Lower Bounds for Linear Degeneracy Testing*

Nir Ailon
Department of Computer Science
Princeton University
nailon@cs.princeton.edu

Bernard Chazelle
Department of Computer Science
Princeton University
chazelle@cs.princeton.edu

ABSTRACT

In the late nineties Erickson proved a remarkable lower bound on the decision tree complexity of one of the central problems of computational geometry: given n numbers, do any r of them add up to 0? His lower bound of $\Omega(n^{\lceil r/2 \rceil})$, for any fixed r, is optimal if the polynomials at the nodes are linear and at most r-variate. We generalize his bound to s-variate polynomials for $s > r$. Erickson's bound decays quickly as r grows and never reaches above pseudo-polynomial: we provide an exponential improvement. Our arguments are based on three ideas: (i) a geometrization of Erickson's proof technique; (ii) the use of error-correcting codes; and (iii) a tensor product construction for permutation matrices.

Categories and Subject Descriptors

F.2.0 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity—*General*

General Terms

Theory

Keywords

Computational Geometry, Linear Decision Trees, Lower Bounds

1. INTRODUCTION

Decision trees have often shown to be realistic and effective models for proving lower bounds on the complexity of fundamental geometric problems [4, 5, 9–12, 15, 16, 21–23]. Testing degeneracy is one such example. The *r-variate degeneracy testing* problem is to decide whether, given a sequence of n reals $x_1, \ldots, x_n$ and a real polynomial f over r variables, there exist distinct indices $i_1, \ldots, i_r$ such that $f(x_{i_1}, \ldots, x_{i_r}) = 0$. Additional constraints might be imposed on the indices. The terminology owes to the problem formulation's suitability for checking the degeneracy of just about any geometric configuration. For example, general position of N points in $\mathbf{R}^d$ can be reduced to d-variate degeneracy testing with respect to dn numbers; in this case, f is a $(d+1)$-by-$(d+1)$ determinant with a row of ones, and constraints on the indices ensure that the entries of the determinant are, indeed, the coordinates of $d+1$ distinct points. Similarly, we can formulate the degeneracy of Voronoi diagrams, power diagrams, algebraic varieties, real semi-algebraic sets, etc. Classical "bichromatic" problems also fall in that category: for example, checking incidence between points and hyperplanes (Hopcroft's problem), rays and triangles, lines and spheres, etc. The list of problems studied in the literature that can be reduced to degeneracy testing is nearly endless.

Even the unconstrained version of r-variate linear degeneracy testing (*r-LDT*) is ubiquitous in the computational geometry landscape. This is looking among n numbers for a zero of $f(y) = a_0 + a_1 y_1 + \cdots + a_r y_r$ ($a_i \neq 0$ for $i > 0$). There is a vast collection of geometric problems known to be 3SUM-hard and 4SUM-hard, all of which are at least as hard as r-LDT (for $r = 3, 4$) via subquadratic reductions [14]. Classical examples are separating line segments by a line, testing if a union of triangles is simply connected, checking for polygon containment under translation, minimizing the Hausdorff distance between segment sets, computing the Minkowski sum of two polygons, sorting the vertices of a line arrangment, etc. [1–3, 6, 7, 18]. Needless to say, the importance of elucidating the complexity of r-LDT can hardly be overstated.

While the problem, being a variant of SUBSET SUM, is clearly NP-complete, its parameterized complexity as a function of r is poorly understood (to put it charitably). The trivial $O(n^r)$ upper bound can be improved to $O(n^{\lceil r/2 \rceil})$ if r is odd and $O(n^{r/2} \log n)$ if r is even [10]. The idea is to write $f = g - h$, where g and h are respectiveley $\lfloor r/2 \rfloor$-variate and $\lceil r/2 \rceil$-variate. We sort all possible values of h and store them in a table. By binary search we look up every possible value of g. Any successful search corresponding to a match with distinct indices is a certificate of degeneracy. In a nonuniform decision tree model, the extra log factor is not needed. In other words, for any n, there exists a binary decision tree for r-LDT of depth $O(n^{\lceil r/2 \rceil})$. Each internal node is associated with a linear polynomial over r variables. What makes this result particularly interesting is the existence of a matching lower bound.

*This work was supported in part by NSF grants CCR-998817, CCR-0306283, ARO Grant DAAH04-96-1-0181, and NEC Research Institute.

STOC'04, June 13–15, 2004, Chicago, Illinois, USA.

The underlying model is the r-linear decision tree: each internal node v is assigned a linear n-variate polynomial q_v with at most r nonzero (real) coefficients; its outgoing edges are labeled $<$, $=$, or $>$. Leaves are labeled *yes* or *no*. To test the degeneracy of an input $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, we evaluate $q_v(x)$ beginning at the root and follow outgoing edges in the obvious way until we reach a leaf, at which point we output its label: *yes* if the input is degenerate and *no* otherwise. Improving on previous work [8,13], Erickson [10] proved that *any* r-linear decision tree for *any* r-LDT problem has depth $\Omega(n^{\lceil r/2 \rceil})$. His proof is quite a tour de force. It is packed with ingenious, tightly coupled arguments, and its only downside is to offer little wiggle room to try out new ideas. In particular, extending the proof to s-linear trees for $s > r$ has long been elusive. Even the case $s = r + 1$, mentioned in Yao's list of major open problems in his 2000 DIMACS lecture [24], has resisted all efforts. The contribution of this paper, while far from closing the book on the problem, represents a significant advance on two fronts: (i) accommodating $s > r$ variables and (ii) allowing for large values of r.

- We prove a lower bound of $\Omega(nr^{-3})^{\lceil r/2 \rceil}$ on the depth of any r-linear decision tree for any r-LDT problem. This improves on Erickson's bound of $\Omega(nr^{-r})^{\lceil r/2 \rceil}$ from pseudopolynomial to exponential for large values of r. Indeed, if $r = r(n) > n^{\varepsilon}$, Erickson's bound can never exceed $n^{\log n / \log \log n}$, while ours is of the form $2^{n^{\Omega(1)}}$. The technical underpinning of this improvement is a new adversarial strategy based on error-correcting codes.

- By using a tensor product construction based on permutation matrices, we are able to generalize the lower bound to the s-linear decision tree model for $s > r$. We show that, for any instance of r-LDT, the tree depth is at least

$$\Omega(nr^{-3})^{\frac{2r-s}{2\lceil (s-r+1)/2 \rceil}(1-\varepsilon_r)},$$

  where $\varepsilon_r > 0$ tends to 0 as $r \to \infty$. Our proof is based on a tensor product for permutation matrices.

The exponential lower bound still holds for $s > r$. For any fixed $\varepsilon > 0$, the depth of an s-linear decision tree is $(nr^{-3})^{r^{\Omega(1)}}$, if $s \leq r + r^{1-\varepsilon}$. In the case $r > n^{\varepsilon}$, this gives a lower bound of $2^{n^{\Omega(1)}}$. Note that our bounds collapse if s is not $O(r)$. This is an obvious limitation of our method, but one must note that a dependency on s is inevitable. Indeed, our lower bound of $n^{\Omega(r)}$ for $s = r + O(1)$ *cannot* hold for arbitrary values of s. By a result of Meyer auf der Heide [19], a decision tree of depth $O(n^4 \log n)$ exists for *any* instance of linear degeneracy testing over r variables, without a restriction on the number of nonzero coefficient of the tree polynomials ($s = n$).

Another contribution of this paper is methodological. To obtain our bounds requires a whole set of new algebraic arguments, but our starting point is essentially a geometrization of Erickson's method. The main benefit is to bypass the complicated machinery of infinitesimals found in [10], obviate the need for Tarski's transfer principle, and more generally do away with analytical arguments.

To make the proof more digestible, we begin our discussion with the geometric framework and then treat the case $s = r$. Next we move on to the case $s = r + 1$, where we introduce the tensor product construction in its simplest form. Finally we cover the general case.

2. A GEOMETRIC FRAMEWORK FOR LOWER BOUNDS

We consider the r-SUM problem: Given a point $x = (x_1, \ldots x_n) \in \mathbf{R}^d$, are there indices $i_1 < \cdots < i_r$ such that $x_{i_1} + \cdots + x_{i_r} = 0$? We wish to prove that any r-linear decision tree used to answer this question is of depth $\Omega(n^{\lceil r/2 \rceil})$. Choosing f to be the symmetric linear function on r variables simplifies the lower bound proof, but our results can be easily extended to *any* r-variate linear function. Each node v is associated with a polynomial q_v whose zeroes define a hyperplane, called a *query*. The set of query hyperplanes is denoted by $\mathcal{Q}$. It is not hard to see that if the decision tree is to be valid, $\mathcal{Q}$ must include every one of the $\binom{n}{r}$ *canonical* hyperplanes $x_{i_1} + \cdots + x_{i_r} = 0$. Indeed, if such a hyperplane h^* is missing in $\mathcal{Q}$, then there exists a pair of points p_1 and p_2 such that p_1 lies on h^* (thus degenerate), p_2 is nondegenerate, and no hyperplane in the finite set $\mathcal{Q}$ separates between p_1 and p_2. Therefore, the decision tree cannot decide r-LDT.

The basic idea of the proof is to identify a "large" face C in the arrangement[1] $\mathcal{A}(\mathcal{Q})$ formed by $\mathcal{Q}$. The face C, called the *chamber*, may not necessarily be full-dimensional but

(C1) it must not be contained in any canonical hyperplane.

We also need a set $\mathcal{H}$ of *critical* hyperplanes. These are canonical hyperplanes tangent to C such that

(C2) each $h^* \in \mathcal{H}$ has a designated point p_h on the boundary of C;

(C3) no two points in the collection $\{p_h\}$ lie in the closure of the same face of the closure of C.

LEMMA 2.1. *Any r-linear decision tree for the r-SUM problem is of depth at least $|\mathcal{H}|$.*

PROOF. The tree must lead to a *no* (resp. *yes*) leaf for any input point $p_0 \in C$ (resp. p_h, where $h^* \in \mathcal{H}$). For this reason, the path followed on input p_h must include a query hyperplane q_h that intersects, but does not contain, the segment $p_0 p_h$. Indeed, the same path would otherwise be followed for input p_0. Since p_h is in the closure of a face of $\mathcal{A}(\mathcal{Q})$ that contains p_0, the hyperplane q_h passes through p_h but does not intersect C. Now the crux is that by (C3) no query hyperplane can pass through more than one point p_h. □

2.1 Critical Hyperplanes via Error-Correcting Codes

By padding the input if necessary, we can always assume that $n = rm$, for some integer m. This allows us to view

[1] *Terminology and Conventions:* Faces of polyhedra and arrangements are disjoint, relatively open sets. The intersection of the closures of any two faces is either empty or the closure of another face. Faces of dimension (codimension) 0 and 1 are called vertices and edges (cells, facets), respectively. Convex polyhedra are assumed to be closed. The hyperplane defined by $h^T x = 0$, where $h, x \in \mathbf{R}^n$, is denoted by h^*.

a vector $h \in \mathbf{R}^n$ (and hence its hyperplane h^* through the origin) as a matrix M^h, whose rows are filled with the coordinates of h; ie, $M^h_{ij} = h_{(i-1)m+j}$. A critical hyperplane being of the form $x_{i_1}+\cdots+x_{i_r} = 0$, its corresponding matrix has r ones and $n-r$ zeroes. We place a single one per row. Where to put the ones is dictated by an error-correcting recipe meant to ensure high "independence." Throughout this section we use the shorthand

$$r_0 = \lceil r/2 \rceil. \tag{1}$$

Let q be the smallest prime greater than r, and let $\mathcal{M}$ be a Reed-Solomon code [17] of length $p = q-1$ and distance $r - r_0 + 1$ over the finite field $\mathbf{F}_q$. This means that any nonzero vector in $\mathcal{M}$ has at least $r - r_0 + 1$ nonzero coordinates. A simple choice is the ideal of $\mathbf{F}_q[X]/(X^p - 1)$ generated by $g(x) = (x-\beta)\cdots(x-\beta^{r-r_0})$, where β is a primitive element of $\mathbf{F}_q$. The code has dimension $k = p-r+r_0$. Now, define $\mathcal{M}_r$ to be the linear subspace of $\mathcal{M}$ defined by adding the contraints $\{\, x_i = 0 \,|\, r < i \le p \,\}$. (Note: this is *not* the same as chopping off the last $p-r$ coordinates.) In this way, we can think of $\mathcal{M}_r$ as a linear code of length r, distance greater than $r - r_0$ and dimension at least $k - (p-r) = r_0$. Let $v_1, \ldots, v_{r_0}$ be an independent set of vectors in $\mathcal{M}_r$. Of course, by permuting coordinates and performing column operations, we can always assume that the set is in column echelon form, ie, the r-by-r_0 matrix $(v_1, \ldots, v_{r_0})$ consists of the r_0-by-r_0 identity matrix on top of some $(r-r_0)$-by-r_0 matrix. Since $\mathbf{F}_q$ is a prime field, we can naturally view the v_i's as vectors in $\mathbf{R}^r$ with coordinates in $\{0, \ldots, q-1\}$. We define $\mathcal{L}$ as the set of vectors $n_1 v_1 + \cdots + n_{r_0} v_{r_0}$ for all nonnegative integers $n_i \le m/qr_0$. The upper bound is chosen so that all coordinates lie in $\{0, \ldots, m-1\}$. (Throughout this paper, the notation *span* $(\mathcal{S})$ refers to the vector space spanned by $\mathcal{S}$ *over the reals*.)

LEMMA 2.2. *Three facts: (i) the set $\mathcal{L}$ consists of at least $(n/r^3)^{r_0}$ vectors in $\mathbf{R}^r$ with coordinates in $\{0, \ldots, m-1\}$; (ii) the first r_0 coordinates of any vector in $\mathcal{L}$ specify it uniquely; (iii) any nonzero vector in* span $(\mathcal{L})$ *has at least $r - r_0 + 1$ nonzero coordinates.*

PROOF. By Nagura's theorem [20], the interval $[x, 6x/5]$ contains a prime for any $x \ge 25$. This shows that $qr_0 \le r^2$; therefore, $|\mathcal{L}| \ge (m/qr_0)^{r_0} \ge (n/r^3)^{r_0}$. Part (ii) comes from the echelon form of the matrix formed by $(v_1, \ldots, v_{r_0})$. To prove (iii), consider a nonzero element $\sum_{i=1}^{r_0} \alpha_i v_i$ of *span* $(\mathcal{L})$. The set of such vectors with at least r_0 zero coordinates can be expressed as a union of linear subspaces, each one defined by a set of homogeneous equations in the α_i's with integer coefficients. Therefore, if the set is nonempty, it must contain a vector v with all its α_i's integral and at least one of them not divisible by q. Reducing v modulo q gives us a nontrivial linear combination of the v_i's. Since these vectors are independent over $\mathbf{F}_q$, it then follows that v is a nonzero vector of the code $\mathcal{M}_r$ with at least r_0 zero coordinates. This contradicts the fact that $\mathcal{M}_r$ has distance greater than $r - r_0$. □

The set $\mathcal{H}$ of critical hyperplanes is in bijection with $\mathcal{L}$. The hyperplane h^* corresponding to $\ell = (\ell_1, \ldots, \ell_r) \in \mathcal{L}$ is defined by its matrix M^h: the coordinate ℓ_i indicates where to place the 1 in the i-th row of the matrix, ie, $M^h_{ij} = 1$ (resp. 0) if $j = \ell_i + 1$ (resp. else). By construction,

$$M^h\,(0, \ldots, m-1)^T \in \mathit{span}\,(\mathcal{L}). \tag{2}$$

The intersection $\cap\mathcal{H}$ of all the hyperplanes h^* in $\mathcal{H}$ is a linear subspace of positive dimension. Indeed, it contains the vector

$$(\underbrace{1, \ldots, 1}_{n-n/r}, \underbrace{1-r, \ldots, 1-r}_{n/r}).$$

Let $\mathcal{K}$ denote the set of query hyperplanes that contain $\cap\mathcal{H}$. Note that $\mathcal{Q} \supseteq \mathcal{K} \supseteq \mathcal{H}$.

LEMMA 2.3. *Given any $q^* \in \mathcal{K}$, (i) $M^q\,(1, \ldots, 1)^T = b\,(1, \ldots, 1)^T$ for some real b, and (ii) $M^q\,(0, \ldots, m-1)^T \in$ span $(\mathcal{L})$.*

PROOF. Since $(\cap\mathcal{H})^\perp$ is the space spanned by the normals of hyperplanes in $\mathcal{H}$ and $q \in (\cap\mathcal{H})^\perp$, $q = \sum_i \lambda_i h_i$, where $h_i^* \in \mathcal{H}$; therefore $M^q = \sum_i \lambda_i M^{h_i}$. But each M^{h_i} has a single one per row, and so $M^{h_i}\,(1, \ldots, 1)^T = (1, \ldots, 1)^T$; hence (i). Similarly, (ii) follows from (2). □

2.2 The Chamber

The query hyperplanes outside of $\mathcal{K}$ intersect $\cap\mathcal{H}$ in lower-dimensional subspaces. Therefore, there exist $c_0 \in \cap\mathcal{H}$ and $\rho > 0$ such that the ball $B(c_0, \rho)$ centered at c_0 of radius ρ intersects none of the hyperplanes of $\mathcal{Q} \setminus \mathcal{K}$. By lying on every critical hyperplane the point c_0 is highly degenerate. Moving it by some vector ψ to be specified next changes all of that (Fig. 1). We define the point

$$p_0 = c_0 + \psi \tag{3}$$

to be safely outside of the critical hyperplanes. To do that, we need a positive convex real function g, meaning one with positive second derivative; eg, $x \mapsto x^2 + 1$. For some fixed, small enough $\gamma > 0$, we define the vector $\psi \in \mathbf{R}^n$ by its matrix M^ψ:

$$M^\psi_{ij} = \begin{cases} \gamma g(j) & \text{if } i \le r_0; \\ \gamma^2 g(j) & \text{else.} \end{cases}$$

Note: γ is a scaling factor that is absolutely needed. The reason we use γ^2, however, is in anticipation of the case $s > r$. We could use γ in this section instead.

LEMMA 2.4. *The point p_0 lies outside of any canonical hyperplane and any hyperplane of $\mathcal{Q} \setminus \mathcal{K}$.*

This implies that the decision tree must output *no* on input p_0. Note, however, that p_0 might still lie on a query hyperplane.

PROOF. Recall that a canonical hyperplane h^* is one with an equation of the form $x_{i_1} + \cdots + x_{i_r} = 0$. By choosing γ small enough, we can ensure that $\|\psi\|_2 \le \rho/2$; therefore, the point p_0 lies inside $B(c_0, \rho)$, safely away from any hyperplane of $\mathcal{Q} \setminus \mathcal{K}$ (Fig. 1). We have already observed that $\mathcal{Q}$ must contain all of the canonical hyperplanes; therefore, the only danger is that p_0 lies on some canonical hyperplane h^* in $\mathcal{K}$. But this is impossible. Indeed, $c_0 \in \cap\mathcal{H} \subseteq h^*$, and so

$$h^T p_0 = h^T \psi = \sum_{j \in J} \gamma g(j) + \sum_{j \in J'} \gamma^2 g(j) > 0,$$

with $|J \cup J'| = r$. □

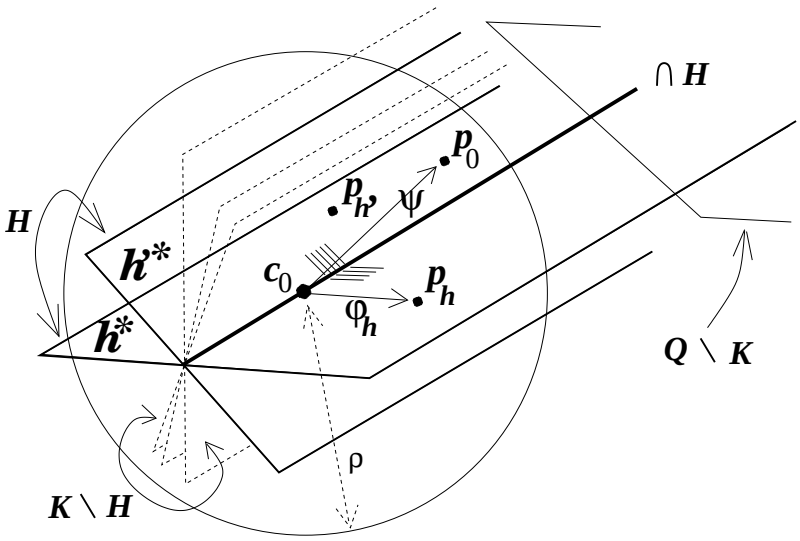

Figure 1: ***Output no if the input is*** p_0 ***but yes if the input is*** p_h ***or*** $p_{h'}$***. Both of these points lie on critical hyperplanes as well as in the flat*** $p_0 + W$ ***.***

The chamber C is the unique face of $\mathcal{A}(\mathcal{Q})$ that contains p_0. To define the map $h^* \in \mathcal{H} \mapsto p_h \in \partial C$, we need to introduce the vector space W spanned by the $2r$ vectors $u_k, w_k \in \mathbf{R}^n$ $(k = 1, \ldots, r)$ whose associated matrices M^{u_k} and M^{w_k} are zero everywhere except at row k, where $M^{u_k}_{ij} = 1$ and $M^{w_k}_{ij} = j$. Given $h^* \in \mathcal{H}$, we define a vector

$$\varphi_h \in \psi + W \tag{4}$$

such that $M^{\varphi_h}_{ij} > 0$ (resp. $= 0$) if $M^h_{ij} = 0$ (resp. else). Note that $\psi + W$ is not necessarily a vector space. One should think of M^{φ_h} as a mask: Its rows mark with zeroes the positions where M^h is 1 and fill the rest with positive entries. To see that such a vector φ_h actually exists, consider the i-th row of the matrix M^{φ_h}. Let $\gamma_i = \gamma$ (resp. γ^2) if $i \le r_0$ (resp. else). It suffices to show that the row can satisfy constraints in t, u of the form $\gamma_i g(j) + t + uj = 0$ if j is equal to the one value j_0 where $M^h_{ij_0} = 1$, and $\gamma_i g(j) + t + uj > 0$ for any $j \neq j_0$. Feasibility is ensured by the fact that

$$\frac{g(j_0) - g(j)}{j_0 - j} < \frac{g(j') - g(j_0)}{j' - j_0}$$

for any $j < j_0 < j'$, which itself is a consequence of the mean-value theorem applied to the convex function g. It is immediate to check that,

$$M^{\varphi_h}_{ij} = \begin{cases} \Theta(\gamma) & \text{if } i \le r_0; \\ \Theta(\gamma^2) & \text{else,} \end{cases} \tag{5}$$

where the Θ notation hides quantities that depend on g and n. These bounds imply that, by scaling down γ if necessary, we can ensure that $\|\varphi_h\|_2 < \rho/2$. We now define

$$p_h = c_0 + \varphi_h. \tag{6}$$

LEMMA 2.5. *The point p_h lies inside the critical hyperplane h^* and outside any hyperplane of $\mathcal{Q} \setminus \mathcal{K}$.*

PROOF. The second part follows directly from the fact that $\|\varphi_h\|_2 < \rho/2$. Recall that φ_h has zero coordinates precisely at the positions where h does not; therefore,

$$h^T p_h = h^T(c_0 + \varphi_h) = h^T \varphi_h = 0.$$

□

Incidentally, note that the p_h's are not strewn all across the boundary of C: By (3–6), they all live in the low-dimensional flat $p_0 + W$.

LEMMA 2.6. *Given any $q^* \in \mathcal{Q}$ and $h^* \in \mathcal{H}$, if $q^T p_0 \neq q^T p_h$ and $q^T p_0 \, q^T p_h \le 0$, then $q^T p_h = 0$ and $q^T p_x \neq 0$ for any $x^* \in \mathcal{H}$ distinct from h^*.*

In other words, collapsing p_0 to any p_h changes at most a single query from non-zero to zero, and leaves the sign of every other query unchanged.

PROOF. Obviously we can assume that $q^* \in \mathcal{K}$, since otherwise q^* would miss $B(c_0, \rho)$ entirely and so $q^T p_0$ and $q^T p_h$ would be nonzero and have the same sign; hence, $q^T p_0 \, q^T p_h > 0$. We distinguish between two cases.

(A) Each row of M^q has at least one nonzero element: Then, since $s = r$, it has exactly one and, by Lemma 2.3 (i), all the nonzero elements are equal to the same number, which without loss of generality we may assume to be 1. Therefore,

$$\begin{aligned} q^T p_0 &= q^T(c_0 + \psi) = q^T \psi \\ &= \sum_{i=1}^{r_0} \gamma g(j_i) + \sum_{i=r_0+1}^{r} \gamma^2 g(j_i) > 0. \end{aligned} \tag{7}$$

By the lemma's assumption, it follows that $q^T p_h \le 0$. For any $x^* \in \mathcal{H}$, $q^T p_x = q^T(c_0 + \varphi_x) = q^T \varphi_x \ge 0$; therefore, $q^T p_h = 0$. Can any other $x^* \in \mathcal{H}$ also satisfy $q^T p_x = 0$? The answer is no. To see why, recall that M^{φ_x} acts as a mask for the 1s in M^x. If M^x does not match M^q in each one of the first r_0 rows then, by (5), $q^T p_x$, which is also $q^T \varphi_x = \operatorname{tr} M^q (M^{\varphi_x})^T$, is

of the form $C\gamma + O(\gamma^2)$, for some $C > 0$, and hence can never be 0 as long as we choose γ small enough. On the other hand, by Lemma 2.2 (ii), if M^x matches M^q in each one of the first r_0 rows, then x is unique, and therefore equal to h. This completes the proof for case (A). Note that we never used the fact that each row $i > r_0$ of M^q has exactly one nonzero element. This will allow us to use the same proof verbatim in the next section, even though the only assurance we will then have on the matrix M^q is each of its first r_0 rows has a single 1.

(B) Some row of M^q is null: By Lemma 2.3 (i),

$$M^q\,(1,\ldots,1)^T = 0,$$

and each row of M^q must then have at least two nonzero elements or none at all. It follows that the number of null rows is at least r_0, and so, by Lemmas 2.2 (iii) and 2.3 (ii), $M^q\,(0,\ldots,m-1)^T = 0$. (This is where error correction kicks in.) As a result, $q^T w = 0$ for any $w \in W$. But, by (3–6), $p_h - p_0 = \varphi_h - \psi \in W$; therefore, $q^T p_h = q^T p_0$, which contradicts the lemma's assumption. □

It is immediate to verify that the chamber C and the points p_h satisfy the requirements C1—C3. If the chamber C lay within a canonical hyperplane, then so would p_0, which would contradict Lemma 2.4; hence C1. Consider a query hyperplane $q^* \in \mathcal{Q}$ and a critical hyperplane $h^* \in \mathcal{H}$. By Lemma 2.6, there are only three possibilities: $p_0 p_h \subset q^*$, $p_0 p_h \cap q^* = \emptyset$, or $p_h \in q^*$ and $p_x \notin q^*$ for any distinct $x^* \in \mathcal{H}$. This proves that p_h lies in the closure of C. Combined with the fact that $p_h \in h^*$ (Lemma 2.5), this establishes condition C2. Finally, if two distinct points p_h and p_x lay in the same face of the closure of C, then some $q^* \in \mathcal{Q}$ would contradict the three possibilities above; hence C3. In view of Lemmas 2.1 and 2.2, we have proven

THEOREM 2.7. *The depth of any r-linear decision tree for r-SUM is $\Omega(nr^{-3})^{\lceil r/2 \rceil}$.*

3. THE CASE $s = r + 1$

What can go wrong with the previous proof if $s = r + 1$? The only place where the number s actually plays a role is in the proof of Lemma 2.6. Case (B) survives almost verbatim. The only problem is that the number of null rows is at least $r - \lfloor s/2 \rfloor$, which can be less than r_0. We fix this by redefining r_0 so that it satisfies

$$1 \le r_0 \le r - \lfloor s/2 \rfloor. \tag{8}$$

In the present case, the setting $r_0 = \lfloor r/2 \rfloor$ will do.

Case (A) is far more difficult to fix. All the rows of M^q have exactly one nonzero element, except for one of them, i_0, which has two nonzeroes (the case of one nonzero in every row having already been handled). Again we can assume that all the nonzero elements are 1, except in row i_0, where the elements are α and $1 - \alpha$, for some real $\alpha \notin \{0, 1\}$. Let $\gamma' = \gamma$ (resp. γ^2) if $i_0 \le r_0$ (resp. else); taking row i_0 into account, we can rewrite (7) as

$$\begin{aligned} q^T p_0 &= q^T \psi \\ &= \gamma' \alpha g(j_{i_0}) + \gamma'(1-\alpha) g(j'_{i_0}) \\ &\quad + \sum_{\substack{i=1 \\ i \ne i_0}}^{r_0} \gamma g(j_i) + \sum_{\substack{i=r_0+1 \\ i \ne i_0}}^{r} \gamma^2 g(j_i). \end{aligned}$$

If $i_0 > r_0$ then all is well. Indeed, by making γ small enough

$$q^T p_0 = \sum_{i=1}^{r_0} \gamma g(j_i) + O_q(\gamma^2) > 0. \tag{9}$$

The remainder of the proof involves only the first r_0 rows of M^q, which happen to be as in case (A), and so it can be repeated verbatim.

The case $i_0 \le r_0$ is a tougher nut to crack. In fact we have not found a way of tackling it directly. Consequently, our strategy is simply to modify $\mathcal{H}$ so that this case cannot happen. Recall that, for the purpose of Lemma 2.6, we can assume that $q^* \in \mathcal{K}$. As we observed in the proof of Lemma 2.3, this implies that $q \in \mathit{span}\,(\mathcal{H}^*)$, where $\mathcal{H}^* = \{\, h \,|\, h^* \in \mathcal{H} \,\}$. Thus, our goal is to redefine a large set $\mathcal{H}$ of critical hyperplanes so that, in addition to all the properties we expect of $\mathcal{H}$, the following should hold: If q is a vector of $\mathbf{R}^n$ such that (i) with the exception of one row $i_0 \le r_0$ each of the first r_0 rows of M^q consists of a single 1 with 0's everywhere else, and (ii) the exceptional row, i_0, is null everywhere except for two entries summing up to 1, then q *cannot* be in the span of $\mathcal{H}^*$.

Recall from the construction of $\mathcal{H}$ that the first r_0 rows of any M^h ($h^* \in \mathcal{H}$) completely determine the remaining ones. Furthermore, each one of the first r_0 rows can be chosen by placing a 1 arbitrarily between positions 1 and $m_0 = \lfloor m/qr_0 \rfloor$ and filling the rest of the row with 0's. So it suffices to concentrate on the first r_0 rows. Once we have the top r_0 rows, we use our Reed-Solomon code to fill in the bottom $r - r_0$ rows just as we did in the previous section.

An r_0-by-a matrix is called *defective* if, with the exception of one row (called *anomalous*), each one consists of a single 1 with 0's everywhere else; furthermore the exceptional row is null everywhere except at two places. We postpone the proof of the next result.

LEMMA 3.1. *There exists a set $\mathcal{P}$ r_0-by-m 0/1 matrices with exactly one 1 per row between positions 1 and m_0 such that no defective r_0-by-m matrix belongs to* span $\mathcal{P}$ *and, for n large enough and any fixed $\varepsilon > 0$,*

$$|\mathcal{P}| \ge (nr^{-3})^{\lfloor r/2 \rfloor (1 - 1/\ln \lfloor r/2 \rfloor)(1-\varepsilon)}.$$

In view of our previous discussion, this automatically implies a lower bound on the depth of $(r+1)$-linear trees. The theorem below does not indicate what happens for small values of r. A careful examination shows that we obtain nontrivial lower bounds for any $r \ge 6$.

THEOREM 3.2. *The depth of any $(r+1)$-linear decision tree for r-SUM is at least $(nr^{-3})^{\lfloor r/2 \rfloor (1-\varepsilon_r)}$, where $\varepsilon_r > 0$ tends to 0 as $r \to \infty$.*

3.1 The Tensor Product Construction

The problem fits into a general class of questions related to codes and combinatorial designs: How to build a large vector space that does not contain a family of forbidden vectors? In the case at hand, we start by building a "core" square matrix that satisfies the desired property and then show how to scale it up into an arbitrarily large rectangular matrix by using a suitable tensor product.

Let A (resp. B) be an r_0-by-a (resp. r_0-by-b) real matrix. Following standard tensor notation, we write the element $A_{i,j}$ as A^i_j instead. The tensor product $P = A \otimes B$ of A and B is defined by the formula $P^i_{j,k} = A^i_j B^i_k$. It is a mixed third-order tensor with two covariant indices and a single contravariant one. This product extends to sets naturally. If $\mathcal{A}$ (resp. $\mathcal{B}$) is a set of r_0-by-a (resp. r_0-by-b) real matrices, then

$$\mathcal{A} \otimes \mathcal{B} = \{\, A \otimes B \,|\, A \in \mathcal{A},\ B \in \mathcal{B} \,\}.$$

Tensor exponentiation for sets is defined by

$$\mathcal{A}^{\otimes k} = \underbrace{\mathcal{A} \otimes \cdots \otimes \mathcal{A}}_{k \text{ times}}.$$

The $|\mathcal{A}|^k$ elements of $\mathcal{A}^{\otimes k}$ belong to the vector space $\mathcal{V}_k$ of mixed $(k+1)$st-order tensors with k covariant indices and 1 contravariant one. By fixing an ordering (say, lexicographic) of the covariant indices, we can interpret the tensors of $\mathcal{V}_k$ as r_0-by-a^k matrices, and vice versa. For our "core," we choose permutation matrices. Let Π denote the set of r_0-by-r_0 0/1 matrices with exactly one 1 per row and column. The lemma below gives our tensor product its raison d'être.

LEMMA 3.3. *No defective r_0-by-r_0^k matrix can belong to the span of $\Pi^{\otimes k}$, for any $k \geq 1$.*

PROOF. In any matrix of *span*(Π) each row and each column sum up to the same number, which can be assumed to be 1. Therefore, the anomalous row of a defective r_0-by-r_0 matrix consists of two entries, $\alpha, \alpha' \neq 0$ summing up to 1. Suppose that the column with the α also includes a set of ℓ ones for $\ell > 0$ (note that these can only be ones). Since the column sum is 1, we have $\alpha + \ell = 1$. This implies that α must be an integer and, since it is nonzero, $\alpha' = 1 - \alpha = \ell \geq 2$. But then the column with α' sums up to more than 1, which gives a contradiction. This implies that neither of the columns with α, α' has any other nonzero element. But then the $r_0 - 2$ other columns sum up to $r_0 - 1$, which exceeds the required count by one. This proves the lemma for $k = 1$.

For $k > 1$, we define the tensor homomorphism $h_l : \mathcal{V}_k \mapsto \mathcal{V}_{k-1}$, where

$$h_l(P)^i_{j_1,\ldots,j_{l-1},j_{l+1},\ldots,j_k} = \sum_{j=1}^{r_0} P^i_{j_1,\ldots,j_{l-1},j,j_{l+1},\ldots,j_k}.$$

Let M be a defective r_0-by-r_0^k matrix. By definition, its anomalous row i_0 contains two nonzero elements: the two corresponding covariant k-tuple indices, being distinct, differ in at least one index l. Since $k > 1$, there exists at least one covariant index $l' \neq l$. We easily verify that $h_{l'}(M)$ is a defective r_0-by-r_0^{k-1} matrix and

$$h_{l'}(\Pi^{\otimes k}) = \Pi^{\otimes (k-1)}.$$

The proof follows by simple linear algebra and induction. □

To maximize its size, we choose the set $\mathcal{P} = \Pi^{\otimes k}$ for the largest k such that $r_0^k \leq m_0 = \lfloor m/qr_0 \rfloor$. Using Stirling's approximation, we find that

$$|\mathcal{P}| = |\Pi|^k = (r_0!)^k \geq (nr^{-3})^{\lfloor r/2 \rfloor (1 - 1/\ln \lfloor r/2 \rfloor)(1-\varepsilon)},$$

for any fixed $\varepsilon > 0$. Filling up each row with 0's to get the proper of length m concludes the proof of Lemma 3.1. □

4. THE CASE $s > r + 1$

We need a new idea to generalize the tensor product construction to higher values of s. We exploit the fact that the (hard part of the) lower bound involves only query hyperplanes whose normal vectors q are spanned by the normals h of the critical hyperplanes $h^* \in \cap\mathcal{H}$. We use this to add combinatorial structure to the matrices M^q by redesigning the set $\cap\mathcal{H}$. We need to redefine r_0 so that the first r_0 rows of M^h can be grouped in equal-sized blocks. Of course, r_0 still needs to satisfy (8). The choice of $r_0 = \lambda\rho_0$ will do, where

$$\lambda = \left\lfloor \frac{s-r}{2} \right\rfloor + 1 \quad \text{and} \quad \rho_0 = \left\lfloor \frac{r - \lfloor s/2 \rfloor}{\lambda} \right\rfloor.$$

Note that this requires that s be not too large, say $s < \lfloor 3r/2 \rfloor$. Divide up the first r_0 rows of M^h into λ blocks of consecutive rows of ρ_0 rows each. To build up a matrix M^h of $\mathcal{H}$ we proceed as follows:

- **Step 1** Use the tensor construction of the previous section (the case $s = r+1$) to produce the top ρ_0 rows of M^h. In carrying out the construction, of course, use permutation matrices of size ρ_0-by-ρ_0 instead of r_0-by-r_0. This gives us a set $\mathcal{P}$ of matrices with the same properties as those of Lemma 3.1, except for the size of $\mathcal{P}$ and the size of the matrices, now ρ_0-by-m.
- **Step 2** For each matrix of $\mathcal{P}$, make λ copies of it and stack them on top of one another to produce an r_0-by-m matrix.
- **Step 3** Complete the bottom $r - r_0$ rows via Reed-Solomon as before.

LEMMA 4.1. *For any $q^* \in \mathcal{K}$, the top r_0 rows of M^q form an r_0-by-m matrix made up of λ copies of the same ρ_0-by-m matrix.*

PROOF. A simple consequence of the fact that

$$q \in \mathit{span}\,\{\, h \,|\, h^* \in \mathcal{H} \,\}.$$

□

There is no need to revisit Lemma 2.6 in detail. Again, only case (A) is worth discussing: Each row of M^q has at least one nonzero element. By analogy with the case $s = r+1$, if all the rows with more than one nonzero have indices greater than r_0 then inequality (9) holds and we are done.

Suppose now that at least one row $i_0 \leq r_0$ contains two or more nonzeroes. By Lemma 4.1, the λ blocks that make up the top r_0 rows of M^q are identical. This shows that no block can have more than $\rho_0 + 1$ nonzeroes. Indeed, any one of

them did, then so would all of the others, and their combined contribution of nonzeroes would be at least $(\rho_0+2)\lambda$. Added to the (at least) $r - r_0$ nonzeroes of the bottom rows, this would give us a total of at least $(\rho_0+2)\lambda+r-r_0 > s$ nonzero coordinates in q, which is ruled out. So, the only possibility left is for each block to have exactly ρ_0 or $\rho_0 + 1$ nonzeroes: the first case was handled in the proof of Lemma 2.6, while the second one was shown to be impossible in the last section because of the tensor product contruction.

THEOREM 4.2. *The depth of any s-linear decision tree for r-SUM is at least*

$$(nr^{-3})^{\frac{2r-s}{2\lceil (s-r+1)/2 \rceil}(1-\varepsilon_r)},$$

where $\varepsilon_r > 0$ tends to 0 as $r \to \infty$.

Note that for any $s \leq r + r^{1-\varepsilon}$, where $\varepsilon > 0$ is arbitrarily small constant, the depth is $(nr^{-3})^{r^{\Omega(1)}}$.

Acknowledgments

We wish to thank Jeff Erickson for enlightening discussions about his lower bound proof.

5. REFERENCES

[1] Arkin, E.M., Chiang, Y.-J., Held, M., Mitchell, J.S.B., Sacristan, V., Skiena, S.S., Yang, T.-C. *On minimum-area hulls*, Algorithmica 21 (1998), 119–136.

[2] Barequet, G., Har-Peled, S. *Polygon containment and translational min-Hausdorff-distance between segment sets are 3SUM-hard*, Int. J. Comput. Geom. and App. 11 (2001), 465–474.

[3] Barrera, A.H. *Finding an $o(n^2 \log n)$ algorithm is sometimes hard*, Proc. 8th Canad. Conf. Comput. Geom., Ottawa, Ontario, Canada, Aug. 1996, 289–294.

[4] Ben-Or, M. *Lower bounds for algebraic computation trees*, Proc. 15th Annual ACM Sympos. Theory Comput. (1983), 80–86.

[5] Björner, A., Lovász, L., Yao, A.C. *Linear decision trees: volume estimates and topological bounds*, Proc. 24th Annual ACM Symp. Theory Comput. (1992), 170–177.

[6] Bose, P., van Kreveld, M., and Toussaint, G. *Filling polyhedral molds*, Proc. 3rd Workshop Algorithms Data Structures, LNCS, Springer-Verlag 709 (1993), 210–221.

[7] de Berg, M., de Groot, M., Overmars, M. *Perfect binary space partitions*, Comput. Geom. Theory Appl. 7 (1997), 81–91.

[8] Dietzfelbinger, M. *Lower bounds for sorting of sums*, Theoret. Comput. Sci. 66 (1989), 137–155.

[9] Dobkin, D.P., Lipton, R.J. *On the complexity of computations under varying set of primitives*, Journal of Computer and Systems Science 18 (1979), 86–91.

[10] Erickson J. *Lower bounds for linear satisfiability problems*, Chicago Journal of Theoretical Computer Science 8, 1999.

[11] Erickson J. *New lower bounds for convex hull problems in odd dimensions*, SIAM J. Comput. 28 (1999), 1198–1214.

[12] Erickson J., Seidel, R. *Better lower bounds on detecting affine and spherical degeneracies*, Disc. Comput. Geom. 13 (1995) 41–57.

[13] Fredman, M.L. *How good is the information theory bound in sorting*, Theoret. Comput. Sci. 1 (1976), 355–361.

[14] Gajentaan, A., Overmars, M. *On a class of $O(n^2)$ problems in computational geometry*, Comput. Geom. Theory Appl. 5 (1995), 165–185.

[15] Grigoriev, D., Karpinski, M., Meyer auf der Heide, F., Smolensky, R. *A lower bound for randomized algebraic decision trees*, Proc. 28th Annual ACM Symp. Theory Comput. (1996), 612–619.

[16] Grigoriev, D., Karpinski, M., Vorobjov, N. *Lower bound on testing membership to a polyhedron by algebraic decision and computation trees*, Discrete Comput. Geom. 17 (1997), 191–215.

[17] MacWillams, F.J., Sloane, N.J.A. *The Theory of Error Correcting Codes*, North Holland, 1977.

[18] Matousek, J. *On geometric optimization with few violated constraints*, Disc. Comput. Geom. 14 (1995), 365–384.

[19] Meyer auf der Heide, F. *A polynomial linear search algorithm for the n-dimensional knapsack problem*, J. ACM 31 (1984), 668–676.

[20] Nagura, J. *On the interval containing at least one prime number*, Proc. Japan Acad. 28 (1952), 177–181.

[21] Steele, M., Yao, A.C. *Lower bounds for algebraic decision trees*, Journal of Algorithms 3 (1982), 1–8.

[22] Yao, A.C. *Decision tree complexity and Betti numbers*, J. Comput. System Sci. 55 (1997), 36–43.

[23] Yao, A.C. *Algebraic decision trees and Euler characteristics*, Theoret. Comput. Sci. 141 (1995), 133–150.

[24] Yao, A.C. *DIMACS Workshop on Intrinsic Complexity of Computation*, April 10–13, 2000.