Probabilistic Analysis of a Greedy Heuristic for Euclidean Matching

By

David Avis[1]
School of Computer Science
McGill University
Montréal, Québec H3 AZ K6

Burgess Davis[2]
Department of Mathematics and Statistics
Purdue University
West Lafayette, IN 47907

J. Michael Steele[3]
Department of Statistics
Princeton University
Princeton, N.J. 08544

Technical Report #84-13

Department of Statistics
Purdue University

March 1984

[1]Research supported by N.S.E.R.C. Grant A3013 and F.C.A.C. Grant EQ-1678

[2]Research supported in part by NSF Grant

[3]Research supported in part by NSF Grant and DOE Grant

Abstract

A heuristic (called GREEDY) for the Euclidean matching problem successively matches the two closest points. We study the behavior of the length G_n of the resulting matching as the number n of points goes to infinity, if the points are random. We prove that if the points are independent two dimensional random variables with a common distribution which has compact support, then $G_n/\sqrt{n}$ approaches a constant almost surely.

I. Introduction

The main goal of this paper is to provide an asymptotic understanding of the greedy matching algorithm as applied to a set of random observations. We will first make this precise and then outline the motivations from computer science and probability theory which lead us to this problem.

For an even integer n and a set $\{x_1, x_2, \ldots, x_n\}$ of n points in the Euclidean plane, a <u>matching</u> is a collection of n/2 disjoint pairs of points. By the <u>weight</u> of a matching we will denote the sum of the Euclidean distances between the element of the pairs in the matching. An object of particular interest is the <u>minimal weight matching</u>.

The best known algorithms for finding a minimal weight matching are implementations of Edmonds' algorithm (see e.g. [6]), and these algorithms have running times proportional to n^3. Because of the relative slowness of the Edmonds algorithm, substantial attention has been given to heuristic methods for obtaining almost optimal matchings. As has been observed by Reingold and Tarjan [8], such heuristics are useful in minimizing the time required to plot a connected graph using a mechanical plotter. This application is also discussed in Iri et al. [4].

One particularly appealing heuristic is the so-called GREEDY algorithm. Here one successively matches the two closest unmatched pairs of points. Despite its simplicity the GREEDY algorithm is surprisingly difficult to analyze both in terms of its running time and in terms of the quality of the solution produced.

The fastest known implementation of GREEDY is due to Bentley and Saxe [3] and the running time of their algorithm is $0\ (n^{3/2} \log n)$. To survey the quality of the GREEDY matching, we let $X = \{x_1, x_2, \ldots, x_n\}$, and we write

$G_n = G(X)$ and $OPT_n = OPT(X)$ for the weight of the GREEDY and the optimal matchings respectively.

Reingold and Tarjan [8] have shown that

$$(1.1) \qquad \frac{G_n}{OPT_n} \leq \frac{4}{3} n^{\log_2 1.5},$$

and they prove also that this bound is achievable for certain configurations. For the plotting application mentioned above, it is convenient to assume that the points are contained in a bounded region, for example the unit square. In this case AVis [1] has shown that

$$(1.2) \qquad G_n \leq 1.074 \sqrt{n} + O(\log n).$$

For additional results on this problem, one can consult the survey paper [2].

Because of the possibility of exploiting probabalistic algorithms in the tradition of Karp's probabilistic algorithm for the traveling salesman problem (Karp [5]), a class of stochastic processes called subadditive Euclidean functions was introduced in Steele [9]. If X_i, $1 \leq i < \infty$ are independent, identically distributed random variables with bounded support, then the theory of subadditive Euclidean functionals establishes that

$$(1.3) \qquad \text{Opt}\{X_1, X_2, \ldots X_n\} \sim C\sqrt{n}$$

with probability one. The theory of subadditive Euclidean functionals does not apply directly to the GREEDY matching, and the main accomplishment of this page is to provide tools which do apply to GREEDY.

Our main result is the following

Theorem 1.1. For each integer $d \geq 2$ there is a positive absolute constant C_d such that if $X_1, X_2, \ldots$ are iid random variables taking values in $\mathbb{R}^d$, and having bounded support, and if G_n denotes the Euclidean edge weight of the matching attained by the GREEDY algorithm applied to $\{X_1, X_2, \ldots X_n\}$ then with probability one

$$G_n \sim C_d n^{(d-1)/d} \int_{\mathbb{R}^d} f(x)^{1/d}\, dx.$$

Here, f is the density with respect to d dimensional Lebesgue measure of the absolutely continuous part of the distribution of the X_i.

If the root density is not integrable the result reads more precisely $G_n/n^{(d-1)/d} \longrightarrow \infty$ with probability one.

The proof of this theorem is pretty much the same for each $d \geq 2$, so it will just be given for the case $d = 2$. We remark in passing that while the asymptotic behavior of the minimal weight matching edge weight for, say, iid uniform (0,1) one dimensional random variables, is trivial to ascertain (it approaches 1/2 almost surely), this behavior is unknown for the GREEDY matching edge weight.

The arguments used by Steele in [9] to handle OPT $\{X_1, \ldots, X_n\}$ are based on the theory of subadditive independent processes, while the arguments here are not. Another difference between this paper and [9] is that in [9] the key inequalities hold for all collections of points, while here they hold only for collections which are likely to occur.

II. Preliminary Lemmas

In this section some nonprobabilistic results about the greedy matching algorithm are proved.

Lemma 2.1. *There is a* c_1 *such that for any* n *points* $\{x_1,x_2,\ldots,x_n\} \subset [0,1]^2$, *there is a pair* $1 \le i < j \le n$ *such that* $|x_i - x_j| \le c_1 n^{-1/2}$.

Proof. If each x_i were covered by non-intersecting discs of radius r each disc would cover at least $\pi r^2/4$ of the square. This implies $n\pi r^2/4 \le 1$ which yields the lemma with $c_1 = 2/\sqrt{\pi}$.

Lemma 2.2. *There is a constant* c_2 *such that any* k *edges of a greedy matching of* $\{x_1,x_2,\ldots x_n\} \subset [0.1]^2$ *will have total weight at most* $c_2\sqrt{k}$.

Proof. If $e_1,e_2,e_3\ldots e_k$ are the edges of the greedy matching we note that by Lemma 2.1, $|e_j| \le c_1(n-2j)^{-1/2}$, $1 \le j \le [n/2]$, where $||$ denotes length. Hence we have

$$\sum_{m=1}^{k} |e_m| \le \sum_{m=1}^{k} c_1 m^{-1/2} \le 2c_1 k^{1/2} ,$$

which proves the lemma with $c_2 \le 2c_1$.

Lemma 2.3. *For any* δ, *the greedy* matching of $\{x_1,x_2,\ldots x_n\} \subset [0,1]^2$ *has at most* $c_3\delta^{-2}$ *edges as great as* δ.

Proof. If there are τ edges as large as δ then the total length of these edges is at least $\tau\delta$ which by Lemma 2.2 satisfies, $\tau\delta \le c_2\tau^{1/2}$, i.e. $\tau \le c_2^2/\delta^2$, and we get $c_3 \le c_2^2$.

Lemma 2.4. $|G(x_1,x_2,\ldots x_n) - G(x_1,x_2,\ldots x_{n+k})| \leq c_4\sqrt{k}$

Proof. We will proceed algorithmically. First let $e_1,e_2,\ldots e_\tau$ (with $\tau = \binom{n+k}{2}$) be the edge list of the complete graph on $\{x_1,x_2,\ldots x_{n+k}\}$ sorted by edge length. An algorithm which simultaneously and inductively constructs the greedy matchings A and B for $\{x_1,x_2,\ldots x_{n+k}\}$ and $\{x_1,x_2,\ldots x_n\}$ respectively is the following:

for i = 1 to τ do

if e_i has no endpoint in A, add e_i to A,

if e_i has no endpoint in B <u>and</u> no endpoint in $D = \{x_{n+1},x_{n+2},\ldots x_{n+k}\}$, add e_i to B.

We have an induction hypothesis:

At iteration i, all non-double edges of AUB are either

(a) edges with an endpoint in D, or

(b) edges connected by at least one alternating path of monotone decreasing edges to a vertex in D.

The hypothesis is valid for i = 1, so consider edge e_{i+1}. If e_{i+1} has an endpoint in D, condition (a) holds and the induction hypothesis still lives. If e_{i+1} is in both A and B, the induction is valid (because of e_{i+1} being a double edge). We are left with two cases:

Case (1). Suppose $e_{i+1} \in A$ but $e_{i+1} \notin B$, and e_{i+1} has no endpoint in D. Since $e_{i+1} \notin B$ one endpoint of e_{i+1} must meet an edge already in B but which is not in A. By the induction hypothesis, we have then found an alternating path of AUB which is monotone and leads to D.

Case (2). This concerns the possibility that $e_{i+1} \in B$ but $e_{i+1} \notin A$ and e_{i+1} does not meed D. The induction step is identical to that of Case (1).

To complete the proof of the lemma, we note that

$$\sum_{e\in A} |e| - \sum_{e\in B} |e| \le \sum_{e_i \text{ meets } D} |e_i| + \sum_{C} \max\{|e| : e \in C\},$$

where the last sum is the sum over all of the alternating chains provided by the induction hypothesis. Since there are at most $2|D|$ such chains (because of having to end in D) we can majorize the two sums by taking three times the sum of the largest $k = |D|$ edges in either A or B. By Lemma 2.2, this provides the required bound.

We now need to consider how breaking the unit squares $Q = [0,1]^2$ into m^2 equal subsquares Q_i, $1 \le i \le m^2$ changes a greedy matching. We let $F = \bigcup_{i=1}^{m^2} \partial Q_i - \partial Q$, i.e. F is the interior grating of the partition of Q given by the Q_i. In just the same way as we established Lemma 4, one can prove the following:

Lemma 2.5. We let A denote the greedy matching of $\{x_1, x_2, \ldots x_n\} \subset Q$ and let B be the union of the m^2 greeding matchings of $\{x_i, x_2, \ldots x_n\} \cap Q_i$, $1 \le i \le m^2$. The union $A \cup B$ consists of (a) double edges, i.e., edges belonging to both A and B (b) edges which intersect F and (c) edges which are joined to F by a monotone alternating path.

Lemma 2.6. Let τ denote the number of edges in the greedy matching of Q which have either (1) an endpoint within $\delta/2$ of F or (2) length as great as δ. We have the bound,

$$|G(Q) - \sum_{i=1}^{m^2} G(Q_i)| \le c_6 \tau^{1/2} \quad .$$

Proof. By Lemma 2.5 we have

$$|G(Q) - \sum_{i=1}^{m^2} G(Q_i)| \leq \sum_P \max\{|e| : e \in P\}$$

where the sum is over all alternating A, B paths P which connect an edge of a Q_i matching to F. The number of such paths is bounded by 2τ since each path must contain an edge of A which hits F, and each edge which hits F can be on at most two alternating paths. Now, if M is the set of maximal edges, one representative chosen for each P, then

$$\sum_P \max\{|e| : e \in P\} = \sum_{e \in M \cap A} |e| + \sum_i \sum_{\substack{e \in M \\ e \subset Q_i}} |e| .$$

If n_A denotes the number of summands of the first sum and if $n_B(i)$ denotes the number of summands of $\sum_{\substack{e \in M \\ e \subset Q_i}} |e|$ then $n_A + \sum_i n_B(i) \leq 2\tau$. We know by Lemma 2.2 that

$$\sum_{e \in M \cap A} |e| \leq c_2 n_A^{1/2}, \text{ and } \sum_{\substack{e \in M \\ e \subset Q_i}} |e| \leq e_2 (n_B(i))^{1/2} m^{-1},$$

using Lemma 2.2 and scaling. So, by Cauchy's inequality, and the bound given by τ, we complete the proof of this lemma.

Remark: It's easy to see that if the X_i were i.i.d. uniform in Q, then $E\tau \leq 2nm\delta + c/\delta^2$. Proper choice of δ will yield $(E\tau)^{1/2} << (nm)^{1/3}$.

III. Uniformly Distributed Case.

We now let X_i, $i \geq 1$, be i.i.d. uniformly distributed in $[0,1]^2 = Q$ and consider N an independent Poisson random variable with parameter λ. We let $\varphi(\lambda) = EG(X_1,X_2,\ldots,X_N)$ and note that, $\varphi(\lambda)$ is a smoothing of the sequence $g_n = EG(X_1,X_2,\ldots X_n)$, i.e.

$$\varphi(\lambda) = \sum_{n=0}^{\infty} g_n(\lambda^n/n!)e^{-\lambda} \quad .$$

If Q_i are defined as before then by scaling and by well known properties of the Poisson process, $G(Q_i) = G(\{X_1,X_2,\ldots X_n\} \cap Q_i), 1 \leq i \leq m^2$, are independent, identically distributed random variables and $EG(Q_i) = \varphi(\lambda/m^2)m^{-1}$. Moreover, by Lemma 2.6,

$$(3.1) \quad G(Q) = \sum_{i=1}^{m^2} G(Q_i) + O(\tau^{1/2})$$

where τ is the number of edges of the greedy matching of Q which are (1) with any end point within $\delta/2$ of the boundary of some Q_i or (2) have length at least δ, (for any $\delta > 0$).

Now, $E(\tau^{1/2}) \leq (E(\tau))^{1/2}$ and

$$E(\tau) \leq 2m\lambda\delta + c_3\delta^{-2}$$

because the area of Q which is within $\delta/2$ of ∂Q_i for some i is bounded by $2m\delta$, and because of the Lemma 3 bound on the number of long edges. Picking δ proportional to $(m\lambda)^{-1/3}$ minimizes $E(\tau)$ and gives

$$(3.2) \quad E(\tau) = O((m\lambda)^{2/3}).$$

Returning to the basic (3.1), and taking expectations gives

$$\varphi(\lambda) = m\varphi(\lambda/m^2) + O(\lambda^{1/3}m^{1/3})$$

or

$$(3.3) \quad \frac{\varphi(m^2\lambda)}{m\sqrt{\lambda}} = \frac{\varphi(\lambda)}{\sqrt{\lambda}} + O(\lambda^{-1/6}) \quad , m = 1,2,\ldots$$

It will be easy to use (3.3) to show $\varphi(\lambda)/\sqrt{\lambda}$ converges.

One route is to choose $a < b$ such that for $\lambda > a$ the error term in (3.3) is less than $\varepsilon > 0$ and such that $\varphi(\lambda)/\sqrt{\lambda} \leq \liminf \varphi(t)/\sqrt{t} + \varepsilon$ for $\lambda \in (a,b)$. Then

$$(3.4) \quad \varphi(m^2\lambda)/m\sqrt{\lambda} \leq \liminf \varphi(t)/\sqrt{t} + 2\varepsilon$$

for all m and $\lambda \in (a,b)$. Since $\bigcup_{m=1}^{\infty} (ma,mb)$ contains a right half line, this proves Lemma 3.1.

$$\lim_{\lambda\to\infty} \varphi(\lambda)/\sqrt{\lambda} = c > 0.$$

(The strict positivity of c is easy to check by elementary considerations.)

We now consider the case where $\lambda = n$ and where m is chosen as a function of n; in particular, $m = [\alpha n^{1/2}]$ when α is a small number to be chosen later. With these choices, we note $G(Q_i)$ are independent random variables, and more pointedly, the random variables converge in distribution and L^P,

$$(3.4) \quad mG(Q_i) \longrightarrow H \quad , \text{ as } n \longrightarrow \infty ,$$

where H is the length of the greedy matching of N' points in $[0.1]^2$ where N' is Poisson with mean $\lambda = \alpha^{-1}$. This convergence is an easy consequence of the fact that the number of points in Q_i is Poisson with mean $n/m^2 \sim \alpha^{-1}$ and the stability lemma, Lemma 2.4.

Next, let $\mu_n = EmG(Q_i)$, $Y_i = mG(Q_i) - \mu_n$, then note

$$EY_i^2 \longrightarrow \sigma^2 = \text{Var } H, \; EY_i^4 \longrightarrow k = E(H - \alpha^{-1})^4, \text{ and}$$

$$(3.5) \quad E\left(\sum_{i=1}^{m^2} Y_i\right)^4 << m^2 k + 6\binom{m^2}{2}\sigma^4$$

because of the independence of the Y_i (and because $EY_i = 0$ so odd terms drop out).

We now begin our basic estimation. By Lemma 3.1, given any $\varepsilon > 0$ we can choose an α such that we have $|EG(Q) - \sum_{i=1}^{m^2} EG(Q_i)| \leq \varepsilon n^{1/2}$ for all $n \geq n(\varepsilon)$. Hence, we have

$$(3.6) \quad P_n = P(n^{-1/2}|G(Q) - EG(Q)| \geq \varepsilon)$$
$$\leq P(n^{-1/2} \; |\sum_{i=1}^{m^2} \{G(Q_i) - EG(Q_i)\}| \geq \varepsilon/2)$$
$$+ P(n^{-1/2}|G(Q) - \sum_{i=1}^{m^2} G(Q_i)| \geq \varepsilon/2).$$

By inequality (3.5), Markov's inequality, and the definition of $m = [\alpha n^{1/2}]$ we see the first probability is bounded by $c\varepsilon^{-4}\alpha^4\sigma^4 n^{-2}$ for some c and n sufficiently large.

Next choosing $\delta = n^{-1/3}m^{-1/3}$ in Lemma 2.6 we see

$$D_n = |G(Q) - \sum_{i=1}^{m^2} G(Q_i)|^2 \leq c_6^2 \, (N' + N''),$$

where N' is the number of points within $\delta/2$ of F, and N'' is the number of edges of length exceeding δ. Now by Lemma 2.3, $N'' \leq c_3\delta^{-2} = c_3 n^{2/3}m^{2/3}$, and it is easily checked that N' is Poisson with mean $2n^{1/3}m^{1/3}$. Since $(nm)^{2/3} \sim \alpha n$, we note elementary bounds on the Poisson tail give

$$(3.7)\qquad \sum_{n=1}^{\infty} P(n^{-1}D_n \geq \varepsilon^4/4) < \infty \quad ,$$

provided $\alpha \leq (\varepsilon^2/8)/(c_6^2\, c_3)$. By these bounds on the two righthand summands (3.6) we see the probabilities P_n are summable.

To relate the Poisson averaged probabilities to the case of fixed n, it will be useful to slightly vary our notation. Let X_i be i.i.d. uniform on $[0,1]^2$ and let N be Poisson with mean n. We have just proved that the probabilities

$$P_n = P\left(\left|\frac{G(X_1,X_2,\ldots X_N)}{\sqrt{n}} - c\right| \geq \varepsilon\right)$$

are summable. Now by Lemma 2.4,

$$|G(X_1,X_2\ldots,X_N) - G(X_1,X_2,\ldots X_n)| \leq c_4|N - n|^{1/2} \text{ , so the bound}$$

$$P\left(\frac{N - n}{n} \geq \beta\right) \leq e^{-n\beta^2/2}$$

shows that also the probabilities

$$P_n{}' = P\left(\left|\frac{G(X_1,X_2,\ldots X_n)}{\sqrt{n}} - c\right| \geq \varepsilon\right)$$

are summable. Thus $n^{-1/2}G(X_1,X_2,\ldots X_n)$ not only converges almost surely, but also converges completely.

IV. General Extension

To aid intuition we first treat the case where the X_i are i.i.d. with compact support in $[0,1]^2$ and with density f. We can partition $[0,1]^2$ into squares of side 1/m and find a step function φ supported on the partition for which $\int|f - \varphi|dx \leq \varepsilon$. We recall Strassen's Lemma, ([10], 1965).

Lemma 4.1. _Let_ μ_1 _and_ μ_2 _be two Borel probability measures on_ $\mathbb{R}^2$ _such that_

$|\mu_1(F) - \mu_2(F)| \leq \delta$ _for all closed sets_ F. _Then there are random variables_ X _and_ Y, _taking values in_ $\mathbb{R}^2$, _and defined on the same probability space,_

<u>such that</u> $P(X \neq Y) \leq \delta$ <u>and with the distribution of</u> X <u>equal to</u> μ_1 <u>and the distribution of</u> Y <u>equal to</u> μ_2.

For f and φ above we can chose Y_i i.i.d. such that Y_i has density φ and $P(X_i = Y_i) < \varepsilon$.

We let $A = \{i: X_i = Y_i\}$ and note by Lemma 2.4

$$\begin{aligned} |G(X_1,X_2,\ldots,X_n) - G(Y_1,Y_2,\ldots Y_n)| &\leq |G(X_i: i\in A) - G(X_1,X_2,\ldots X_n)| \\ &\quad + |G(Y_i: i\in A) - G(Y_1,Y_2,\ldots Y_n)| \\ &\leq 2c_4|A^c|^{1/2}. \end{aligned}$$

Since this last term is a.s. majorized by $2c_4\varepsilon^{1/2}n^{1/2}$, it suffices to show

$G(Y_1,Y_2,\ldots,Y_n) \sim cn^{1/2} \int\varphi^{1/2}dx$ in order to prove $G(X_1,X_2,\ldots X_n) \sim cn^{1/2} \int f^{1/2}dx$.

To prove the former, note

$$G(Y_1,Y_2,\ldots Y_n) = \sum_{j=1}^{m^2} G(Y_i\in Q_j) + c_6\tau^{1/2}$$

where $\tau \leq c_3\delta^{-2} + \sum_{i=1}^{n} 1(X_i \in \{x_i: |x - \bigcup_{i=1}^{m^2} \partial Q_i| \leq \delta/2\})$. By the main result of Section III

$$G(Y_i\in Q_j) \sim \frac{c}{m}\Big(\sum_{i=1}^{n} 1(Y_i\in Q_j)\Big)^{1/2}$$

but, since $P(Y_i\in Q_j) = m^{-2}h_j$, ($h_j$ the height of φ on Q_j) we have,

$$G(Y_i\in Q_j) \sim cn^{1/2} \int_{Q_j} \varphi^{1/2}dx.$$

Chosing $\delta = n^{-1/3}$ is good enough to show $\tau^{1/2} = O(n^{1/2})$ a.s., so we have proved.

Lemma 4.2 <u>If</u> X_i <u>are i.i.d. with density</u> f <u>and compact support then</u> $G(X_1,X_2,\ldots X_n) \sim Cn^{1/2} \int f^{1/2}dx$.

Now we treat the case where the distribution of X_i, while still concentrated in the unit square, may have a singular component, which we denote μ. Let K, a set of Lebesgue measure zero, be the support of μ, and let g be the density, with respect to Lebesgue measure, of the absolutely continuous part of the distribution of X_i, so that

$$P(X_i \in A) = \int_A g\,dx + \mu(A \cap K).$$

Let $\varepsilon > 0$. Let $S_1, S_2, \ldots, S_r$ be closed squares each of side length α, such that, if $3S_i$ denotes the square with the same center as S_i and side length 3α, $\Gamma = \bigcup_{i=1}^{r} S_i$, and $3\Gamma = \bigcup_{i=1}^{r} 3S_i$, we have

i) $\mu(K - \Gamma) < \varepsilon$

ii) $\alpha^2 r < \varepsilon$

iii) $\int_{3\Gamma} g < \varepsilon$.

Now let $A_n = \{X_i: X_i \in K \cap \Gamma,\ i \leq n)$ and

$$B_n = \{X_i: X_i \notin K,\ i \leq n).$$

It is easily proved using Lemma 4.2 that $G(B_n)/\sqrt{n} \longrightarrow c\int\sqrt{g}\,dx$ a.s. as $n \longrightarrow \infty$. We note that every edge in the greedy matching of $\{X_1, X_2, \ldots X_n\}$ which has an endpoint in A_n must have the other endpoint in 3Γ, with at most r exceptions, one for each of the squares S_i. This is true since each point outside 3Γ is farther from the points of $K \cap \Gamma$ than the diameter of the S_i, so a point in S_i would be matched with any other point in S_i before being matched with a point outside 3Γ. Let D_n be the collection of these at most r exceptions, and E_n denote all the points in $A_n - D_n$ together with the points that the points in $A_n - D_n$ are matched with in the greedy matching of $\{X_1, \ldots X_2, \ldots X_n\}$.

Next, observe that the edges in the greedy matching of $\{X_1, X_2, \ldots, X_n\}$ which do not connect the points of E_n are precisely the edges in the greedy matching of $\{X_1, \ldots, X_n\} - E_n = H_n$, so that $G(\{X_1, \ldots, X_n\}) = G(H_n) + G(E_n)$.

To complete the proof we will show that $G(E_n)/\sqrt{n}$ is small and then that $G(H_n)/\sqrt{n}$ and $G(B_n)/\sqrt{n}$ are close together if n is large and ε is small.

Now all the edges of the greedy matching of E_n fall into one of the r squares $3S_1, 3S_2, \ldots, 3S_r$. Further, if n_i is the number of edges in $3S_i$, the total length of these edges is no more than $3\alpha c_2\sqrt{n_i}$ using the argument of Lemma 2.2. Now $\sum_{i=1}^{r} n_i \leq 100n$ (the 100 is conservative) since the S_i are disjoint, so that each X_i in E_n is in at most 100 of the $3S_i$. Thus using Cauchy's inequality we have

$$\begin{aligned} G(E_n) &\leq \sum_{i=1}^{r} 3\alpha c_2\sqrt{n_i} \leq \sum_{i=1}^{r} 3\alpha c_2 \sqrt{\frac{100n}{r}} \\ &= 30c_2\sqrt{\alpha^2 r}\;\sqrt{n} \\ &= 30c_2\;\sqrt{\varepsilon}\,\sqrt{n}. \end{aligned}$$

Finally, we note that the points in B_n which are not in H_n are those points which are matched with points of $A_n - D_n$. By definition each of these points falls in 3Γ, so by iii) we have cardinality $(B_n - H_n) \ll \varepsilon n$. The points in $H_n - B_n$ consist of the at most r points in D_n together with the points in $K - \Gamma$, so by i), cardinality $(H_n - B_n) \ll \varepsilon n$. Now Lemma 2.4 gives

$$\begin{aligned} |G(H_n) - G(B_n)| &\leq |G(H_n) - G(H_n \cup B_n)| + |G(B_n) - G(H_n \cup B_n)| \\ &\ll 2c_2\sqrt{\varepsilon n}, \end{aligned}$$

finishing the proof of Theorem 1.1 in the case of $d = 2$. As has been mentioned, the proof for higher dimensions is similar.

References

[1] D. Avis, Worst case bounds for the Euclidean matching problem, Computers and Math. Appl. 7 (1981) 251-257.

[2] D. Avis, A survey of heuristics for the weighted matching problem, Networks 13 (1983).

[3] J. L. Bentley and J. B. Saxe, Decomposable searching problems 1: Static to dynamic transformations, J. Algorithms 1 (1980) 301-358.

[4] M. Iri, K. Murota, and S. Matsui, Linear time heuristics for the minimum weight perfect matching problem on a plane, Information Proc. Letters 12 (1981) 206-209.

[5] Karp, R. M. The probabilistic analysis of some combinatorial search algorithms. In Proc. Symp. New Directions and Recent Results in Algorithms and Complexity. (F. Taub, ed.) Academic, San Francisco.

[6] C. H. Papadimitrion and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[7] C. H. Papadimitrion, The probabilistic analysis of matching heuristics, Proc. 15 Allerton Conf. on Comm. Control and Computing (1977) 368-378.

[8] E. M. Reingold and R. E. Tarjan, On a greedy heuristic for complete matching, SIAM J. Computing 10 (1981) 676-681.

[9] J. M. Steele, Subadditive Euclidean functionals and non-linear growth in geometric probability, Ann. Probability (1981) 365-376.

[10] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Statist. 36 423-439.