On the Average Case Performance of Some Greedy Approximation Algorithms For the Uncapacitated Facility Location Problem

Abraham D. Flaxman
abie@cmu.edu

Alan M. Frieze[*]
alan@random.math.cmu.edu

Juan C. Vera
jvera@andrew.cmu.edu

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA, 15213, USA

ABSTRACT

In combinatorial optimization, a popular approach to NP-hard problems is the design of approximation algorithms. These algorithms typically run in polynomial time and are guaranteed to produce a solution which is within a known multiplicative factor of optimal. Unfortunately, the known factor is often known to be large in pathological instances. Conventional wisdom holds that, *in practice*, approximation algorithms will produce solutions closer to optimal than their proven guarantees. In this paper, we use the rigorous-analysis-of-heuristics framework to investigate this conventional wisdom.

We analyze the performance of 3 related approximation algorithms for the uncapacitated facility location problem (from [Jain, Mahdian, Markakis, Saberi, Vazirani, 2003] and [Mahdian, Ye, Zhang, 2002]) when each is applied to an instances created by placing n points uniformly at random in the unit square. We find that, with high probability, these 3 algorithms do not find asymptotically optimal solutions, and, also with high probability, a simple plane partitioning heuristic does find an asymptotically optimal solution.

Categories and Subject Descriptors:

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*routing and layout*; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms:

Algorithms

Keywords:

Approximation Algorithms, Probabilistic Analysis of Algorithms, Uncapacitated Facilty Location Problem

1. INTRODUCTION

Many optimization problems are NP-hard. This is an unfortunate fact of life. There are a variety of approaches to dealing with this problem. One approach is to find approximation algorithms with provably good worst-case performance guarantees. Another approach is to design heuristics which work well "on average". In this paper we will take the novel approach of analyzing an approximation algorithm in a probabilistic setting. The aim is to investigate the notion that such algorithms will "typically" do better than their worst-case guarantees. To the best of our knowledge this is the first attempt at such an analysis of an approximation algorithm. (In a typical probabilistic analysis, the algorithm is designed with the probability distribution of inputs in mind).

In the uncapacitated facility location problem (UFLP) we are given a set of facilities $\mathcal{F}$ and a set of cities $\mathcal{C}$. For every facility $i \in \mathcal{F}$ there is a cost f_i for opening that facility, and for every facility-city pair $(i, j) \in \mathcal{F} \times \mathcal{C}$ there is a cost $c_{i,j}$ for connecting facility i to city j. There are no bounds on the number of cities that can be connected to a facility. Thus, if we open the set of facilities $F \subseteq \mathcal{F}$ then each city j will connect to the open facility with cheapest connection cost, and the total cost will be

$$c(F) = \sum_{i \in F} f_i + \sum_{j \in \mathcal{C}} \min_{i \in F} c_{i,j}.$$

The goal is to find a set of facilities F that will minimize the total cost $c(F)$.

Unfortunately, the problem is NP-hard, as it contains set-cover as a special case. It has been the focus of a great deal of attention from many perspectives. In the 1980's, the Operations Research community focused on branch and bound algorithms for solving it, which led to some considerable success, see for example [9]. From that period, there is also some worst-case analysis of the performance of greedy heuristics [5] and a probabilistic analysis of the related k-median problem [1]. More recently, the Theoretical Computer Science community has placed a significant emphasis on finding approximation algorithms for NP-hard problems and one of its most notable successes has been in finding constant factor approximations for this problem when the connection costs obey the triangle inequality. The first algorithm to obtain a constant factor approximation was based on LP rounding [13] and subsequent approaches based on LP rounding improved the constant to $1 + 2/e$ [4] and then to

---

[*] Supported in part by NSF grant CCR0200945

STOC'05, May 22-24, 2005, Baltimore, Maryland, USA.

1.58 [15]. Alternative approaches to approximating the solution are based on local search techniques [10], primal-dual schema [8] and combinations of these [3]. At the present time the best approximation guarantee that is obtainable in polynomial time is 1.52, due to Mahdian, Ye and Zhang [11]. This is a greedy augmentation algorithm, and in the present paper, we will focus our attention on it and on 2 related greedy algorithms [7].

It is likely that approximation algorithms will find solutions closer to optimal than their guarantees guarantee. How much closer? One way to provide some answer to this question is via an experimental study, which is exactly the approach of [2, 6] and is also considered in Section 7 of [7]. Another way, which we will follow in this paper, is to consider theoretically the result of applying the algorithms to an appropriate random instance. Since the constant factor approximation algorithms are only supposed to work on metric instances, we rule out one common random model, in which all distances are chosen independently and uniformly from $[0, 1]$. Another random model we do not study comes from choosing all distances from a discrete distribution that takes only the values 1 and 2. The random model we use will be geometric in nature, formed by placing points uniformly at random in the unit square. For additional reference on combinatorial optimization over instances derived from random points, see [12, 14, 16]. Although it is possible to design algorithms to take advantage of the special structure of these instances, that is not the focus of the current investigation. Instead of first choosing a distribution over instances and then designing an algorithm to work **whp** over this distribution, we begin by choosing the algorithms to study and then choose an interesting (but tractable) distribution of instances on which to run them.

1.1 Random model

We will study random instances formed by choosing n points $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ uniformly at random in the unit square $[0, 1]^2$. We assume that each point represents a city and also the possible location of a facility. For simplicity we will use the ℓ_∞ distance between each facility-city pair as the connection cost.

Let m be a positive integer satisfying $m = o((n/\log n)^{1/2})$. Then let $\alpha = m^{-1}$ and define $\omega = m^{-1}(n/\log n)^{1/2}$, so that $\omega \to \infty$ with n.

We will give every facility the same opening cost,

$$f = \frac{1}{6}\alpha^3 n.$$

We have selected these values for later convenience in notation, and summarize it in the following table.

$$\omega \to \infty \qquad m = \omega^{-1}\sqrt{\frac{n}{\log n}} \qquad \alpha = \omega\sqrt{\frac{\log n}{n}}$$

$$f = \frac{1}{6}\omega^3\frac{(\log n)^{3/2}}{\sqrt{n}}$$

We denote the ℓ_∞ distance between two points X_i and X_j by $d(X_i, X_j)$. All logarithms are base e.

We initially expected to prove that the algorithm of [11], which has worst-case approximation ratio 1.52, was asymptotically optimal i.e. that **whp**[1], as $n \to \infty$, the ratio of the cost of the solution found by the approximation algorithm and the optimum tends to 1. Instead we give a proof of the following: Let OPT denote the value of a minimum cost solution. The algorithm of [11] is similar in spirit to the 2 algorithms given in [7], which have worst-case approximation ratios of at most 1.861 and 1.61. We denote these approximation algorithms by H_1, H_2, H_3, and recall their descriptions in detail in Section 2. We let Z_i denote the value of the solution found by H_i.

THEOREM 1. *There exists a positive constant $\epsilon > 0$ such that for $i = 1, 2, 3$,* **whp**

$$\frac{Z_i}{OPT} \geq 1 + \epsilon.$$

On the other hand it is not difficult to describe a "trivial heuristic" which *is* asymptotically optimal and so it is disappointing that these sophisticated approximation algorithms are in fact beaten by triviality **whp**.

1.2 Outline

In the next section we describe the greedy approximation algorithms and the trivial heuristic in detail, and give a non-rigorous explanation of "what goes wrong" to prevent the approximation algorithms from finding an asymptotically optimal solution.

Since our non-rigorous explanation will rely heavily on the asymptotic optimality of the trivial heuristic, we prove that the heuristic is asymptotically optimal in Section 3. The proof has 2 parts. First we obtain an upper bound that holds **whp** on the value of the solution found by the heuristic. Since the heuristic is so simple, this only requires us to consider basic probabilistic arguments. Some of these recur frequently enough to merit little lemmas, which are stated and proved in Section 3.1. Then we obtain an asymptotically matching lower bound that holds **whp** on the value any solution. We do this by constructing a solution to the dual of the LP-relaxation which is feasible **whp**.

The remainder of the paper proves Theorem 1. To do so, in Section 4.1, we state and prove some lemmas which show that the structure of *any* near optimal solution must take a certain form; it must choose facilities to open so that, for most open facilities, the region of the plane which is closer to that facility than any other is approximately a square of a certain size and is approximately centered on the facility. Lemma 5 from Section 4.1 is a quantitative version of this. Roughly, it says that if there are ϵn facilities opened which violate these conditions then the solution will be a $1 + \delta$ factor away from optimal.

To complete the proof of Theorem 1, in Section 4.2 we show that the approximation algorithms from Section 2 open too many facilities which do not meet the requirements for a close to optimal solution.

2. APPROXIMATION ALGORITHMS

The approximation algorithms we consider are all similar. We first recall Algorithm 1 of [7] (which is most convenient for us in its restated form).

---

[1] A sequence of events $\mathcal{E}_n$ occurs *with high probability* (**whp**), if $\lim_{n\to\infty} \mathbf{Pr}(\mathcal{E}_n) = 1$

Approximation Algorithm 1

(a) The algorithm starts at time 0. Initially, each city is defined to be *unconnected*. The set of unconnected cities is denoted by U. All facilities are considered to be *unopened* and $\delta_i = 0$ for $i \in C$, the set of cities.

(b) While $U \neq \emptyset$, increase the time and simultaneously for every city $j \in U$ increase the parameter δ_i at the same rate, until one of the following events occurs:

1. For some unconnected city i, and some open facility j, $\delta_i = d(i,j)$. In this case, connect city i to facility j and remove j from U.
2. For some unopened facility j, $\sum_{i \in U} \max\{0, \delta_i - d(i,j)\} = f_j$. In this case open this facility and for every unconnected city with $\delta_i \geq d(i,j)$, connect i to j and remove it from U.

Now we recall Algorithm 2 of [7], which is very similar to Algorithm 1, but allows connected cities to contribute funds towards opening additional facilities.

Approximation Algorithm 2

(a) The algorithm starts at time 0. Initially, each city is defined to be *unconnected*. The set of unconnected cities is denoted by U. All facilities are considered to be *unopened* and $\delta_i = 0$ for $i \in C$, the set of cities. We denote by π the mapping from connected cities to open facilities.

(b) While $U \neq \emptyset$, increase the time and simultaneously for every city $j \in U$ increase the parameter δ_i at the same rate, until one of the following events occurs:

1. For some unconnected city i, and some open facility j, $\delta_i = d(i,j)$. In this case, connect city i to facility j and remove j from U.
2. For some unopened facility j, we have

$$\sum_{i \in U} \max\{0, \delta_i - d(i,j)\} \\ + \sum_{i \notin U} \max\{0, c_{i,j} - c_{i,\pi(i)}\} = f_j.$$

   In this case open this facility and for every unconnected city with $\delta_i \geq d(i,j)$, connect i to j and remove it from U, and for every connected city with $c_{i,j} < c_{i,\pi(i)}$ change the facility to which i connects from $\pi(i)$ to j.

Now, we recall Algorithm 3, which appears in [11] and currently has the best proven bound on worst-case approximation ratio.

Approximation Algorithm 3

(a) In the first phase, the algorithm scales up the opening costs of all facilities by a constant $\delta = 1.504$, and uses Algorithm 2 to find a solution to the problem with these new costs.

(b) In the second phase, the algorithm considers the unmodified costs and performs a greedy augmentation to the solution found in phase 1. Let C denote the total connection cost in the phase 1 solution. For each unopened facility j, let C_j denote the total connection cost when j is also opened. If the maximum over unopened facilities of the ratio $(C - C_j - f_j)/f_j$ is positive, then open the facility that maximizes this ratio.

Finally, we describe the plane partitioning heuristic, which is *not* guaranteed to produce a solution within any constant factor. Figure 1 provides a visual reference.

Trivial Heuristic

(a) We partition the square into an $m \times m$ grid Γ of subsquares $S_{p,q}$, $1 \leq p, q \leq m$ of side length α, and then open the facility $F_{p,q}$ closest to the center of each subsquare, assuming that there is one within distance $\alpha/\omega = \left(\frac{\log n}{n}\right)^{1/2}$ of its center.

(b) If any subsquare $S_{p,q}$ has no facility within distance α/ω of its center, then open each X_i in $S_{p,q}$ as a facility.

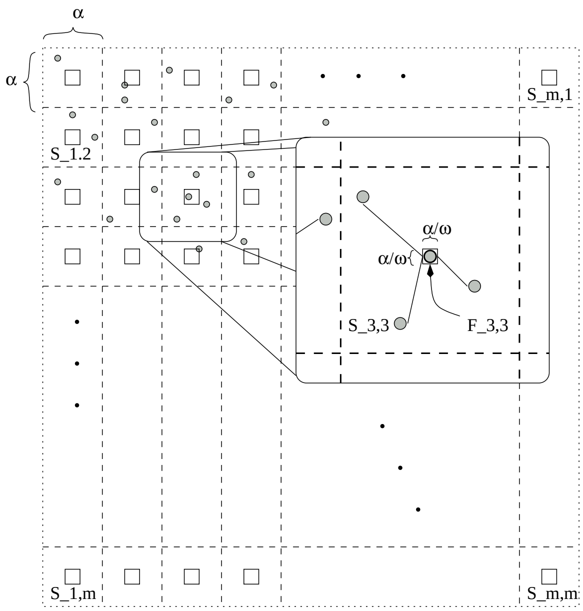

Figure 1: A schematic representation of the asymptotically optimal solution.

The Trivial Heuristic pays little attention to the structure of the instance, but, as we will prove in Section 3, it produces a solution which is asymptotically optimal **whp**. In fact, in some sense, it is *because* it does not pay attention to the instance that it out-performs the approximation algorithms. All of the greedy algorithms are distracted by local deviations in city density, and (at least at first) they will open facilities at what amount to random points in the plane. This results in non-uniform coverage and requires some unlucky cities to suffer excessive connection costs.

3. AN ASYMPTOTICALLY OPTIMAL SOLUTION

In this section, we prove that the solution found by the Trivial Heuristic is asymptotically optimal. To do so, we obtain an upper bound on the cost of this solution and a matching lower bound on the dual of the LP-relaxation.

Let HEU denote the total cost of the solution found by the Trivial Heuristic.

An intuition which explains the near optimality of this solution is that the cities and facilities are roughly uniformly distributed in the square, so the advantage of using the special structure of the instance is negligible.

To make this intuition rigorous, in the following 2 subsections, we obtain an upper bound on HEU which holds **whp**, and a lower bound on OPT which also holds **whp** and asymptotically matches the upper bound on HEU. But first we state and prove 2 lemmas that will aid in our analysis.

3.1 Some simple lemmas

The following 2 lemmas will help us in analyzing the heuristic and the dual lower bound.

LEMMA 1. *Let $A_1, \dots, A_k$ be subsets of $[0,1]^2$ each of area a, let $\mathcal{X}$ be a set of n random points distributed uniformly and independently in $[0,1]^2$, and let λ be a positive real with $\lambda \le 1/3$. Then*

$$\mathbf{Pr}[\exists i\colon A_i \cap \mathcal{X} = \emptyset] \le k \cdot e^{-an} \tag{1}$$

$$\mathbf{Pr}[\exists i\colon |A_i \cap \mathcal{X}| \notin (1 \pm \lambda)an] \le k \cdot 2e^{-\lambda^2 an/3} \tag{2}$$

Proof (1) follows because the probability that a single point avoids A_i is $1-a$ and $1-x \le e^{-x}$ and the union bound.

(2) follows from Chernoff's bound and the union bound. □

LEMMA 2. *Let t be a positive real, let $F_1, \dots, F_k$ be points in $[t, 1-t]^2$, let $\mathcal{X}$ be a set of n random points distributed uniformly and independently in $[0,1]^2$, and let λ be a positive real with $\lambda \le 1/6$. For $i = 1, \dots, k$, let $Z_i = \sum_{\substack{X \in \mathcal{X}\\ d(X,F_i) \le t}} d(X, F_i)$. Then*

$$\mathbf{E}[Z_i] = \frac{n(2t)^3}{3} \tag{3}$$

$$\mathbf{Pr}\left[\exists i\colon Z_i \notin (1 \pm \lambda)\frac{n(2t)^3}{3}\right] \le k \cdot 4e^{-\lambda^2(2t)^2 n/12}. \tag{4}$$

Proof

We begin by considering the contribution of a particular point X to Z_i. Conditioning on $d(X, F_i) \le t$, the expected distance is

$$\mathbf{E}[d(X,F_i) \mid d(X,F_i) \le t] = t^{-2}\int_{u=0}^{t} u \cdot 2u \, du = \frac{2t}{3}.$$

We define N_i to be the number of points within distance t of F_i,

$$N_i = |\{X \in \mathcal{X}\colon d(X,F_i) \le t\}|.$$

It follows from the linearity of expectations that

$$\mathbf{E}[Z_i \mid N_i] = N_i \frac{2t}{3}.$$

And, since $\mathbf{E}[N_i] = (2t)^2 n$, we have established (3),

$$\mathbf{E}[Z_i] = \left((2t)^2 n\right)\frac{2t}{3}.$$

Conditioning on N_i, Z_i is a sum of N_i independent random variables in the range $[0,t]$. So Hoeffding's inequality gives

$$\begin{aligned}\mathbf{Pr}\left[Z_i \notin (1 \pm \lambda)N_i\frac{2t}{3} \;\middle|\; N_i\right] &\le 2e^{-2(\lambda N_i 2t/3)^2/(N_i t^2)}\\ &= 2e^{-8\lambda^2 N_i/9}.\end{aligned}$$

Now, we apply Lemma 1 with $A_i = \{X\colon d(X,F_i) \le t\}$ and (2) shows that the probability that some N_i does not contain $(1 \pm \lambda)(2t)^2 n$ points is at most $k \cdot 2e^{-\lambda^2(2t)^2 n/3}$. Combining this with the conditional upper bound on the large deviation probability of Z_i and the union bound gives

$$\begin{aligned}&\mathbf{Pr}\left[\exists i\colon Z_i \notin (1 \pm \lambda)\left((1 \pm \lambda)(2t)^2 n\right)\frac{2t}{3}\right]\\ &\qquad \le k \cdot 2e^{-\lambda^2(2t)^2 n/3} + k \cdot 2e^{-8\lambda^2(1-\lambda)(2t)^2 n/9}.\end{aligned}$$

Since $\lambda \le 1/3$, this simplifies to

$$\mathbf{Pr}\left[\exists i\colon Z_i \notin (1 \pm \lambda)(2t)^3 n/3\right] \le 4ke^{-\lambda^2(2t)^2 n/3}.$$

□

3.2 An upper bound on HEU

To achieve this goal, we define several events and random variables and bound probabilities related to them.

Let $\hat{F}_{p,q}$ be the point in the center of subsquare $S_{p,q}$.

We begin by showing that in each subsquare, there is likely to be a facility within distance α/ω of $\hat{F}_{p,q}$ that we will open. To do this, we apply Lemma 1 with $k = m^2$ and A_{pm+q} equal to the square within distance α/ω of $\hat{F}_{p,q}$. Then, since $\text{area}(A_{pm+q}) = (2\alpha/\omega)^2 = \frac{4\log n}{n}$, (1) shows that

$$\mathbf{Pr}[\exists p,q:\ A_{pm+q} \cap \mathcal{X} = \emptyset] \le m^2 \cdot e^{-4\log n} = o(n^{-3}). \tag{5}$$

Now we bound the transportation costs. We define a mapping π so that for each X_i with $X_i \in S_{p,q}$ and $F_{p,q} = X_j$ we have $\pi(i) = j$ to indicate that facility j services city i. In the unlikely event that A_{pm+q} is empty, we open all the facilities in $S_{p,q}$ and set $\pi(i) = i$ for each of them, which results in transportation cost 0.

Note that, since $F_{p,q}$ is within α/ω of $\hat{F}_{p,q}$, we have

$$\sum_{X_i \in \mathcal{X}} d(X_i, X_{\pi(i)}) \le \sum_{X_i \in \mathcal{X}} d(X_i, F_{p,q}) + n\alpha/\omega. \tag{6}$$

We apply Lemma 2 with $t = \alpha/2$, $k = m^2$, $F_{pm+q} = \hat{F}_{p,q}$, and $\lambda = \omega^{-1}$. Then (3) and (6) together imply that

$$\mathbf{E}\left[\sum_{X_i \in \mathcal{X}} d(X_i, X_\pi(i))\right] \le m^2\frac{n\alpha^3}{3} + n\alpha/\omega$$

and (4) and (6) imply that

$$\begin{aligned}\mathbf{Pr}\left[\sum_{X_i \in \mathcal{X}} d(X_i, X_\pi(i)) \ge m^2 \cdot (1 + 4\omega^{-1})\frac{n\alpha^3}{3} + \frac{n\alpha}{\omega}\right]&\\ \le m^2 \cdot 4e^{-16\omega^{-2}\alpha^2 n/12}&\\ = 4m^2 e^{-4\log n/3}.&\end{aligned}$$

Since there are m^2 facilities opened with probability at least $1 - n^{-3}$, and there are at most n facilities opened in even the most pathological point set, we may the bound expected total cost of the solution by

$$\mathbf{E}[HEU] = \frac{n\alpha}{3} + n\alpha/\omega + m^2 f + nfn^{-3} = \frac{1}{2}\alpha n(1+o(1)).$$

Finally, we observe that the probability that HEU exceeds this bound tends to 0; the transportation cost is at most $\frac{n\alpha}{3}(1+O(\omega^{-1}))$ with probability $1-o(1)$ and the probability that more than m^2 facilities open is $o(1)$. So we conclude that

$$HEU \le \frac{n\alpha}{2}(1+o(1)) \qquad \mathbf{whp}. \tag{7}$$

3.3 Lower bound on OPT

To show this solution is asymptotically optimal, we will construct a solution to the dual of the strong LP relaxation:

$$\begin{array}{llll}
 & (\textbf{LP-RELAX}) & & \\
\min & \sum_{j=1}^{n} f y_j + \sum_{i=1}^{n}\sum_{j=1}^{n} d(X_i,X_j)x_{i,j} & & \\
\text{subj. to} & \sum_{j=1}^{n} x_{i,j} \;=\; 1 & & 1 \le i \le n \\
 & 0 \;\le\; x_{i,j} \le y_j & & 1 \le i,j \le n. \\
 & (\textbf{DUAL}) & & \\
\max & \sum_{i=1}^{n} u_i & & \\
\text{subj. to} & \sum_{i=1}^{n} v_{i,j} \;\le\; f & & 1 \le j \le n \\
 & -v_{i,j} + u_i \;\le\; d(X_i,X_j) & & 1 \le i,j \le n \\
 & v_{i,j} \;\ge\; 0 & & 1 \le i,j \le n.
\end{array}$$

We get a good solution to **DUAL** as follows:

$$u_i = \begin{cases} \frac{\alpha}{2}(1-3\omega^{-1}) & X_i \in [\alpha, 1-\alpha]^2. \\ 0 & \text{otherwise.} \end{cases}$$

$$v_{i,j} = \max\{u_i - d(X_i,X_j), 0\}.$$

The fact that this solution is feasible **whp** follows from Lemma 1 and Lemma 2. We take $t = \frac{\alpha}{2}(1-3\omega^{-1})$, $k = n$, $F_i = X_i$, and $\lambda = 4\omega^{-1}$. Then (4) shows that

$$\begin{aligned}
\mathbf{Pr}\left[\exists i \colon Z_i \le (1-4\omega^{-1})n(\alpha(1-3\omega^{-1}))^3/3\right] & \\
&\le n \cdot 4e^{-16\omega^{-2}(\alpha(1-3\omega^{-1}))^2 n/12} \\
&= 4ne^{-16(1-3\omega^{-1})^2 \log n/12} \\
&= o(1).
\end{aligned}$$

Taking A_i to be the $\alpha\left(1-3\omega^{-1}\right) \times \alpha\left(1-3\omega^{-1}\right)$ square centered at X_i, (2) shows that

$$\begin{aligned}
\mathbf{Pr}[\exists i \colon |A_i \cap \mathcal{X}| \ge (1+4\omega^{-1})(1-3\omega^{-1})^2\alpha^2 n] & \\
&\le n \cdot 2e^{-16\omega^{-2}\left(1-3\omega^{-1}\right)^2\alpha^2 n/3} \\
&= 2ne^{-16\left(1-3\omega^{-1}\right)^2 \log n/3} \\
&= o(1).
\end{aligned}$$

So **whp** for all j we have

$$\begin{aligned}
\sum_{i=1}^{n} v_{i,j} &= \sum_{X_i \in \mathcal{X}} \max\left\{\frac{\alpha}{2}\left(1-3\omega^{-1}\right) - d(X_i,X_j), 0\right\} \\
&< \frac{n\alpha^3}{6} = f.
\end{aligned}$$

Since the objective value of this solution asymptotically matches that of (7), we conclude that our "heuristic" is asymptotically optimal.

4. PROOF OF MAIN THEOREM

To prove Theorem 1, in Section 4.1 we state and prove some lemmas which show that the structure of *any* near optimal solution must take a certain form. In particular, the solution must choose facilities to open so that, for most open facilities, the region of the plane which is closer to that facility than any other (the Voronoi cell) is approximately a square of a certain size and is approximately centered on the facility. Lemma 5 from Section 4.1 gives a quantitative version of this fact: it says roughly that if there are ϵn facilities opened which violate the conditions then the solution will be a $1+\delta$ factor away from optimal.

To complete the proof of Theorem 1, in Section 4.2 we show that the approximation algorithms from Section 2 open too many facilities which do not meet the requirements given in Lemma 5 for a close to optimal solution **whp**.

4.1 Properties of close to optimal solutions

4.1.1 Refining Γ to super-grid Γ_1

Now let $m_1 = \lfloor \omega^{1/2} \rfloor m$ and let Γ_1 be the $m_1 \times m_1$ super-grid of Γ where each subsquare has side $\alpha_1 = m_1^{-1}$. If we fix a subsquare S of Γ_1 then the number of points ν_S of $\mathcal{X}$ which fall in S is distributed as $\mathrm{B}(n, \alpha_1^2)$. Thus $\mathbf{E}(\nu_S) = \alpha_1^2 n = \omega \log n(1+o(1))$. It follows from Lemma 1, part (2) that

$$\begin{aligned}
\mathbf{Pr}\left[\exists S \in \Gamma_1 :\ \nu_S \notin (1 \pm \omega^{-1/3})\alpha_1^2 n\right] & \\
&\le m_1^2 \cdot 2e^{-\omega^{-2/3}\alpha_1^2 n/3} \\
&< n \cdot 2e^{-\omega^{1/3}\log n/3}
\end{aligned}$$

We use the term quite surely (**qs**) to describe a sequence of events which occurs with probability exceeding $1 - O(n^{-k})$ for any constant k. In this notation, we may say that

$$|\nu_S - \alpha_1^2 n| \le \omega^{2/3}\log n,\ \forall S \in \Gamma_1, \qquad \mathbf{qs}. \tag{8}$$

4.1.2 An assignment which respects super-grid Γ_1

For a set of facilities $\mathcal{F}$ and an assignment of cities to facilities $\phi \colon \mathcal{X} \to \mathcal{F}$ we let

$$\kappa(\mathcal{F}, \phi) = f|\mathcal{F}| + \sum_{X \in \mathcal{X}} d(X, \phi(X)).$$

The assignment which maps points to their closest facility in $\mathcal{F}$ will be denoted $\phi^\star_{\mathcal{F}}$ so that

$$c(\mathcal{F}) = \kappa(\mathcal{F}, \phi^\star_{\mathcal{F}}).$$

Consider a particular facility set $\mathcal{F} = \{F_1, F_2, \ldots, F_k\} \subseteq \mathcal{X}$. For each F_i let V_i be the Voronoi cell associated with F_i, which is to say V_i is the set of points in $[0,1]^2$ which are as close (in ℓ_∞ norm) to F_i as to any other member of $\mathcal{F}$.

We say an assignment ϕ respects Γ_1 if all the cities in a common subsquare of Γ_1 are assigned to the same facility by ϕ.

The next lemma says that there is an assignment which respects Γ_1 and is not much worse than $\phi^\star_{\mathcal{F}}$.

Lemma 3. *There exists an assignment $\tilde{\phi}_{\mathcal{F}}$ that respects Γ_1 and has $|\kappa(\mathcal{F}, \tilde{\phi}_{\mathcal{F}})| - \kappa(\mathcal{F}, \phi^\star_{\mathcal{F}})| \le 2\alpha_1 n$.*

Proof The proof of the lemma is a shifting argument. For any assignment ϕ, if there exists some $S \in \Gamma_1$ and $i \in [k]$ such that $V_i \cap S \cap \mathcal{X} \neq \emptyset$ and $S \setminus V_i \neq \emptyset$ then we make a slightly different assignment $\tilde{\phi}$ which assigns all cities in S to the same facility. Let i be the smallest index in $[k]$ such that cities in S are assigned to F_i. Then we re-assign all $X_j \in S \setminus V_i$ to facility F_i. We claim that this adds at most $2\alpha_1$ in transportation cost for each city. Indeed, suppose that $X_j \in S \cap V_{i'}$ for $i' \neq i$. Then $d(X_j, F_i) \leq d(X_j, X) + d(X, F_i)$. If $X \in V_i \cap S$, then we also have that $d(X, F_i) \leq d(X, X_j) + d(X_j, F_{i'})$, since X is in V_i and not V_j. So $d(X_j, F_i) \leq d(X_j, F_{i'}) + 2d(X, X_j)$. Since X and X_j are both in S, $d(X, X_j) \leq \alpha_1$.

By starting with $\phi^\star_{\mathcal{F}}$ and repeating this shifting we eventually arrive with an assignment $\tilde{\phi}_{\mathcal{F}}$ (since assignments to cities in each cell are adjusted at most once). This assignment respects Γ_1 by construction, and (again because each city is reassigned at most once) we have

$$\kappa(\mathcal{F}, \tilde{\phi}_{\mathcal{F}}) \leq \kappa(\mathcal{F}, \phi^\star_{\mathcal{F}}) + 2\alpha_1 n. \tag{9}$$

□

4.1.3 The likely cost per facility under $\tilde{\phi}_{\mathcal{F}}$

For $F_i \in \mathcal{F}$, let the $\tilde{V}_i$ be the union of the subsquares in Γ_1 which contain cities which are mapped to F by $\tilde{\phi}_{\mathcal{F}}$ (we think of $\tilde{V}_i$ as the "quantized Voronoi cell" of F_i). Let η_i denote the number of subsquares in $\tilde{V}_i$. Let $\mathcal{X}_i = \mathcal{X} \cap \tilde{V}_i$ and let

$$c_i = \sum_{X \in \mathcal{X}_i} d(X, F_i).$$

Note that, because of the way $\tilde{\phi}_{\mathcal{F}}$ was constructed, for any Γ_1-subsquare S, if $S \subseteq V_i$ then $S \subseteq \tilde{V}_i$.

We say that $\tilde{V}_i$ is an *ϵ-quasi-square* if there exists a square S centered at F_i such that $\max\{\mathrm{area}(S \setminus \tilde{V}_i), \mathrm{area}(\tilde{V}_i \setminus S)\} \leq \epsilon\, \mathrm{area}(\tilde{V}_i)$.

LEMMA 4. *Assume that* (8) *holds. Assume that* $\epsilon \gg \alpha_1$. *Then the following hold* **whp**

(i) $c_i \geq \frac{1}{3} n (1 - \omega^{-1/3})\, \mathrm{area}(\tilde{V}_i)^{3/2}$.

(ii) *If* $\tilde{V}_i$ *is not an ϵ-quasi-square then* $c_i \geq \frac{1+\epsilon^2/4}{3} n (1 - \omega^{-1/3})\, \mathrm{area}(\tilde{V}_i)^{3/2}$.

Proof In light of (8), this lemma reduces to a pair of geometric facts about collections of squares. However, it is convenient for us to prove the facts via linear programming.

We begin by establishing part (i) of the lemma. Fix i. For every j define $U_j = \{S \in \Gamma_1 : S \subseteq \tilde{V}_i \text{ and } j\alpha_1 \leq d(S, F_i) < (j+1)\alpha_1\}$. We have $|U_j| \leq 8j + 4$. Let k be such that $U_j = \emptyset$ for every $j > k$. Such k exists because $\tilde{V}_i$ is compact. By counting the number of Γ_1-squares in $\tilde{V}_i$ we get

$$\sum_{j=0}^{k} |U_j| = \eta_i = \mathrm{area}(\tilde{V}_i)/\alpha_1^2.$$

Now,

$$\begin{aligned}
c_i &= \sum_{X \in \mathcal{X}_i} d(X, F_i) \\
&= \sum_{S \in \Gamma_1 : S \subseteq \tilde{V}_i} \sum_{X \in S} d(X, F_i) \\
&\geq \sum_{S \subseteq \tilde{V}_i} \nu_S d(F_i, S) \\
&\geq (\alpha_1^2 n - \omega^{2/3} \log n) \sum_{S \subseteq \tilde{V}_i} d(F_i, S) \\
&= (\alpha_1^2 n - \omega^{2/3} \log n) \sum_{j=0}^{k} \sum_{S \in U_j} d(F_i, S) \\
&\geq (\alpha_1^2 n - \omega^{2/3} \log n) \alpha_1 \sum_{j=0}^{k} j |U_j|
\end{aligned}$$

As we want a lower bound for c_i we consider the primal-dual pair

$$\begin{array}{lllll}
 & \textbf{(P.i)} & & & \\
\min & \sum_{j=0}^{k} j x_j & & & \\
\text{subj to} & x_j & \leq & 8j+4 & j = 0, 1, \ldots, k \\
 & \sum_{j=0}^{k} x_j & = & \eta_i & \\
 & x_j & \geq & 0 & j = 0, 1, \ldots, k \\
 & \textbf{(D.i)} & & & \\
\max & \eta_i z - \sum_{j=0}^{k} (8j+4) y_j & & & \\
\text{subj. to} & z - y_j & \leq & j & j = 0, 1, \ldots, k \\
 & y_j & \geq & 0 & j = 0, 1, \ldots, k
\end{array}$$

A feasible solution for **D.i** is to take $z = \eta_i^{1/2}/2$ and $y_j = \max(\eta_i^{1/2}/2 - j, 0)$, $j = 0, \ldots, k$ with dual value $\geq \eta_i^{3/2}/3$, and then $\sum_{j=0}^{k} j|U_j| \geq \eta_i^{3/2}/3 = \mathrm{area}(V_i)^{3/2}/3\alpha_1^3$.

(The expression $\sum_{j=0}^{\ell} (8j+4)(A-j) = 4A(\ell+1)^2 - \left(\frac{8}{3}\ell^3 + 6\ell^2 + \frac{4}{3}\ell\right)$ will no doubt help the reader to verify the above claim.)

Now we show that part (ii) of the lemma holds. We introduce extra constraints in the linear program above in order to enforce the condition that $\tilde{V}_i$ is not an ϵ-quasi-square. For this, assume that $\tilde{V}_i$ is not an ϵ-quasi-square, let $\ell = \lfloor \eta_i^{1/2}/2 \rfloor$ and let S be the square of side $2\ell\alpha_1$ centered a F_i. Then $\mathrm{area}(\tilde{V}_i) \geq \mathrm{area}(S) \geq (1-\epsilon)\, \mathrm{area}(\tilde{V}_i)$ and therefore $\mathrm{area}(S \cap \tilde{V}_i) < (1-\epsilon)\, \mathrm{area}(\tilde{V}_i)$, otherwise $\mathrm{area}(\tilde{V}_i \setminus S) = \mathrm{area}(\tilde{V}_i) - \mathrm{area}(\tilde{V}_i \cap S) \leq \epsilon\, \mathrm{area}(\tilde{V}_i)$ and $\mathrm{area}(S \setminus \tilde{V}_i) = \mathrm{area}(S) - \mathrm{area}(S \cap \tilde{V}_i) \leq \epsilon\, \mathrm{area}(\tilde{V}_i)$. Then $\sum_{j=0}^{\ell} |U_j| = \mathrm{area}(S \cap \tilde{V}_1)/\alpha_1^2 \leq (1-\epsilon)\, \mathrm{area}(\tilde{V}_i)/\alpha_1^2 = (1 -$

$\epsilon)\eta_i$, so we consider the primal-dual pair

(P.ii)

$$\begin{array}{llcll} \min & \sum_{j=0}^{k} jx_j & & & \\ \text{subj to} & x_j & \leq & 8j+4 & j=0,1,\ldots,k \\ & \sum_{j=0}^{k} x_j & = & \eta_i & \\ & \sum_{j=0}^{\ell} x_j & \leq & (1-\epsilon)\eta_i & \\ & x_j & \geq & 0 & j=0,1,\ldots,k \end{array}$$

(D.ii)

$$\begin{array}{lrcll} \max & \multicolumn{4}{l}{\eta_i z-(1-\epsilon)\eta_i z_1-\sum_{j=0}^{k}(8j+4)y_j} \\ \text{subj to} & z-z_1-y_j & \leq & j & j=0,1,\ldots,\ell \\ & z-y_j & \leq & j & j=\ell+1,\ldots,k \\ & z_1 & \geq & 0 & \\ & y_j & \geq & 0 & j=0,1,\ldots,k \end{array}$$

A feasible solution for **D.ii** is $z=(1+\epsilon)\eta_i^{1/2}/2$, $z_1=\epsilon\eta_i^{1/2}/2$, $y_j=(1-\epsilon/2)\eta_i^{1/2}/2-j$, $j=0,\ldots,\ell$ and $y_j=\max((1+\epsilon/2)\eta_i^{1/2}/2-j,0)$, $j=\ell+1,\ldots,k$ with dual value $\geq (1+\epsilon^2/4)\eta_i^{3/2}/3$, and then $\sum_{j=0}^{k} j|U_j| \geq (1+\epsilon^2/4)\eta_i^{3/2}/3 \geq (1+\epsilon^2/4)\,\mathrm{area}(V_i)^{3/2}/3\alpha_1^3$. □

4.1.4 *The structure of any near optimal solution*

We continue by proving a property of any near optimal solution to the UFLP.

LEMMA 5. *Assume that* (8) *holds. Let ϵ be a sufficiently small constant, and let $\mathcal{F} \subseteq \mathcal{X}$ with $\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \leq (1+\epsilon)\alpha n/2$. Then for $\epsilon_1 = 5\epsilon^{1/2}$,*

(a) $|\mathcal{F}| \in [(1-\epsilon_1)m^2,(1+\epsilon_1)m^2]$.

(b) *Suppose that $\theta_1 = 4\epsilon^{1/3}$ and $\theta_2 = 4\epsilon^{1/3}$ and $\epsilon_0 = 3\epsilon^{1/3}$. Then at least $(1-2\theta_2)m^2$ of the points $F_i \in \mathcal{F}$ are such that $\tilde{V}_i$ is an ϵ_0-quasi-square of area in the range $[(1-\theta_1)\alpha^2,(1+\theta_1)\alpha^2]$.*

Proof Let $\mathcal{F}=\{F_1,F_2,\ldots,F_k\}$ and let $a_i=|\tilde{V}_i|$ for $1 \leq i \leq k$. Let $J=\{j : \tilde{V}_j$ is not a ϵ_0-quasi-square$\}$. Applying Lemma 4 and equation (9) we see that

$$\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq kf+\frac{1-\omega^{-1/3}}{3}n\left(\sum_{i=1}^{k} a_i^{3/2}+\frac{\epsilon_0^2}{12}\sum_{j\in J} a_j^{3/2}\right) - 2\alpha_1 n. \quad (10)$$

Now let $a_j=\frac{1+x_j}{k}$, where $-1 \leq x_j$ and $\sum_{j=1}^{k} x_j = 0$.

By examining the power series for $(1+x)^{3/2}$ when $|x| \leq 1$ and using elementary calculus for $x > 1$ we see that

$$(1+x)^{3/2} \geq 1+\frac{3}{2}x+\min\left\{1,\frac{1}{4}x^2\right\} \qquad x \geq -1. \quad (11)$$

It follows from (10) that

$$\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq kf+\left(\frac{1-\omega^{-1/3}}{3}n\right) \times \left(k^{-1/2}+k^{-3/2}\sum_{i=1}^{k}\min\left\{1,\frac{1}{4}x_i^2\right\}+\frac{\epsilon_0^2}{12}\sum_{j\in J} a_j^{3/2}\right) - 2\alpha_1 n \quad (12)$$

Now, let $k=(1+\theta)\alpha^{-2}$ for some $\theta \geq -1$ and assume wlog that $|\theta| \gg \omega^{-1/6}$. Notice that from (10) that we can assume $\theta < 3$, otherwise $kf \geq \frac{4}{6}\alpha n$. If $\theta \in [-1,3]$ then $\frac{1}{(1+\theta)^{1/2}} \geq 1-\frac{1}{2}\theta+\frac{1}{16}\theta^2$, and we get

$$kf+\frac{1-\omega^{-1/3}}{3}nk^{-1/2}-2\alpha_1 n \geq \frac{(1+\theta)}{6}\alpha n+\frac{(1-\omega^{-1/3})}{3}\alpha n\left(1-\frac{1}{2}\theta+\frac{1}{16}\theta^2\right)-2\alpha_1 n \geq \frac{\alpha n}{2}\left(1+\frac{\theta^2}{25}\right). \quad (13)$$

And using (12) we get

$$\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq \frac{\alpha n}{2}\left(1+\frac{\theta^2}{25}\right) + \frac{n}{4}\left(k^{-3/2}\sum_{i=1}^{k}\min\left\{1,\frac{1}{4}x_i^2\right\}+\frac{\epsilon_0^2}{12}\sum_{j\in J} a_j^{3/2}\right). \quad (14)$$

Part (a) follows from (14): $(1+\epsilon)\alpha n/2 \geq \kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq \frac{\alpha n}{2}\left(1+\frac{\theta^2}{25}\right)$ and so $|\theta| \leq \epsilon^{1/2}/5$.

Using (14) again we get

$$\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq \frac{1}{2}\alpha n+\frac{n}{4k^{3/2}}\sum_{j=1}^{k}\min\left\{1,\frac{1}{4}x_j^2\right\},$$

So if $B=\{j\colon |x_j| \geq \theta_1\}$ and $|B| \geq \beta k$ for $\theta_1,\beta \leq 1$, we have $\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) \geq \frac{1}{2}\alpha n+\frac{\theta_1^2\beta}{16(1+\epsilon_1)^{1/2}}\alpha n$. Setting $\theta_1 = 2\epsilon^{1/3}$ we get $\beta \leq \theta_2 = 4\epsilon^{1/3}$. Returning once again to (14) we write

$$\begin{aligned}\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) &\geq \frac{1}{2}\alpha n+\frac{\epsilon_0^2}{48}n\sum_{j\in J} a_j^{3/2} \\ &\geq \frac{1}{2}\alpha n+\frac{3\epsilon^{2/3}}{16}n(|J|-\theta_2 k)\left(\frac{1-\theta_1}{k}\right)^{3/2}.\end{aligned}$$

Thus, if $|J| \geq 2\theta_2 m^2$ then

$$\begin{aligned}\kappa(\mathcal{F},\phi^\star_{\mathcal{F}}) &\geq \frac{1}{2}\alpha n+\frac{3\epsilon^{2/3}}{16}n\theta_2(2m^2-k)\left(\frac{1-\theta_1}{k}\right)^{3/2} \\ &\geq \frac{1}{2}\alpha n+\frac{12}{16}\epsilon n(1-\epsilon_1)m^2\left(\frac{1-2\epsilon^{1/3}}{(1+\epsilon_1)m^2}\right)^{3/2} \\ &\geq \frac{1}{2}\alpha n+\frac{11}{16}\epsilon\alpha n.\end{aligned}$$

□

4.2 Properties of Solutions Found by Greedy Approximation Algorithms

Let $pf(X,t)$ be the *potential funds* at point X at time t. Let $T(X)=\min\{\min\{t : pf(X,t)=f\},\alpha\}$ be the *earliest* *opening time* of point X (truncated at time α, because we want $T(X)$ to only depend on the position of nearby points).

We note that $\mathbf{E}(pf(X,\alpha)) = f$ and $pf(X,\alpha)$ is the sum of n independent bounded random variables and so the Central Limit Theorem implies that

$$\mathbf{Pr}(pf(X,\alpha) \geq f) = \frac{1}{2} - o(1). \tag{15}$$

Consider concentric squares, $S_1, S_2, \ldots,$ where S_i is an $i\alpha \times i\alpha$ square. Some facility $X^\star$ in S_5 has the minimum value of $T(X)$ among all facilities in S_5, and which one it is only depends on the configuration of points in S_7.

Note that if $X^\star$ is in S_1, (and $T(X^\star) < \alpha$) then (in all 3 of the greedy approximation algorithms) $X^\star$ *actually* opens at time $T(X^\star)$, because no cities within distance α of S_1 are connected (because no facilities within 2α of S_1 are open; in other words, no facilities besides $X^\star$ are open in S_5.) Since nothing within α of $X^\star$ is connected, all the funds potentially being offered to $X^\star$ are actually being offered.

We will partition S_1 into subsquares of size $\alpha/4$, and obtain a constant lower bound on the probability $X^\star$ appears in one of these subsquares. Let $Q_{p,q}$ denote such a subsquare, where $(p,q) \in [4]^2$. Figure 2 provides a visual reference.

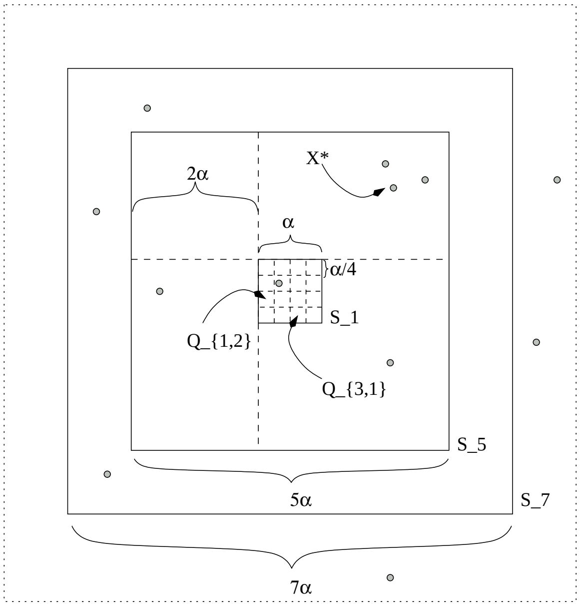

Figure 2: A schematic representation of the points and sets used in the proof.

Lemma 6. *Let $X^\star$ be the facility in S_5 which minimizes $T(X)$ over $X \in \mathcal{X} \cap S_5$. There exists an absolute constant γ_0 such that for any $(p,q) \in [4]^2$,*

$$\mathbf{Pr}[X^\star \in Q_{p,q} \text{ and } pf(X^\star,\alpha) \geq f] \geq \gamma_0.$$

Proof We begin by considering an analogous question on S_7 where the edges have been identified to "wrap-around" (making it a torus topologically). In this case, some subsquare contains the the facility with minimum $T(X)$ and, by symmetry, each subsquare is equally likely to contain it. So the probability that it is in $Q_{p,q}$ is exactly $(7 \cdot 4)^{-2}$.

Now, we remove the wrap-around on S_7 but ignore all the points of $\mathcal{X}$ that lie outside S_7. This can only affect the earliest opening time of points within distance α of the boundary of S_7, and it can only makes their earliest opening times increase. So every configuration of points in which $Q_{p,q}$ contains the point which opens first on the torus is also a configuration in which $Q_{p,q}$ contains the point which opens first here. So the probability that $Q_{p,q}$ contains the point which opens first here is at least $(7 \cdot 4)^{-2}$.

Now, note that considering the contributions of points outside of S_7 does not affect the earliest opening time of any point inside S_5. So the probability that $Q_{p,q}$ contains the point which opens first in S_5 with respect to $\mathcal{X}$ is at least the probability that $Q_{p,q}$ contains the point which opens first in S_7 with respect to $\mathcal{X} \cap S_7$. Since the previous paragraph showed that this is at least $(7 \cdot 4)^{-2}$ we can multiply this by $1/3$ (see (15)) to complete the proof of the lemma. □

Consider 2 side-by-side copies of S_7. Let $\mathcal{B}_1$ be the event that in the first copy $X^\star$ appears in $Q_{1,q}$ for some q. Let $\mathcal{B}_2$ be the event that in the second copy $X^\star$ appears in $Q_{3,q'}$ for some q'. $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent, so $\mathbf{Pr}[\mathcal{B}_1\mathcal{B}_2] \geq \gamma_0^2$.

Suppose now that $\mathcal{B}_1$ and $\mathcal{B}_2$ occur. Let Σ be the $(1+\epsilon)\alpha \times 8\alpha$ strip containing S_1 and S_1', where S_1, S_1' are located symmetrically at distance $\epsilon\alpha/2$ from the horizontal borders of Σ which are of length 8α (and ϵ is some sufficiently small positive constant). Let I be the index set of those open facilities whose quantized Voronoi cells $\tilde{V}_i$ meet the strip Σ. We show that there must be some facility $i \in I$ for which $\tilde{V}_i$ is not ϵ^3-quasi-square with area in $(1 \pm \epsilon^3)\alpha^2$. Assume for the sake of contradiction that this is not the case. Each such Voronoi region $\tilde{V}_i$ can therefore be associated with a square W_i of side in the range $(1 \pm \epsilon^3)\alpha$. Furthermore, any two such squares have a common area of at most $\epsilon^3\alpha^2$. **Whp** there is no open facility j at distance 2α or more from Σ for which the quantized Voronoi region $\tilde{V}_j$ intersects Σ (every point in $\mathcal{X} \cap \Sigma$ is connected to a closer open facility). Thus $|I| \leq \frac{8(5+\epsilon)}{1-\epsilon/2} < 50$. It follows that all but an area of at most $50\epsilon^3\alpha^2$ of Σ is covered by the $W_i, i \in I$. Now let Σ_1 denote a strip of length 8α and thickness $\epsilon\alpha/4$ running across the middle of Σ. Any sub-strip of Σ_1 which is of length $\epsilon\alpha$ is of area $\epsilon^2\alpha^2/4$ and so will contain members of $\mathcal{X}$ which are covered by some W_i, $i \in I$.

If the center of this W_i is outside Σ then W_i has side at least $(1+\epsilon/4)\alpha$, which contradicts our assumption. So let J be the set of facilities j with center in Σ for which there is a member of Σ_1 contained in W_j. If any of these facilities is not ϵ^3-quasi-square then we are done, so we may assume that they all are. There is an open facility in $S_{1,q}$ and in $S_{3,q'}'$, and these facilities cover squares of side at least α. Thus the other members of J appear in a substrip with length between 7.25α and 7.75α. If there are 6 or fewer open facilities in this the strip bounding the 2 copies, then some pair of facilities are at least 1.04α apart. Therefore, one of them, call it F_i has a W_i with side at least $(1.04 - 100\epsilon^3)\alpha$, contradiction. On the other hand, if there are 7 or more facilities in the strip, then some pair are at most $.96\alpha$ apart, and so some F_i has a W_i with side at most $(.96 + \frac{2\epsilon^3}{.48})\alpha$, contradiction.

By considering copies of the event $\mathcal{B}_1\mathcal{B}_2$ on sufficiently separated disjoint regions of the square and conditioning on there being enough points in an appropriate region for each copy, **whp** we will have $\Omega(m^2)$ facilities for which $\tilde{V}_i$ is not an

ϵ^3-quasi-square with area $(1 \pm \epsilon^3)\alpha$. Then Lemma 5 finishes the proof of the Theorem. □

5. REFERENCES

[1] S. Ahn, C. Cooper, G. Cornuéjols and A. M. Frieze, Probabilistic analysis of a relaxation for the k-median problem, *Mathematics of Operations Research* 13 (1988) 1-31.

[2] F. Barahona and F. A. Chudak, Solving large scale uncapacitated facility location problems, in *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, Kluwer Academic Publishers, (1999).

[3] M. Charikar and S. Guha, Improved combinatorial algorithms for facility location and k-median problems, *Proc. of the 40th Annual IEEE Symposium on Foundations of Computer Science*, (1999), 378-388.

[4] F. A. Chudak and D. B. Shmoys, Improved approximation algorithms for the uncapacitated facility location problem, to appear in SIAM Journal of Computing.

[5] G. Cornuéjols, G. L. Nemhauser and L. A. Wolsey, The Uncapacitated Facility Location Problem, *Discrete Location Theory*, (1990).

[6] M. Hoefer. Experimental comparison of heuristic and approximation algorithms for uncapacitated facility location. *Proc. 2nd Intl. Workshop on Experimental and Efficient Algorithms (WEA 2003)*, LNCS 2647, (2003) 165-178.

[7] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani, Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP, *Journal of the ACM*, 50 (6), (2003) 795-824.

[8] K. Jain and V. V. Vazirani, Primal-dual approximation algorithms for metric facility location and k-median problems, *Proc. of the 40th Annual IEEE Symposium on Foundations of Computer Science* (1999) 2-13.

[9] J. Krarup and P. M. Pruzan, The simple plant location problem: Survey and synthesis,*European Journal of Operational Research* 12 (1983) 36-81.

[10] M. R. Korupolu, C. G. Plaxton, and R. Rajaraman, Analysis of a local search heuristic for facility location problems, *Proc. of the 9th Annual ACM-SIAM symposium on discrete algorithms* (1998) 1-10.

[11] M. Mahdian, Y. Ye, and J. Zhang, Improved Approximation Algorithms for Metric Facility Location Problems, *Proc. of the 5th Intl. Workshop on Approximation Algorithms for Combinatorial Optimization* (2002) 229-242.

[12] M. Penrose, *Random Geometric Graphs*, Oxford University Press, (2003).

[13] D. B. Shmoys, E. Tardos, and K. I. Aardal, Approximation algorithms for facility location problems, *Proc. of the 29th Annual ACM Symposium on the Theory of Computation* (1997) 265-274.

[14] J. M. Steele, *Probability Theory and Combinatorial Optimization CBMS-NSF Regional Conference Series in Applied Mathematics – Volume 69*, SIAM, (1997).

[15] M. Sviridenko, An Improved Approximation Algorithm for the Metric Uncapacitated Facility Location Problem, *Proc. of the Conference on Integer Programming and Combinatorial Optimization (IPCO02)*, LNCS 2337, (2002) 230-239.

[16] J. E. Yukich, *Probability Theory of Classical Euclidean Optimization Problems*, Springer-Verlag, (1998).