

Improved Lower and Upper Bounds for Universal TSP in Planar Metrics

Mohammad T. Hajiaghayi*

Robert Kleinberg*[†]

Tom Leighton*[‡]

Abstract

A universal TSP tour of a metric space is a total ordering of the points of the space such that for any finite subset, the tour which visits these points in the given order is not too much longer than the optimal tour. There is a vast literature on the TSP problem, and universal TSP tours have been studied since the 1980's when Bartholdi and Platzman [29] introduced the spacefilling curve heuristic for the Euclidean TSP problem and conjectured that there exists a constant-competitive universal TSP tour based on this heuristic. Here, we settle this conjecture negatively by proving an $\Omega\left(\sqrt[6]{\log n / \log \log n}\right)$ lower bound for universal TSP tours of the $n \times n$ grid; this is the first known example of a family of finite metrics with no constant-competitive universal tour.

Generalizing from the $n \times n$ grid to arbitrary weighted planar graph metrics, and more generally H -minor-free metrics, we improve the best known upper bound for universal tours of such metrics from $O(\log^4 n / \log \log n)$ to $O(\log^2 n)$.

1 Introduction

Consider a city whose roads form a grid (e.g. Manhattan), or more generally an arbitrary planar graph (e.g. Boston). Each day, a delivery van must deliver packages to several places in the city, and it is desirable to visit these locations in an order which incurs the minimum total distance. This is of course the famous *traveling salesman problem*, restricted to the shortest-path metric of a planar graph.

Since the problem must be solved with a different set of delivery locations each day, the simplest heuristic for the van's driver to use would be to have access to a *universal ordering* of the delivery locations (pre-computed offline), and to plan the van's route each day by simply visiting that day's locations in the order in which they appear in this universal listing. If

the length of this tour is guaranteed to be no more than ρ times longer than the optimal tour, we say that the universal tour achieves *competitive ratio* ρ . Our goal in this paper is to study such *planar universal TSP tours*, improving on the best known upper and lower bounds for their competitive ratio.

The first work on this question dates from the late 1980's, when Bartholdi and Platzman [29] proposed the spacefilling curve heuristic for Euclidean TSP (a universal ordering of the points of the plane based on the order in which they are visited by a spacefilling curve) and they proved that their universal tour returns a sub-tour within an $O(\log n)$ factor of optimal length. They conjectured that the competitive ratio of their spacefilling curve is in fact $O(1)$. This conjecture was disproved by Bertsimas and Grigni [6], who proved that the $O(\log n)$ competitive ratio for the Bartholdi-Platzman curve is tight. It remained an open problem whether there exists a constant-competitive universal tour for the unit square. Bertsimas and Grigni conjectured that the answer is no; in fact they made the stronger conjecture that for any ordering of the vertices of the $n \times n$ grid there is some subset of vertices such that the length of their induced tour is an $\Omega(\log n)$ factor off from optimal, and they observed that even a $\omega(1)$ lower bound would be interesting. The main result of this paper achieves progress toward this conjecture by demonstrating a lower bound of $\Omega\left(\sqrt[6]{\log n / \log \log n}\right)$ for the $n \times n$ grid. Such an $\omega(1)$ lower bound is known for the related problem of universal Steiner Tree [1, 21] in the Euclidean plane, but the Steiner Tree lower bound does not imply a $\omega(1)$ lower bound for universal TSP tours. In fact, our approach for universal TSP lower bound is quite different from (and more complicated than) these previous approaches.

Recently, Jia, Lin, Noubir, Rajaraman and Sundaram studied the universal TSP problem for arbitrary metrics. More precisely, they constructed a universal TSP tour with competitive ratio $\rho = O(\log^4 n / \log \log n)$ for arbitrary metrics. Furthermore, for the special case of doubling metrics, which includes both constant-dimensional Euclidean and

*Department of Mathematics and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 200 Technology Square, Cambridge, MA 02139, U.S.A., Emails: {hajiagh,rdk,ftl}@theory.csail.mit.edu.

[†]Department of Computer Science, Cornell University, Ithaca, NY 14853.

[‡]Akamai Technologies, Eight Cambridge Center, Cambridge, MA 02139, U.S.A.

growth-restricted metrics, they obtain competitive ratio $\rho = O(\log n)$. In this paper, we obtain improved competitive ratio $\rho = O(\log^2 n)$ for planar metrics and more generally H -minor-free metrics. (Note that planar metrics and H -minor-free metrics are not necessarily doubling metrics, hence the best previously-known bound was $O(\log^4 n / \log \log n)$.) It is worth mentioning that even for arbitrary metrics, no lower bound better than $\Omega(1)$ was known; our $\Omega\left(\sqrt[\epsilon]{\log n / \log \log n}\right)$ for the $n \times n$ grid is the first such bound.

1.1 History and prior work The traveling salesman problem (TSP) [25] has been a testbed for almost every new algorithmic idea during the past half-century: linear programming, cutting planes and polyhedral combinatorics, probabilistic algorithms, local search, Boltzmann machines, genetic algorithms, simulated annealing, neural nets, dynamic programming, Lagrangian relaxation, approximability, and even NP-completeness. The TSP was shown early to be non-approximable in its general case (see e.g. [13]), and Christofides’ 3/2 approximation algorithm [9] for the *metric TSP* problem (in which the pairwise distances form a metric space) is well-known. However, metric TSP is MAX SNP-hard even when all distances are one or two [28]. For three decades it has been an open question whether the 3/2 approximation factor can be improved. One of the most important special cases, first considered by Grigni, Koutsoupias, and Papadimitriou [15], is *planar TSP* in which we are given a planar graph and we are looking for the shortest closed walk that visits all nodes at least once (equivalently, the metric TSP for the shortest-path metric of a planar graph). In this groundbreaking work, the authors obtain a PTAS for planar TSP, thereby ruling out the possibility of MAX SNP-hardness for this case. This work later led to the seminal work of Arora [2] who obtained PTASs for Euclidean TSP and related problems (see also Mitchell [27]). Arora et al. [3] generalized this work to allow distances defined by non-negative edge weights and Grigni [14] extended this work from planar graphs to bounded genus graphs and graphs excluding a fixed graph as a minor (but he obtained a quasi-polynomial time approximation scheme instead of a PTAS). Very recently, Klein [23] improved the running time of the PTAS for weighted planar graphs to linear time.

The concept of “universal” solutions has also been used many times before, notably in the context of hash functions [8] and routing [35] which use randomization to give an oblivious solution to the input. Apart from the papers on spacefilling curve heuristics for Euclidean TSP cited above [6, 29], the

works which are most closely related to the concept of universal solution considered in this paper are the recent results on oblivious routing [4, 17, 18, 19, 30] and universal approximation algorithms [21]. It is worth mentioning that in oblivious routing, the optimum for an instance is exactly computable in polynomial time, whereas universal approximation algorithms are usually applied when the optimum solution itself is intractable. Some other related concepts are:

- online algorithms, in which the algorithm, upon receiving an input, must take an action without the knowledge of future inputs [5, 7, 12, 34],
- settings where a certain distribution over the space of inputs is assumed [16, 17, 20, 22],
- stochastic optimization problems, in which we commit on some initial (first-stage) actions and some further recourse actions are performed once the actual data is realized [10, 16].

Our Results and Techniques. The main result of this paper is a lower bound of $\Omega\left(\sqrt[\epsilon]{\log n / \log \log n}\right)$ for the competitive ratio of universal TSP on a (planar) $n \times n$ grid. This lower bound represents progress toward proving the conjectured lower bound $\Omega(\log n)$ of Bertsimas and Grigni [6]. In fact, this is the first $\omega(1)$ lower bound for the competitive ratio of universal TSP in *any* metric. The same lower bound trivially implies that the spacefilling curve heuristic for Euclidean TSP is not constant-competitive for *any* spacefilling curve. Intuitively, our approach is based on proving that any total ordering of the $n \times n$ grid must possess one of the following substructures: a *zigzag* — i.e. a large set of points having diameter d , such that the optimum tour has length $O(d)$ but at least half the steps of the universal tour have length $\Omega(d)$ — or a large set of nearly-collinear points such that the universal tour performs many *backtracks* while visiting these points. In fact, in the absence of the first type of substructure (the zigzag), a *random* line is likely to produce the second type of substructure.

We complement our lower bound for the universal TSP problem on an $n \times n$ grid by constructing a universal TSP with competitive ratio of $O(\log^2 n)$ in arbitrary planar metrics and more generally H -minor-free metrics. This improves upon the competitive ratio $O(\log^4 n / \log \log n)$ in these kinds of metrics [21]. Our approach is based on demonstrating a novel connection between the concept of *sparse partitions* introduced by Jia et al. [21], and the partitioning technique for planar (more generally, H -minor-free) graphs developed by Klein, Plotkin, and Rao [24] which has been further extended by Rao [31] for embedding planar (H -minor-free) metrics into ℓ_2 .

2 Lower Bound for Universal TSP in Grids

DEFINITION 1. Given a total ordering \preceq of a metric space (X, d) and a finite subset $S \subseteq X$, suppose the elements of S are numbered s_1, s_2, \dots, s_n so that $s_1 \preceq s_2 \preceq \dots \preceq s_n$. Adopting the convention that $s_{n+1} = s_1$ and that S_n denotes the permutation group of the set $\{1, 2, \dots, n\}$, we define

$$\begin{aligned} \text{utsp}(S) &= \sum_{i=1}^n d(s_i, s_{i+1}) \\ \text{opt}(S) &= \inf_{\pi \in S_n} \left\{ \sum_{i=1}^n d(s_{\pi(i)}, s_{\pi(i+1)}) \right\} \\ \rho(S) &= \text{utsp}(S) / \text{opt}(S) \\ \rho_n(X) &= \sup\{\rho(S) : S \subseteq X, |S| \leq n\}. \end{aligned}$$

We aim to prove lower bounds on $\rho_n(X)$ when X is the unit square $[0, 1]^2$ with the ℓ_2 metric, or when X is the vertex set of the N -by- N grid graph, with its shortest-path metric. Both lower bounds are easy corollaries of Theorem 2.1 below, in which \mathbb{G}_N denotes the set

$$\mathbb{G}_N = \left\{ \left(\frac{2i-1}{2N}, \frac{2j-1}{2N} \right) : 1 \leq i, j \leq N \right\},$$

regarded as a subspace of the metric space $[0, 1]^2$ with the ℓ_2 metric. (If $[0, 1]^2$ is subdivided into an N -by- N grid of congruent squares, the points of \mathbb{G}_N are the centers of these squares.)

THEOREM 2.1. There exist positive constants c_1, c_2 such that for all sufficiently large integers $N \geq k > 0$, and for all total orderings of \mathbb{G}_N , at least one of the following is satisfied.

1. There exists a set S of at most k points in \mathbb{G}_N , such that $\rho(S) \geq c_1 k$.
2. There exists a set S of $O(N)$ points in \mathbb{G}_N such that

$$\text{opt}(S) < 5 + \frac{18 \text{utsp}(S)}{k},$$

and

$$\text{utsp}(S) > \left(\frac{c_2}{k^5} \right) \log_k N.$$

COROLLARY 2.1. If $G = (V, E)$ is the N -by- N grid graph and (V, d) is its shortest-path metric, then $\rho(V) = \Omega\left(\sqrt[6]{\frac{\log N}{\log \log N}}\right)$

Proof. The obvious bijection from V to \mathbb{G}_N has distortion $\sqrt{2}$ because V is isometric (up to rescaling) to the point set \mathbb{G}_N in the ℓ_1 metric, and the natural mapping from $(\mathbb{R}^2, \|\cdot\|_1)$ to $(\mathbb{R}^2, \|\cdot\|_2)$ has

distortion $\sqrt{2}$. Therefore, it suffices to prove that $\rho(\mathbb{G}_N) = \Omega\left(\sqrt[6]{\frac{\log N}{\log \log N}}\right)$. This lower bound follows immediately from Theorem 2.1 by taking k to be $\left\lceil \sqrt[6]{\frac{\log N}{\log \log N}} \right\rceil$.

COROLLARY 2.2. For any total ordering of $[0, 1]^2$, $\rho_N([0, 1]^2) = \Omega\left(\sqrt[6]{\frac{\log N}{\log \log N}}\right)$.

As stated in the introduction, the proof of Theorem 2.1 relies on proving that every total ordering of \mathbb{G}_N has at least one of the following two substructures, which correspond to the first and second cases in the conclusion of Theorem 2.1:

- a *zigzag*, i.e. a set S of cardinality $\theta(k)$ and diameter d , such that $\text{opt}(S) = O(d)$ and a constant fraction of the hops in the universal tour of S have length $\Omega(d)$.
- a set S of $O(N)$ nearly-collinear points containing many *backtracks*.

The precise definition of a “backtrack” is formulated as follows. Here and throughout this paper, we adopt the following notation: if A, B are subsets of a totally ordered set (S, \preceq) then we write $A \preceq B$ if A completely precedes B in the ordering, i.e. $\forall a \in A, b \in B, a \preceq b$.

DEFINITION 2. A strip in \mathbb{R}^2 is the region between two parallel lines; the width of the strip is the distance between these two lines. Given a strip σ bounded by lines L_1, L_2 , a rectangular interval of σ is a rectangle $R \subset \mathbb{R}^2$ such that two sides of R are segments of L_1 and L_2 , respectively. The length of either of these two sides is called the length of the rectangular interval R .

If S is a subset of a strip σ and \preceq is a total ordering of S , a backtrack of length s in S is a triple of disjoint rectangles R_1, R_2, R_3 , along with points $p_i \in R_i \cap S$ for $i = 1, 2, 3$, such that:

- R_1, R_2, R_3 are rectangular intervals of σ having length s .
- R_2 lies between R_1 and R_3 , i.e. any line which intersects both R_1 and R_3 also intersects R_2 .
- Either $\{p_2\} \preceq (R_1 \cup R_3) \cap S$ or $\{p_1, p_3\} \preceq R_2 \cap S$.

The following lemma is the main technical tool for demonstrating the existence of zigzags. In the statement of the lemma, and throughout this paper, we use the notation $m(\cdot)$ to denote the Lebesgue measure (i.e. area) of a measurable subset of \mathbb{R}^2 and we use the symbol \oplus to denote the symmetric difference of two sets, i.e.

$$A \oplus B = (A \setminus B) \cup (B \setminus A).$$

LEMMA 2.1. Suppose given measurable sets $A_1 \subseteq A_2 \subseteq \dots \subseteq A_j \subseteq [0, 1]^2$ and sets H_1, H_2, \dots, H_j satisfying the following for some constant c :

- $\frac{c}{j} \leq m(A_{i+1} \setminus A_i) \leq \frac{2c}{j}$ for $1 \leq i < j$.
- Each set H_i is the intersection of a halfspace with the unit square.
- $m(A_i \oplus H_i) \leq \frac{c}{100j}$ for $1 \leq i \leq j$.
- $m(A_1) \geq c$.
- $m([0, 1]^2 \setminus A_j) \geq c$.

Then there exist point sets $U = \{u_1, u_2, \dots, u_{j-1}\}$ and $V = \{v_1, v_2, \dots, v_{j-1}\}$ such that:

- $u_i, v_i \in A_{i+1} \setminus A_i$ for at least $(j-1)/2$ distinct values of i .
- $d(u_i, v_i) = \Omega(1)$ for $1 \leq i < j$
- $\text{opt}(U \cup V) < 10$.

For space reasons, we present the proof of Lemma 2.1 in the Appendix.

The following lemma accounts for the basic reason why every total ordering of the grid must contain either zigzags or backtracks. It explains this dichotomy in terms of a more fundamental dichotomy: for every 2-coloring of the grid, there is either a halfspace which nearly separates the two color classes, or a strip containing three fairly wide monochromatic intervals such that the middle interval is colored differently from the outer ones.

LEMMA 2.2. Suppose that a square Q of side length s is partitioned into a k -by- k grid of congruent squares Q_{ij} ($1 \leq i, j \leq k$) and that each of these squares is colored either white or black. Let W denote the union of the white squares and let B denote the union of the black squares, and call a set $S \subseteq \mathbb{R}^2$ monochromatic if it is disjoint from at least one of W, B . (If so, the color of S is white if $S \cap B = \emptyset$, black otherwise.) Then, for sufficiently large k , at least one of the following holds:

1. There is a halfspace H such that $m(W \oplus (H \cap Q))/m(Q) \leq 40/k$.
2. There is a line L and a strip σ containing all the squares Q_{ij} which intersect L . This strip has three rectangular intervals $R_1, R_2, R_3 \subseteq Q \cap \sigma$ of length s/k , such that R_2 lies between R_1 and R_3 , and such that the sets R_2 and $R_1 \cup R_3$ are monochromatic and oppositely colored.

For space reasons, we present the proof of Lemma 2.2 in the Appendix.

The proof of Theorem 2.1 relies on the notion of sampling a random line which intersects $[0, 1]^2$. Let us start by specifying a probability measure on such lines and proving some useful properties of this measure.

DEFINITION 3. Let S be a bounded convex subset of \mathbb{R}^2 . A random line through S will refer to a line L which is randomly sampled according to the following procedure: choose θ uniformly at random from the interval $[0, 2\pi]$ and choose a number b uniformly at random from the interval $\{y \cos \theta - x \sin \theta : (x, y) \in S\}$, then take L to be the solution set of the linear equation $y \cos \theta - x \sin \theta = b$. Informally, one first chooses the direction of L uniformly at random and then one chooses L uniformly at random among the set of lines which face in the specified direction and intersect S . The probability measure on the set of lines in \mathbb{R}^2 defined by this sampling procedure will be denoted by Λ_S . When $S = [0, 1]^2$ we will abbreviate this to Λ .

LEMMA 2.3. Suppose $S_0 \subseteq S$ is homothetic to S , i.e. there exists a point $v \in S$ and a positive constant $r \leq 1$ such that

$$S_0 = \{v + r \cdot (u - v) : u \in S\}.$$

The probability that a random line through S intersects S_0 is r , and the distribution of a random line through S conditional on intersecting S_0 is the same as the distribution of a random line through S_0 . More formally, if \mathcal{L} is a set of lines intersecting S_0 and $\Lambda_S(\mathcal{L})$ is defined, then $\Lambda_{S_0}(\mathcal{L}) = r\Lambda_S(\mathcal{L})$.

Proof. If (θ, b) is an ordered pair such that the line $y \cos \theta - x \sin \theta = b$ intersects S_0 , then the probability density assigned to (θ, b) by measure Λ_{S_0} is $1/r$ times the probability assigned to (θ, b) by measure Λ_S . (The probability density of sampling θ is the same under both distributions, and the probability density of sampling b is $1/r$ times greater under distribution Λ_{S_0} because the interval $\{y \cos \theta - x \sin \theta : (x, y) \in S_0\}$ is r times shorter than the interval $\{y \cos \theta - x \sin \theta : (x, y) \in S\}$.)

Proof. [Proof of Theorem 2.1] If $N < k$ the theorem is trivial since $\log_k N = 1$ and this implies that case 2 is always satisfied. Otherwise, replacing N with $k^{\lceil \log_k(N) \rceil}$ if necessary, we may assume that $N = k^r$. For $t = 0, 1, 2, \dots, r-1$, we may partition $[0, 1]^2$ into a $k^t \times k^t$ grid of congruent squares $\{Q_{ij}^{(t)} : 1 \leq i, j \leq k\}$, each of which intersects \mathbb{G}_N in a sub-grid which is isometric to \mathbb{G}_{N/k^t} , up to rescaling. We can subdivide $Q_{ij}^{(t)}$ into a k -by- k grid of congruent squares Q_{ab} ($1 \leq a, b \leq k$). Each of the sets Q_{ab} has nonempty intersection with \mathbb{G}_N , since $N/k^t \geq k$. Hence for each a, b we may choose a point $p_{ab} \in Q_{ab} \cap \mathbb{G}_N$ which precedes all other points of $Q_{ab} \cap \mathbb{G}_N$ in the total ordering \preceq , and we may totally order the squares Q_{ab} according to the relative ordering of the points p_{ab} . For any number ℓ between 1 and k^2 , we may

color the first ℓ squares in this ordering black, and the rest of the squares white. This coloring must satisfy either case 1 or case 2 of Lemma 2.2. If case 1 is satisfied for all ℓ , then define a set $A_i \subseteq Q_{ab}$ for $i = 1, 2, \dots, j = \lceil k/16,000 \rceil$ to be the union of the first $\ell = \frac{1}{4}k^2 + 8000ik$ squares in the ordering. By assumption, there is a set H_i which is the intersection of a halfspace with A_i such that $m(A_i \oplus H_i)/m(Q_{ij}^{(t)})$ is at most $\frac{40}{k} < \frac{1}{400j}$. Scaling up $Q_{ij}^{(t)}$ if necessary so that it is congruent to $[0, 1]^2$, we may apply Lemma 2.7 to establish that there exists a set $S = U \cup V$ satisfying the conclusion of case 1 of Theorem 2.1.

We are left with proving Theorem 2.1 under the following assumption: for every square $Q_{ij}^{(t)}$, if we subdivide it into a k -by- k grid $\{Q_{ab}\}_{i,j=1}^k$ and order the squares as above, there exists an ℓ such that case 2 of Lemma 2.2 applies when the first ℓ squares of the ordering are colored black and the remaining $k^2 - \ell$ are colored white. This means that:

- There is a strip σ of width at least $1/k^{t+1}$ with three rectangular subintervals R_1, R_2, R_3 of length $1/k^{t+1}$;
- There are squares $Q_{ab}, Q_{a'b'}, Q_{a''b''}$ contained in σ , which intersect R_1, R_2, R_3 , respectively;
- $Q_{a'b'}$ does not come between Q_{ab} and $Q_{a''b''}$ in the ordering of the squares of the k -by- k grid.

Let p, p', p'' denote the points $p_{ab}, p_{a'b'}, p_{a''b''}$. If $Q_{ab}, Q_{a''b''} \preceq Q_{a'b'}$, then let L_0 denote the line through p, p'' . If $Q_{a'b'} \preceq Q_{ab}, Q_{a''b''}$, let L_0 denote the line through p' parallel to the boundary of σ . In either case, the reader may check that L_0 is contained in a strip of width $1/k^{t+2}$ which contains a backtrack of length $1/k^{t+1}$. Call this strip $\sigma(Q_{ij}^{(t)})$. Let R_1, R_2, R_3 be the three rectangles defining the backtrack in $Q_{ij}^{(t)}$ and for $a = 1, 2, 3$ let $p_a(Q_{ij}^{(t)})$ be a point of $\mathbb{G}_N \cap R_i$ which precedes all other points of $\mathbb{G}_N \cap R_i$ in the ordering \preceq .

Now let L denote a random line through $[0, 1]^2$ and let \mathcal{Q} denote the set of all squares $Q = Q_{ij}^{(t)}$ ($1 \leq t \leq r, 1 \leq i, j \leq k^t$) such that $L \cap Q$ is contained in $\sigma(Q) \cap Q$. Let S denote the union of all the sets $\{p_1(Q), p_2(Q), p_3(Q)\}$ as Q runs through the elements of \mathcal{Q} . Note that

$$|S| = 3|\mathcal{Q}| \leq 3 \sum_{t=1}^r k^t \leq 6 \cdot k^r = O(N).$$

For each $Q = Q_{ij}^{(t)}$ let $\beta(Q) = 1/k^{t+1}$ if $Q \in \mathcal{Q}$, $\beta(Q) = 0$ otherwise. We claim that

1. $3\text{utsp}(S) + 1 \geq \sum_{i,j,t} \beta(Q_{ij}^{(t)})$,
2. $\text{opt}(S) < 4 + \sum_{i,j,t} 6\beta(Q_{ij}^{(t)})/k$, and

3. $\mathbf{E}(\beta(Q_{ij}^{(t)})) > (c/k^5)k^{-2t}$ for some constant c .

The theorem then follows by choosing a line L such that the random variable $\sum_{i,j,t} \beta(Q_{ij}^{(t)})$ assumes a value at least as great as its expectation.

We prove (1) by assigning a charge of $\beta(Q)$, for each $Q \in \mathcal{Q}$, to a segment of the UTSP tour in such a way that the total charge assigned to the segment s_n, s_1 is at most 1, and the total charge assigned to every other segment s_i, s_{i+1} is at most $3d(s_i, s_{i+1})$. The charging scheme is defined as follows. If $s_1 \in Q$ then we charge $\beta(Q)$ to the segment s_n, s_1 , and we call this a charge of type A. Otherwise, the UTSP tour begins outside of the square Q . Let R_1, R_2, R_3 be the three rectangles which define the backtrack of length $\beta(Q)$ inside Q . If $p_2(Q) \preceq \{p_1(Q), p_3(Q)\}$ then by the definition of a backtrack, it follows that there is a segment s_i, s_{i+1} of the tour such that $d(s_i, s_{i+1}) \geq \beta(Q)$ and $s_{i+1} \in Q$, namely the first segment s_i, s_{i+1} such that s_{i+1} belongs to the portion of $\sigma(Q)$ between R_1 and R_3 . We charge $\beta(Q)$ to this segment, and we call this a charge of type B. Finally, if $\{p_1(Q), p_3(Q)\} \preceq p_2(Q)$, then assume without loss of generality that $p_1(Q) \preceq p_3(Q)$. We assign a charge of $\beta(Q)$ to the first segment s_i, s_{i+1} of the UTSP tour such that $p_1(Q) \preceq s_i$ and the line segment ℓ with endpoints s_i, s_{i+1} intersects R_2 . By the definition of a backtrack, we know that neither endpoint of ℓ belongs to R_2 , and that the length of $\ell \cap R_2$ is at least $\beta(Q)$. We call this a charge of type C, and we label the charge with the ordered pair (R_2, Q) .

Let us now prove that the total of all charges is less than $3\text{utsp}(S) + 1$. Observe that for each t there is at most one $Q_{ij}^{(t)}$ containing s_1 , hence the total charge of type A is less than $1/k + 1/k^2 + \dots = 1/(k-1) \leq 1$. Similarly, for each $i < n$ and each t there is at most one $Q_{ij}^{(t)}$ containing s_{i+1} so the total charge of type B to segment s_i, s_{i+1} is dominated by a geometric progression in which each term is k times the following term. Moreover, no term in this series exceeds $d(s_i, s_{i+1})$ since our scheme never assigns a type-B charge $\beta(Q)$ to segment s_i, s_{i+1} unless $\beta(Q) \leq d(s_i, s_{i+1})$. Therefore the total charge of type B assigned to segment s_i, s_{i+1} is bounded above by

$$d(s_i, s_{i+1}) \sum_{i=0}^{\infty} k^{-i} \leq 2d(s_i, s_{i+1}),$$

hence the sum of all charges of type B is at most $2\text{utsp}(S)$. Finally, to bound the sum of the charges of type C, consider two distinct charges of type C assigned to the same segment s_i, s_{i+1} and suppose that their labels are (R, Q) and (R', Q') . We claim that R and R' have disjoint interiors. This is obvious if Q and Q' have disjoint interiors; otherwise, one of

the squares Q, Q' is contained in the other, so let us assume without loss of generality that $Q' \subseteq Q$. By the definition of a charge of type C, the following properties must be satisfied:

- For any square $Q_{ij}^{(t)}$ which is contained in Q and intersects R , $\{s_i, s_{i+1}\} \preceq S \cap Q_{ij}^{(t)}$ and $\{s_i, s_{i+1}\}$ is disjoint from $Q_{ij}^{(t)}$.
- $S \cap Q'$ contains a point p such that $p \preceq s_i$.

Together, these two properties establish that Q' is disjoint from R , hence R' is disjoint from R as claimed. Now recall that when a charge of type C is assigned to segment s_i, s_{i+1} and labeled with a pair (R, Q) , it means that the line segment ℓ joining s_i and s_{i+1} intersects R in an interval of length at least $\beta(Q)$. As we have now confirmed that distinct type-C charges are labeled with disjoint rectangles, it implies that the sum of the type-C charges to segment s_i, s_{i+1} is bounded above by the sum of the lengths of a set of disjoint subintervals of the line segment ℓ . Hence the sum of the type-C charges assigned to segment s_i, s_{i+1} is at most $d(s_i, s_{i+1})$ and the sum of all charges of type C is at most $\text{utsp}(S)$.

We prove claim 2 by observing that all the points of S are joined together by a Steiner tree consisting of $L \cap [0, 1]^2$ together with a segment joining each $p \in S$ to the closest point of L . The length of the tour $\text{opt}(S)$ is at most twice the length of this Steiner tree. Observe that the length of $L \cap [0, 1]^2$ is less than 2. Also observe that for each $Q \in \mathcal{Q}$ the strip $\sigma(Q)$ has width $\beta(Q)/k$, so Q contributes three points to S and these points contribute at most $3\beta(Q)/k$ to the total length of the Steiner tree. It follows that the total length of the Steiner tree is less than $2 + \sum_{i,j,t} 3\beta(Q_{ij}^{(t)})/k$ which confirms claim 2.

Finally to prove claim 3, it suffices to prove that $\Pr(Q_{ij}^{(t)} \in \mathcal{Q}) > (c/k^4)k^{-t}$ for some constant c . Recalling Lemma 2.3, we see that $\Pr(L \cap Q_{ij}^{(t)} \neq \emptyset) = k^{-t}$ and that, conditional on the event $L \cap Q_{ij}^{(t)} \neq \emptyset$, the distribution of L is the same as that of a random line through $Q_{ij}^{(t)}$. It remains to show that if L is a random line through $Q_{ij}^{(t)}$, then the line segment $L \cap Q_{ij}^{(t)}$ is contained in $\sigma(Q_{ij}^{(t)}) \cap Q_{ij}^{(t)}$ with probability $\Omega(1/k^4)$. Let $[\theta_0, \theta_1]$ be the interval consisting of all angles θ such that there exists a line at angle θ whose intersection with $Q_{ij}^{(t)}$ is contained in $\sigma(Q_{ij}^{(t)}) \cap Q_{ij}^{(t)}$. The fact that the width of $\sigma(Q_{ij}^{(t)})$ is $\Omega(1/k^2)$ times the side length of $Q_{ij}^{(t)}$ has the following consequences: first, $\theta_1 - \theta_0 = \Omega(1/k^2)$; and second, if L is a random line through $Q_{ij}^{(t)}$, the probability that $L \cap Q_{ij}^{(t)} \subseteq \sigma(Q_{ij}^{(t)}) \cap Q_{ij}^{(t)}$, conditional on the event that the

direction of line L belongs to the middle third of the interval $[\theta_0, \theta_1]$, is $\Omega(1/k^2)$. Combining these two consequences, we see that $L \cap Q_{ij}^{(t)} \subseteq \sigma(Q_{ij}^{(t)}) \cap Q_{ij}^{(t)}$ with probability $\Omega(1/k^4)$; as explained earlier, this establishes claim 3.

3 Upper Bound for Universal TSP in Planar Graphs and H -minor-free Graphs

Our approach to obtain $O(\log^2 n)$ competitive ratio for universal TSP in planar graphs relies on a partitioning technique for planar graphs developed by Klein, Plotkin, and Rao [24] which has been further extended by Rao [31]. We also use the concept of *sparse partitions* introduced by Jia et al. [21]. Let's first start by introducing sparse partitions.

DEFINITION 4. ([21]) *A (r, σ, I) -partition of a metric space (V, d) is a partition $\{S_i\}$ of V such that (i) the diameter of every set S_i in the partition is at most $r \cdot \sigma$ and (ii) for every node $v \in V$, the ball $B_r(v) = \{u \in V \mid d(u, v) \leq r\}$ intersects at most I sets in the partition. A (σ, I) -partition scheme is a procedure that computes (r, σ, I) -partition for any $r > 0$.*

Using a partition scheme, Jia et al. obtain a universal TSP as follows.

LEMMA 3.1. ([21]) *Given a metric space (V, d) with n vertices and a polynomial-time (σ, I) -partitioning scheme, there is a polynomial-time algorithm which computes a universal TSP with competitive ratio $O(\sigma^2 I \log_\sigma n)$.*

A *planar metric* is the shortest path metric of a planar graph. More generally, we define *H -minor-free metric* to be the shortest path metric of an H -minor-free graph, i.e., a graph which does not contain the fixed graph H as a minor. We begin with graphs where all edges have unit weight, and then we indicate how graphs with arbitrary edge weights can be treated. Rao [31] was the first who considered planar (H -minor-free) metrics and proved that such metrics with n points can be embedded in ℓ_2 with distortion only $O(\sqrt{\log n})$, as opposed to $O(\log n)$ for general metrics. We use Rao's approach which relies on a partitioning scheme specified in [24]. This allows us to prove the following lemma.

LEMMA 3.2. *There is a $(O(1), O(\log n))$ -partition scheme for planar metrics and more generally for H -minor-free metrics. This partition scheme is computable by a randomized algorithm in polynomial time.*

Combining Lemmas 3.1 and 3.2, we obtain the main result of this section which is an improvement on the competitive ratio $O(\log^4 n / \log \log n)$ of [21] in planar (H -minor-free) metrics.

THEOREM 3.1. *There is a polynomial-time randomized algorithm which returns a universal TSP with competitive ratio $O(\log^2 n)$ in planar or more generally H -minor-free metrics.*

The rest of this section is devoted to the proof of Lemma 3.2.

First we define an r -cover \mathcal{C} to be a collection of subsets of V such that (i) for each $v \in V$, $B_r(v)$ is contained in at least one set in \mathcal{C} , (ii) every vertex is contained in at most $O(\log n)$ sets in \mathcal{C} , and finally (iii) each set in \mathcal{C} has diameter at most $O(r)$.

Before showing the existence of an r -cover \mathcal{C} for planar (H -minor-free) graphs we show that existence of an r -cover \mathcal{C} implies existence of a $(r, O(1), O(\log n))$ -partition \mathcal{P} . We construct a partition \mathcal{P} from \mathcal{C} as follows. For each v , we select an arbitrary set $C_v \in \mathcal{C}$ which contains $B_r(v)$ and then we partition V such that all vertices v with the same C_v are in one set of this partition. Clearly this forms a partition and the diameter of each set in \mathcal{P} is in $O(r)$. Finally, we need to show that for every node v , $B_r(v)$ intersects at most $O(\log n)$ sets in \mathcal{P} . Suppose $B_r(v)$ intersects a set A in \mathcal{P} and consider an $a \in A \cap B_r(v)$. It means node v also belongs to $B_r(a)$ and thus it is in C_a . Thus for distinct sets A that $B_r(v)$ intersects there are distinct sets C_a in \mathcal{C} which contain v . Thus the number of sets in \mathcal{P} that intersect $B_r(v)$ is at most the number of sets in \mathcal{C} that contain v , which is bounded by $O(\log n)$ from the definition of an r -cover.

To construct an r -cover \mathcal{C} for planar (H -minor-free) metrics, we use the following decomposition. Let G be our graph and let Δ be an integer parameter. We take an arbitrary node $v_0 \in V(G)$ and build a breadth first tree. Let $d(v, u)$ denote the distance of u and v in G . We choose an integer $x \in \{0, 1, \dots, \Delta - 1\}$ uniformly at random, and we let $B_1 = \{v \in V(G) : d(v, v_0) \bmod \Delta = x\}$. By deleting the vertices of B_1 from G , the remaining vertices are partitioned into connected components; this is the first level of the decomposition. For each of these components of $G - B_1$, we repeat the same procedure; Δ remains unchanged and r is chosen anew at random but we can use the same r for all the components. Let B_2 be the set of vertices deleted from G in this second round, taken together for all components. The second level of the decomposition consists of the connected components of $G - B_1 - B_2$, and decompositions of levels 3, 4, \dots can be produced similarly. For planar graphs, it suffices to use a 3-level decomposition, and for every fixed graph H , there is a suitable $k = O(|V(H)|)$ such that a k -level decomposition is appropriate for an H -minor-free graph G .

Let $B = B_1 \cup B_2 \cup \dots \cup B_k$ which can be considered as the boundary of the components in

the k -level decomposition. We use the following two properties proved by Rao [31] (see also a survey by Matousek [26][Chapter 15]).

- For each vertex $v \in V(G)$, we have $\min_{b \in B} d(v, b) \geq c_1 \Delta$ with probability at least c_2 , for some constants $0 < c_1, c_2 < 1$, where the probability is with respect to the random choices of the parameters x at each level of the decomposition.
- Each component in the resulting decomposition has diameter at most $O(\Delta)$ in the planar (H -minor-free) metric.

Now, we are ready to construct an r -cover \mathcal{C} . First, we fix $\Delta = r/c_1$ and $\mathcal{C} = \emptyset$. Now we construct the decomposition as above and we add all connected components C_1, C_2, \dots, C_p of $G - B$ to \mathcal{C} . We repeat this process for $5(1/c_2) \log n$ times. We claim that \mathcal{C} which is the union of all connected components during these repetitions is an r -cover with high probability. First we note that each vertex is contained in at most $5(1/c_2) \log n = O(\log n)$ sets of \mathcal{C} . Each set in \mathcal{C} has diameter at most $O(\Delta) = O(r)$, by property (2) of the decomposition above. Finally with high probability, for each vertex $v \in V$, in at least one repetition $\min_{b \in B} d(v, b) \geq c_1 \Delta = r$ (see property (1) of the decomposition above) which means the ball $B_r(v)$ is contained in at least one of the components of $G - B$ and thus in \mathcal{C} . This algorithm results in an r -cover \mathcal{C} with high probability. Since we can check whether a collection of subsets of V is an r -cover or not, we can run the algorithm several times and obtain a real r -cover.

Finally, we note that when we have arbitrary nonnegative weights on edges, we can add artificial vertices by subdividing the appropriate edges. In this way, though the distances can be in a very wide range and not just from 1 to n , still via simulations of artificial nodes, we can find the decomposition mentioned above. However, in constructing the r -cover, we just put the actual (and not artificial) nodes of each connected component in each set. In this way the diameter of each set is still in $O(r)$. Also, since we care about the fact that $B_r(v)$ is contained in one set of the r -cover only when v is an actual (and not artificial) node, still $O(\log n)$ random samples of the decomposition are sufficient.

4 Open Problems

What is the best possible competitive ratio for universal TSP in general metrics? It is at least plausible that the answer should be $\Theta(\log n)$, and we conjecture that a lower bound of $\Omega(\log n)$ can be proven using the shortest-path metric of a constant-degree expander graph.

What is the best possible competitive ratio for universal TSP in the $n \times n$ grid? Our lower bound is still quite far from the upper bound of $O(\log n)$ derived using either [29] or [21], and Bertsimas and Grigni [6] have conjectured that this upper bound is optimal. A related but perhaps easier problem is to obtain a tight bound (up to constant factors) for $\rho_n([0, 1]^2)$. Here, as in the case of the grid, the best known upper bound is $O(\log n)$ and this is conjecturally optimal.

Can the current upper bound of $O(\log^2 n)$ for planar (H -minor-free) metrics be improved? For instance, can one prove an upper bound of $O(\log n \log \log n)$ by combining our construction of r -covers with the proof of Lemma 3.1 via the divide-and-conquer graph decomposition approach of Seymour [33], which has been used in various contexts before (see e.g. [5, 11, 19, 32]).

5 Acknowledgments

We thank Guolong Lin, Rajmohan Rajaraman, and Ravi Sundaram for fruitful discussions.

References

- [1] N. ALON AND Y. AZAR, *On-line steiner trees in the euclidean plane*, in Proceedings of the eighth annual symposium on Computational geometry (SCG '92), New York, NY, USA, 1992, ACM Press, pp. 337–343.
- [2] S. ARORA, *Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems*, J. ACM, 45 (1998), pp. 753–782.
- [3] S. ARORA, M. GRIGNI, D. KARGER, P. KLEIN, AND A. WOLOSZYN, *A polynomial-time approximation scheme for weighted planar graph TSP*, in Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms (SODA '98), Philadelphia, PA, USA, 1998, SIAM, pp. 33–41.
- [4] Y. AZAR, E. COHEN, A. FIAT, H. KAPLAN, AND H. RÄCKE, *Optimal oblivious routing in polynomial time*, in Proceedings of the 35th ACM Symposium on Theory of Computing (STOC '03), ACM Press, 2003, pp. 383–388.
- [5] Y. BARTAL, *On approximating arbitrary metrics by tree metrics*, in Proceedings of the thirtieth annual ACM symposium on Theory of computing (STOC '98), New York, NY, USA, 1998, ACM Press, pp. 161–168.
- [6] D. BERTSIMAS AND M. GRIGNI, *Worst-case examples for the spacefilling curve heuristic for the Euclidean traveling salesman problem*, Oper. Res. Lett., 8 (1989), pp. 241–244.
- [7] A. BORODIN AND R. EL-YANIV, *Online computation and competitive analysis*, Cambridge University Press, New York, 1998.
- [8] J. L. CARTER AND M. N. WEGMAN, *Universal classes of hash functions*, J. Comput. System Sci., 18 (1979), pp. 143–154.
- [9] N. CHRISTOFIDES, *Worst-case analysis of a new heuristic for the traveling salesman problem*, in Symposium on New Directions and Recent Results in Algorithms and Complexity, New York, NY, USA, 1976, Academic Press, p. 441.
- [10] B. C. DEAN, M. X. GOEMANS, AND J. VONDRÁK, *Approximating the stochastic knapsack problem: The benefit of adaptivity.*, in Proceedings of the 45th Symposium on Foundations of Computer Science (FOCS '04), 2004, pp. 208–217.
- [11] G. EVEN, J. NAOR, S. RAO, AND B. SCHIEBER, *Divide-and-conquer approximation algorithms via spreading metrics*, J. ACM, 47 (2000), pp. 585–616.
- [12] A. FIAT AND G. J. WOEGINGER, eds., *Online algorithms*, vol. 1442 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, 1998.
- [13] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability: A Guide to the Theory of NP-completeness*, W. H. Freeman and Co., San Francisco, Calif., 1979.
- [14] M. GRIGNI, *Approximate TSP in graphs with forbidden minors.*, in Proceedings of the 27th International Colloquium on Automata, Languages and Programming, (ICALP '00), 2000, pp. 869–877.
- [15] M. GRIGNI, E. KOUTSOPIAS, AND C. PAPANITRIOU, *An approximation scheme for planar graph TSP*, in Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS'95), Washington, DC, USA, 1995, IEEE Computer Society, p. 640.
- [16] A. GUPTA, M. PAL, R. RAVI, AND A. SINHA, *Boosted sampling: approximation algorithms for stochastic optimization*, in Proceedings of the thirty-sixth annual ACM symposium on Theory of computing (STOC '04), New York, NY, USA, 2004, ACM Press, pp. 417–426.
- [17] M. HAJIAGHAYI, J. H. KIM, T. LEIGHTON, AND H. RäCKE, *Oblivious routing in directed graphs with random demands*, in Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC '05), New York, NY, USA, 2005, ACM Press, pp. 193–201.
- [18] M. HAJIAGHAYI, R. D. KLEINBERG, T. LEIGHTON, AND H. RÄCKE, *Oblivious routing on node-capacitated and directed graphs*, in Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA '05), ACM Press, 2005, pp. 782–790.
- [19] C. HARRELSON, K. HILDRUM, AND S. RAO, *A polynomial-time tree decomposition to minimize congestion*, in Proceedings of the 15th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA '03), 2003, pp. 34–43.
- [20] N. IMMORLICA, D. KARGER, M. MINKOFF, AND V. S. MIRROKNI, *On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems*, in Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '04), Philadelphia, PA, USA, 2004, SIAM, pp. 691–700.
- [21] L. JIA, G. LIN, G. NOUBIR, R. RAJARAMAN, AND

- R. SUNDARAM, *Universal approximations for TSP, steiner tree, and set cover*, in Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC '05), New York, NY, USA, 2005, ACM Press, pp. 386–395.
- [22] D. R. KARGER AND M. MINKOFF, *Building steiner trees with incomplete global knowledge*, in Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS '00), Washington, DC, USA, 2000, IEEE Computer Society, p. 613.
- [23] P. KLEIN, *A linear-time approximation scheme for planar weighted tsp*, in Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS '05), Washington, DC, USA, 2005, IEEE Computer Society. To appear.
- [24] P. KLEIN, S. A. PLOTKIN, AND S. RAO, *Excluded minors, network decomposition, and multicommodity flow*, in Proceedings of the twenty-fifth annual ACM symposium on Theory of computing (STOC '93), New York, NY, USA, 1993, ACM Press, pp. 682–690.
- [25] E. L. LAWLER, J. K. LENSTRA, A. H. G. R. KAN, AND D. B. SHMOYS, *The Traveling Salesman Problem : A Guided Tour of Combinatorial Optimization*, John Wiley and Sons, New York, 1985.
- [26] J. MATOUSEK, *Lectures on Discrete Geometry*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.
- [27] J. S. B. MITCHELL, *Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems*, SIAM J. Comput., 28 (1999), pp. 1298–1309.
- [28] C. H. PAPADIMITRIOU AND M. YANNAKAKIS, *The traveling salesman problem with distances one and two*, Math. Oper. Res., 18 (1993), pp. 1–11.
- [29] L. K. PLATZMAN AND I. JOHN J. BARTHOLDI, *Spacefilling curves and the planar travelling salesman problem*, Journal of the ACM, 36 (1989), pp. 719–737.
- [30] H. RÄCKE, *Minimizing congestion in general networks*, in Proceedings of the 43rd Symposium on Foundations of Computer Science (FOCS '02), IEEE Computer Society, 2002, pp. 43–52.
- [31] S. RAO, *Small distortion and volume preserving embeddings for planar and euclidean metrics*, in Proceedings of the fifteenth annual symposium on Computational geometry (SCG '99), New York, NY, USA, 1999, ACM Press, pp. 300–306.
- [32] S. RAO AND A. W. RICHA, *New approximation techniques for some ordering problems*, in Proceedings of the 9th annual ACM-SIAM symposium on Discrete algorithms (SODA '98), SIAM, 1998, pp. 211–218.
- [33] P. D. SEYMOUR, *Packing directed circuits fractionally*, Combinatorica, 15 (1995), pp. 281–288.
- [34] D. D. SLEATOR AND R. E. TARJAN, *Amortized efficiency of list update and paging rules*, Comm. ACM, 28 (1985), pp. 202–208.
- [35] L. G. VALIANT AND G. J. BREBNER, *Universal schemes for parallel communication*, in Proceedings

of the thirteenth annual ACM symposium on Theory of computing (STOC '81), New York, NY, USA, 1981, ACM Press, pp. 263–277.

A Proof of Lemma 2.1

Proof. Let L_i denote the boundary of the halfspace which defines H_i . Divide the middle third of the line segment $L_i \cap [0, 1]^2$ into $j - 1$ equal segments with endpoints $x_{i,1}, x_{i,2}, \dots, x_{i,j}$. For $1 \leq i, \ell \leq j - 1$, the points $x_{i,\ell}, x_{i+1,\ell}, x_{i+1,\ell+1}, x_{i,\ell+1}$ are the vertices of a convex quadrilateral $Q_{i,\ell}$ with area at least $\frac{1}{10j^2}$ and diameter at most $3/j$. These quadrilaterals fit together into a $(j - 1)$ -by- $(j - 1)$ checkerboard pattern. Color a square of this checkerboard red if the corresponding $Q_{i,\ell}$ does not intersect $A_{i+1} \setminus A_i$. Let n_i denote the number of red squares in column i of the checkerboard, i.e. the number of distinct ℓ such that $Q_{i,\ell}$ is disjoint from $A_{i+1} \setminus A_i$. We have

$$\begin{aligned} \frac{n_i}{10j^2} &\leq \sum_{\ell=1}^{j-1} m(Q_{i,\ell} \setminus (A_{i+1} \setminus A_i)) \\ &\leq m\left(\left(\bigcup_{\ell=1}^{j-1} Q_{i,\ell}\right) \setminus (A_{i+1} \setminus A_i)\right) \\ &\leq m((H_{i+1} \oplus H_i) \setminus (A_{i+1} \oplus A_i)) \\ &\leq \frac{c}{50j} \end{aligned}$$

which implies that $n_i \leq j/5$, hence the number of red squares in the entire checkerboard is at most $(j - 1)j/5$. It follows that at least $(j - 1)/5$ rows have fewer than $j/4$ red squares. Among the indices ℓ such that row ℓ has fewer than $j/4$ red squares, choose two such indices ℓ, ℓ' satisfying $\ell' - \ell \geq (j - 1)/5$. Now for $1 \leq i < j$, choose points $u_i \in Q_{i,\ell}, v_i \in Q_{i,\ell'}$, such that each of u_i, v_i is chosen to belong to the set $A_{i+1} \setminus A_i$ if possible. It is an exercise to check that the conclusions of the lemma are satisfied by these points $\{u_i, v_i\}$.

B Proof of Lemma 2.2

Proof. Without loss of generality, assume that $Q = [0, 1]$. If any row or column satisfies case (2) then we are done. Otherwise, in every row or column, all the white squares are contiguous and all the black squares are contiguous. (We will henceforth call this property of the coloring *monotonicity*.) The reader may check that this has the following consequence, which will be useful later on: in any nine-by-nine sub-grid of the k -by- k grid, there is a monochromatic three-by-three sub-grid.

Suppose now that one edge of the k -by- k grid has two endpoints which are the same color. Without loss of generality, say these are the upper left and right corners and they are both white. The monotonicity property implies that there are numbers i_1, i_2 ($1 \leq$

$i_1, i_2 \leq k$) such that in the left column, the top i_1 squares are white and the rest are black, and in the right column, the top i_2 squares are white and the rest are black. Without loss of generality, assume $i_1 \leq i_2$. We may assume that $i_1 < k$, as otherwise all squares are colored white and the coloring satisfies case 1 in the statement of the lemma. The monotonicity property now implies that:

1. In rows $1, \dots, i_1$, every square is white.
2. In rows $i_1 + 1, \dots, i_2$, there is a block of black squares on the left followed by a block of white squares on the right.
3. In rows $i_2 + 1, \dots, k$, all squares are black.

Let Q be the union of the squares in the $(k - 2)$ -by- $(k - 2)$ grid obtained by deleting all the squares Q_{ij} which lie on the boundary of $[0, 1]^2$. If Q is monochromatically white or monochromatically black, then the coloring satisfies case 1 and we are done. Otherwise, properties (1)-(3) above imply the following: along the boundary of Q all the white squares are contiguous and all the black squares are contiguous. Let p, q be the points on the boundary of Q where the coloring changes from white to black and vice-versa. Note that p is on the left edge of Q by our assumptions (1)-(3). Let σ be a strip of width $10/k$ centered on line pq . The complement of σ consists of two halfspaces H_1, H_2 . Since line pq doesn't intersect the top edge of $[0, 1]^2$, this edge is disjoint from at least one of H_1, H_2 (say H_1). If both halfspaces are monochromatic, we are done. Otherwise, there is either a white square Q_{ij} which intersects H_1 or a black square which intersects H_2 . We will assume the first case (a white square Q_{ij} intersects H_1); the second case is handled by a similar argument. By our assumptions (1)-(3) we know that there is a monochromatically white rectangle R containing Q_{ij} and the upper right corner of $[0, 1]^2$. Now let L be a line parallel to line pq and the boundary line of H_1 , halfway between these two lines, and let σ be the union of all lines L' parallel to L such that there exists a square $Q_{i'j'}$ intersecting both L and L' . The reader may now check that:

- σ has a monochromatically white rectangular interval R_2 of length $1/k$ which is contained in the rectangle R .
- σ has two monochromatically black rectangular subintervals R_1, R_3 such that the two intersection points of L with the boundary of Q are contained in R_1 and R_3 , respectively.
- R_2 is between R_1 and R_3 .

This completes the proof of the lemma in the case that one edge of the k -by- k grid has two endpoints

of the same color. In the remaining case, we may assume without loss of generality that the upper left and lower right corners are white and the upper right and lower left corners are black. Now there are two cases:

1. The 13 -by- 13 sub-grid in each corner of the k -by- k grid is monochromatic.
2. At least one of these sub-grids is not monochromatic.

In case 1, it follows that the diagonal joining the upper left and lower right corners is the mid-line of a strip σ of width $16/k$ which may be divided into 16 parallel strips $\sigma_1, \dots, \sigma_{16}$ of width $1/k$, each containing monochromatically white rectangular subintervals R_1, R_3 of length $1/k$ within distance $O(1/k)$ of the upper left and lower right corners, respectively. Similarly, the diagonal joining the lower left and upper right corners is the mid-line of a strip σ' of width $16/k$ which may be divided into eight parallel strips $\sigma'_1, \dots, \sigma'_{16}$ of width $1/k$, each containing monochromatically black rectangular subintervals R_1, R_3 of length $1/k$ within distance $O(1/k)$ of the lower left and upper right corners, respectively. The intersection of σ_i and σ'_j is a square \tilde{Q}_{ij} which is a rectangular subinterval of length $1/k$ in both of the strips σ_i, σ'_j . If any of the 256 squares \tilde{Q}_{ij} ($1 \leq i, j \leq 16$) is monochromatic, we have confirmed that case 2 in the statement of the lemma holds, and we are done. But the union of these 256 squares is a $(16/k)$ -by- $(16/k)$ square containing a nine-by-nine sub-grid of the k -by- k grid. Earlier we noted that the monotonicity of the coloring implies that this nine-by-nine sub-grid contains a monochromatic three-by-three sub-grid, and at least one of the squares \tilde{Q}_{ij} must lie completely inside the union of the squares in this three-by-three sub-grid.

In case 2, we may assume without loss of the generality that the 13 -by- 13 sub-grid in the upper left corner is not monochromatic, and that there is a black square Q_{ij} with $1 \leq i \leq 13, 1 < j \leq 13$ such that $Q_{i,j-1}$ is white. Let Q be the union of the squares in the $(k - 12)$ -by- $(k - 12)$ sub-grid whose upper left corner is Q_{ij} . Inside Q we have a monotonic coloring, and the top edge of Q has two endpoints which are the same color, i.e. black. We have already confirmed that in the case, there is either a halfspace H satisfying the conclusion in case 1 of the Lemma, or a strip σ with three rectangular intervals $R_1, R_2, R_3 \subseteq Q$ satisfying the conclusion in case 2. In the first case, we are done because enlarging Q to the unit square $[0, 1]^2$ increases the measure of $W \oplus (H \cap [0, 1]^2)$ by no more than $24/k$. In the second case, it is trivial to see that the conclusion of the Lemma continues to hold when we enlarge Q to $[0, 1]^2$.