9

COMPLEXITY BOUNDS VIA ROTH'S METHOD OF ORTHOGONAL FUNCTIONS

BERNARD CHAZELLE

To Klaus F. Roth, with deep gratitude for his beautiful, inspiring work.

1. Introduction

It is the holy grail of theoretical computer science to find algorithms that are provably optimal with respect to some complexity measure; usually the time they take to run or the storage they require. While the field has had great success in designing fast algorithms for all sorts of problems, proving complexity lower bounds has been the weak link. In 1954, K.F. Roth's work on the discrepancy of boxes [7] led him to introduce a proof technique, sometimes referred to as the "method of orthogonal functions"; see also [1, 4]. Almost half a century later, the method played a key role in the derivation of a complexity lower bound for a classical computer science problem. We explain this serendipitous connection.

2. Background

Computational geometry is the branch of theoretical computer science concerned with the complexity of computing over geometric inputs. A classical problem in the field, *orthogonal range searching*, has the following formulation: given n points $p_1, \ldots, p_n \in \mathbf{R}^2$ and n boxes (i.e. axis-parallel rectangles) $R_1, \ldots, R_n \subset \mathbf{R}^2$, count how many points lie in each box. The problem statement is usually generalized by associating a variable $x_i \in \mathbf{R}$, sometimes called a *weight*, with each point p_i. Therefore, the input $\mathcal{I}$ to the problem consists of the three sets $\{p_i\}$, $\{x_i\}$, $\{R_i\}$, and the desired output $\mathcal{O}$ is the set of numbers

$$\sum_{\substack{j=1 \\ p_j \in R_i}}^{n} x_j, \quad i = 1, \ldots, n.$$

The algorithms considered for solving this problem obey a specific format. Each one consists of a list of instructions $I_1, \ldots, I_\ell$, where each I_k is of the form $z_k \leftarrow x_k$, for $1 \leqslant k \leqslant n$; and $z_k \leftarrow \alpha_k z_i + \beta_k z_j$, where $i, j < k$ and $\alpha_k, \beta_k \in \mathbf{C}$, for $n < k \leqslant \ell$. In other words, each instruction consists of acquiring a new variable z_k and then either initializing it with the weight x_k or assigning it a linear combination of previously computed variables, with real or complex coefficients. The requirement should be that, for any choice of weights $\{x_i\}$, the output $\mathcal{O}$ should be a subset of $\{z_1, \ldots, z_\ell\}$. The *complexity* of the algorithm is ℓ. This computing model is called a *straightline program* in the literature, or a *linear circuit*. One key point is that a given algorithm must work for *any* assignment of weights x_i. Different algorithms can be used for different inputs,

however; in other words, ℓ as well as all of the α_k, β_k may be functions of the points and boxes but *not* of the weights.

It has been a longstanding open question to determine how big ℓ must be, in the worst case, as a function of n. For each input $\mathcal{I}$, consider the shortest program (i.e. smallest ℓ) and then the longest such minimal program over all inputs with n points and n boxes. The corresponding number $\ell = \ell(n)$ denotes the *complexity* of the problem. Let A be the incidence matrix of the points and boxes, i.e. $A_{ij} = 1$ if the i-th box contains the j-th point and $A_{ij} = 0$ otherwise. The output is the vector Ax, where $x = (x_1, \ldots, x_n)^T$. Thus the problem can be seen as that of multiplying a square matrix from a particular family by an arbitrary vector. For this reason, the function ℓ is often called the linear[1] complexity of the matrix family.

Obviously, $n \leqslant \ell(n) \leqslant n^2$. Unfortunately, no better bounds are known. It is therefore customary to make the assumption that the moduli of all the coefficients α_k, β_k are bounded by an absolute constant. While relaxing this assumption leaves us with one of the foremost open questions in complexity theory, enforcing it can be justified in a number of ways – besides the feeble excuse that it allows us to prove something. All of the algorithms actually used in practice for computing Ax in real-life applications, which abound, satisfy the bounded-coefficient condition.[2] The bounded-coefficient model will be assumed henceforth.

Most lower bound results relate the linear complexity of a matrix to combinatorial and algebraic properties such as its rigidity (see [8]) or its spectrum; typically the median singular value of A (see [2]) or the traces of A^TA and A^TAA^TA (see [5]). The simplest such result, known as the *Morgenstern bound* [6], was also the first one historically: it states that $\ell = \Omega(\log|\det A|)$. The question then is how to design an input set of n points and n boxes whose incidence matrix has a large determinant. We will exhibit a set system whose incidence matrix A satisfies

$$|\det A| = \Omega(\log n)^{n/2}. \tag{1}$$

By Morgenstern's bound, this implies the complexity of orthogonal range searching is $\Omega(n \log\log n)$.

3. The Proof

The points $p_1, \ldots, p_n$ are defined by using the standard Halton–Hammersley (bit reversal) construction. We define a slightly bigger set S from which we extract the points p_i. Assume without loss of generality that n is a large power of 2. Let $m = 8n$ and

$$S = \left\{ \left(\frac{1}{2m} + c(k), \frac{1}{2m} + \frac{k}{m} \right) : 0 \leqslant k < m \right\},$$

where

$$c(k) = \sum_{i \geqslant 0} \frac{b(i)}{2^{i+1}}$$

[1]The algorithm involves only linear functions.

[2]One of the most common examples is the FFT algorithm, which is used to compute the Fourier transform over one's favourite abelian group.

and $\{b(i)\}$ is the binary expression for the running index k, i.e.

$$k = \sum_{i \geqslant 0} b(i) 2^i.$$

For any $1 \leqslant k \leqslant \log m$, let $\mathcal{G}_k$ be the grid obtained by dividing $[0,1]^2$ into m axis-parallel rectangles of size $2^{-k} \times (2^k/m)$. Figure 1 illustrates the point set S and the grids for $m = 8$.

One can easily verify that each cell σ of $\mathcal{G}_k$ is a rectangle of area $1/m$ that contains precisely one point p of S. We say that p is *well-centred for* $\mathcal{G}_k$ if it lies near the centre c_σ of σ; specifically, within the box $(\sigma + c_\sigma)/2$. A simple examination reveals that at least a quarter of the points in S are well centred for $\mathcal{G}_k$; therefore, at least $n = m/8$ of them are each well-centred for at least $\frac{1}{8} \log m$ grids $\mathcal{G}_k$. This defines the set $\{p_1, \ldots, p_n\}$.

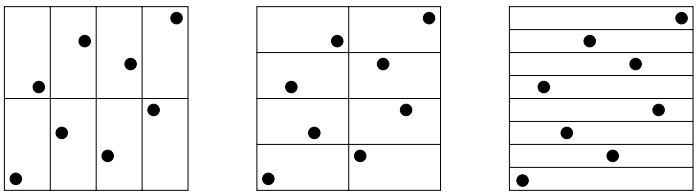

Figure 1. *Halton–Hammersley points and the grids* $\mathcal{G}_k$

Let $\mathcal{G}$ be the $(\sqrt{N}-1) \times (\sqrt{N}-1)$ square grid covering $[0,1]^2$, where $N = (m^2+1)^2$. We define an $N \times n$ matrix B as follows: each column corresponds to a distinct point p_i; each row is associated with the southwest quadrant cornered at a distinct grid point. For each grid point (u, v) there is a distinct row in B that forms the characteristic vector of the set

$$\{p_1, \ldots, p_n\} \cap ((-\infty, u] \times (-\infty, v]).$$

In this way, the N rows are not all distinct. Our next result rather "oddly" relates the L^2 norm of Bx to the L^1 norm of x. We postpone its proof, which is where Roth's method of orthogonal functions kicks in and therefore deserves special treatment.

Lemma 3.1. *For any* $x \in \mathbf{R}^n$, $\|Bx\|_2 = \Omega(n^{-1}\sqrt{N \log n}\, \|x\|_1)$.

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of $B^T B$. We establish a lower bound for the eigenvalues λ_i and then for the determinant of $B^T B$.

Lemma 3.2. *For any* k, *the* k*-th largest eigenvalue* λ_k *of* $B^T B$ *satisfies*

$$\lambda_k = \Omega(n^{-2}(n-k+1)N \log n).$$

Proof. Let $\{v_i\}$ be an orthonormal eigenbasis for $B^T B$, where v_i is a unit eigenvector associated with λ_i. If $Q = (q_{ij})$ denotes the orthonormal matrix whose rows are the eigenvectors v_i, then the column vector ξ obtained by expressing x in the basis $\{v_i\}$ satisfies $\xi = Qx$. What makes the use of Lemma 3.1 a little awkward is that the L^2 norm of Bx is bounded as a function of the L^1 norm of x. To get around this difficulty, we show that the invariant subspace spanned by $\{v_k, \ldots, v_n\}$ contains a unit vector x whose L^1 norm is as large as $\sqrt{n-k+1}$. The lower bound will then follow from the variational characterization of eigenvalues.

We use an Erdős-style probabilistic argument. Let $R = (r_{ij})$ be the matrix obtained by replacing each of the first $k-1$ rows of Q by a row of zeros, and let $y = (y_1, \dots, y_n)$ be a random vector chosen uniformly in $\{-1, 1\}^n$. Then

$$\begin{aligned}\mathbf{E}\|Ry\|_2^2 &= \sum_{i=1}^{n} \mathbf{E}\left(\sum_{j=1}^{n} r_{ij}y_j\right)^2 = \sum_{i\geqslant k}\sum_{j=1}^{n} q_{ij}^2 \mathbf{E}y_j^2 + \sum_{i\geqslant k}\sum_{j\neq j'} q_{ij}q_{ij'}\mathbf{E}y_j y_{j'} \\ &= \sum_{i\geqslant k}\sum_{j=1}^{n} q_{ij}^2 = n-k+1.\end{aligned}$$

This proves the existence of a vector $y \in \{-1,1\}^n$ such that $\|Ry\|_2^2 \geqslant n-k+1$; and therefore, the $(n-k)$-flat defined by the equations

$$\begin{cases} \xi_i = 0, \quad 1 \leqslant i \leqslant k-1, \\ (Qy)^T\xi = \sqrt{n-k+1}, \end{cases}$$

intersects the ball of unit radius centred at the origin. If x is a point of the intersection, then it follows from $\xi = Qx$ that $\|x\|_1 \geqslant y^T x = (Qy)^T\xi = \sqrt{n-k+1}$, and, from Lemma 3.1,

$$\begin{aligned}\lambda_k = \lambda_k\|x\|_2^2 &\geqslant \sum_{i=1}^{n} \lambda_i\xi_i^2 = \|Bx\|_2^2 = \Omega(n^{-2}\|x\|_1^2 N\log n) \\ &= \Omega(n^{-2}(n-k+1)N\log n).\end{aligned}$$

□

By Binet–Cauchy,

$$\det B^T B = \sum_{1\leqslant j_1<\ldots<j_n\leqslant N} \left|\det B \begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ 1 & 2 & \ldots & n \end{pmatrix}\right|^2;$$

therefore, there exists an $n\times n$ submatrix A of B such that

$$\det A^T A = \left|\det B \begin{pmatrix} j_1^* & j_2^* & \cdots & j_n^* \\ 1 & 2 & \ldots & n \end{pmatrix}\right|^2 \geqslant \binom{N}{n}^{-1} \det B^T B.$$

By Lemma 3.2,

$$\det B^T B = \prod_{i=1}^{n} \lambda_i = \Omega(n^{-2}N\log n)^n n!$$

and so

$$\det A^T A = \Omega(1)^n \left(\frac{n}{\mathrm{e}N}\right)^n \left(\frac{n}{\mathrm{e}}\right)^n \left(\frac{N\log n}{n^2}\right)^n = \Omega(\log n)^n,$$

which establishes (1).

4. Roth's Method of Orthogonal Functions

We prove Lemma 3.1 by using (a discrete version of) Roth's method of orthogonal functions [7]. The idea is simple but very clever: it consists of manufacturing a custom-made family of orthogonal functions and use its sum as an auxiliary function[3] whose inner product with Bx can be bounded from below. We can then apply Cauchy–Schwarz to derive the desired lower bound.

[3]Here we consider vector rather than function, since we discretize everything.

By reversing signs if necessary, we can always assume that

$$\|x\|_1 \leqslant 2 \sum_{x_i > 0} x_i. \tag{2}$$

Fix $1 \leqslant k \leqslant \log m$; a cell σ of $\mathcal{G}_k$ is called *k-heavy* if it contains a well-centred point p_i such that $x_i > 0$. We assign a weight to each grid point q of the $(\sqrt{N}-1) \times (\sqrt{N}-1)$ square grid $\mathcal{G}$ as follows: Let σ be any cell of $\mathcal{G}_k$ that contains q.

- If σ is not uniquely defined because q lies on its boundary, or if σ is not k-heavy, then assign q a weight of 0.
- Otherwise, subdivide σ into four equal-size quadrants similar to σ. Assign q a weight of 1 if it lies in the interior of the northeast or southwest quadrant; assign a weight of -1 if it lies in the interior of the northwest or southeast quadrant. If q lies elsewhere, assign it a weight of 0.

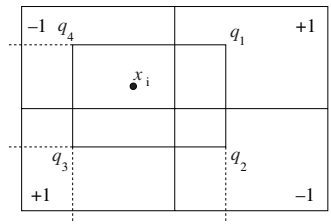

Figure 2. *Cancellation at work*

Using the same ordering as the rows of B, let $g_k \in \mathbf{R}^N$ be the column vector of weights. It is easy to see that the $\log m$ vectors g_k form an orthogonal family. Let G be the matrix $(g_1, \ldots, g_{\log m})$ and let u be the column vector made of $\log m$ entries of 1. By the orthogonality of G,

$$\|Gu\|_2^2 = \sum_{k=1}^{\log m} \|g_k\|_2^2 \leqslant N \log m. \tag{3}$$

Now, if we sum separately over each k-heavy cell σ, we obtain

$$g_k^T Bx = \Omega\left(\frac{N}{m} \sum_{\substack{i=1 \\ p_i \in \text{ some } k\text{-heavy cell of } \mathcal{G}_k}}^{n} x_i \right). \tag{4}$$

Figure 2 illustrates how, in a k-heavy cell, the weights of 1 at q_1, q_3 and -1 at q_2, q_4 produce the cancellations that contribute x_i to $g_k^T Bx$, for any q_1 higher and to the right of x_i. Since the cell in question is k-heavy, the set of such q_1's covers at least a fraction of the cell, hence the N/m factor in (4). Now since each p_i is well centred for at least a fraction of the grids $\mathcal{G}_k$, we know from (2) and $m = 8n$ that

$$(Gu)^T Bx = \Omega(n^{-1} \|x\|_1 N \log n).$$

Finally, by Cauchy–Schwarz and (3),

$$\frac{N \log n}{n} \|x\|_1 = O((Gu)^T Bx) = O(\|Gu\|_2 \|Bx\|_2) = O(\sqrt{N \log n}\, \|Bx\|_2).$$

This completes the proof of Lemma 3.1.

Acknowledgment. This work was supported in part by NSF grants CCR-0306283 and ARO Grant DAAH04-96-1-0181.

References

[1] J. Beck, W.W.L. Chen. *Irregularities of distribution* (Cambridge Tracts in Mathematics 89, Cambridge University Press, 1987).

[2] B. Chazelle. A spectral approach to lower bounds with applications to geometric searching. *SIAM J. Comput.*, 27 (1998), 545–556.

[3] B. Chazelle. Lower bounds for off-line range searching. *Discrete Comput. Geom.*, 17 (1997), 53–65.

[4] B. Chazelle. *The discrepancy method: randomness and complexity* (Cambridge University Press, 2000; paperback version, 2001).

[5] B. Chazelle, A. Lvov. A trace bound for the hereditary discrepancy. *Discrete Comput. Geom.*, 25 (2001), 519–524.

[6] J. Morgenstern. Note on a lower bound of the linear complexity of the fast Fourier transform. *J. ACM*, 20 (1973), 305–306.

[7] K.F. Roth. On irregularities of distribution. *Mathematika*, 1 (1954), 73–79.

[8] L. Valiant. Graph-theoretic arguments in low-level complexity. *Mathematical Foundations of Computer Science 1977* (J. Gruska, ed.), pp. 162–176 (Lecture Notes in Computer Science 53, Springer-Verlag, 1977).