Halving Point Sets

Artur Andrzejak and Emo Welzl

Abstract. Given n points in $\mathbb{R}^d$, a hyperplane is called halving if it has at most $\lfloor n/2 \rfloor$ points on either side. How many partitions of a point set (into the points on one side, on the hyperplane, and on the other side) by halving hyperplanes can be realized by an n-point set in $\mathbb{R}^d$?

1991 Mathematics Subject Classification: 52C10, 52B05, 52B55, 68R05, 68Q25, 68U05

Keywords and Phrases: combinatorial geometry, computational geometry, k-sets, k-levels, probabilistic method, matroid optimization, oriented matroids, Upper Bound Theorem.

Consider the following algorithmic problem first. Given n points in $\mathbb{R}^d$, we want to find a hyperplane that minimizes the sum of Euclidean distances to these n points. A glimpse of reflection tells us that an optimal hyperplane cannot have a majority ($\lfloor n/2 \rfloor + 1$ or more) of the points on either side; otherwise a parallel motion towards this side will improve its quality [YKII, KM]. A hyperplane with at most $\lfloor n/2 \rfloor$ points on either side is called *halving*. How many partitions of a point set (into the points on one side, on the hyperplane, and on the other side) by halving hyperplanes can be realized by an n-point set in $\mathbb{R}^d$? The notions and results mentioned below are closely related to this question. Emphasis in the presentation is on techniques that may be useful elsewhere, and on interconnections to other topics in discrete geometry and algorithms. A more complete treatment is in preparation [AW].

Halving edges and a crossing lemma

Let P be a set of n points in the plane, n even, and no three points on a line. A *halving edge* is an undirected edge between two points, such that the connecting line has the same number of points on either side. Around 1970 L. Lovász [Lo] and P. Erdős et al. [ELSS] were the first to investigate the geometric graph of halving edges of a point set, and proved that there cannot be more than $O(n^{3/2})$ such edges. Except for a small improvement to $O(n^{3/2}/\log^* n)$ [PSS], there was no progress on the problem until T. Dey [De] recently gave an upper bound of $O(n^{4/3})$. He shows that the graph of halving edges cannot have more than $O(n^2)$ pairs of crossing edges. Then he employs a *crossing lemma* (due to M. Ajtai et al. [ACNS] and T. Leighton [Le]), which has a number of other applications: A geometric

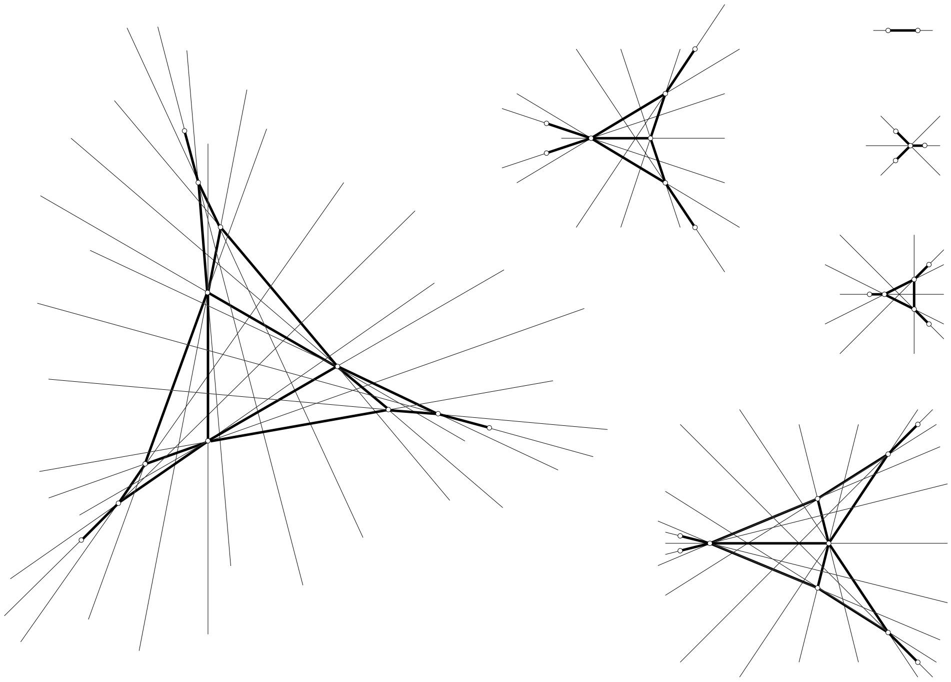

Figure 1: Graphs of halving edges. The configurations maximize the number of halving edges for the given number of points [AAHSW]. Note that, in general, graphs of halving edges are not plane!

graph with n vertices and c pairs of crossing edges has at most $O(\max(n, \sqrt[3]{cn^2}))$ edges.

A variation of Dey's proof ([AAHSW]) goes via the following identity.

Lemma 1

$$C + \sum_{p \in P} \binom{(\deg p + 1)/2}{2} = \binom{n/2}{2}$$

where $\deg p$ *is the number of halving edges incident to* p *(this number is always odd), and* C *is the number of pairwise crossings of halving edges.*

The lemma shows that the number of pairwise crossings in a graph of halving edges is bounded by $\binom{n/2}{2} < n^2/4$. We will now prove the implication on the number, m, of halving edges of P. Recall that a geometric graph without crossings of edges has at most $3n-6$ edges. Now we choose a random induced subgraph G_x of the graph of halving edges of P by taking each point with probability[1] $x = 2/\sqrt[3]{n}$, independently from the other points. Let P_x be the resulting point set, let m_x be the number of halving edges of P with both endpoints in P_x, and let C_x be the

[1]Here we have to assume that $n \geq 8$.

number of pairwise crossings among such edges. We know that $m_x - C_x \leq 3n_x - 6$, since all crossings in G_x can be removed by deletion of C_x edges. The expected value for n_x is xn, for m_x it is $x^2 m$, and for C_x it is $x^4 C$. Hence, due to linearity of expectation, $x^2 m - x^4 C \leq 3xn - 6$, which gives[2]

$$m \leq x^2 C + 3n/x - 6/x^2 \leq \frac{4}{\sqrt[3]{n^2}} \frac{n^2}{4} + 3n \frac{\sqrt[3]{n}}{2} = \frac{5}{2} n^{4/3} \ .$$

For a proof of Lemma 1, we first observe that the identity holds if P is the set of vertices of a regular n-gon. Then the halving edges are connecting the antipodal vertices of the polygon. We have $n/2$ halving edges, $\deg p = 1$ for all points, and any pair of halving edges crosses. An alternative example is given by the vertices of a regular $(n-1)$-gon together with its center. Then the halving edges connect this center with the other points, with no crossing of halving edges. In a second and final step one verifies that the identity remains valid under continuous motion of a point set. We will not go through this argument, but we mention here a lemma due to L. Lovász [Lo], which is essential for this argument and for most proofs in this context.

LEMMA 2 *Let line ℓ contain a unique point p in P. Assume there are x halving edges incident to p emanating into the side of ℓ which contains less points from P than the other side of*[3] *ℓ. Then there are $x+1$ halving edges emanating into the other side of ℓ.*

The lemma can be proven by rotating a line λ about point p starting in position ℓ until it coincides with ℓ again. The halving edges incident to p are encountered in alternation on the large and small side of ℓ, starting and ending on the large side.

It is remarkable, that the graph of halving edges is the *unique* graph that satisfies Lemma 2, i.e., it completely characterizes the graph of halving edges of a point set. Simple implications of the lemma are that the number of halving edges incident to a point in P is always odd, and that there is exactly one halving edge incident to each extreme point of P. Moreover, we have the following implication, which, in fact, is equivalent to Lemma 2.

COROLLARY 1 *Let ℓ be a line disjoint from P with x points from P on one side and y points on the other side, $x+y=n$. Then ℓ crosses $\min(x,y)$ halving edges of P.*

The corresponding problem of bounding the number of *halving triangles* of n points in $\mathbb{R}^3$, n odd, has also been investigated in a sequence of papers with a currently best bound of $O(n^{8/3})$ due to T. Dey and H. Edelsbrunner [DE]. Building blocks of the proof are a probabilistic argument similar to the one given above, and a counterpart of Corollary 1: No line crosses more than $n^2/8$ halving triangles.

While the bound in $\mathbb{R}^3$ still allows for a simple proof, the situation gets more involved in dimensions 4 and higher, where the best bounds due to P. Agarwal et al. [AACS] are based on a colored version of Tverberg's Theorem [Tv] by R. T. Živaljević and S. T. Vrećica [ZV].

[2]The general bound of $O(\max(n, \sqrt[3]{cn^2}))$ in the crossing lemma [ACNS, Le] is obtained with $x = \min(1, \sqrt[3]{n/c})$. The best known constant in the asymptotic bound can be found in [PT].

[3]This side is unique, since ℓ contains a point, and $|P|$ is even.

k-LEVELS AND PARAMETRIC MATROID OPTIMIZATION.

Let H be a set of n non-vertical lines in $\mathbb{R}^2$. For $0 \leq k \leq n-1$, the k-level of the arrangement of H is the set of all points which have at most k lines below and at most $n-k-1$ above. Clearly, points on the k-level must lie on at least one line. Moreover, the k-level can be easily seen to be an x-monotone polygonal curve from $-\infty$ to $+\infty$, since it intersects every vertical line in exactly one point.

We will now show how the halving edges of a planar point set P, $|P|$ even, correspond to vertices of the $(n/2-1)$- and $(n/2)$-level of some line arrangement. To this end we consider the mapping $p = (a, b) \mapsto p^* : y = ax + b$ from points to non-vertical lines, and the mapping $h : y = kx + d \mapsto h^* = (-k, d)$ from non-vertical lines to points. This mapping preserves incidences and relative position: p lies on h iff p^* contains h^*, and p lies above h iff h^* lies below p^*. Set $P^* = \{p^* | p \in P\}$. Now a pair of points p and q is connected by a halving edge iff the intersection[4] of p^* and q^* lies both on the $(n/2-1)$- and the $(n/2)$-level of the arrangement of P^*.

The results in [De] imply an upper bound of $O(n\sqrt[3]{k+1})$ on the number of vertices on the k-level. k-levels have a number of applications in the analysis of algorithmic problems in geometry. We briefly outline here a connection where the methods for analyzing k-levels proved useful.

A *matroid* of rank k consists of a set of n elements and a non-empty family of k-element subsets, called *bases*. The family of bases is required to fulfill the basis exchange axiom: for two bases B_1, B_2 and an element $x \in B_1 \setminus B_2$ we can always find $y \in B_2 \setminus B_1$ such that $(B_1 \setminus \{x\}) \cup \{y\}$ is again a basis.

Typical examples of matroids are the set of edges of a graph with its spanning trees as bases, or a set of vectors with its bases. If we equip the elements e with weights $w(e)$, we can ask for the minimal weight basis (i.e., the basis with minimal sum of weights). The matroid property ensures that the greedy method finds such an optimal basis.

Assume now that the weights are linear functions $w(e) = k_e\lambda + d_e$ depending on some real number λ [Gu, KI]. While λ ranges from $-\infty$ to ∞, we obtain a sequence of minimal weight bases. How long can this sequence be?

By plotting the weights of the elements along the λ-axis, we obtain an arrangement of n lines. The changes of the minimal weight basis occur at vertices of this arrangement. In the special case of a uniform matroid, i.e., where each set of k elements forms a basis, the changes of minimal weight basis occur at the vertices of the $(k-1)$-level of the line arrangement. N. Katoh was the first to notice this connection.

For general rank k matroids it is known [Ep] that the length of the minimal base sequence is bounded by the total number of vertices of k convex polygons whose edges do not overlap and are drawn from n lines. T. Dey [De] has shown an upper bound $\mathrm{O}(nk^{1/3}+n^{2/3}k^{2/3})$ (which is $\mathrm{O}(nk^{1/3})$ for $k \leq n$) on this quantity by a modification of his proof for the complexity of a k-level. This bound is optimal, due to a lower bound $\Omega(nk^{1/3})$ obtained by D. Eppstein [Ep].

[4] This intersection may vanish to infinity, if the halving edge is vertical.

For graphs the problem looks at the number of different minimal spanning trees for edge weights parameterized by some linear fuction in a parameter λ. The best lower bound for this quantity is $\Omega(n\alpha(k))$ [Ep], where n is the number of edges, $k+1$ is the number of vertices, and α is a slowly growing inverse of the Ackermann function. The known upper bound is the same as for general matroids.

Lower bounds and oriented matroids

The upper bounds mentioned may be far from optimal. In the plane several constructions of n-point sets with $\Omega(n \log n)$ halving edges are known [ELSS, EW, EVW]. If we consider the corresponding problem for oriented matroids (cf. [BLSWZ]) of rank 3 (or pseudoline arrangements in the dual), then there is an unpublished lower bound of $n2^{\Omega(\sqrt{\log n})}$ due to M. Klawe, M. Paterson, and N. Pippenger, inspired by a connection to sorting networks (cf. [AW]). It is open whether this construction is realizable (stretchable) or not.

j-Facets and the Upper Bound Theorem

The following notion generalizes halving edges and triangles. Let P be a set of $n > d$ points in $\mathbb{R}^d$ in general position, i.e., no $d+1$ points on a common hyperplane. A *j-facet of P* is an oriented $(d-1)$-simplex spanned by d points in P that has exactly j points from P on the positive side of its affine hull. The 0-facets correspond to the facets of the convex hull of P. Hence, the Upper Bound Theorem due to P. McMullen [McM] (cf. [Zi]) gives us a tight upper bound on the number of 0-facets, which is attained by the vertices of cyclic polytopes: $2\binom{n-\lfloor d/2 \rfloor -1}{\lfloor d/2 \rfloor}$ for d odd, and $2\binom{n-\lfloor d/2 \rfloor}{\lfloor d/2 \rfloor} - \binom{n-\lfloor d/2 \rfloor -1}{\lfloor d/2 \rfloor}$ for d even. Below we will use the fact that these expressions are upper bounded by $2\binom{n}{\lfloor d/2 \rfloor}$. For d fixed, K. L. Clarkson and P. W. Shor [CS] derive an asymptotically tight bound of $O(n^{\lfloor d/2 \rfloor}(j+1)^{\lceil d/2 \rceil})$ for the number of $(\leq j)$-facets (i.e., i-facets with $0 \leq i \leq j$) by an argument along the following lines.

We use g_j for the number of j-facets of P and G_j for the number of $(\leq j)$-facets, i.e., $G_j = \sum_{i=0}^{j} g_i$. Now fix some j, $0 \leq j \leq n-d$ and x, $0 < x \leq 1$. We take a random sample P_x of P by selecting each point in P with probability x, independently from the other points. Let $n_x = |P_x|$ and let F_x be the number of 0-facets of P_x.

On the one hand, the Upper Bound Theorem implies $F_x \leq 2\binom{n_x}{\lfloor d/2 \rfloor}$ and so

$$\mathrm{E}(F_x) \leq 2\binom{n}{\lfloor d/2 \rfloor} x^{\lfloor d/2 \rfloor} , \tag{1}$$

since $\mathrm{E}(\binom{X}{i}) = \binom{N}{i}x^i$ for a random variable X following the binomial distribution of N Bernoulli trials with success probability x. On the other hand, an i-facet of P appears as a 0-facet of P_x with probability $x^d(1-x)^i$ – we have to select the d points that determine the i-facet, but none of the i points on its positive side.

Hence,

$$\mathrm{E}(F_x) = \sum_{i=0}^{n-d} x^d(1-x)^i g_i \geq x^d(1-x)^j \sum_{i=0}^{j} g_i = x^d(1-x)^j G_j \ . \qquad (2)$$

Combining (1) and (2), we have $G_j \leq 2(1-x)^{-j}\binom{n}{\lfloor d/2 \rfloor}x^{-\lceil d/2 \rceil}$. By setting $x = \lceil d/2 \rceil/(j+\lceil d/2 \rceil)$,

$$G_j \leq 2\binom{n}{\lfloor d/2 \rfloor}\frac{(j+\lceil d/2\rceil)^{j+\lceil d/2\rceil}}{j^j\lceil d/2\rceil^{\lceil d/2\rceil}} \leq 2\left(\frac{e}{\lceil d/2\rceil}\right)^{\lceil d/2\rceil}\binom{n}{\lfloor d/2\rfloor}(j+\lceil d/2\rceil)^{\lceil d/2\rceil}$$

and the claimed asymptotic bound follows.

Except for dimensions 2 and 3, no exact upper bounds for the number of $(\leq j)$-facets are known. In particular, it is not known whether the *exact* maximum is attained for sets in convex position or not. It is still possible that the exact maximum can be obtained for points on the moment curve, where the number of $(\leq j)$-facets can be easily counted.

We summarize the known bounds for the number of j-facets.

Proposition 1 *Let P be a set of $n > d$ points in $\mathbb{R}^d$ in general position, i.e., no $d+1$ points on a common hyperplane. Let $0 \leq j \leq n-d$.*
(0) *There is a constant $\varepsilon_d > 0$ dependent on d only, such that*

$$g_j = O(n^{\lfloor d/2 \rfloor}(j+1)^{\lceil d/2 \rceil - \varepsilon_d})$$

[AACS]. *There are point sets with $g_{\lfloor (n-d)/2 \rfloor} = \Omega(n^{d-1}\log n)$* [Ed].

$$G_j = O(n^{\lfloor d/2 \rfloor}(j+1)^{\lceil d/2 \rceil})$$

which, for d fixed, is asymptotically tight for points on the moment curve [CS].
(1) *If $d = 2$ then*

$$g_j = O(n\sqrt[3]{j+1})$$

[De]. *$G_j \leq n(j+1)$ for $j < n/2 - 1$* [AG, Pe], *which is tight for points in convex position.*
(2) *If $d = 3$ then*

$$g_j = O(n(j+1)^{5/3})$$

[AACS].

$$G_j \leq (j+1)(j+2)n - 2(j+1)(j+2)(j+3)/3$$

for $j \leq n/4 - 2$, which is tight if P is in convex position [AAHSW].

And k-sets?

We have met halving edges and triangles, k-levels and j-facets, but if the reader inspects the references, she will repeatedly encounter the term 'k-set.' In fact, many people think of the problem in the following setting (although proofs and applications go via the notions we have discussed above):

Let P be a set of n points in $\mathbb{R}^d$. A subset S of P is called *k-set*, if $|S| = k$ and S can be separated from $P \setminus S$ by a hyperplane. The maximum possible number of k-sets of n-point sets in $\mathbb{R}^d$ is related to the maximum possible number of k-facets, although the connection is somewhat subtle [AAHSW, AW].

REFERENCES

[AACS] Pankaj K. Agarwal, Boris Aronov, Timothy M. Chan, and Micha Sharir, On levels in arrangements of lines, segments, planes, and triangles, *Discrete Comput Geom* (1998), to appear

[ACNS] Miklós Ajtai, Vasek Chvátal, Monty M. Newborn, and Endre Szemerédi, Crossing-free subgraphs, *Ann Discrete Math* 12 (1982) 9–12

[AG] Noga Alon and Ervin Győri, The number of small semispaces of a finite set of points in the plane, *J Combin Theory Ser A* 41 (1986) 154–157

[AAHSW] Artur Andrzejak, Boris Aronov, Sariel Har-Peled, Raimund Seidel, and Emo Welzl, Results on k-sets and j-facets via continuous motion, *in* "Proc 14th Ann ACM Symp on Comput Geom" (1998) 192-199

[AW] Artur Andrzejak and Emo Welzl, k-Sets and j-facets – A tour of discrete geometry, in preparation (1998)

[BLSWZ] Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter Ziegler, *Oriented Matroids*, Cambridge University Press, Cambridge UK (1993)

[CS] Kenneth L. Clarkson and Peter W. Shor, Applications of random sampling in computational geometry, II, *Discrete Comput Geom* 4 (1989) 387–421

[De] Tamal K. Dey, Improved bounds for planar k-sets and related problems, *Discrete Comput Geom* 19 (1998) 373-382

[DE] Tamal K. Dey and Herbert Edelsbrunner, Counting triangle crossings and halving planes, *Discrete Comput Geom* 12 (1994) 281–289

[Ed] Herbert Edelsbrunner, *Algorithms in Combinatorial Geometry*, Springer Verlag, Berlin, Heidelberg, New York (1987)

[EVW] Herbert Edelsbrunner, Pavel Valtr, and Emo Welzl, Cutting dense point sets in half, *Discrete Comput Geom* 17 (1997) 243–255

[EW] Herbert Edelsbrunner and Emo Welzl, On the number of line separations of a finite set in the plane, *J Combin Theory Ser A* 38 (1985) 15–29

[ELSS] Paul Erdős, László Lovász, A. Simmons, and Ernst G. Straus, Dissection graphs of planar point sets, *in* A Survey of Combinatorial Theory (Eds. Jagdish N. Srivastava et al.), North Holland Publishing Company (1973) 139–149

[Ep] David Eppstein, Geometric lower bounds for parametric matroid optimization, *in* "Proc 27th Symp Theory Comput" (1995) 662–671

[Gu] Dan Gusfield, Bounds for the parametric spanning tree problem, *in* "Proc Humboldt Conf on Graph Theory, Combinatorics, and Computing", Utilitas Mathematica (1979) 173–183

[KI] Naoki Katoh and Toshihide Ibaraki, On the total number of pivots required for certain parametric problems, Tech. Report Working Paper 71, Inst. Econ. Res., Kobe Univ. Commerce (1983)

[KM] Nikolai M. Korneenko and Horst Martini, Hyperplane approximations and related topics, *in* "New Trends in Discrete and Computational Geometry" (János Pach, Ed.), *Algorithms and Combinatorics* 10 (1993) 135–161

[Le] Tom Leighton, *Complexity Issues in VLSI*, Foundation of Computing Series, MIT Press, Cambridge (1983)

[Lo] László Lovász, On the number of halving lines, *Ann Universitatis Scientarium Budapest, Eötvös, Sectio Mathematica* 14 (1971) 107–108

[McM] Peter McMullen, The maximum number of faces of a convex polytope, *Mathematika* 17 (1971) 179–184

[PSS] János Pach, William Steiger, and Endre Szemerédi, An upper bound on the number of planar k-sets, *Discrete Comput Geom* 7 (1992) 109–123

[PT] János Pach and Géza Toth, Graphs drawn with few crossings per edge, *Lecture Notes in Comput Sci* 1190 (1997) 345–354

[Pe] G. W. Peck, On k-sets in the plane, *Discrete Math* 56 (1985) 73–74

[Sh] Micha Sharir, k-Sets and random hulls, *Combinatorica* 13 (1993) 483–495

[Tv] Helge Tverberg, A generalization of Radon's Theorem, *J London Math Soc* 41 (1966) 123–128

[YKII] Peter Yamamoto, Kenji Kato, Keiko Imai, and Hiroshi Imai, Algorithms for vertical and orthogonal L_1 linear approximation of points, *in* "Proc 4th Ann ACM Symp on Comput Geom" (1988) 352–361

[Zi] Günter Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, Spinger-Verlag (1995)

[ZV] Rade T. Živaljević and Sinisa T. Vrećica, The colored Tverberg's problem and complexes of injective functions, *J Combin Theory Ser A* 61 (1992) 309–318

Artur Andrzejak
Institute of
Theoretical Computer Science
ETH Zürich
CH-8092 Zürich
Switzerland
artur@inf.ethz.ch

Emo Welzl
Institute of
Theoretical Computer Science
ETH Zürich
CH-8092 Zürich
Switzerland
emo@inf.ethz.ch