MIT/LCS/TM-153

WORST-CASE AND PROBABILISTIC ANALYSIS OF A GEOMETRIC LOCATION PROBLEM

Christos H. Papadimitriou

February 1980

WORST-CASE AND PROBABILISTIC ANALYSIS OF A GEOMETRIC LOCATION PROBLEM

Christos H. Papadimitriou
Laboratory for COmputer Science
Cambridge, Mass. 02139

December 1979

This Research was supported by N.S.F. Grant MCS77-01193
and a Miller Fellowship

KEY WORDS: Location problems, K-median problem, NP-complete problems, probabilistic analysis of algorithms.

Worst-Case and Probabilistic Analysis of a Geometric Location Problem

Christos H. Papadimitriou*

Laboratory for Computer Science
Massachusets Institute of Technology

Abstract

We consider the problem of choosing K "medians" among n points on the Euclidean plane such that the sum of the distances from each of the n points to its closest median is minimized. We show that this problem is NP-complete. We also present two heuristics that produce arbitrarily good solutions with probability going to 1. One is a partition heuristic, and works when K grows linearly -- or almost so -- with n. The other is the "honeycomb" heuristic, and is applicable to rates of growth of K of the form $K \sim n^{\epsilon}$, $0<\epsilon<1$.

1. Introduction

In this paper we study a classical location problem: Suppose that we are given n points on the plane, and an integer $K<n$. We are asked to choose K of these n points and proclaim them to be centers or medians in such a way that if we add the distances from each point to its closest median, this sum is as small as possible. We call this optimization problem the K-median problem.

In [FH] it is conjectured that this problem is NP-complete (see [Ka1], [GJ], [PS] for definitions concerning NP-completeness). It was already known [KH] that the K-median problem, with a metric not Euclidean but induced by a graph, is indeed NP-complete. The performance of heuristics for the problem with the general metric was analyzed both deterministically and probabilistically in [CFN] and [CNW]. Furthermore, a continuous version of the problem was of concern for a long time in economic location theory [St], [Bo], [FT1].

In Section 2 we show that the Euclidean metric version of the K-median problem is NP-complete, thus proving the Conjecture of [FH]. The result and its proof follow in style the analagous result about the traveling salesman problem[Pa1], [GGJ].

Once an optimization problem is shown NP-complete, the interest of researchers is usually shifted towards the analysis of efficient heuristics that, hopefully, produce good

---

Research supported by N.S.F. Grant MCS77-01193 and a Miller Fellowship

-- though suboptimal -- solutions (see [GJ], Chapter 6). In fact, Karp [Ka2], [Ka3] has initiated research on a probabilistic refinement of this approach: He gave heuristics for several hard combinatorial optimization problems that were efficient (sometimes on the average) and produced solutions which, with probability arbitrarily close to one, were arbitrarily close to the optimum. Such an approach to the K-median problem was taken in [FH].

Before explaining the results of [FH], we need to make an observation about the K-median problem. Any instance of the K-median problem with n points can be solved exhaustively in time proportional to n^{c+1}, where $c = \min(K,n-K)$. Thus, although the problem is NP-complete when K is not fixed but comes as a part of the input, it is polynomial for any fixed K. In fact, if we restrict K to grow extremely slowly with n -- say, $K=\log\log n$ -- then the exhaustive algorithm is not polynomial any more, but it certainly is subexponetial. It therefore makes sense to subdivide the instances of the K-median problem to classes, according to the rate of growth of K with n. [FH] give an "aggregation" heuristic, which is polynomial and has favorable error analysis when K grows slower than $\log n$. (Notice that, for this growth, the problem is most probably not NP-complete since it is solvable by a subexponential algorithm.) As a Lemma, they show that there are constants c_1, c_2 such that the cost of the optimum is almost certainly between $\frac{c_1 n}{\sqrt{K}}$ and $\frac{c_2}{\sqrt{K}}$ when n/K goes to infinity. We improve this result in two ways: We prove it for bounded n/K (Lemma 2), and we find the exact limit $\frac{c_3 n}{\sqrt{K}}$ for n/K going to infinity faster than $\log n$ (Corollary to Theorem 5)

In Sections 3 and 4 we give probabilistic algorithms for fast growths of K. Section 3 is concerned with the case in which K grows faster than $n(\log n)^{-1/3}$. We give a partitioning algorithm for this problem, and we show that when the points are drawn from a Poisson distribution with mean N, then this algorithm has $O(N^3/\log N)$ average execution time, and has a relative error smaller than any $\epsilon>0$ with probability going to 1 as N goes to infinity. Our main tool for proving this is a combinatorial lemma (Lemma 1) which shows that in the optimal solution with probability going to 1, no point is "much" further from its closest median.

In Section 4 we study the case in which K grows slower than $n/\log n$, but faster than $\log n$. We notice that the continuous location problem [Sta], [Bo] becomes relevant. We give a proof that the continuous location problem is asymptotically optimized when the area is divided up into hexagonal cells (this result was apparently known to L. Fejes Toth [FT1], as quoted by Bollobas [Bo]; an independent proof was found by Mordocai Haimovich [Ha]). We then use this result to analyze a very simple "honeycomb" heuristic which, in time $O(n\log n)$ constructs a solution that has relative error smaller than $\epsilon>0$ with probability going to 1. Our probabilistic assumptions are that the n points are n independently and uniformly distributed variables on the unit square.

Finally, in Section 5 we discuss our results, a related recent development that may simplify our approach for the case in which K grows exactly as n, as well as several related open problems.

2. NP-completeness

In order to show that the K-median problem is NP-complete, we have to formulate it first in a more suitable manner. We assume that the points are in the integral lattice, and are given as pairs of integer coordinates.

A familiar problem arises -- see, for example, [Pa2], [GGJ], [Pa1], [PS]: In order to be able to argue that the problem is in $\mathcal{NP}$, we must round the distances down to the closest integer -- i.e., if $p_1 = (x_1,y_1)$ and $p_2 = (x_2,y_2)$, then $\text{dist}(p_1,p_2) = \lfloor\sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}\rfloor$. This is done in order to avoid the difficulty of comparing sums of radicals, a problem of rather mysterious complexity.

We define our problem thus:

K-MEDIAN

"Given a multiset $P = \{p_1,...,p_n\}$ of points with integer coordinates and integers K and L, is there a subset $M = \{m_1,...,m_K\} \subseteq P$ such that $\sum_{j=1}^{n} d_j \leq L$, where $d_j = \min_{m \in M} dist(p_j,m)$?"

This strict definition only serves the purposes of the present Section. In order to apply probabilistic techniques, we will have to make the problem continuous. Even in the constructions of the present Section, we shall allow fractional -- and even irrational -- coordinates. The assumption is that all coordinates -- as well as the limit L -- will be eventually multiplied by a sufficiently large integer and rounded, so that any required precision can be accomplished.

We shall also occasionally define a point in p with a weight w. This will mean that there are w points in P with exactly the same coordinate. If a fractional weight is used, we are assuming that all weights (including unit weights) will be eventually multiplied by a sufficiently (yet polynomially) large integer, so that all weights become integers. In the sequel we shall use weights and non-integral coordinates without further explanation.

Theorem 1 The K-MEDIAN problem is NP-complete

Proof That K-MEDIAN, as defined above, is in $\mathcal{NP}$ is immediate. To prove NP-completeness, we shall reduce to K-MEDIAN the following problem:

EXACT COVER

"Given a set $U = \{U_1, U_2, ..., U_{3n}\}$ and a family $F = \{S_1, ..., S_k\}$ *of subsets of* U with $|S_j| = 3$ $j=1,...,k$, is there a cover $C \subseteq F$ such that $|C| = n$ and $\bigcup_C S_j = U$?"

This problem is known to be NP-complete [GJ].

Before proceeding to the actual reduction, we shall discuss the properties of the configuration of points R (Figure 1a). R is called a row of length m. It has $6m+4$ points, and the two extreme points b,b' have weight m^2 -- this will imply that they have to be medians in any optimal solution. Suppose that we must allocate $m+2$ medians to R. Then the two best solutions are shown in Figure 1a. They both designate b,b' as medians, plus m more points. Either $\{p_{2j_1}, p_{5j_2}, ..., p_{3m-1,j_m}\}$ for some choices of $j_l = 1$ or 2, $l=1,...,m$; or $\{p_{3j_1}, p_{6j_2}, p_{9j_3}, ..., p_{3m,j_m}\}$ again for some choices of j_l. The former is called solution 1 -- it is really a family of solutions -- and the second solution 2. Solution 1 induces to the points of R the partition shown in solid lines in Figure 1a, whereas solution 2 the one with broken lines. Among each resulting group of 6 points the median can be chosen either as in Figure 1b (called an upper median) or in Figure 1c (lower median). Notice that solution 1 is cheaper by 2ϵ where $\epsilon = m^{-4}$. For our reduction, given any instance $U = \{u_1, ..., u_{3n}\}$ and $F = \{S_1, ..., S_k\}$ of the EXACT COVER problem, we shall construct a point set P (weighted) and integer K, as well as a limit L such that P has K medians with cost L or less iff there is an exact cover $C \subseteq F$ of U. P consists of k rows $R_1, ..., R_k$, each of length $3n$, arranged parallel to each other. (Figure 2, schematically). Thus, we can distinguish $3n$ columns of this formation, corresponding to the elements of U.

We shall examine in detail the "window" W of the Figure 2. It is shown in detail in Figure 3.

The spots x,y,w,z of Figure 3 are not points of P, but only possible positions of points. For each window, one of x,y and one of w,z positions is occupied with points of weight n^{-2}. x is occupied iff $U_i \notin S_{j-1}$; y iff $U_i \in S_{j-1}$. Similarly w is occupied iff $U_i \in S_j$; z iff $U_i \notin S_j$.

We now define $K = k(3n+2) + 3n(k-1)$. The first term provides enough medians for all k rows, and the second one median for the q-q' pair in each window W.

L consists of 3 components $L = L_1 + L_2 + L_3$ $L_1 = k(2*1.5 + 3n(2.2 + 2\sqrt{1.04})) - 2n\epsilon$. This cost comes from the k rows. In order to be achieved, all rows must be grouped according to solution 1 or 2, and, because of the $-2n\epsilon$ term, at least n of them must be grouped by solution 1.

$L_2 = 3n(k-1)$, and comes from the cost due to the q *or* q' points. Only one in each

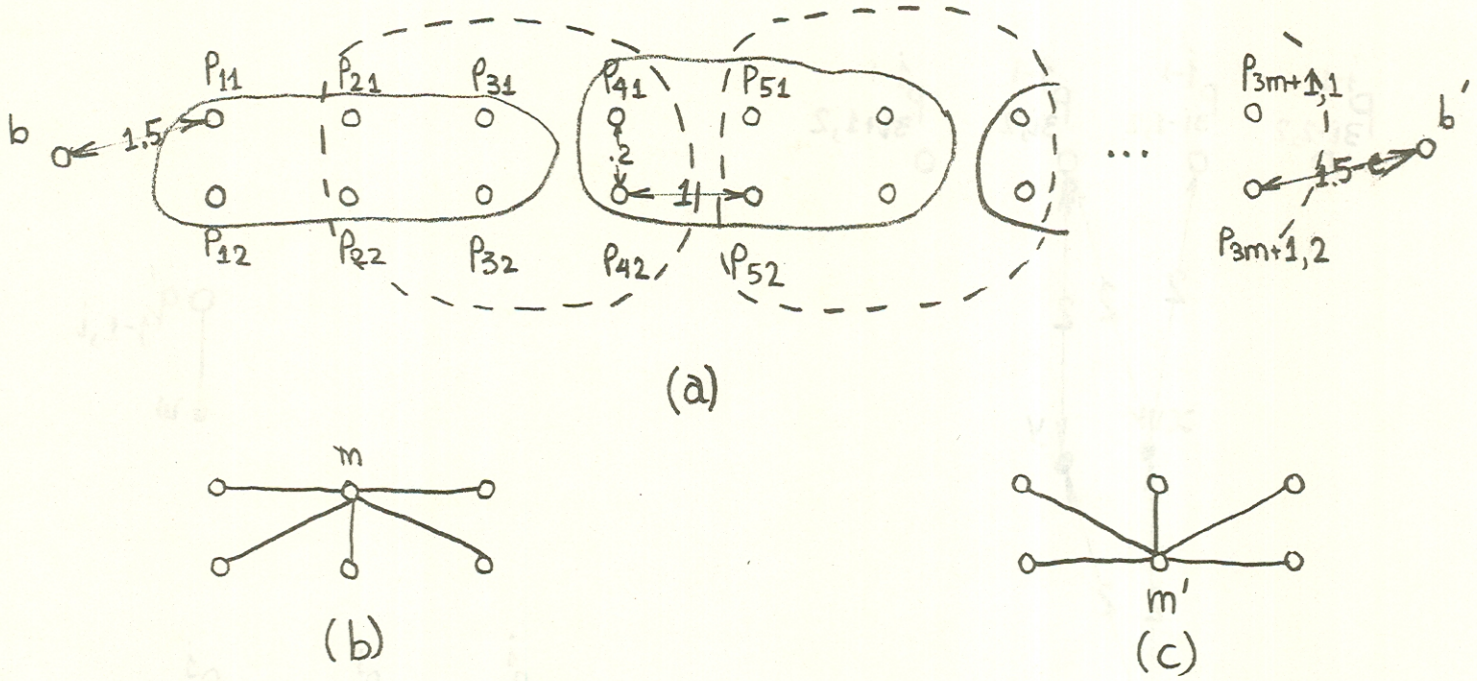

Figure 1

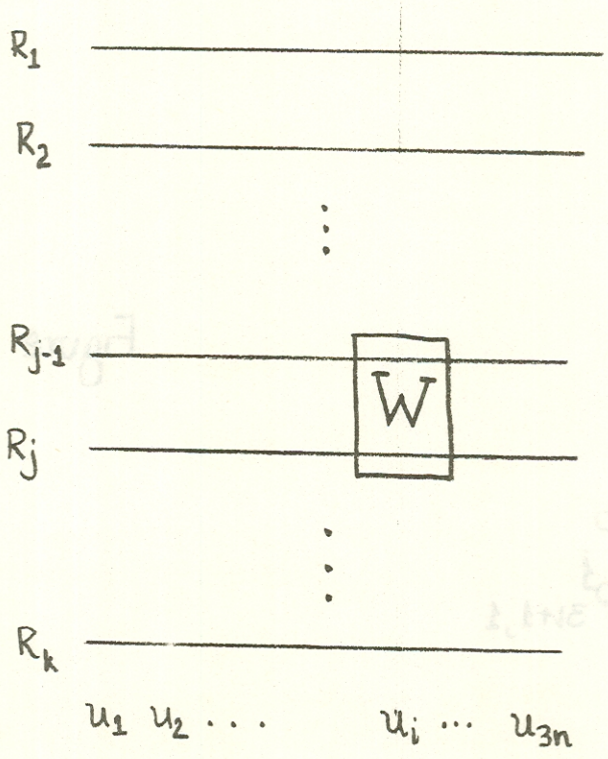

Figure 2

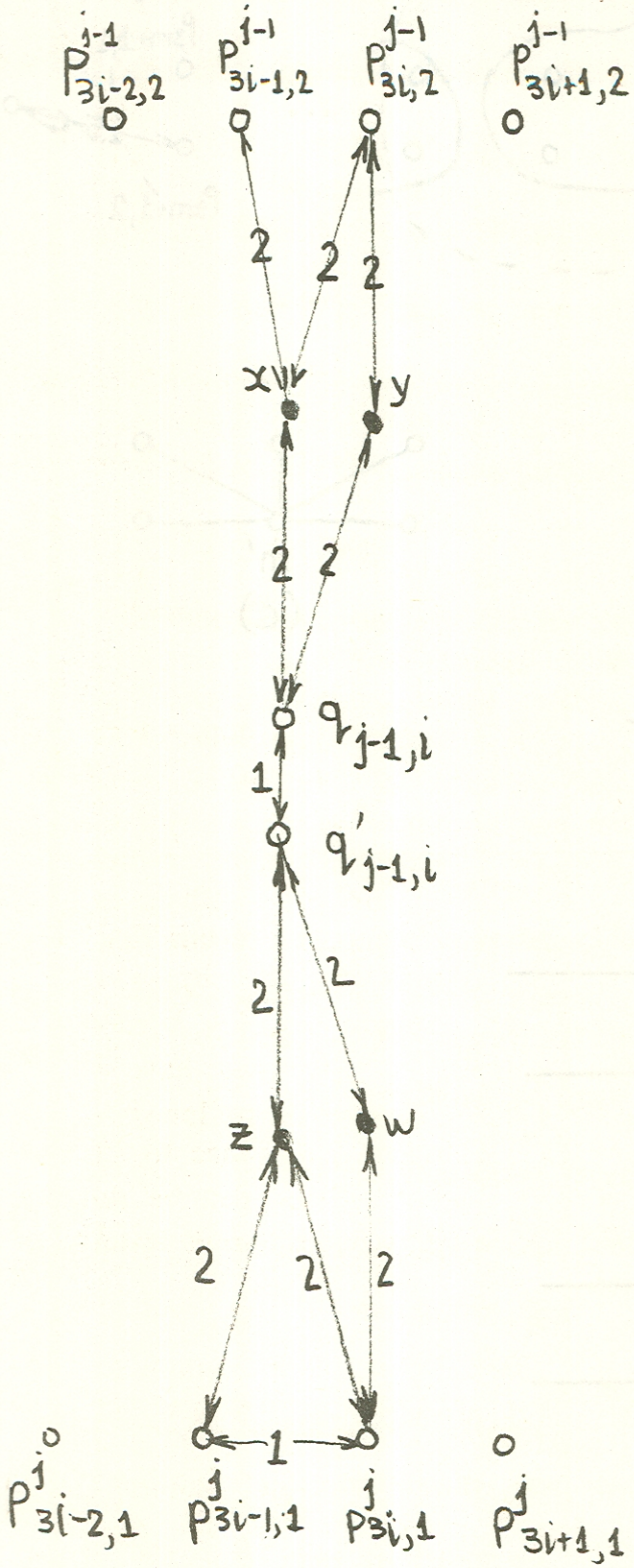

Figure 3

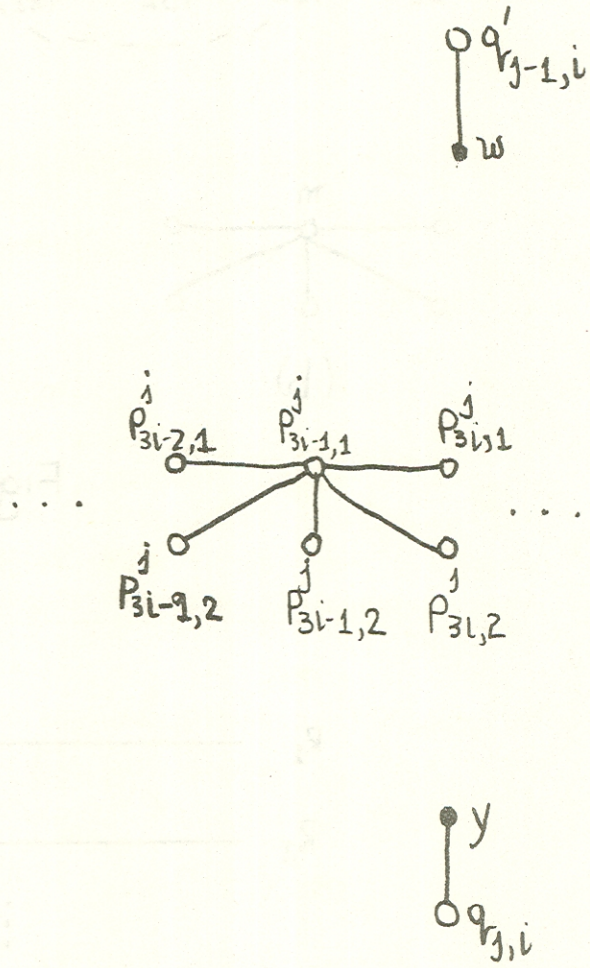

Figure 4

pair will become a median, at a cost of 1 per pair.

$L_3=\frac{12m(k-1)}{n^2}$ is the cost of connecting each of the $6m(k-1)$ points x,y,w,z to the closest q, q' or p point, always 2 away.

Claim There exists $M \subseteq P$ with $|M| = K$ with cost L or less iff F contains an exact cover C of U.

Proof of Claim Suppose that such an M exists. It is not hard to see that $3n+2$ medians must be allocated to each row and one median to each q-q' pair for the cost to be L or less -- to see this it suffices to consider the incremental cost or gain associated with adding or taking away one median. Each row is therefore grouped by solution 1 or 2 . Take the fact that R_j is grouped by solution 1 to mean that $S_j \in C$, where C is the claimed exact cover. In fact, at least n rows must be grouped by solution 1 for L to be achieved, and hence C must contain at least n sets.

Suppose that R_j is grouped by solution 1 (i.e., $S_j \in C$). Consider the i^{th} group, where $U_i \in S_j$. It looks like Figure 4 (or the corresponding lower median configeration).

Since $U_i \in S_j$ both w (above) and y (below) positions are occupied by a point. These two points cannot therefore be connected to their p-median with a link of length 2, as required. So they are connected to the corresponding q-medians. But this means that the x or y point of $q_{j-1,i}$ (respectively, the w or z point of $q'_{j,i}$) must be picked by their corresponding lower (respectively, upper) p-medians in R_{j+1} (respectively, R_{j-1}). By induction, therefore, the i-th group of any row R_k $k<j$ (respectively, $k>j$) must have a lower (respectively, upper) median. Hence this change in this kind -- upper vs. lower -- of medians can occur at most once per column. However, R_j causes this change to all three columns corresponding to the three elements $U_i \in S_j$, and thus there can be no overlaps in the sets S_j of C. So, C contains at least n sets without overlaps: it is an exact cover.

Conversely, suppose that the given instance of EXACT COVER has a solution C. Then we can infer a solution M of the K-MEDIAN problem by allocating $3n+2$ medians to each row, and 1 to each q-q' pair, having each R_j grouped by solution 1 if $S_j \in C$, and by solution 2 otherwise. Finally, let $j(i)$ be the index j of the unique $S_j \in C$ such that $i \in S_j$. We group the i^{th} of R_j by an upper median if $j \geq j(i)$, and a lower median if $j<j(i)$. It follows that the solution has cost L.□

It is more meaningful to consider the special cases of the K-MEDIAN problem, for which K is related to n in a prespecified way. Let $K:\mathbb{N}\rightarrow\mathbb{N}$ be a (polynomially computable) function. By $K(n)$- MEDIAN we mean the set of all instances of

K-MEDIAN for which $K=K(n)$. We can show, from Theorem 1, the following stronger result.

Corollary Suppose that $\min(K(n),n-K(n)) = \Omega(n^{\epsilon})$ for some $\epsilon>0$. Then the $K(n)$-MEDIAN problem is NP-complete.

Proof Standard "padding" arguments.□

Corrollary 1 is in a sense the strongest possible result, with the present state of our understanding of complexity theory, because if either $n-K(n)$ or $K(n)$ grows slower than n^{ϵ} for all ϵ, the $K(n)$-median problem can be solved by a subexponential algorithm.

3 The Linear Case

In this Section we consider the case in which $K(n) = \lfloor \alpha n \rfloor$ for some $\alpha<1$. Informally, this means that a fixed fraction of the customers are proclaimed medians, and therefore each median will be, on the average, responsible for a constant number of customers. We consider point sets $P = \{p_1,...,p_n\}$ drawn from a Poisson process of intensity N on the unit square. As a result, the distributions of points in any two prespecified non-overlapping subregions of the unit square are independent n is a random variable with expected value N.

We divide the unit square into $Q = \lceil \sqrt{N}/\log N \rceil^2$ equal smaller squares $S_1,...,S_Q$ each of side $\lceil \sqrt{N}/\log N \rceil^{-1}$ and containing approximately $\log N$ points on the average. If $M \subseteq P$, we let $f_M(p_j) = m \in M$ iff $\text{dist}(p_j, m) < \text{dist}(p_j, m')$ for all $m' \in M-\{m\}$. With probability 1, f_M is well-defined for all $M \subseteq P$. We let $d_j^M = \text{dist}(p_j, f_M(p_j))$ and $C(M) = \sum_{j=1}^{n} d_j^M$, the cost of the set M of medians. Once we have fixed $S_1,...,S_Q$ we shall define the separable cost of M. Let $S(p_j)$ denote the square S_i among $S_1,...,S_Q$ for which $p_j \in S_i$. We define for $p_j \in P$, $g_j^M = \min_{\substack{m \in M \\ S(m)=S(p_j)}} dist(p_j, m)$. Finally, the separable cost of M is defined as $C'(M) = \sum_{j=1}^{n} g_j^M$. Thus, informally, $C'(M)$ is the cost of M under the additional restriction that customers must go to medians in the same square S_i.

We shall prove the following:

Theorem 2 Let $\hat{M}$ be the optimal solution of P. Then $C'(\hat{M}) - C(\hat{M}) = o(N^{1/2})$ with probability $1-o(1)$

All asymptotic statements are meant as N goes to infinity. Thus, "with probability

1-o(1)" means "with probability going to 1 as N goes to infinity".

We first need the following Lemma:

Lemma 1 There is a constant $c_1>0$ such that $\max_j d_j^{\hat{M}} < c_1(\alpha^3 N)^{-1/2}$ with probability *1-o(1)*.

Proof It is clear that, with probability *1-o(1)*, $n \geq \frac{N}{2}$, and thus $|\hat{M}| \geq \lfloor \frac{\alpha N}{2} \rfloor$. If $m \in \hat{M}$, let $A(m) = \{p_j : f_{\hat{M}}(p_j) = m\}$. It follows that, with probability *1-o(1)*, there exist at least $\lfloor \frac{\alpha N}{4} \rfloor$ medians $m \in M$ with $|A(m)| \leq \frac{2}{\alpha}$. It is therefore obvious that, for some constant $c_2>0$, two of these medians -- say, m and m' -- are closer to each other than $c_2(\alpha N)^{-1/2}$ with probability *1-o(1)* (To see this, divide the unit square into $\lfloor \sqrt{\lfloor \frac{\alpha N}{4} \rfloor} - 1 \rfloor^2$ equal squares; two of the $\lfloor \frac{\alpha N}{4} \rfloor$ medians are bound, by the pigeonhole principle, to fall in the same of these squares, and hence they cannot be further apart than the diameter of the square. This argument gives $c_2 = \sqrt{8}$; $c_2 = \frac{4}{3}\sqrt{12}$ is possible).

Suppose now that one of the points $p_k \in P$ has $d_k^{\hat{M}} > 2c_2(N\alpha^3)^{-1/2}$. Then we claim that $C'(\hat{M} - \{m\} \cup \{p_k\}) < C(\hat{M})$ -- absurd, since $\hat{M}$ is optimum. To prove our claim, we shall construct a mapping

$f: P \to \hat{M} - \{m\} \cup \{p_k\}$ such that $\sum_{j=1}^{n} dist(p_j, f(p_j)) < C(\hat{M})$. f is defined as follows:

$f(p_k) = p_k$.
$f(p_j) = m'$ if $p_j \in A(m)$ and $p_j \neq p_k$.
$f(p_j) = m''$ if $p_j \in A(m'')$, $m'' \neq m$ and $p_j \neq p_k$.

Thus $C(\hat{M}) - \sum_{j=1}^{n} dist(p_j, f(p_j)) = d_k^{\hat{M}} - \sum_{p_j \in A(m)} (dist(p_j, m') - dist(p_j, m)) \geq$

(by the triangle inequality)

$\geq d_k^{\hat{M}} - |A(m)|\, dist(m, m') \geq$

(since $|A(m)| < \frac{2}{\alpha}$ and $dist(m, m') \leq c_2(\alpha N)^{-1/2}$)

$\geq d_k^{\hat{M}} - 2c_2(N\alpha^3))^{-1/2} > 0$

This proves the Lemma with $c_1 = 2c_2$.□

Proof of Theorem 2 Lemma 1 implies that with probability *1-o(1)*, all points $p_j \in P$

such that $S(p_j) \neq S(f_{\hat{M}}(p_j))$ lie in a "corridor" of width $2c_1(\alpha^3 N)^{-1/2}$ around the perimeters of the small squares (Figure 5). Using this, we shall show how to modify the optimal solution so as to make it separable, with a total increase in cost which is $(o(N^{1/2}))$.

The main idea is the following: It is clear that the total number of points in P that lie within these "corridors" is $o(N)$ with probability $1-o(1)$ --since each square has side asymptotically $\sim\sqrt{\frac{\log N}{N}}$, whereas the width of the corridor is $\sim\sqrt{\frac{1}{N}}$. We shall show that we can assign each of these points to a median $\hat{M}$ in its own square which is, on the average, $O(N^{-1/2})$ away.

The details are as follows: Let us divide each square S_1 into $4\lceil\sqrt{\log N}\rceil$ triangular "slices" as shown in figure 6. The unit square is now divided into $R = 4\lceil\sqrt{\log N}\rceil Q$ triangles $t_1,\ldots,t_R$ and the corresponding trapezoids $r_1,\ldots,r_R$ *(see Figure 6).* If $p_j \in P$, we denote by $r(p_j)$ the trapezoid it is in, or by $t(p_j)$ the triangle it is in -- exactly one is well-defined. Let b_j be a random variable denoting $|P \cap r_j|$; $j=1,\ldots,R$; and let $q_j \in P$ be the point in t_j that is closest to the basis of t_j parallel to the side of the square, while h_j is the distance of q_j from the basis (see Figure 6).

It is immediate that each t_j has area J/N, where $J = c_7\sqrt{\log N}$, and thus the probability that $P \cap t_j = \emptyset$ *is* e^{-J}. Thus we may assume that no $P \cap t_j$ is empty (i.e., q_j exists) with probability $(1-e^{-J})^R \geq 1-Re^{-J} = 1-o(1)$. We now claim that, with probability $1-o(1)$, we can assign the points of $P \cap r_j$ to the median $f_{\hat{M}}(q_j) \in \hat{M}$, and repeat this for all r_j's, at a total incremental cost that does not exceed $c_3\sqrt{\frac{N}{\log N}}$, for some $c_3 > 0$. This would settle the theorem.

The cost of connecting each point p of the b_j points in $r_j \cap P$ to $f_{\hat{M}}(q_j)$, $dist(p, f_{\hat{M}}(q_j))$ can be bounded from above as follows (with probability $1-o(1)$).

$$dist(p, f_{\hat{M}}(q_j)) \leq dist(p,q_j) + dist(q_j, f_{\hat{M}}(q_j))$$

$$\leq \sqrt{2}(c_1(\alpha^3 N)^{-1/2} + h_j) + c_1(\alpha^3 N)^{-1/2}$$ Thus the total increase in cost is bounded by

$$C'(\hat{M}) - C(\hat{M}) \leq c_4(\alpha^3 N)^{-1/2} \sum_{j=1}^{R} b_j + \sqrt{2} \sum_{j=1}^{R} b_j h_j$$

The area of each r_j is equal to $c_5\alpha^{-3/2}/N$ for some constant c_5; thus the b_j's are

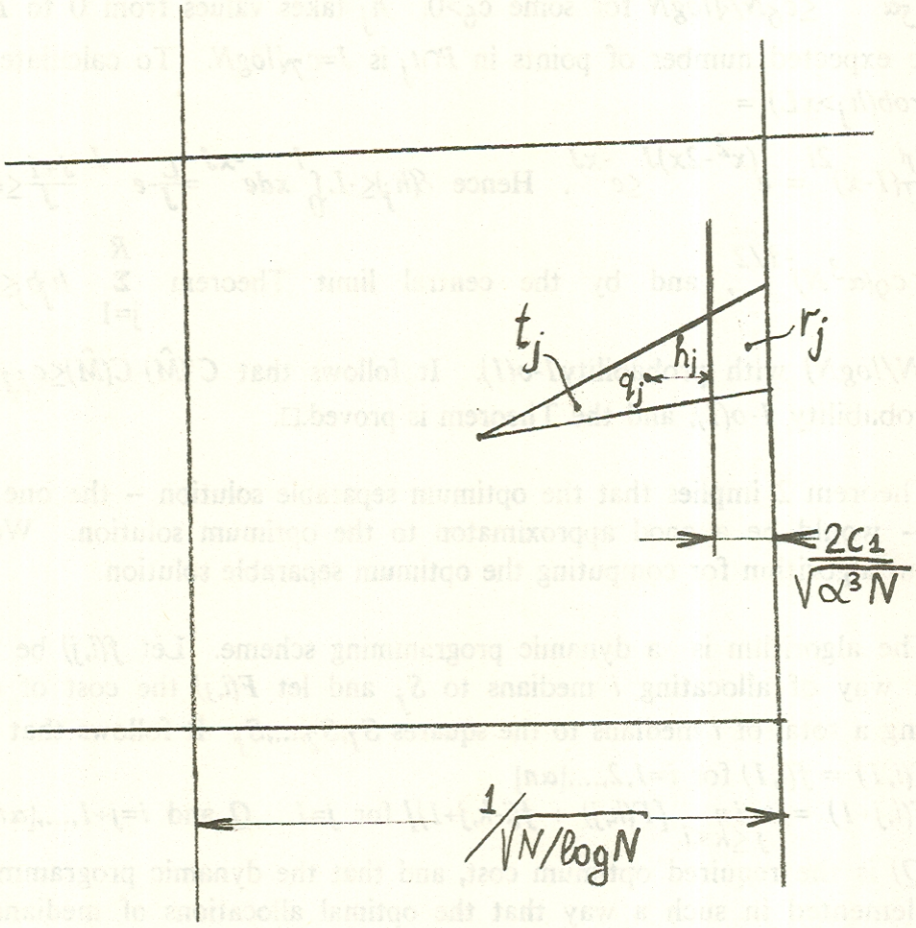

Figure 6

identically and independently distributed random variables with mean $c_5\alpha^{-3/2}$, and hence, by the central limit Theorem, with probability $1-o(1)$ $\sum_{j=1}^{R} b_j \le 2Rc_5\alpha^{-3/2} \le c_6 N/\sqrt{\log N}$ for some $c_6>0$. h_j takes values from 0 to $L=1/2\lfloor\sqrt{\frac{N}{\log N}}\rfloor^{-1}$, and the expected number of points in $P\cap t_j$ is $J=c_7\sqrt{\log N}$. To calculate $\mathcal{E}(h_j)$, we note that, $prob(h_j>xL) =$

$e^{-J}\sum_{i=0}^{\infty}\frac{J^i}{i!}(1-x)^{2i} = e^{(x^2-2x)J} \le e^{-xJ}$. Hence $\mathcal{E}(h_j) \le -L\int_0^1 x de^{-xJ} = \frac{L}{J}-e^{-J}\frac{J+1}{J} \le c_8 N^{-1/2}$. Thus $\mathcal{E}(h_j b_j) \le c_9(\alpha^3 N)^{-1/2}$, and by the central limit Theorem $\sum_{j=1}^{R} h_j b_j \le 2RC_9(\alpha^3 N)^{-1/2}$ $\le c_{10}\sqrt{(N/\log N)}$ with probability $1-o(1)$. It follows that $C'(\hat{M})-C(\hat{M}) \le c_{11}(\alpha^3 \log N/N)^{-1/2}$ with probability $1-o(1)$, and the Theorem is proved.□.

Theorem 2 implies that the optimum separable solution -- the one that minimizes $C'(M)$ -- would be a good approximaton to the optimum solution. We shall describe below an algorithm for computing the optimum separable solution.

The algorithm is a dynamic programming scheme. Let $f(i,j)$ be the cost of the optimal way of allocating i medians to S_j and let $F(i,j)$ the cost of optimal way of allocating a total of i medians to the squares $S_1, S_2, \ldots, S_j$. It follows that

$F(i,1) = f(i,1)$ for $i=1,2,\ldots,\lfloor\alpha n\rfloor$

$F(i,j+1) = \min_{j \le k < i} [F(k,j) + f(i-k,j+1)]$ for $j=1,\ldots,Q$ and $i=j+1,\ldots,\lfloor\alpha n\rfloor$. Notice that $F(\lfloor\alpha n\rfloor, Q)$ is the required optimum cost, and that the dynamic programming scheme can be implemented in such a way that the optimal allocations of medians are recovered after the determination of the optimal cost. This dynamic programming scheme requires $Q\lfloor\alpha n\rfloor$ evaluations of the function $f(i,j)$. Notice that $\mathcal{E}(Q\lfloor\alpha n\rfloor) = O(\frac{N^2}{\log N})$.

For any i and fixed j, evaluating $f(i,j)$ can clearly be done in a number of operations bounded by $d2^k$, where $d>0$ and $k=|P\cap S_j|$. Thus the expected number of operations for evaluating $f(i,j)$ is

$\sum_{k=0}^{\infty}\frac{d*2^k(\log N)^k}{k!}e^{-\log N} = de^{\log N} = d*N$ Therefore, this algorithm has an expected total number of steps $O(\frac{N^3}{\log N})$.

To evaluate the relative error of the algorithm, we need the following Lemma.

Lemma 2 If $\hat{M}$ is the optimum solution, then, with probability $1-o(1)$, $C(\hat{M}) \geq c_1\sqrt{N}$ for some $c_1>0$.

Proof Let us fix a value for n; the points in P_j are thus uniformly distributed in the unit square. Divide the unit square into $\lceil\sqrt{\frac{n}{p}}\rceil^2$ squares, where $p<<1$ is to be determined. How many squares contain exactly one point? This is at least as much as a binomial variable with n trials and probability $1-p$. So, by Chebyshef's inequality, at least $(1-2p)n$ squares have one point, with probability $1-o(n^{-1})$.

Consider these $(1-2p)n$ squares, each of side $a=\lceil\sqrt{\frac{n}{p}}\rceil^{-1}$. In how many of these squares the point is further than pa from the boundary? By Chebyshef, at least $(1-3p)^3 n$ with probability $1-o(n^{-1})$. Choose $(1-3p)^3=(1-\frac{\alpha}{2})$. Suppose that a_j is the distance from p_j to the point in $P-p_j$ closest to p_j and assume that P has been ordered in decreasing a_j's. The above argument suggests that, with probability $1-o(1)$, $a_j \geq \sqrt{p/n}$ for $j \leq (1-\frac{\alpha}{2})n$. Thus $\sum_{j=n-\lfloor\alpha n\rfloor+1}^{n} a_j \geq \sqrt{\frac{\alpha p}{2}}\sqrt{n}$. It is, however, clear that $\sum_{j=n-\lfloor\alpha n\rfloor+1}^{n} a_j$ is a lower bound for $C(\hat{M})$. Since $n \geq \frac{N}{2}$ with probability $1-o(1)$, the Lemma follows.□

Combining the proceeding observations with Theorem 2 and Lemma 2 above, we obtain

Theorem3 The dynamic programming algorithm above requires an expected number of operations $O(\frac{N^3}{\log N})$ and produces a set M of medians with relative error $\frac{C(M)-C(\hat{M})}{C(\hat{M})} = o(1)$ with probability $1-o(1)$.□.

We note that the same approach works for slightly sublinear growths of $K(n)$ in particular $K(n)=\omega(n/\log^{1/3} n)$. In the next Section we handle in a very different way more sublinear growths of $K(n)$, namely $K(n)=n^{\epsilon}$ for some $0<\epsilon<1$.

4. The Honeycomb Heuristic

In this Section we consider the case in which $K(n)$ grows slower than $n/\log n$ but faster than $\log n$. For simplicity, we shall only deal explicitly with the family of growth $k(n)=\lfloor n^{\epsilon}\rfloor$ for some ϵ $0<\epsilon<1$; generalization to the above mentioned range is immediate from our proofs.

A simple consequence of this growth is that both $K(n)$ and $n/K(n)$ go to infinity quite fast -- as $n^{\delta}, \delta>0$. That $n/K(n)$ grows quite fast means, intuitively, that the average median would, asymptotically, be responsible for a "continuum" of points. It is

therefore natural to consider the following continuous, deterministic version of the problem.

"Place K points $m=\{m_1,...,m_K\}$ in the unit square so that

$$C^*(M)=\sum_{j=1}^{K}\int_{D_j} dist(m_j,A)dA \text{ is minimized",}$$

where $D_j=\{x\in[0,1]^2: dist(x,m_j)\le dist(x,m_i)$ for all $i\neq j\}$ are the Dirichlet cells (Voronoi polygons of Shamos [Sh])of the point m_j with respect to the point set M (see Figure 7 for an example). D_j is then the locus of all points that are closest to m_j; it is easy to see that D_j is always a convex polygon (since it is the non-empty intersection of the half-planes $dist(x,m_j)\le dist(x,m_i)$, $i\neq j$).

Let R_n be the regular n-gon of unit area, and let c be its center. We let $\gamma(n)=\int_{R_n} dist(c,A)dA$. A simple calculation yields $\gamma(n)=[arcsinh(t)+t\sqrt{1+t^2}]/6\sqrt{t^3n}$, where $t=tan(\frac{\pi}{n})$. Some values of $\gamma(n)$ are given in Table 1. The following Lemma is shown in [FT]

Lemma 3 Let S_n be a convex n-gon with unit area, and let $p\in S_n$. Then $\int_{S_n} dist(p,A)dA\ge\gamma(n)$□ .

A tedious calculation yields

Lemma 4 The function $(\gamma(x))^{-2}$ is concave for $x\ge3$.□

The following Lemma says, essentially, that if we partition the unit square into many polyogons, then each polygon must, on the average, have fewer than 6 sides. It is a rather surprising application of Euler's formula, and can be found, for example, in Heawood's 5-color proof [He].

Lemma 5 Let $\{S_1,...,S_k\}$ be a partition of the unit square in k convex polygons, with $m_1,...,m_k$ sides, respectively. Then $\sum_{j=1}^{k} m_j\le6k-2$.□

Using these Lemmas, we can show the following Theorem:

Theorem 4 If $|M|=K$, then $C^*(M)\ge K^{-1/2}\gamma(6)$.

Proof By definition, $C^*(M)=\sum_{j=1}^{K}\int_{D_j} dist(m_j,A)dA$ where D_j is the Dirchlet cell of m_j with respect to M. If D_j is and n_j-gon, we have by Lemma 3 $C^*(M) \geq \sum_{j=1}^{K} |D_j|^{3/2}$ $\gamma(n_j) = \sum_{k=3}^{\infty} \gamma(k) \sum_{j\in G_k} |D_j|^{3/2}$, where $G_k = \{j: n_j = k\}$. Recall that $\sum_{i=1}^{n} x_i^{3/2}$ with $\sum_{i=1}^{n} x_i$ fixed is minimized when all x_i's are equal.

Thus, $C^*(M) \geq \sum_{k=3}^{\infty} \gamma(k)|G_k|(\frac{\Sigma|D_j|}{|G_k|})^{3/2}$.

The latter expression can be written as

$$C^*(M) \geq \sum_{k=3}^{\infty} \frac{|G_k|}{\gamma(k)^2} \left(\sum_{j\in G_k} |D_j| \frac{\gamma^2(k)}{|G_k|}^{3/2} \right)$$

By the same argument, the right hand side of the inequality above can be bounded from below

$C^*(M) \geq \{\sum_{k=3}^{\infty} (|D_j|\frac{\gamma^2(k)}{|G_k|}) (\frac{|G_k|}{\gamma^2(k)})^{3/2} / \{\sum_{k=3}^{\infty} |G_k|/\gamma^2(k)\}^{1/2} = (\sum_{j=1}^{K} \gamma^{-2}(n_j))^{-1/2}$. However, since $\gamma^{-2}(x)$ is concave, (Lemma 4) we have, by Jensen's inequality

$C^*(M) \geq K^{-1/2} \gamma(\frac{\sum_{j=1}^{K} n_j}{k})$ which, by Lemma 5 and since γ is non-increasing, gives:

$C^*(M) \geq K^{-1/2} \gamma(6)$.□

Theorem 4 implies that asymptotically as K grows the optimal partition of the square into polygons with respect to the valuation C^* is the one that consists of regular hexagons -- ignoring the effects of the boundaries of the unit square. A very simple and efficient heuristic for solving the $K(n)$-median problem for this range of growths of $K(n)$ is immediately suggested by Theorem 4.

Given $P=\{p_1,...,p_n\}$

a. Find $K=K(n)$

b. Tile the plane with hexagons $H_1,H_2,...$ each of area $1/K$. Choose those hexagons H_j for which $H_j \subseteq [0,1]^2$. Let the set of their centers be $H=\{h_1,...,h_k\}$, $k\leq K$.

c. Define the set of medians $M=\{m_1,...,m_k\}\subseteq P$ by $dist(m_j,h_j)\leq dist(p,h_j)$ for all $p\in P$.

The remaining part of this Section is a probabilistic analysis of this honeycomb heuristic. Our probabilistic assumptions are that n is fixed and $P=\{p_1,...p_n\}$ consists of points independently and uniformly distributed over the unit square. The following

n	$\gamma(n)$	$[1/\gamma(n)]^2$
3	.545	3.37
4	.478	4.37
5	.444	5.07
6	.424	5.55
7	.412	5.88
8	.404	6.12
9	.398	6.30
10	.394	6.43
15	.384	6.77
20	.381	6.90
30	.378	6.99
40	.377	7.03
50	.377	7.04
100	.376	7.06
1000	.376	7.07
∞	$\frac{2}{3\sqrt{\pi}} \simeq .376$	$\frac{9\pi}{4} \simeq 7.07$

Table 1 - Values of $\gamma(n)$

Lemma is shown in [FH]

Lemma 6 If $\hat{M}$ is the optimum solution, there is a constant c such that $C(\hat{M}) \geq cnK(n)^{1/2}$ with probability $1-o(1)$.□

We also need the following Lemma, which is a specialized result about multinomial distributions. Suppose that we have divided the unit square into 2^{2m} equal small squares, where $n2^{-2m}=n^{\delta}$ *for some* $\delta>0$.

Lemma 7 With probability $1-o(1)$ each small square will contain N_j points of P, where $|N_j-n^{\delta}|=o(n^{\delta})$.

To prove Lemma 7 we need a purely probabilistic fact:

Lemma 8 Let b be a binomially distributed random variable with probability 1/2 and n trials. Then $prob(|b-n/2|>n^{.5+\delta} \leq e^{-n\delta}$ for large enough n.

Proof It is well-known that

$prob(b=j) = B_{j,n} = \binom{n}{j}2^{-n}$. By Stirling's formula

$$B_{j,n}=\frac{E(n,j)}{\sqrt{\pi n/2}}\binom{n}{j}^{j+1/2}\binom{n}{n-j}^{n-j-1/2}2^{-n}$$

where $E(n,j)$ is an error term

$$E(n,j) = 1+\frac{1}{1/12}\left(\frac{n^2-nj+j^2}{nj(n-j)}\right)+\ldots = O(1).$$

Let $j=\frac{n}{2}-x$, $x\geq 0$ (for $x\leq 0$, similarly). Then for some $c_1>0$,

$$B_{j,n}\leq c_1\left(n^{1/2}\left(\frac{n/2-x}{n/2}\right)^{-n/2+x}\left(\frac{n/2+x}{n/2}\right)^{-x-n/2}\right)$$

$$=c_1\left(n^{1/2}\left(1-\frac{x^2}{(n/2)^2}\right)^{-n/2}\left(\frac{1-\frac{x}{n/2}}{1+\frac{x}{n/2}}\right)^{x}\right.$$

Letting $z=\frac{x}{n/2}$

$$B_{j,n}\leq c_1\left(n^{1/2}(1-z^2)^{-n/2}\left(\frac{1-z}{1+z}\right)^{x}\right)$$

$$let\ A(j,n) = (1-z^2)^{-n/2}\left(\frac{1-z}{1+z}\right)^{x}$$

$$logA(j,n) = \frac{n}{2}\left(z^2+\frac{z^4}{z}+\frac{z^6}{3}+\ldots\right)$$

$$+x\left(-z-\frac{z^2}{z}-\frac{z^3}{3}\ldots-z+\frac{z^2}{z}-\frac{z^3}{3}+\ldots\right)$$

$$= -\frac{n}{2}\sum_{i=0}^{\infty} a_i z^{zi}, \text{ with } a_j=\frac{1}{j(2j-1)}$$

Hence $logA(j,n) \leq -\frac{n}{2}z^2 = -\frac{2x^2}{n}$, and

$B(j,n) \leq c_1 n^{1/2} e^{-\frac{2x^2}{n}}$. Hence

$prob(|b-\frac{n}{2}| \geq n^{.5+\delta}) \leq nB(\lceil \frac{n}{2} - n^{.5+\delta} \rceil, n) \leq c_1 n^{3/2} e^{-n^{2\delta}} \leq e^{-n^{\delta}}$ for large enough n.□

Proof of Lemma 7 Fix an $\alpha > 0$, and define

$$p_j = je^{-n^{\alpha\delta/2}}$$

$$f_j = \sum_{i=1}^{j} 2^{-i}(n2^{-(j-1)})^{.5+\alpha}$$

We shall prove by induction on k that each of the 2^k subregions -- squares if k even, rectangles is k odd -- contain between $n2^{-k}+f_k$ points, with probability $1-p_k$. Notice that this settles the Lemma.

To start the induction, the assertion holds for $k=0$. Suppose that it holds for $k=j$. Let A_{j+1} be one of the 2^{j+1} subregions, and let A_j be the unique subregion among the 2^j that contains it (if j is even, A_j is a square and A_{j+1} is its half rectangle; if j is odd, A_j is a rectangle consisting of two squares, and A_{j+1} is one of them). For N large enough, we have by Lemma 8:

$prob(A_{j+1}$ contains $\frac{N}{2} \pm N^{.5+\alpha}$ pts $|$ A_j contains $N) \geq 1-e^{-N\alpha}$ Hence

$prob(A_{j+1}$ has $n2^{-(j+1)} \pm (\frac{1}{2}f_j + (n2^{-j}+f_j)^{.5+\alpha}$ pts $|A_j$ has $n2^{-j}+f_j) \geq 1-e^{-(n2^{-j}-f_j)^{\alpha}}$.

But this can be rewritten as

$prob(A_{j+1}$ has $n2^{-(j+1)} \pm f_{j+1}$ pts $|$ A_j has $n2^{-j} \pm f_j) \geq 1-e^{-(n2^{-j}-f_j)^{\alpha}}$.

$prob($all A_{j+1}'s have $n2^{-(j+1)} \pm f_{j+1}$ pts $|$ all A_j's have $n2^{-j} \pm f_j) \geq$

$(1-e^{-(n2^{j}-f_j)^{\alpha}})^{2^{j+1}} \geq 1-2^{j+1}e^{-(n2^{j}-f_j)^{\alpha}} \geq 1-e^{-n^{\alpha\delta/2}}$ Thus $prob($All A_{j+1}'s have $n2^{-(j+1)} \pm f_{j+1}$ pts)

$\geq prob($All A_j's have $n2^{-j} \pm f_j$ pts$)(1-e^{-n^{\alpha\delta/2}})$

$\geq$ (by induction hypothesis) $(1-p_j)(1-e^{-n^{\alpha\delta/2}})$

$\geq 1-e^{-n^{\alpha\delta/2}} - p_j \geq 1-p_{j+1}$.□

Lemma 7 implies that with probability $1-o(1)$ every subdivision of the unit square into not more than $n^{1-\epsilon}$ equal squares will contain only squares that have n^{ϵ} points,

plus or minus a lower order term. It will be very useful in evaluating how well the continuous deterministic problem approximates the K-median problem.

The error analysis is rather simple, though tedious, since it involves the numbers $C(\hat{M})$, $nC^*(\hat{M})$,$C(M)$,$nC^*(M)$,$nC^*(H)$ and $\gamma(6)n/\sqrt{K}$ -- here $\hat{M}$ is the optimal solution, M is the solution found by the heuristic, and H is the set of hexagonal centers. C is our ordinary cost valuation, whereas C^* is its continuous counterpart. Our strategy is shown in Figure 9. A solid undirected line between A and B means that we shall show -- in the Lemma whose number is indicated on the line -- that $|A-B|=o(n/\sqrt{k(n)})$ with probability $1-o(1)$. A broken directed line from A to B means that $A \geq B$. Once we establish all this, it is immediate that $C(M)-C(\hat{M})=o(n/\sqrt{K(n)})$.

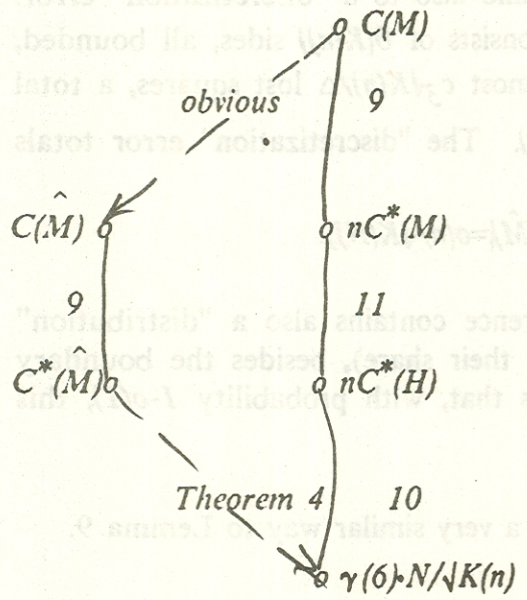

Figure 9.

Lemma 9 $|C(\hat{M}-nC^*(\hat{M})| = 0(n/\sqrt{k(n)}$ with probability (1-o(1).

Proof Recall that $K(n) = \lfloor n^{\epsilon} \rfloor$. Let us divide the unit square into 2^{2m} equal squares, where $m = \left\lfloor \frac{\log_2(n^{\delta})}{2} \right\rfloor$ for some $\epsilon < \delta < 1$. Let each such square have a side of length Δ.

We shall show that, with probability $1-o(1)$ no point $p_j \in P$ has $d_{\hat{M}}(p_j) \geq c_1/\sqrt{K(n)}$ for some constant c_1. Suppose that such a point p_j existed. Then, with probability $1-o(1)$, the disc with center p_j and radius $c_1/3\sqrt{2K(n)}$ has at least $\frac{\pi c_1^2 n}{18K(n)}$ points of P in it. This is established by considering the small $\Delta \times \Delta$ squares falling within this disc, and applying Lemma 7. However, one could then make p_j into an additional median, at a

savings of at least $\frac{\pi n}{54}\left(\frac{c_1}{\sqrt{K(n)}}\right)^3$. Therefore, one can argue in a manner identical to Lemma 1, that there exist two medians in $\hat{M}$ within distance $(2K(n))^{-1/2}$, each having at most $2n/K(n)$ points. By choosing δ appropriately, we establish that $\max_j d_{\hat{M}}(p_j) \le c_1/\sqrt{K(n)}$ with probability $1-o(1)$.

Let us define now yet another valuation $\bar{C}$ of any set of medians. $\bar{C}$ is a discretized version of C^*: we assume that one point of weight $n\Delta^2$ is in the center c_j of each Δ Δ square S_j and calculate

$\bar{C}(M) = \sum_{j=1}^{K} n\Delta^2 \sum_{S_i \subseteq D_j} dist(m_j, c_i)$. Let us calculate $|\bar{C}(\hat{M}) - nC^*(\hat{M})|$. This difference is due to "lost" squares along the perimeters of the D_j's, and also to a "discretization" error. The perimeter does not exceed $c_2\sqrt{K(n)}$ -- since it consists of $O(K(n))$ sides, all bounded, by the above remark, by $c_1/\sqrt{K(n)}$; so there are at most $c_3\sqrt{K(n)}/\Delta$ lost squares, a total error $\le n\Delta^2 c_3(\sqrt{K(n)}/\Delta)\max_j d_{\hat{M}}(p_j) = c_4 n\Delta = o(n/\sqrt{K(n)})$. The "discretization" error totals to, at most, $n\Delta\frac{\sqrt{2}}{2}$, also $o(n/\sqrt{K(n)})$. Thus $|\bar{C}(\hat{M}) - nC^*(\hat{M})| = o(n/\sqrt{K(N)})$.

Let us now evaluate $|\bar{C}(\hat{M}) - C(\hat{M})|$. This difference contains also a "distribution" error (squares that have more or fewer points than their share), besides the boundary and discretization errors. However, Lemma 7 says that, with probability $1-o(1)$, this new error is $o(n)\max_j d_{\hat{M}}(p_j) = o(n/\sqrt{K(n)})$.□

That $|n\bar{C}(\hat{M}) - C^*(\hat{M})| = o(n/\sqrt{K(n)}$ can be shown in a very similar way to Lemma 9.

We now turn to

Lemma 10 $n|C^*(H) - \gamma(6)/\sqrt{K(n)}| = o(n/\sqrt{K(n)}$ with probability $1-o(1)$

Proof Each of the hexagons in the tiling $\{H_1, H_2, \ldots\}$ has area 1/K(n), and therefore side $c_4/\sqrt{K(n)}$. Thus, there are at most $c_5\sqrt{K(n)}$ hexagons that cross the boundary of the unit square. It is contributions from these squares that increase $C^*(H)$ away from $\gamma(6)/\sqrt{K(n)}$. However, each hexagon on the boundary adds at most $c_6(K(n))^{-3/2}$ to $C^*(H)$ and thus the total deviation is $n|C^*(H) - \gamma(6)/\sqrt{K(n)}| \le c_7 n/K(n) = o(n/\sqrt{K(n)})$.□

Lemma 11 $n|C^*(M) - C^*(H)| = o(n/\sqrt{K(n)})$ with probability $1-o(1)$.

Proof The difference between $C^*(M)$ and $C^*(H)$ is due to the "displacement" of the medians from the centers of the hexagons to the points closest to the centers. Now,

each center of hexagon falls in one of the $\Delta*\Delta$ squares ($\Delta=n^{-1/2+\delta}$, arbitrary $\delta>0$), and we know that, with probability *1-o(1)*, there is at least one point from *P* in each square (Lemma 7). Thus this desplacement is for no center greater than $\sqrt{2}\Delta$, with probability *1-o(1)*. The total error due to displacement is therefore no greater than $n\sqrt{2}\Delta$. Taking $\delta<\frac{1-\epsilon}{2}$, we prove the Lemma. □

We can finally show:

Theorem 5 The honeycomb heuristic constructs in time 0(nlogn) a set *M* of medians having relative error
$(C(M)-C(\hat{M})/C(M\hat{)}) = o(1)$ with probability *1-o(1)*.

Proof The error analysis follows from Lemmata 9 through 11 and Theorem 5. For the time bound, we have to show that in time 0(nlogn) we can find for each point in a set *H*, $|H|\leq N$, the closest to it from another point set *P*, $|P|\leq n$. This, however, is possbile by the Voronoi techniques of Shamos [Sh].□

By Theorem 5, we can explicitly calculate the exact limit of the optimal cost for this range of *K(n)*:

Corollary For $K(n)=\omega(\log n)$, $K(n)=o(n/\log n)$, we have $prob(|\frac{C(\hat{M})\sqrt{K(n)}}{n} - \gamma(6)|>\epsilon)=1-o(1)$ for all $\epsilon>0$.□

5. Discussion

We note that no NP-completeness results are known for the following two variants of the K-median problem:

(a)One does not restrict *M* to be a subset of *P*. That is, the *K* medians can be chosen to be totally new points.

(b)The min-max version of the *K*-median problem.

We conjecture that both problems are NP-complete.

As far as probabilistic analysis of heuristics is concerned, our results leave open two regions of the spectrum of growth of *K(n)*:

(a) $K(n)=c\log n$

(b) $\frac{c_1 n}{\log n}\leq K(n)\leq \frac{c_2 n}{(\log n)^{1/3}}$.

For the case that $K(n)=\lfloor \alpha n \rfloor$, a very interesting recent result by J.Michael Steele [Ste] may simplify our approach considerably. Steele proved that, if any valuation f mapping finite point sets to the reals, satisfies the following properties

(a) f is Euclidean, i.e., linear and invariant under translations
(b) f is monotone, i.e., $f(P \cup \{p_{n+1}\}) \geq f(P)$
(c) f has bounded variance, under the uniform distribution
(d) f is subadditive, i.e. if $\{S_i\}_{i=1}^{m}$ is a partition of the unit square into squares of total perimeter L, $f(P) \leq \sum_{i=1}^{m} f(P \cap s_i) + 0(L)$. Then with probability 1

$\lim_{n \to \infty} \frac{f(P_1,\ldots,P_n)}{\sqrt{n}} = B_f$ a constant. Using this Theorem, Steele gives a simple derivation of the Beardwood, Halton, and Hammersley Theorem [BHH]. Notice that the valuation $f_\alpha(P)=min\{C(M): |M|=\lfloor \alpha |P| \rfloor\}$ for some α $0<\alpha<1$, does not satisfy conditions *(b)* and *(d)* above; however, Steele claims that the conclusion of the Theorem still holds for F_α [Ste]. Explicit proofs of this fact have actually appeared [Ha,Ho] This suggests the following simple partition heuristic for the $K(n)=\lfloor \alpha n \rfloor$ case.

1)Partition the unit square into $\sim n/\log n$ smaller squares
2)Solve the *K(n)-median* problem for each of the smaller squares

As a result of Steele's Theorem, the optimum such restricted separable solution gives a solution that is asymptotically very close to the exact optimum. We note, however, that our approach is still necessary for $K(n)=o(n)$.

Acknowledgment

Much of the motivation for this work, as well as more technical influence, came from discussions with Dorit Hochbaum and Dick Karp. The conjecture which eventually became Theorem 4 is due to the insights of Gerard Cornuejols. Some of the calculations were made using the symbolic manipulation system MACSYMA at M.I.T.

References

[AHU] A.V.Aho, J.E. Hopcroft, J.D. Ullman "The Design and Analysis of Computer Algorithms, Addison-Wesley, 1978.

[BHH] J.Beardwood, J.H. Halton, J.M. Hammersley "The Shortest Path Through Many Points" Proc. Cambridge Philo. Soc., 55, pp 29-327,1959.

[Bo] B. Bollobas "The Optimal Arrangement of Producers" J. London Math. Society(2) 6, pp 417-421, 1973.

[CFN] G. Cornuejols, M.L. Fischer, G.L. Nemhauser "Location of Bank Accounts to Optimize Float:An Analytic Study of Exact and Approximate Algorithms", Management Science, 23, 8, pp 789-810, 1978.

[CNW] G. Cornuejols, G.L. Nemhauser, L.A. Wolsey "Worst-case and Probabilistic Analysis of Algorithms for a Location Problem", TR375, IEOR, Cornell University, 1978.

[FH] M.L. Fisher, D.S. Hochbaum "Probabilistic Analysis of the Euclidean K-median Problem", to appear in Mathematics of Operations Research, 1979.

[FT1] L. Fejes Toth Lagerungen in der Ebene, auf der Kugel und im Raum, Berlin, 1953.

[FT2] G. Fejes Toth "Sum of Moments of Convex Polygons", Acta Math. Acad. Scient. Hungaricae, Vol. 24 (3-4), pp 417-421, 1973.

[GGJ] M.R. Garey, R.L. Graham, D.S. Johnson "Some NP-complete Geometric Problems", Proc. 8th Annual Symposium on Theory of Computing, pp. 10-27, 1976.

[GJ] M.R. Garey, D.J. Johnson Computers & Intractability: a Guide to the Theory of NP-completeness, Freeman, 1979.

[Ha] M.Haimovich manuscript, M.I.T. May 1979.

[He] P.J. Heawood, "Map Colour Theorem" Quart. J. Math. 24, pp. 332-338, 1890

[Ho] D.S. Hochbaum, "The Probabilistic Asymptotic Properties of Some Combinatorial Geometric Problems", manusctipt, Carnegie-Mellon University, November 1979.

[Ka1] R.M. Karp "Reducibilities Among Combinatorial Problems" in Complexity of Computer Computations R.E. Miller, J.W. Thatcher (eds.), Plenum, pp. 85-104, 1972.

[Ka2] R.M. Karp "The Probabilistic Analysis of some Combinatorial Search Algorithms", in J.F. Traub (ed.) Algorithms and Complexity: New Directions and Recent Results, Academic Press, 1977.

[Ka3] R.M. Karp "Probabilistic Analysis of Partitioning Algorithms for the Travelling Salesman Problem" Math. Operations Research 2, pp. 209-224, 1977.

[KH] O.Kariv, S.L. Hakimi "An Algorithmic Approach to Network Location Problems Part 2: the p-medians" Manuscript, 1976.

[Pa1] C.H. Papadimitriou, "The Euclidean Traveling Salesman Problem is NP-complete" J. Theor. Computer Science 4, pp. 237-244, 1977.

[Pa2] C.H. Papadimitriou, The Complexity of Combinatorial Optimization Problems, Ph.D. Thesis, Princeton University, 1976.

[PS] C.H. Papadimitriou, K. Steiglitz Combinatorial Optimization: Algorithms & Complexity, in preparation, 1980.

[Sh] M.I. Shamos Computational Geometry Ph.D. Thesis, Yale University, 1978.

[Sta] D.A. Starret "Principles of Optimal Location in a Large Homogeneous Area", J. of Economic Theory,9, pp. 418-448, 1974.

[Ste] J. Michael Steele "Subadditive Euclidean Functionals and Non-linear Growth in Geometric Probability", submitted to the Annals of Probability, 1979.