Generalized Ham-Sandwich Cuts for Well Separated Point Sets

William Steiger * Jihui Zhao‡

Abstract

Bárány, Hubard, and Jerónimo recently showed that for given well separated convex bodies $S_1, \dots, S_d$ in $\mathbb{R}^d$ and constants $\beta_i \in [0, 1]$, there exists a unique hyperplane h with the property that $\mathrm{Vol}(h^+ \cap S_i) = \beta_i \cdot \mathrm{Vol}(S_i)$; h^+ is the closed positive transversal halfspace of h, and h is called a "generalized ham-sandwich cut". We give a discrete analogue for n points in general position in $\mathbb{R}^d$ which are partitioned into a family $S = P_1 \cup \cdots \cup P_d$ of well separated sets. The combinatorial proof inspires an $O(n(\log n)^{d-3})$ algorithm which, given positive integers $a_i \leq |P_i|$ and $|P_1| + \cdots + |P_d| = n$, finds the unique hyperplane h incident with a point in each P_i and having $|h^+ \cap P_i| = a_i$.

1 Introduction.

Given d sets $S_1, S_2, ..., S_d \in \mathbb{R}^d$, a ham-sandwich cut is a hyperplane h that simultaneously bisects each S_i. "Bisect" means that $\mu(S_i \cap h^+) = \mu(S_i \cap h^-) < \infty$, h^+, h^- the closed halfspaces defined by h and μ a suitable, "nice" measure on Borel sets in $\mathbb{R}^d$, e.g., the volume. The well known ham-sandwich theorem guarantees the existence of such a cut. As with other consequences of the Borsuk-Ulam theorem [9] there is a discrete version that applies to sets $P_1, \dots, P_d$ of points in general position in $\mathbb{R}^d$. For example Lo et. al [8] gave a direct proof of a discrete version of the ham-sandwich theorem which inspired an efficient algorithm to compute a cut. More recently Bereg [4] studied a discrete version of a result of Bárány and Matoušek [2] showing the existence of wedges that simultaneously equipartition three measures on $\mathbb{R}^2$ (they are called equitable two-fans). By seeking a direct, combinatorial proof of a discrete version (for counting measure on points sets in $\mathbb{R}^2$) he was able to strengthen the original result and to obtain a beautiful, nearly optimal algorithm to construct an equitable two-fan. Finally Roy and Steiger [11] followed the same path to obtain complexity results for several other consequences of Borsuk-Ulam.

The present paper is in the same spirit. The starting point is a recent, interesting result about generalized ham-sandwich cuts.

*Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854-8004; ({steiger,zhaojih}@cs.rutgers.edu.)

Definition: *Given $k \leq d+1$, sets $S_1, \dots, S_k$ in $\mathbb{R}^d$ are well-separated if, for every choice of $x_i \in S_i$, the affine hull of $x_1, \dots, x_k$ is a $(k-1)$-dimensional flat in $\mathbb{R}^d$.*

Bárány et.al. [1] proved

Proposition 1 *Let $K_1, ..., K_d$ be well separated convex bodies in $\mathbb{R}^d$ and $\beta_1, \dots, \beta_d$ given constants with $0 \leq \beta_i \leq 1$. Then there is a unique hyperplane $h \subset \mathbb{R}^d$ with the property that $Vol(K_i \cap h^+) = \beta_i \cdot Vol(K_i)$, $i = 1, \dots, d$.*

Here h^+ denotes the closed, positive transversal halfspace defined by h: that is the halfspace where, if Q is an interior point of h^+ and $z_i \in K_i \cap h$, the $d-$simplex $\Delta z_1, ..., z_d, Q$ is *negatively* oriented [1]. Specifying this choice of halfspaces is what forces h to be uniquely determined. Bárány et. al. give analogous results for such *generalized ham-sandwich cuts* for other kinds of well separated sets that support suitable measures.

Here we are interested in a version of Theorem 1 for sets $P_1, ..., P_d$ of points in general position in $\mathbb{R}^d$. We will use

Definition: *A hyperplane h is an a_i-cut for the set P_i if*

- $h \cap P_i \neq \emptyset$ and
- $|h^+ \cap P_i| = a_i$.

Definition: *Given positive integers $a_i \leq |P_i|$, an $(a_1, ..., a_d)-$cut is a hyperplane h which, for each $1 \leq i \leq d$, is an a_i-cut for P_i.*

Such a cut is the discrete analogue of the generalized ham-sandwich cut in Theorem 1. In the next section we give a direct combinatorial proof of

Theorem 2 *Let $P_1, \dots, P_d$ be sets of points in general position in $\mathbb{R}^d$ which are well separated, and where $|P_1| + \cdots + |P_d| = n$. Given positive integers $a_i \leq |P_i|$, there is a unique hyperplane h that is an $(a_1, \dots, a_d)-$cut.*

This could be proved using the results of [1] via a standard argument that takes the average of n probability measures, one centered at each point. The variance of the measures is decreased to zero, and one argues about the limit (see [7]). However there is also interest in the algorithmic problem where, given n points distributed among d well separated sets in $\mathbb{R}^d$, the object is to find the cut for given $a_1, \ldots, a_d$. The combinatorial proof of Theorem 2 that we give leads to the formulation of an efficient, $O(n(\log n)^{d-3})$ algorithm for generalized cuts. The proof is in the next section. Some details are omitted because of the space limitations. We explain the algorithm in Section 3.

2 The Discrete Version.

We mention several equivalent forms of "well separated" that are known [3], or properties that hold for well separated sets.

1. Sets $S_1, \ldots, S_k, k \leq d+1$ are well separated if and only if, when I and J are disjoint subsets of $1, \ldots, d+1$, there is a hyperplane separating the sets $S_i, i \in I$ from the sets $S_j, j \in J$.
2. Sets S_i are well separated if and only if the sets $\text{Conv}(S_i)$ are well separated, $i \leq d+1$.
3. $S_1, \ldots, S_d$ are well separated in $\mathbb{R}^d$ if and only if there is no $(d-2)$-dimensional transversal flat that meets all of them.

We will assume that we have d well separated sets $P_1, \ldots, P_d$ of points in $\mathbb{R}^d$ with $|P_i| = n_i$ and $n_1 + \cdots + n_d = n$. The n points in $S = P_1 \cup \cdots \cup P_d$ are supposed to be in general position, so no $(k-1)-$dimensional flat in $\mathbb{R}^d$ contains more than k points of S, $k \leq d$.

Under these assumptions, given d points $p_1, \ldots, p_d$, $p_i \in P_i$, the hyperplane $h \equiv \text{aff}\{p_1, \ldots, p_d\}$ is a transversal hyperplane of dimension $d-1$. If a unit vector c satisfies $\langle c, p_i \rangle = t$ for some fixed constant t and for all i, the unit normal vector v of h can be chosen as either c or $-c$. The *positive transversal hyperplane* arises when v is chosen so that, [1]

$$\det \begin{vmatrix} p_1 & p_2 & \cdots & p_d & v \\ 1 & 1 & \cdots & 1 & 0 \end{vmatrix} > 0.$$

In this case we write h as $\{p \in \mathbb{R}^d : \langle p, v \rangle = t\}$, and h^+, the *positive transversal halfspace*, as

$$h^+ = \{p \in \mathbb{R}^d : \langle p, v \rangle \leq t\}.$$

It is not hard to show that if p is in h^+, no matter how we rotate or translate the coordinate system, p will always be in h^+. This property is important and simplifies the arguments in the detailed proofs.

Here we (only) sketch the main ideas for the combinatorial proof of Theorem 2. This proof gives geometric and combinatorial insights that lead to an efficient algorithm.

Proof of Theorem 2: The proof is by induction. For $d = 2$ the statement is folklore (see [10]). Points in P_1 may be dualized to lines having positive slopes and those in P_2, to lines having negative slope. Then each level in the first arrangement has a unique intersection with every level of the second. The unique intersection can be found in linear time (see Lo et. al. [8] or Megiddo [10]).

Next suppose that the claim holds for dimension $j < d$. We fix one of the sets, say P_1 and a point $x \in \text{Conv}(P_1)$.

Definition: *Given a point $x \in \text{Conv}(P_1)$, a hyperplane h_x a semi-cut if*

- *$x \in h_x$ and*
- *If $|(P_1 \cup \{x\}) \cap h_x^+| = m$, then h_x is an $(m, a_2, \ldots, a_d)$-cut for $(P_1 \cup \{x\}), P_2 \ldots, P_d$, where h_x^+ is the positive transversal halfspace defined by h_x.*

It is easy to establish

Lemma 3 *For every $x \in Conv(P_1)$ there is a unique semi-cut h_x.*

Proof: Because $P_1, \ldots, P_d$ are well separated, we can find a hyperplane π that separates P_1 from $P_2 \cup \cdots \cup P_d$. Choose a point $x \in \text{Conv}(P_1)$ and project each $z \in P_2 \cup \cdots \cup P_d$ onto π. Let P_i' denote the set of images in π of the points $z \in P_i$. It follows that

Fact: $P_2', \ldots, P_d'$ are $d-1$ well-separated sets in π.

By way of contradiction suppose not. Then there exists a $(d-3)$ dimensional flat $\rho \subset \pi$ that meets all these projections, and the span of x and ρ is a $(d-2)-$dimensional flat in $\mathbb{R}^d$ that meets all $\text{Conv}(P_i), i = 1, \ldots, d$. But this contradicts the fact that $P_1, \ldots, P_d$ are well-separated, and thus proves the Fact.

By the induction hypotheses, there exists a unique $(a_2, \ldots, a_d)-$cut ρ in π. Let h be the span of x and ρ. It is easy to show that for any given data point $z \in P_2 \cup \cdots \cup P_d$, we have $z \in h^+$ if and only if $z' \in \rho^+$, where z' is the projection of z on π. The statement in the lemma now follows since h is a semi-cut and because ρ is unique. □

Now for each point $p \in P_1$, there is a unique semi-cut h_p. And for different $p, q \in P_1$, the cuts h_p and h_q do not meet in $\text{Conv}(P_1)$, since otherwise, there would be a point $z \in \text{Conv}(P_1)$ that is in both cuts. Because z must have a unique semi-cut, both p and q would have to be in the same semi-cut, a fact that would contradict the general position assumption. Thus, for $p, q \in P_1$, $p \in h_q$ if and only if $q \notin h_p$.

Finally, consider the semi-cuts h_x for each $x \in P_1$. The previous property easily implies that the integers $a_x \equiv |h_x^+ \cap P_1|$, $x \in P_1$, form a permutation of $1, \ldots, n_1$. Thus one data point $x \in P_1$ has $|h_x^+ \cap P_1| = a_1$ and therefore, it determines an $(a_1, \dots, a_d)$-cut; in fact x is the unique point in P_1 with this property. This advances the induction and thus proves the theorem. □

3 An Algorithm for Generalized Cuts

The statement of Theorem 2 implies that there is a unique set of data points $p_1, \dots, p_d$, $p_i \in P_i$, for which $\text{aff}(p_1, \dots, p_d)$ is an $(a_1, \dots, a_d)-$cut. So we could use a brute force enumeration and find it in $O(n^{d+1})$, $O(n)$ being the cost to test each $d-$tuple..

A small improvement can be obtained by resorting to the following algorithmic result of [8] (slightly restated to reflect new upper bounds on k-sets [6], [12]).

Proposition 2. *Given n points in $\mathbb{R}^d$ which are partitioned into d sets $P_1, \dots, P_d$ in $\mathbb{R}^d$, a ham-sandwich cut can be computed in time proportional to the (worst-case) time needed to construct a given level in the arrangement of n given hyperplanes in $\mathbb{R}^{d-1}$. The latter problem (i) requires at least $\Omega(n^{d-2})$ time; (ii) is easy to solve in $O(n^{d-1})$ time; (iii) can be solved within the following bounds:*

$$\begin{array}{ll} O(n^{4/3}\log^2 n/\log^* n) & \text{for } d = 3, \\ O(n^{5/2}\log^{1+\delta}) & \text{for } d = 4, \\ O(n^{d-1-a(d)}) & \text{for } d \geq 5. \end{array}$$

$\delta > 0$ is an appropriate constant and $a(d) > 0$ a small constant; also $a(d) \to 0$ and $d \to \infty$.

It is not difficult to verify that the ham-sandwich algorithms given in [8] may be extended to find generalized cuts and in this way, the complexity of generalized cuts may be reduced to $O(n^{d-1-a(d)})$.

Finally, we describe a much more practical algorithm, applying ideas from the proof in Section 2. We showed there that for each data point $x \in P_1$, there is a unique semi-cut that contains x. Furthermore, for each j, $1 \leq j \leq n_1$, there is an $x \in P_1$ whose semi-cut has exactly j points of P_1 in its positive transversal halfspace. We will search for the data-point $z \in P_1$ whose semi-cut has a_1 points of P_1 in its positive transversal halfspace. To do this we effect a binary-like search by constructing an approximate center point μ of P_1 and finding its semi-cut h. We then compute $|h^+ \cap P_1|$ and can eliminate all points in one of the open halfspaces defined by h, *a fixed fraction of the* n_1 points in P_1. Iterating in this fashion on the remaining candidates in P_1 we will find the $z \in P_1$ after $O(\log n)$ such search steps. The cost for each step of the search is dominated by the cost to find a generalized cut in dimension $d - 1$.

ALGORITHM GEN-CUT

- **choose $c > 0$, a small, fixed integer (say 10)**
- **Find a hyperplane π that separates P_1 from $P_2 \cup \cdots \cup P_d$**
- $C \leftarrow P_1$
- $a \leftarrow a_1$
- **WHILE $|C| > c$ DO**
  1. **Find a point μ in $\text{Conv}(C)$ with Tukey depth at least $\frac{|C|}{2d}$**
  2. **Project each $z \in P_2 \cup \cdots \cup P_d$ onto π; let P_i' denote the projections of the points in P_i**
  3. **Find the $(a_2, ..., a_d)$-cut ρ in π for the projections P_2', P_3', ..., P_d' by solving a $(d-1)$ dimensional problem**
  4. **Get h, the hyperplane that spans μ and ρ.**
  5. **Compute the number of points of C in the positive transversal hyperplane h^+, and compare it with a**
  6. **Remove the points above or below h from C and adjust C and a**
- **END WHILE**
- **For each remaining data point in C, check if it is the point $z \in P_1$ whose semi-cut has a_1 points of P_1 in the positive transversal halfspace, stopping when z is found.**

Finding a separating hyperplane π can be formulated as a linear programming problem and can be solved in time $O(n)$, for fixed dimension d. C is the set of candidates for the sought point $z \in P_1$; initially $C = P_1$. a denotes the number of undeleted points in the positive transversal halfspace of z's semicut; initially $a = a_1$.

In Step 1, we need to find a point μ with large Tukey depth. We do it by constructing an approximate centerpoint: The system $(C, \mathcal{A})$, where $\mathcal{A}$ is the set of all halfspaces in $\mathbb{R}^d$ that contains some points in C, has VC dimension $d + 1$. So we can construct A, an ϵ-approximation of size $O(\frac{d+1}{\epsilon^2}\log\frac{d+1}{\epsilon})$ in time $O((d+1)^{3(d+1)}(\frac{d+1}{\epsilon^2}\log\frac{d+1}{\epsilon})^{d+1}|C|)$ time (see e.g., [5]. If $\mu \in A$ is the Tukey median of the points in A, then its depth (in C) is at least $(\frac{1}{d+1} - \epsilon)|C|$. If we take $\epsilon = \frac{1}{d+1} - \frac{1}{2d}$, this guarantees that μ has depth at least $|C|/(2d)$, and any positive transversal halfspace containing μ will have at least this many points of C. The cost to compute A and μ is $O(|C|)$.

In Step 5, suppose $|h^+ \cap C| = t < a$. Then the point $z \in P_1$ we seek is not in h^+, so we remove from C all points in $h^+ \cap C$, and adjust $a \leftarrow a - t$. Otherwise, we can remove from C all points that are not in h^+. In both cases at least a fixed fraction $1/(2d)$ is deleted.

The geometric decrease in $|C|$ implies that the number of iterations of the loop is bounded by $O(\log n_1) = O(\log n)$. Therefore Step 3 contributes $O(B_{d-1} \log n)$ to the total cost of the loop, where B_k denotes the complexity of the present algorithm in dimension k. This dominates the total cost of the loop because all other steps have cost either $O(n)$ or $(O|C|)$ and contribute a total of $O(n \log n)$ to the loop.

When the loop terminates, each remaining point in C is treated in $O(B_{d-1})$ by executing Step 2 through Step 4; then, instead of Step 5, we test whether $|h^+ \cap P_1| = a_1$. Since the base case for dimension $d = 2$ has linear running time, the present algorithm will find a generalized cut in $O(n(\log n)^{d-2})$.

Finally, for $d = 3$, Lo, et. al. [8] showed how to find a ham-sandwich cut for well separated point sets in linear time. That algorithm is easily adapted to generalized cuts. Using this as the base case, the algorithm just described now has running time $O(n(\log n)^{d-3})$.

References

[1] Bárány, I., Hubard, A., Jerónimo, J. “ Slicing Convex Sets and Measures by a Hyperplane”. *Discrete and Computational Geometry 39*, 67-75 (2008).

[2] I. Bárány and J. Matoušek. “Separating Measures by Two-Fans”. *Discrete and Computational Geometry*

[3] Bárány, I., Valtr, P.: A positive fraction Erdős-Szekeres theorem. Discrete and Computational Geometry 19, 335-342 (1998)

[4] S. Bereg. “Equipartitions of Measures by 2-Fans”. *Discrete and Computational Geometry 34 (1)*, 87-96 (2005).

[5] Chazelle, B.: The Discrepancy Method: Randomness and Complexity, Cambridge University Press (2000).

[6] T. Dey. ”Improved Bounds for Planar k-Sets and Related Problems”. *Discrete and Computational Geometry 19 (3)* 373- 382 (1998).

[7] Edelsbrunner, H.: Algorithms in Combinatorial Geometry. Springer, New York (1987)

[8] C.-Y. Lo, J. Matoušek, and W. Steiger. “Algorithms for Ham-Sandwich Cuts”. *Discrete and Computational Geom. 11.* 433-452 (1994).

[9] J. Matoušek. *Using the Borsuk-Ulam Theorem.* Springer-Verlag (2003).

[10] N. Megiddo. “Partitioning with two lines in the plane”. *J. Algorithms 6*, 430-433 (1985)

[11] S. Roy and W. Steiger. “Some Combinatorial and Algorithmic Aspects of the Borsuk-Ulam Theorem”. *Graphs and Combinatorics 23*, 331-341, (2007).

[12] M. Sharir, S. Smorodinsky, and G. Tardos. “An Improved Bound for k-Sets in Three Dimensions”. *Discrete and Computational Geometry 26*, 195-204 (2001).