CS368: Geometric Algorithms Design and Analysis Handout #11
Stanford University Wednesday, 4 June 2003

Lecture #15: 28 May 2003
Topics: Geometric Divide and Conquer
Scribe: Eftychios Sifakis

1 Introduction

Geometric divide and conquer, in analogy to similar techniques commonly used in combinatorial algorithms, is a process used to reduce a given problem to solving a set of simpler similar problems and then combining their results into the solution to the full problem. Nevertheless, in contrast to its combinatorial counterpart, geometric divide and conquer does not split the set of input primitives into disjoint subsets; instead it works by partitioning the *ambient space* of the problem into certain cells we will call *parts*, naturally inducing a partitioning on the individual objects embedded in it. In contrast to the combinatorial setting, a given object may appear in several of the newly generated parts. Because of this, an important goal is to limit the complexity of the data input to the subproblems, while using a reasonable number of parts. The following discussion looks at an instance of that technique applied to the Euclidean plane (the ambient space) for an input consisting of a line arrangement.

2 Problem statement

We will focus our description on partitions of the Euclidean plane E^2 populated with a set of geometric objects, in our case the set of n lines $L = \{l_1, l_2, \dots, l_n\}$.

Given a parameter r we seek a partitioning of the plane into a set of t *generalized triangles* $\Delta_1, \Delta_2, \dots, \Delta_t$ (intersections of three half-planes, possibly unbounded, as illustrated in figure 1) such that the interior of each triangle Δ_i is intersected by at most n/r lines of L. We shall call a partitioning of E^2 with this property a $1/r$-*cutting* of the arrangement $\mathcal{A}(L)$.

Theorem (Cutting lemma) There exists a $1/r$-cut of the arrangement $\mathcal{A}(L)$ that uses $t = \Theta(r^2)$ generalized triangles.

The bound implied by the cutting lemma is in fact optimal in the sense that any partitioning of E^2 (not just with triangles) that achieves a $1/r$-cutting of the arrangement $\mathcal{A}(L)$ has to contain $\Omega(r^2)$ parts. This follows from the fact that any non-degenerate arrangement $\mathcal{A}(L)$ of n lines has $\Theta(n^2)$ intersection vertices. Any part in a potential $1/r$-cutting can be intersected by at most n/r lines, therefore can contain at most $O(n^2/r^2)$

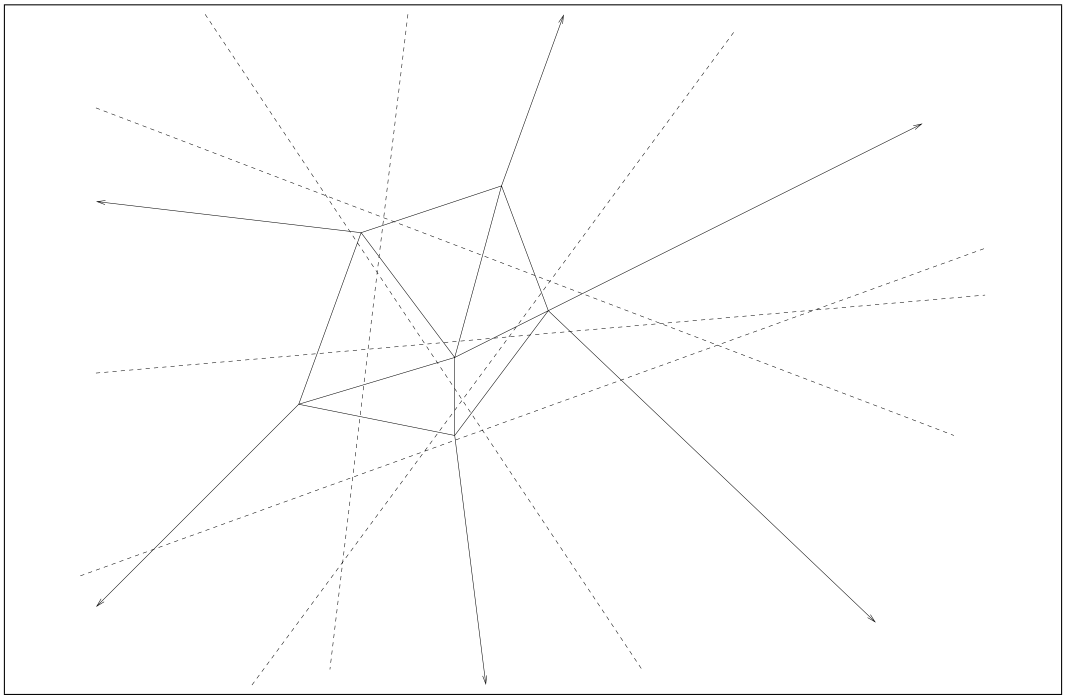

Figure 1: A 1/2-cut on an arrangement of 6 lines

intersection vertices. Therefore, we need at least $\Theta(n^2)/O(n^2/r^2) = \Omega(r^2)$ parts to achieve the desired cut.

We will proceed with a probabilistic argument that yields a weaker $O(r^2 \log^2 n)$ bound but also provides a very simple randomized algorithm for the cutting creation. Subsequently we will illustrate a deterministic construction of a cutting achieving the asymptotically optimal $\Theta(r^2)$ number of parts.

2.1 Probabilistic proof for the $O(r^2 \log^2 n)$ bound

The existence of a $1/r$-cutting with $O(r^2 \log^2 n)$ generalized triangles can be established through a probabilistic argument which also suggests a randomized algorithm for the computation of the particular cutting. This algorithm is based on the method of *self-sampling*, that is using a subset of the data of the problem at hand to construct a solution. In the following proof, some of the edges of the triangles used in the cut may be parts of arrangement lines, which is a valid practice since our definition of a $1/r$-cut imposes bounds on the number of lines intersecting the *interior* of the triangles used.

Given the arrangement L we select a random sample $S \subseteq L$ of the lines by performing s random draws with replacement from the pool of n lines (a line may be selected several times). Subsequently we create a partitioning of the plane into the generalized triangles $\Delta_1, \Delta_2, \ldots, \Delta_t$ by triangulating the polygonal faces of the sub-arrangement $\mathcal{A}(S)$. We have $t = \Theta(s^2)$ since $\mathcal{A}(S)$ contains $\Theta(s^2)$ vertices. If the size of our selected sample is

$s = 6r \ln n$ we can show that the triangulation thus derived has a *positive probability* of being a $1/r$-cut. This shows that there actually exists an appropriate sample S leading to the desired result.

Proof Let us describe a given triangle Δ as *fat* if its interior is intersected by $k \geq n/r$ lines of L. Obviously a partitioning is a $1/r$-cutting if and only if contains no fat triangles. Fix a choice of a fat triangle T. The probability that the interior of T is not intersected by any of the lines in the sample S (which is drawn randomly with replacement) is

$$\left(1 - \frac{k}{n}\right)^s < e^{-\frac{ks}{n}} \leq e^{-\frac{s}{r}} = e^{-6\ln n} = n^{-6} \tag{1}$$

We will call a triangle T *interesting* for L if it can appear as part of a triangulation induced by some sample S. The number of potential interesting triangles is obviously bounded by the number of all triangles that can be formed using 3 of the $\binom{n}{2}$ vertices of the overall arrangement L, and that is less than n^6 triangles. A triangulation induced by some sample S fails to be a $1/r$-cutting if any of its triangles is fat. The probability of that happening is trivially bounded by the possibility that *any* of the interesting triangles for L is fat, that is

$$(\#\text{ of interesting triangles})\,\Pr\{\text{An interesting triangle is fat}\} < n^6 \cdot n^{-6} = 1$$

Thus the case that our triangulation is indeed a $1/r$-cutting has a positive probability, being the complement of the event just described whose probability is less than 1.

The role of the constant factor of 6 in our choice of the sample size $s = 6r \ln n$ becomes clear in the derivation of inequality (1). Apparently, if we were to choose a far greater constant (say $s = 100r \ln n$) the probability our triangulation is not a $1/r$-cutting would be made practically zero (n^{-94}). This leads to the conclusion that any sample S of such a size would almost definitely result in an acceptable cutting, which provides a particularly simple randomized algorithm for the $1/r$-cutting problem (just pick a random sample of adequate size, triangulate the faces of the corresponding arrangement and verify whether it is an acceptable cutting or not).

A substantially more sophisticated argument leads to the stronger result that a sample S of size only $s = O(r)$ would suffice and in fact that particular result extends to higher dimensions as well.

2.2 A deterministic construction for a $O(r^2)$-size cutting

Here we will provide a deterministic argument for the existence of a $1/r$-cut of size $O(r^2)$ for an arrangement L of lines in E^2, assuming that L exhibits no degeneracies. Central to this argument is the notion of the *level* of a given point :

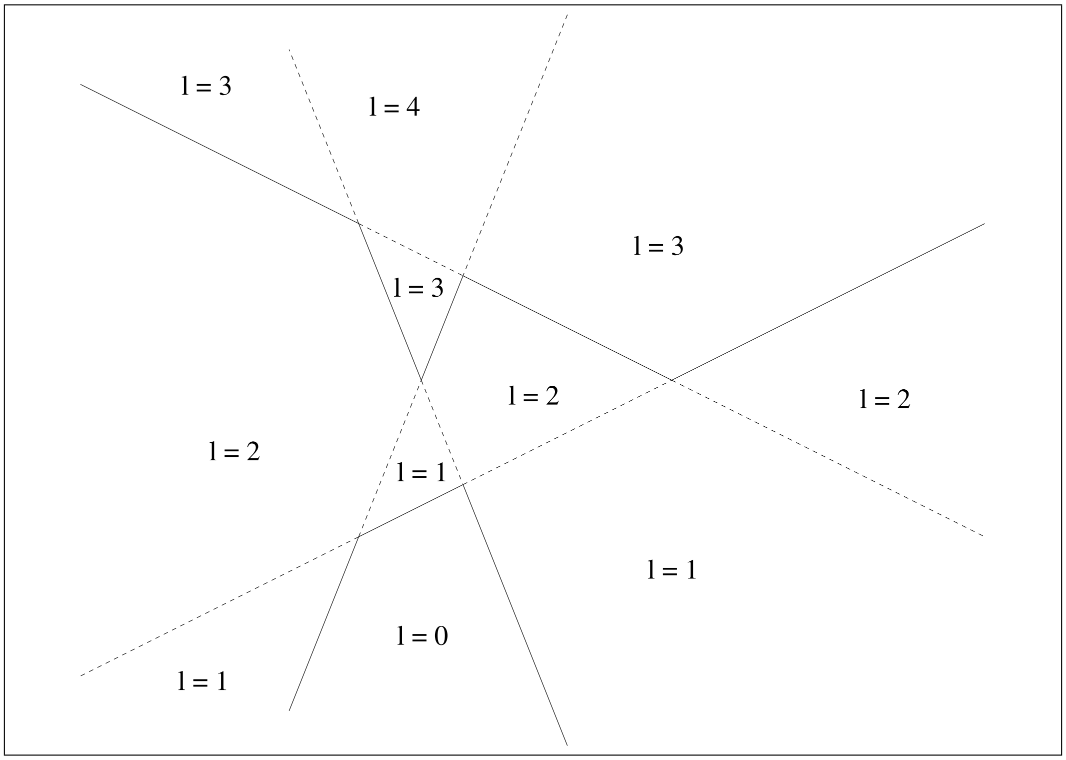

Figure 2: Level boundaries induced by a line arrangement

Definition The *level* of a point $p \in E^2$ is the number of lines of L that are strictly below p. That is the number of lines that are intersected by a vertical open half line connecting p to $-\infty$ in the y - direction.

This definition induces a partitioning of E^2 into $n+1$ *level regions* as illustrated in figure 2. Apparently, all points within a given cell of $\mathcal{A}(L)$ belong to the same level. For $0 \leq k < n$ we shall define the *k-level* of L as the set of all edges of the arrangement having level exactly equal to k (the four distinct levels in figure 2 correspond to the alternating solid/dashed polylines). Each of those level polylines is an x-monotone chain of segments which "turns" at each endpoint of an arrangement edge.

We define a *q-simplification* of a level as a chain of segments that connects every q-th vertex on the level and also includes the extreme (infinite) edges of the level (figure 3). If a given level V consists of t edges, its q-simplification is easily verified to consist of $\lfloor t/q \rfloor + 3$ edges. Let $p_1, p_2, \ldots, p_{t-1}$ be the left-to-right enumeration of the vertices on a given level and $p_i, p_{i+1}, \ldots, p_{i+q}$ a subsequence of vertices connected by a chain of q segments that is shortened into the single edge (p_i, p_{i+q}). We note that the original chain meets with at most $q+1$ distinct lines of the arrangement (including those met at the two extreme vertices), since the chain takes a turn at each line it encounters. Furthermore, each line intersecting the shortened edge must intersect the original chain as well. Thus the shortened edge is intersected by at most $q+1$ lines itself. Since at each intersecting

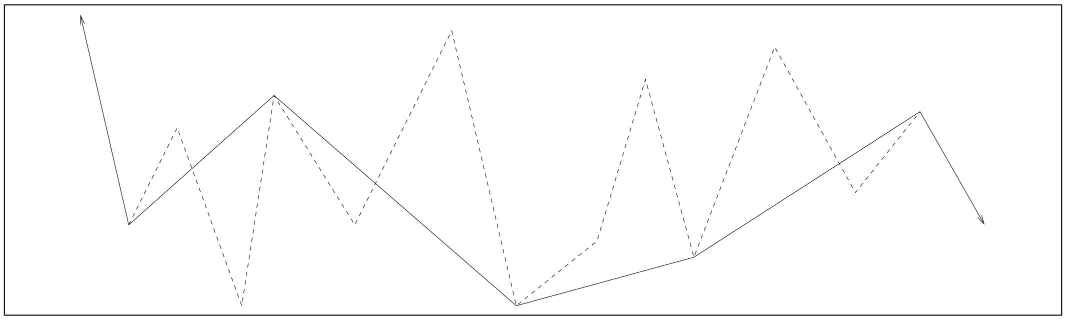

Figure 3: A 3-simplification (solid line) of an arrangement level (dashed line)

line the shortened edge makes a transition into the directly higher or directly lower level (intersecting a line at an arrangement vertex is a degenerate case) and the initial level, say k, at p_i is also the final level at p_{i+q}, the level of any cell the shortened edge crosses into cannot deviate from k by more than $\lceil q/2 \rceil$. Thus, the q-simplification of the k-level V is entirely contained between levels $k - \lceil q/2 \rceil$ and $k + \lceil q/2 \rceil$.

Proof of the Cutting Lemma We can assume that $r \leq n/10$ since if this is not the case we have $n = \Theta(r)$ and we can simply triangulate the entire arrangement $\mathcal{A}(L)$ using $\Theta(n^2) = \Theta(r^2)$ triangles. Set $q = \lceil \frac{n}{10r} \rceil$. We divide the n levels $E_0, E_1, \ldots, E_{n-1}$ of the arrangement into q groups so that the i-th group contains all levels E_j such that $j \equiv i \bmod q$, $i = 0, 1, \ldots, q-1$. The entire arrangement has n^2 edges, therefore some group $E_i, E_{i+q}, E_{i+2q}, \ldots$ must contain no more than n^2/q edges.

Now define P_j to be the q-simplification of level E_{i+jq}. If the number of edges in E_{i+jq} is m_j the simplified level P_j has $\lfloor m_j/q \rfloor + 3$ edges. Thus, the total number of edges of all simplified levels P_j is

$$\sum \left(\frac{m_j}{q} + 3 \right) \leq \frac{1}{q} \sum m_j + 3 \left(\frac{n}{q} + 1 \right) \leq \frac{n^2}{q^2} + 3 \left(\frac{n}{q} + 1 \right) = O \left(\frac{n^2}{q^2} \right) .$$

Moreover, none of the simplified levels can properly intersect, since level P_j is at most allowed to span levels $i + qj - \lceil \frac{q}{2} \rceil$ through $i + qj + \lceil \frac{q}{2} \rceil$. Consider the trapezoidal decomposition (figure 4 on the arrangement of all polylines P_j. Since the P_j's are non-intersecting, the only vertices present are the $O(n^2/q^2)$ already existing on the individual polylines. Thus the trapezoidal decomposition has $O(n^2/q^2)$ cells. The top and bottom boundaries of such a cell are parts of the edges of some P_j's, therefore can be intersected by at most $q+1$ lines each. Similarly, each of the vertical boundaries connects two points that each can deviate no more than $\lceil q/2 \rceil$ from their original levels, which are q apart. Overall, the level difference of two endpoints of a vertical cell boundary cannot be more than $2q + 1$. In total, at most

$$6q + 4 \;\leq\; 10q \;\leq\; \frac{n}{r}$$

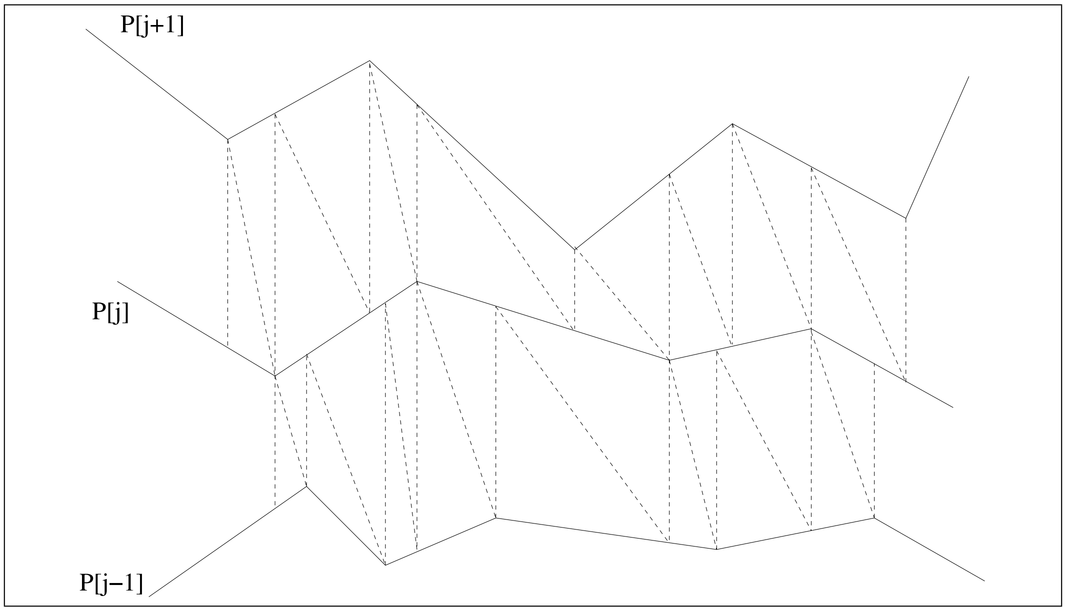

Figure 4: Cut construction by triangulation of the trapezoidal decomposition of the P_j's

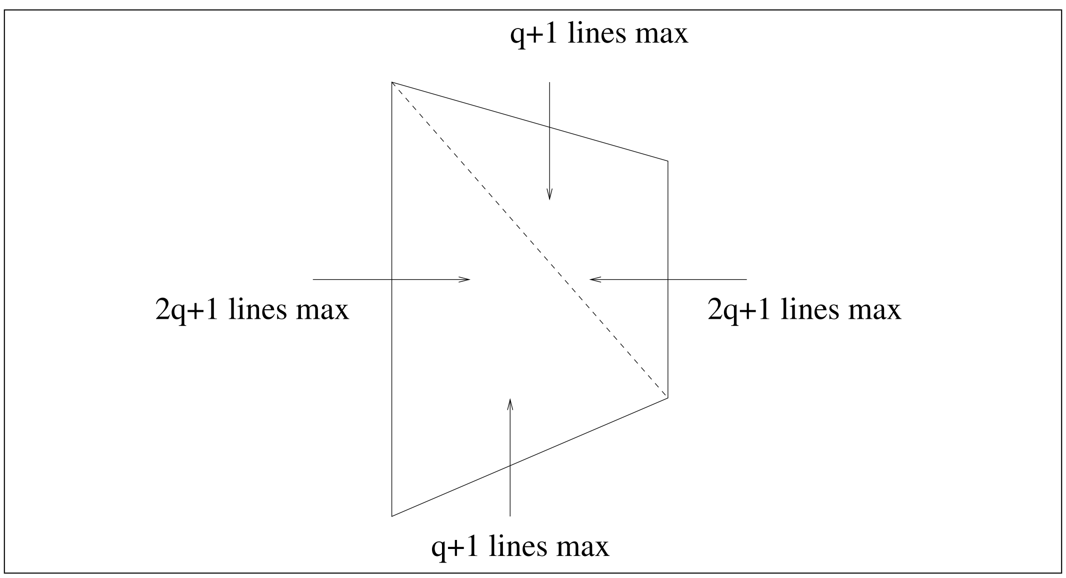

Figure 5: Maximum lines intersecting a trapezoidal cell. Bottom vertices lie in level range $[i+qj-\lceil q/2\rceil, i+qj+\lceil q/2\rceil]$, top ones in $[i+q(j+1)-\lceil q/2\rceil, i+q(j+1)+\lceil q/2\rceil]$.

lines can possibly intersect each cell.

Splitting the trapezoidal cell into two triangles we have the guarantee that they cannot be intersected by more than n/r lines. Therefore, the triangulation thus created is a $1/r$-cut of size $O(r^2)$.

3 An application : The complexity of m faces in the arrangement of n lines

As before, consider an arrangement $\mathcal{A}(L)$ of the n-line set $L = \{l_1, l_2, \ldots, l_n\}$ in E^2. Let us suppose we are interested in m specific of the $\Theta(n^2)$ faces of the arrangement. We will provide a series of bounds for the total number of edges $K(m,n)$ contained in all of those faces of interest

Theorem (Canham's lemma) $K(m,n) \leq 4\left(\begin{array}{c} m \\ 2 \end{array}\right) + n.$

Proof Denote by d_i the number of faces of interest that have an edge on line l_i. Apparently, $K(m,n) = \sum_i d_i$. Furthermore, any pair of faces can lie on at most four common lines, since they are convex cells and have exactly four common tangents. Thus, counting each pair of faces at each line they possibly share, we have

$$\sum_i \left(\begin{array}{c} d_i \\ 2 \end{array}\right) \leq 4\left(\begin{array}{c} m \\ 2 \end{array}\right),$$

which gives

$$K(m,n) = \sum_i d_i \leq \sum_i \left[\left(\begin{array}{c} d_i \\ 2 \end{array}\right) + 1\right] \leq 4\left(\begin{array}{c} m \\ 2 \end{array}\right) + n\,.$$

This bound is optimal when $m = O(\sqrt{n})$ since the n term is dominant. With a slightly different manipulation we can show the following improved result

Theorem $K(m,n) = O(m\sqrt{n} + n)$.

Proof Using the Cauchy-Schwartz inequality we have

$$\sum_i (d_i - 1)\cdot 1 \leq \sqrt{\sum_i (d_i-1)^2 \cdot \sum_i 1^2} \leq \sqrt{\sum_i \left(\begin{array}{c} d_i \\ 2 \end{array}\right)\cdot n} \leq \sqrt{4\left(\begin{array}{c} m \\ 2 \end{array}\right)\cdot n} = O(m\sqrt{n})$$

Thus

$$K(m,n) = \sum_i d_i = \sum_i (d_i - 1) + n = O(m\sqrt{n} + n)\,.$$

Using this intermediate result and the $1/r$-cut divide and conquer technique previously described we can prove the following bound which is optimal without the need for any hypotheses on the values on m, and n

Theorem $K(m,n) = O(m^{2/3}n^{2/3} + n)\,.$

Proof Create a $1/r$-cut of the arrangement using a total of $t = \Theta(r^2)$ generalized trianges. Assign each of the m faces to one of the triangles it intersects (there might be more than one) and denote by m_Δ, n_Δ respectively the number of faces and lines that appear within a given triangle Δ. We have $m = \sum_\Delta m_\Delta$ and by definition of the cut $n_\Delta \leq n/r$. We count each edge of a face that is fully contained within a triangle using the previous bound on that triangle. We account for the edges present in a given triangle that form parts of faces assigned to neighboring trianges by adding a term equal to the total lines n_Δ intersecting the current triangle Δ. With these points considered we have

$$\begin{aligned} K(m,n) &= \sum_\Delta (K(m_\Delta, n_\Delta) + O(n_\Delta)) + O(n) \\ &= \sum_\Delta (O(m_\Delta\sqrt{n_\Delta} + n_\Delta) + O(n_\Delta)) + O(n) \\ &= O(m\sqrt{\frac{n}{r}} + \left(\frac{n}{r}\right) r^2 + \left(\frac{n}{r}\right) r^2 + n) \\ &= O(m\sqrt{\frac{n}{r}} + nr + n)\,. \end{aligned}$$

We minimize the term $m\sqrt{\frac{n}{r}} + nr$ by selecting $r = m^{2/3}n^{-1/3}$. Substituting into the above relation yields

$$K(m,n) = O(m^{2/3}n^{2/3} + n)$$

which is in fact an optimal bound for any choice of m and n.

References

[1] J. Matoušek, *Lectures on discrete geometry*, Springer, 2002.

[2] K.L. Clarkson, H. Edelsbrunner, L.J. Guibas, M. Sharir and E. Welzl, *Combinatorial complexity bounds for arrangements of curves and spheres*, Disc. Comp. Geom., 5:99-160, 1990.