We then find a matching with maximum edge length  $\alpha$  in the overlap of the original square and the shifted square. The  $-$  points in this matching come from the lower left corner of the original square and the  $+$  points from the upper right corner of the shifted square (See Figure 1.4.1). The  $+$  points and the  $-$  points in this new square are still uniformly distributed. By removing points at random from this square (either  $+$  or  $-$  points, depending on which are excess), we can ensure an equal number of  $+$  and  $-$  points in this square. These points are still distributed uniformly in the square. We can then with high probability match these points with maximum edge length  $\alpha$ . Shifting the  $+$  points back to their original positions, all the edges become up-right edges. The number of unmatched points is the maximum of the number of  $+$  points along the bottom and left borders and the number of  $-$  points along the top and right borders. This has expected value  $\Theta(n\alpha)$ .

## 1.5. Duality

All the problems we have described are special cases of bipartite matching. The formulation of the bipartite matching problem that we use is: given a complete bipartite graph  $G$  with weights on the edges, find a minimum perfect matching. Bipartite matching has a dual problem, with the maximum solution to the dual problem equal to the minimum weight matching. This dual problem is just the linear programming dual of bipartite matching. It is often easier to prove theorems by using the dual problem than by working directly with the original problem. [PS]

We now look at the dual problem for minimum weighted perfect matching in a complete bipartite graph  $G$ . Let  $P_1, P_2, \dots, P_n$  and  $Q_1, Q_2, \dots, Q_n$  be the vertices of a bipartite graph. Let  $f(P_i Q_j)$  be the weight on edge  $P_i Q_j$ . The dual problem for the minimum perfect matching is: give weights  $w(P_i)$ ,  $w(Q_i)$  to the vertices of  $G$  such that for any edge  $P_i Q_j$ ,

$$w(Q_j) - w(P_i) \leq f(P_i Q_j).$$

The value of this solution to the dual problem is

$$\sum_i w(Q_i) - \sum_i w(P_i).$$

The maximum solution to the dual problem is the weight of the minimum perfect matching.

It is easy to see that the maximum solution to the dual problem is at most the minimum perfect matching. Let  $M$  be a matching. Suppose  $P_i$  is matched to  $Q_{\sigma(i)}$  in  $M$ , where  $\sigma$  is a permutation of  $n$ . Summing over the edges of  $M$ , we get

$$\sum_{i=1}^n f(P_i Q_{\sigma(i)}) \geq \sum_{i=1}^n w(Q_{\sigma(i)}) - w(P_i) = \sum_{i=1}^n w(Q_i) - \sum_{i=1}^n w(P_i).$$



This shows that any solution to the dual problem has value less than any matching. Showing that the maximum solution to the dual has value equal to the weight of the minimum matching is more difficult. It can be done by showing the the optimum solution to the corresponding linear program is a matching and appealing to the theorem for linear programming. It can also be done by giving an algorithm that finds equal weight solutions to the primal and dual problems. For a proof of this result, see [PS].

An especially simple case of this theorem is that of non-weighted bipartite matching. For this problem, a bipartite graph  $G$  is given. The object is to find the largest matching contained in  $G$ . This graph  $G$  can be considered a complete bipartite graph with weights of 0 and 1 on the edges. A weight of 1 is given to the edges in  $G$ . The goal is now to maximize the weight of a matching in the complete bipartite graph. An optimal solution of the dual problem in this case needs weights of only 0 and 1. The dual is stated in a more easily used form by Hall's Theorem below.

**Hall's Theorem:** In a bipartite graph  $G$  between two sets of points  $\mathcal{P}^+$  and  $\mathcal{P}^-$ , the number of unmatched  $+$  points in a maximal matching is

$$\max_{A \subseteq \mathcal{P}^+} |A| - |R(A)|,$$

where  $R(A)$  is the set of vertices of  $\mathcal{P}^-$  that are adjacent to the vertices of  $A$ .

For the planar matching problems we are looking at, the dual functions have nice properties. For the up-right and the maximum edge length matching problems, we are dealing with unweighted matchings. Thus, Hall's theorem for maximum matching applies to these cases. For the rightward matching and the average distance matching, one can take the weight function to be a function mapping the unit square into  $[-1, 1]$  that has certain constraints on its slope. This has two nice properties: the same function applies to the  $+$  and the

– points, and the dual problem is geometrically meaningful, and so is easier to work with.

For the up-right matching problem, the dual problem is simple. Consider a subset  $A$  of the + points. Let  $B$  be the set of – points that can be matched to a point in set  $A$  (i.e., the set of – points below and to the left of a point of  $A$ ). We wish to find a set  $A$  which maximizes  $|A| - |B|$ . We will show in Chapter 3 how to find a set such that  $|A| - |B| = \Omega(\sqrt{n} \log^{3/4} n)$ . This shows that upward right matching leaves at least  $\Omega(\sqrt{n} \log^{3/4} n)$  points unmatched.

For maximum distance matching, we need to specify a distance  $d$  to obtain the dual problem. To obtain a perfect matching with edge length  $d$ , we need a  $d$  such that for all subsets  $A$  of + points, the set  $B$  of – points within distance  $d$  of a point in  $A$  contains at least as many points as  $A$  does. In Chapter 3, we show that for  $d = \Omega(\log^{3/4} n / \sqrt{n})$ , this holds with high probability. This implies that the maximum distance in an optimal matching is  $O(\log^{3/4} n / \sqrt{n})$ . Combined with the result on up-right matching above, this shows that an optimal up-right matching leaves  $\Theta(\sqrt{n} \log^{3/4} n)$  points unmatched and an optimal maximum distance matching has edges of length  $\Theta(\log^{3/4} n / \sqrt{n})$ .

For the average edge length matching problem, the constraints are that if a + point  $x_+$  and a – point  $x_-$  are separated by distance  $d(x_+, x_-)$ , then  $w(x_+) - w(x_-) \leq d(x_+, x_-)$ . We will show that given a dual function satisfying this, we can get a function giving a dual solution at least as large which maps the unit square into  $[-1, 1]$  and has slope  $(\max_{x \neq y} |w(x) - w(y)| / d(x, y))$  at most 1. We do this by only using the values for the – points. We will increase the values of the + points as far as possible, consistent with the given – points. We will also decrease the values of any – points which are “dominated” by other – points. We claim that this gives a dual function mapping the unit square into  $\mathbb{R}$  which has slope at most 1. The function we obtain will clearly give at least as

good a dual solution as the original function, since we decrease the values of the  $-$ 's and increase the values of the  $+$ 's. All we must show is that this function has a slope of at most 1.

We start with the set  $\mathcal{P}^-$  of  $-$  points and a weight function  $w$  on them. The function  $w'$  will be

$$w'(X) = \min_{P \in \mathcal{P}^-} (d(P, X) + w(P)).$$

Clearly,  $w'(P) \leq w(P)$  if  $P \in \mathcal{P}^-$ , since one of the terms we take the minimum of is  $w(P)$ . Also,  $w'(Q) \geq w(Q)$  if  $Q$  is a  $+$  point since  $w(Q) \leq w(P) + d(P, Q)$  for any  $-$  point  $P$ . We wish to show that the slope of  $w'$  is at most 1, that is, that if two points are separated by distance  $d$ , then their values differ at most  $d$ . Suppose there are two points,  $X_1$  and  $X_2$ , that do not satisfy this condition, with  $w'(X_1) < w'(X_2)$ . Then

$$w'(X_2) - w'(X_1) > d(X_1, X_2).$$

Since  $w'(X) = \min_{P \in \mathcal{P}^-} w(P) + d(P, X)$ , there is a  $-$  point  $P$  such that

$$w'(X_1) = w(P) + d(P, X_1).$$

We also have

$$w'(X_2) \leq w(P) + d(P, X_2).$$

Subtracting the above equations, we obtain

$$\begin{aligned} d(P, X_2) - d(P, X_1) &\geq w'(X_2) - w'(X_1) \\ &\geq d(X_1, X_2), \end{aligned}$$

which violates the triangle inequality. Thus, the points of the optimal dual solution satisfy the condition that the slope of the function between any two of them is at most 1.

We have shown that we can assume our dual function is a function mapping the unit square into  $\mathbf{R}$  with slope at most 1. Since no two points are farther than  $\sqrt{2}$  and the slope of  $w'$  is at most 1, the most that two values of  $w'$  can differ by is  $\sqrt{2}$ . We can thus normalize  $w'$  by adding some constant to it so that  $w'$  maps the square into  $[-1, 1]$ . ( $[-\sqrt{2}/2, \sqrt{2}/2]$  to be exact.)

The proof above actually works in a more general setting. We will need this generalization for the dual problems to  $M'_a$ ,  $M_r$ , and  $M'_r$ . In the general setting, the weight  $f(X, Y)$  of an edge from  $X$  to  $Y$  is not necessarily the same as the weight  $f(Y, X)$  of an edge from  $Y$  to  $X$ . All we will need is the triangle inequality:  $f(X, Y) + f(Y, Z) \geq f(X, Z)$ . This generalization is stated in the following lemma.

**Lemma 1.5.1:** Suppose we are given a set of points  $\mathcal{X}$  and a function  $f$  mapping  $(\mathcal{X}, \mathcal{X})$  into the non-negative reals satisfying  $f(X, X) = 0$  for  $X \in \mathcal{X}$  and the triangle inequality  $f(X, Y) + f(Y, Z) \geq f(X, Z)$  for  $X, Y, Z \in \mathcal{X}$ . Suppose we are given a set of  $+$  points  $\mathcal{P}^+ \subseteq \mathcal{X}$  and a set of  $-$  points  $\mathcal{P}^- \subseteq \mathcal{X}$ . Let  $w$  be a weight function mapping  $\mathcal{P}^+ \cup \mathcal{P}^-$  into  $\mathbf{R}$  such that  $w(Q) - w(P) \leq f(P, Q)$  for  $Q \in \mathcal{P}^+$  and  $P \in \mathcal{P}^-$ . Then there is a weight function  $w'$  such that  $w'(Q) \geq w(Q)$  for  $Q \in \mathcal{P}^+$ ,  $w'(P) \leq w(P)$  for  $P \in \mathcal{P}^-$ , and  $w'(X_2) - w'(X_1) \leq f(X_1, X_2)$  for  $X_1, X_2 \in \mathcal{X}$ .

**Proof:** As before, we let

$$w'(X) = \min_{P \in \mathcal{P}^-} (w(P) + f(P, X)).$$

Suppose there are two points  $X_1$  and  $X_2$  such that

$$w(X_2) - w(X_1) > f(X_1, X_2).$$

By the choice of  $w'$ , we can find a  $P \in \mathcal{P}^-$  such that

$$w'(X_1) = w(P) + f(P, X_1)$$

and

$$w'(X_2) \leq w(P) + f(P, X_2).$$

From these equations, we get

$$\begin{aligned} f(P, X_2) &\geq w'(X_2) - w(P) \\ &> f(X_1, X_2) + f(P, X_1), \end{aligned}$$

a contradiction since it violates the triangle inequality. Thus, for any two points,

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

We still need to show we have reduced the values of the  $-$  points and increased the values of the  $+$  points. We have  $w'(P) \leq w(P)$  for  $P \in \mathcal{P}^-$  since one of the terms in the expression for  $w'(P)$  is  $w(P) + f(P, P) = w(P)$ . We have  $w'(Q) \geq w(Q)$  for  $Q \in \mathcal{P}^+$  since  $w(Q) \leq w(P) + f(P, Q)$  for all  $P \in \mathcal{P}^-$ , and to obtain  $w'(Q)$  we take the minimum of these. This proves Lemma 1.5.1. ■

Now we will show that if points may be matched to the boundary of the square as well as to points of the opposite kind, the optimal dual function can be made to satisfy the conditions that the slope is at most 1 and that the function is 0 on the boundary of the square. Here, the conditions on the values of a  $+$  point  $Q$  and a  $-$  point  $P$  is that they satisfy

$$w(Q) - w(P) \leq \min(d(P, Q), d(P, S) + d(Q, S)),$$

where  $d(P, S)$  is the distance from point  $P$  to the nearest boundary of the square. This is because we can take the unmatched points, and consider them to be matched in pairs. Matching both  $P$  and  $Q$  to the boundary gives edge lengths  $d(P, S) + d(Q, S)$ . This weight function satisfies the triangle inequality, since it is the shortest distance from  $P$  to  $Q$ , when you are allowed to go through the boundary from any point on the boundary to any other point on it.

By Lemma 1.5.1 above, we obtain a dual function  $w'$  such that any two points  $X_1, X_2$  satisfy

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

We claim that functions  $w$  satisfying this equation are exactly those functions with slope at most 1 which are constant on the boundary. We have  $|w(X_2) - w(X_1)| \leq d(X_1, X_2)$ , which is the condition for slope 1. We have  $w(X_2) - w(X_1) \leq d(X_1, S) + d(X_2, S)$ . Since for  $X$  on the boundary of the square,  $d(X, S) = 0$ , we have that  $w(X_2) = w(X_1)$  for  $X_1, X_2$  on the boundary of the square.

Furthermore, any function which has slope at most 1 and is constant on the boundary satisfies the equation. If the slope is at most 1,  $|w(X_1) - w(X_2)| \leq d(X_1, X_2)$  and  $|w(X) - w(S)| \leq d(X, S)$ , where  $w(S)$  is the value of  $w$  on the boundary of the square. Thus,

$$\begin{aligned} w(X_2) - w(X_1) &\leq |w(X_1) - w(S)| + |w(S) - w(X_2)| \\ &\leq d(X_1, S) + d(X_2, S). \end{aligned}$$

We can normalize the function so that  $w'$  is 0 on the boundary of the square. Any function which is 0 on the boundary and has slope at most 1 satisfies the above constraints for the dual problem of  $M'_a$ , so we can take these as the conditions for the dual function.

We now use the dual problems given above to prove the following theorem.

**Theorem 1.5.2:**  $\mathcal{D}_a = \Theta(\mathcal{D}'_a)$ .

**Proof:** Clearly,  $\mathcal{D}_a \geq \mathcal{D}'_a$ , since being permitted to match to the boundary can only decrease the expected sum of the edge lengths. We thus only need to prove that  $\mathcal{D}'_a = O(\mathcal{D}_a)$ . We construct a dual function for the problem  $M'_a$  by taking a dual function for the problem  $M_a$  on the middle ninth of the square and extending it to make the boundary of the square 0. (See figure 1.5.1)

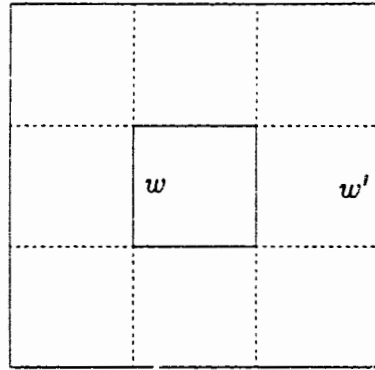


Figure 1.5.1: Extending the function  $w$ .

We can construct a dual function on the middle ninth so as to map the middle ninth of the square into the interval  $[-\frac{1}{6}, \frac{1}{6}]$  and so as never to have slope more than  $\frac{1}{2}$ . We can then extend the function to the rest of the square by using a linear function between the edge of the middle square and the edge of the outside square. This function has a slope of at most 1, since it has a slope of at most  $\frac{1}{2}$  both in the horizontal and in the vertical directions. The points outside the middle square add an expected value of 0 to the function, because they are as likely to be + points as - points, and they had no effect on the function we constructed. Thus, the expected value of the dual solution on the whole square is the expected value on the middle square, which is  $\Theta(D'_a)$ . ■

We now give a similar argument to show that the two rightward matching problems are equivalent. Recall  $D_r$  is the expected value of the vertical edge length of a rightward matching (where matching to the top and bottom of the square is allowed) and  $D'_r$  is the expected value of the average edge length of a rightward matching (where matching to any side of the square is allowed.) We will use the Manhattan metric to measure  $D'_r$ . This changes the edge lengths at most by a factor of  $\sqrt{2}$ , and makes the proof easier. We show that the constraints on the dual function for the  $M_r$  problem are that this dual function is 0 on the top and bottom of the square, that its vertical slope is less than 1 and