The Clarkson-Shor Technique Revisited and Extended[*]

Micha Sharir[†]

ABSTRACT

We provide an alternative simpler and more general derivation of the Clarkson-Shor probabilistic technique [4] and use it to obtain in addition several extensions and new combinatorial bounds.

1. INTRODUCTION

In abstract and general setting, the Clarkson-Shor technique [4] deals with the following type of problems. Let S be a set of n objects, and C a set of *configurations*, each defined by d objects of S, for some constant integer parameter d. We are also given a *conflict relationship* between objects $a \in S$ and configurations $c \in C$, where it is assumed that the d objects that define c are not at conflict with it. The *weight* of a configuration c is the number of objects that are in conflict with c.

As a concrete example, let S be a set of n lines (the objects) in the plane. A configuration is a vertex of the arrangement $\mathcal{A}(S)$, defined by $d = 2$ lines. A line is at conflict with a vertex if it passes below the vertex. The zero-weight vertices are those that appear on the lower envelope of the lines, and their number is at most $n - 1$. Vertices of weight k belong to the k-th level of the arrangement.

Let $C_0(S)$ denote the set of 0-weight configurations, let $C_k(S)$ denote the set of configurations of weight exactly k, and let $C_{\leq k}(S)$ denote the set of configurations of weight at most k, where k is any integer between 0 and $n - d$. Put $N_0(S) = |C_0(S)|$, $N_k(S) = |C_k(S)|$, and $N_{\leq k}(S) = |C_{\leq k}(S)|$. We also denote by $N_0(n)$ (resp. $N_k(n)$, $N_{\leq k}(n)$)

the maximum of $N_0(S)$ (resp. of $N_k(S)$, $N_{\leq k}(S)$), over all sets S of n objects of this kind.

The Clarkson-Shor technique provides the following upper bound for $N_{\leq k}(S)$:

THEOREM 1.1 (CLARKSON AND SHOR [4]).

$$N_{\leq k}(n) = O(k^d N_0(n/k)).\tag{1}$$

We derive a somewhat different bound, which can then be manipulated to yield the Clarkson-Shor's bound—see below.

THEOREM 1.2.

$$N_{\leq k}(S) \leq \frac{\mathbf{E}(N_0(R_p))}{p^d(1 - pk)},\tag{2}$$

for any $0 < p < 1/k$, where R_p is a random sample of objects from S, where each object is chosen independently with probability p, and $N_0(R_p)$ is the number of configurations, all of whose defining objects are in R_p and none of whose conflicting objects are in R_p.

Remark: Theorem 1.1 can in fact also be stated in terms of expectation: It asserts that $N_{\leq k}(S)$ is upper bounded by $O(k^d)$ times the expected value of $N_0(R'_{n/k})$, where $R'_{n/k}$ is a random sample of n/k objects from S, where all such samples are chosen with equal probability.

By now, the Clarkson-Shor technique needs no praise—it has become a cornerstone of many of the developments in computational and combinatorial geometry in the past decade, for example, in the analysis of randomized incremental algorithms [9], and in the derivation of sharp bounds on the complexity of lower envelopes and other substructures in higher-dimensional arrangements [11]. This paper serves two main purposes:

(i) It puts the Clarkson-Shor's technique in a somewhat different context, leading to a simpler proof and perhaps also to a better understanding of the technique.

(ii) It facilitates the extension of the context in which the technique can be applied. We obtain in fact several results which do not seem to be derivable by the original technique.

Our proof technique is a simple extension of the probabilistic proof of the *Crossing Lemma* of Leighton [8] and of Ajtai et al. [2]. (This technique is so elegant that it has achieved a status of 'a proof from the *Book*' in [1].) On one hand, this technique is quite general, but, on the other hand, in almost all applications to date it has been administered to

crossings between edges of graphs drawn in the plane (see [3, 5] for some exceptions). It is our hope that the results reported here constitute a first step towards a wider range of applications of the technique.

2. ANALYSIS

Before proceeding to the proof of Theorem 1.2, we first establish the following more general result.

Let C' be any subset of configurations. Let X be the number of conflicts with the configurations of C'. That is,

$$X = \left| \{(a,c) \mid c \in C', a \text{ is at conflict with } c \} \right|$$

We then have

$$X \geq |C'| - N_0(S). \qquad (3)$$

Indeed, if $|C'| > N_0(S)$ then C' contains at least one configuration c that has at least one conflict, contributing at least 1 to the count X. Remove c from C' and repeat this argument, and keep doing so as long as $|C'| > N_0(S)$.

Draw a random sample R of objects from S, by drawing each $a \in S$ independently, with the same probability p. Let C'_R denote the set of configurations in C' that appear in R, that is, those configurations for which all d defining objects are chosen in R, and let X_R denote the number of conflicts of these configurations in R, that is, the number of pairs (a,c) such that $c \in C'_R$, $a \in R$ and a is at conflict with c.

By (3), we have

$$X_R \geq |C'_R| - N_0(R),$$

so the same holds for the expectation of this inequality:

$$\mathbf{E}(X_R) \geq \mathbf{E}(|C'_R|) - \mathbf{E}(N_0(R)).$$

It follows from the construction that

$$\mathbf{E}(|C'_R|) = |C'|p^d \quad \text{and} \quad \mathbf{E}(X_R) = Xp^{d+1}.$$

Hence we obtain

THEOREM 2.1. *For any set C' of configurations, which have a total of X conflicts, and for any probability $0 < p \leq 1$, one has*

$$Xp \geq |C'| - \frac{\mathbf{E}(N_0(R_p))}{p^d}, \qquad (4)$$

where R_p is as in the statement of Theorem 1.2.

Proof of Theorem 1.2: Let us specialize Theorem 2.1 to the case $C' = C_{\leq k}(S)$. In this case we have, by definition, $X \leq kN_{\leq k}(S)$, which implies:

$$(1 - pk)N_{\leq k}(S) \leq \frac{\mathbf{E}(N_0(R_p))}{p^d},$$

and therefore completes the proof. □

A quick illustration. Consider the case of lines and vertices mentioned in the introduction. We have $\mathbf{E}(N_0(R_p)) = \mathbf{E}(|R_p| - 1) \leq np$, so

$$N_{\leq k}(S) \leq \frac{np}{p^2(1 - pk)} \leq 4nk,$$

by choosing $p = 1/(2k)$.

The Clarkson-Shor bound. Let us turn (2) into the more familiar bound given in Theorem 1.1. Suppose that $N_0(S) \leq A\binom{n}{\gamma}$, for any set S of n objects and for some integer parameter $\gamma \geq 1$ and constant $A > 0$. Using the fact

that $|R_p|$ has an (n, p)-binomial distribution, it is routine to verify that

$$\mathbf{E}\left[\binom{|R_p|}{\gamma} \right] = \binom{n}{\gamma} p^\gamma.$$

Hence (2) becomes

$$N_{\leq k}(S) \leq \frac{A\binom{n}{\gamma}}{p^{d-\gamma}(1 - pk)}.$$

Now choose $p = \beta/k$, for some parameter $\beta \in (0,1)$, to obtain

$$N_{\leq k}(S) \leq \frac{A\binom{n}{\gamma}k^{d-\gamma}}{\beta^{d-\gamma}(1 - \beta)}.$$

The denominator is maximized when $\beta = (d-\gamma)/(d-\gamma+1)$, which is easily seen to yield

$$N_{\leq k}(n) \leq A(d - \gamma + 1)e\binom{n}{\gamma}k^{d-\gamma}. \qquad (5)$$

This is essentially the Clarkson-Shor result.

3. EXTENSIONS—SIMPLE CONSTRAINTS

Recall that the inequality (4) of Theorem 2.1 is fairly general, and does not impose any specific assumptions on the set C'. We next apply this theorem to several other problems involving sets of configurations C' whose overall number of conflicts can be upper bounded by some other simple argument, which will lead to various new upper bounds on the size of such sets C'. In the following section, we will extend the technique further, by considering conflicts between a configuration and *several* objects.

Configurations for which no object is at conflict with many. Here we consider an 'inverse problem', where we wish to bound the maximum possible size of a set $C^*_{\leq k}$ of configurations so that no object is at conflict with more than k configurations in $C^*_{\leq k}$.

For such sets we have trivially $X \leq nk$. Assume also, as above, that $N_0(n) \leq A\binom{n}{\gamma}$, for some integer $\gamma \geq 1$ and constant $A > 0$. Then choose p in (2) so that

$$|C^*_{\leq k}| = \frac{2A\binom{n}{\gamma}p^\gamma}{p^d},$$

or

$$p = \left(\frac{2A\binom{n}{\gamma}}{|C^*_{\leq k}|} \right)^{1/(d-\gamma)},$$

which makes sense only when $|C^*_{\leq k}| > 2A\binom{n}{\gamma}$. Assuming this to be the case, (4) implies that

$$|C^*_{\leq k}| \leq 2nkp = 2nk \left(\frac{2A\binom{n}{\gamma}}{|C^*_{\leq k}|} \right)^{1/(d-\gamma)},$$

from which it follows that

$$|C^*_{\leq k}| = O\left(n^{d/(d-\gamma+1)} k^{(d-\gamma)/(d-\gamma+1)} \right).$$

We thus have

THEOREM 3.1. *Assume the above abstract setup of objects, configurations, and conflicts, and suppose that the number of configurations with no conflicts in any set of n objects*

is $O(n^\gamma)$. Let $1 \le k \le n$ be a given parameter. Then the maximum cardinality of any set of configurations with the property that no object is at conflict with more than k of them is

$$O\left(n^\gamma + n^{d/(d-\gamma+1)}k^{(d-\gamma)/(d-\gamma+1)}\right). \tag{6}$$

Examples. (1) Suppose that, as in the introduction, the objects are n lines in the plane, configurations are vertices of their arrangement, and a vertex is at conflict with a line if the vertex lies above the line. In this case we have $d = 2$, $\gamma = 1$, and we obtain

COROLLARY 3.2. *The maximum number of vertices in an arrangement of n lines in the plane such that none of the lines passes below more than k of them is $O(nk^{1/2})$. This bound is tight in the worst case. Dually, the maximum number of lines connecting pairs of points in an n-element point set in the plane, such that none of the given points lies below more than k connecting lines is $O(nk^{1/2})$, which again is worst-case tight.*

The lower bound is obtained as follows. Take $n/k^{1/2}$ lines, all appearing along their lower envelope. replace each line by a bundle of $k^{1/2}$ parallel lines sufficiently close to each other. Each vertex of the lower envelope of the original lines is replaced by k new vertices. Collecting all these vertices, we obtain a set of $nk^{1/2}$ vertices, and it is clear that no line passes below more than $2k$ of them.

(2) A similar problem can be stated and analyzed for hyperplanes in $\mathbb{R}^d$, where the configurations are vertices in an arrangement of n hyperplanes, with the parameter d equal to the dimension, and $\gamma = \lfloor d/2 \rfloor$:

COROLLARY 3.3. *The maximum number of vertices in an arrangement of n hyperplanes in $\mathbb{R}^d$ for which no hyperplane passes below more than k of them is*

$$O\left(n^{\lfloor d/2 \rfloor} + n^{d/(\lceil d/2 \rceil+1)}k^{\lceil d/2 \rceil/(\lceil d/2 \rceil+1)}\right),$$

and this bound is tight in the worst case. The same bound holds for the maximum number of hyperplanes spanned by a set of n points in $\mathbb{R}^d$ so that no point lies below more than k of them.

The lower bound follows from a construction similar to that for the planar case, which is based on the upper bound theorem for convex polytopes.

In $d = 3$ dimensions, we get the bound $O(nk^{2/3})$. By using a standard lifting transform from the plane to three dimensions [7] and by specializing the preceding corollary to $d = 3$, we also obtain:

COROLLARY 3.4. *The maximum number of circles spanned by n points in the plane, such that none of the given points lies in more than k circles is $O(nk^{2/3})$, and this bound is tight in the worst case.*

4. EXTENSIONS—MORE GENERAL CONFLICTS

So far, we have only considered conflicts, each involving one configuration and one single object, but the technique is sufficiently powerful to allow us to consider more elaborate types of conflicts, each involving one configuration and *several* objects. We illustrate this in two examples:

In the first example, the objects are lines in the plane, the configurations are triangles bounded by triples of the lines, and a triangle Δ is at conflict with two other lines ℓ_1, ℓ_2 iff the vertex $\ell_1 \cap \ell_2$ lies in the interior of Δ.

In the second example, the objects are points in 3-space, the configurations are triangles spanned by triples of the points, and a triangle Δ is at conflict with two other points u, v iff the segment uv crosses the relative interior of Δ.

Handling conflicts of this kind can be done by a straightforward modification of the method of Section 2. Specifically, suppose that a conflict involves one configuration and b objects. Starting from the inequality $X \ge |C'| - N_0(S)$, and passing to a random sample R_p as above, we have

$$\mathbf{E}(|C'_R|) = |C'|p^d \quad \text{and} \quad \mathbf{E}(X_R) = Xp^{d+b},$$

so we obtain:

THEOREM 4.1. *For any set C' of configurations, which have a total of X conflicts, each involving one configuration and b objects, and for any probability $0 < p \le 1$, one has*

$$Xp^b \ge |C'| - \frac{\mathbf{E}(N_0(R_p))}{p^d}, \tag{7}$$

where R_p is as in the statement of Theorem 1.2.

We next apply this theorem to the examples mentioned above.

4.1 Triangles and vertices in a line arrangement

THEOREM 4.2. *In an arrangement of n lines in the plane, there are at most $O(n^2k^{1/2})$ triangles whose edges lie on three of the given lines and which contain at most k vertices of the arrangement in their interiors. This bound is tight in the worst case.*

Proof: Let L be the given set of lines. Here the objects are the lines of L and the configurations are triangles bounded by triples of lines in L. Conflicts are more involved: A triangle is at conflict with a vertex of $\mathcal{A}(L)$ if it contains the vertex in its interior. Thus a conflict is defined in terms of 5 lines: three defining the triangle and two the vertex.
Claim: The number $N_0(L)$ of triangles spanned by three lines of L and containing no vertex of $\mathcal{A}(L)$ in their interior is $O(n^2)$.

Indeed, let T_0 be the set of these triangles that are bounded by three lines of L and do not contain any vertex in their interior. The number of triangles in T_0 that are not crossed by any line of L is clearly $O(n^2)$—they are faces of the arrangement.

Suppose then that a triangle Δ of T_0 is crossed by at least one line ℓ but contains no vertex in its interior. Charge Δ to an intersection point v of ℓ with $\partial\Delta$, say with some line ℓ_0 bounding Δ. We claim that there does not exist any other triangle Δ' in T_0 that is bounded by ℓ_0, lies on the same side of ℓ_0 as Δ, and contains v on its boundary. Indeed, suppose that such a Δ' exists. The intersection $\Delta \cap \Delta'$ is a convex polygon that contains v on its boundary and is crossed by ℓ. Hence ℓ must intersect the boundary of this region at a second point w. Without loss of generality, assume that w lies on $\partial\Delta$. But then w is a vertex of $\mathcal{A}(L)$ that lies in the interior of Δ', contradicting the fact that Δ' is in T_0. This implies that the number of triangles in question is at most

proportional to the number of vertices v of the arrangement, so their number is $O(n^2)$. This completes the proof of the claim.

Let $N_{\leq k}(L)$ denote the number of triangles with weight at most k, and let X denote the total number of conflicts that they have, which is clearly upper bounded by $X \leq kN_{\leq k}(L)$. Using Theorem 4.1, we thus obtain

$$kp^2 N_{\leq k}(L) \geq N_{\leq k}(L) - \frac{\mathbf{E}(N_0(R_p))}{p^3},$$

so

$$N_{\leq k}(L) \leq \frac{\mathbf{E}(N_0(R_p))}{p^3(1 - kp^2)}.$$

Choosing $p = \beta/k^{1/2}$, for an appropriate $\beta < 1$, and using the fact that $N_0(n) = O(n^2)$, the upper bound of the theorem follows readily. The lower bound is obtained for an arrangement consisting of $n/3$ equally-spaced horizontal lines, $n/3$ equally-spaced vertical lines, which together form part of the integer grid, and $n/3$ additional equally-spaced lines of slope 1, passing very near the grid points formed by the first two subfamilies. We leave the easy verification of the lower bound to the reader. $\square$

THEOREM 4.3. *In an arrangement of n lines in the plane, there are at most $O(n^2k^{1/3})$ triangles whose edges lie on three of the given lines and for which no vertex of the arrangement is contained in the interiors of more than k of the triangles. This bound is tight in the worst case.*

Proof: Here $X \leq k\binom{n}{2}$, so, using Theorem 4.1, the number of such triangles is at most $O(n^2kp^2 + n^2/p)$. The theorem follows by choosing $p = 1/k^{1/3}$. The lower bound can be obtained from the same construction used in the preceding proof. $\square$

Remark: The claim in the proof of Theorem 4.2 does not seem to extend to vertical trapezoids in the arrangement.

4.2 Triangles and crossing segments in a 3-dimensional point set

THEOREM 4.4. *Given a set of n points in $\mathbb{R}^3$, the maximum number of triangles spanned by the points of S that are crossed by at most k segments connecting the points is $O(n^2k^{1/2})$.*

Proof: Let $T_{\leq k}$ denote the set of these triangles, and let X denote the number of conflicts between triangles in $T_{\leq k}$ and segments. We have $X \leq k|T_{\leq k}(S)|$. By the results of [5, 13], we have $N_0(S) = O(n^2)$. Hence, by Theorem 4.1, we have

$$kp^2 |T_{\leq k}(S)| \geq |T_{\leq k}(S)| - \frac{\mathbf{E}(N_0(R_p))}{p^3},$$

so

$$|T_{\leq k}(S)| \leq \frac{\mathbf{E}(N_0(R_p))}{p^3(1 - kp^2)}.$$

Choosing $p = \beta/k^{1/2}$, for an appropriate $\beta < 1$, and using the fact that $N_0(n) = O(n^2)$, the theorem follows readily. $\square$

THEOREM 4.5. *Given a set of n points in $\mathbb{R}^3$, the maximum number of triangles spanned by the points of S, so that no segment connecting two points crosses more than k of them is $O(n^2k^{1/3})$.*

Proof: Here we have $X \leq k\binom{n}{2}$, so the proof proceeds as in the proof of Theorem 4.3. $\square$

COROLLARY 4.6. *The number of halving triangles in an n-element point set in 3-space is $O(n^{8/3})$.*

Proof: By Lovász lemma, a segment crosses at most $O(n^2)$ halving triangles. $\square$

Note that this bound is weaker than the best known bound $O(n^{5/2})$ [13].

COROLLARY 4.7. *The maximum number of distinct triangles that lie on the boundaries of k convex polytopes spanned by a set of n points in 3-space is $O(n^2k^{1/3})$.*

Proof: Clearly, no segment crosses more than $2k$ of these triangles. $\square$

This result has been obtained by Aronov and Dey [3] using a more involved argument. A simple alternative proof is given in [12].

5. DISCUSSION

Clearly, this paper only scratches the surface of the realm of applications of this (extended) technique. For example, in the original setup of the Crossing Lemma of [2, 8], a conflict occurs between two configurations (a conflict is a crossing between two edges, that is, between two configurations), a situation that we haven't considered at all here, but one that should be amenable to the new technique just as the other cases studied here.

We believe that the ideas developed here will have additional applications. For example, our next planned step in this research is to find algorithmic applications for the new bounds, extending similar applications of the standard Clarkson-Shor's bounds, e.g., to the analysis of randomized incremental algorithms.

We end the paper by presenting an alternative interpretation of the analysis employed in this paper. Let S be a set of objects, and let A be a subset of S^q, for some q. Put $d_A = q$, and refer to it as the *dimension* of A. For any $R \subseteq S$, let $A_R = A \cap R^q$.

Suppose now that we have a finite collection $A_1, \ldots, A_\nu$ of such sets of ordered tuples, possibly with different dimensions d_{A_i}, with corresponding (positive or negative) constants $c_1, \ldots, c_\nu$, so that the linear relation

$$\sum_{i=1}^{\nu} c_A |A_R| \leq f(|R|)$$

holds for any $R \subseteq S$, where $f(\cdot)$ is some function of $|R|$.

Then, for any $p \in (0,1)$, we have

$$\sum_{i=1}^{\nu} c_i p^{d_{A_i}} |A_i| \leq \mathbf{E}[f(|R_p|)]. \qquad (8)$$

The proof of the Crossing Lemma of [2, 8] is an instance of this observation, using three sets A_1, A_2, A_3 of respective dimensions $1, 2, 4$, where A_1 is S, the set of vertices of the given graph, A_2 is the set of its edges, and A_3 is the set of edge crossings (each represented by the quadruple of the vertices incident to the pair of crossing edges). The linear relation is $|A_3| - |A_2| + 3|A_1| \geq 0$, and (8), with an appropriate choice of p, yields the lemma.

The derivation of, say, Theorem 2.1 can also be interpreted as an application of (8) to the inequality (3), as the

reader can easily verify. This interpretation applies also to other theorems derived in this paper.

This view of the analysis presented here shows that it is indeed strongly related to the original Crossing Lemma.

Acknowledgments

6. REFERENCES

[1] M. Aigner and G.M. Ziegler, *Proofs from The Book*, Springer-Verlag, Berlin 1998.

[2] M. Ajtai, V. Chvátal, M.M. Newborn, and E. Szemerédi, Crossing-free subgraphs, *Ann. Discrete Math.* 12 (1982), 9–12.

[3] B. Aronov and T. Dey, Polytopes in arrangements, *Discrete Comput. Geom.* 25 (2001), 51–63.

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

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

[6] T. Dey and J. Pach, Extremal problems for geometric hypergraphs, *Discrete Comput. Geom.* 19 (1998), 473–484.

[7] H. Edelsbrunner, *Algorithms in Combinatorial Geometry*, Springer Verlag, Heidelberg 1987.

[8] F.T. Leighton, *Complexity Issues in VLSI, Foundations of Computing Series*, MIT Press, Cambridge, Mass., 1983.

[9] K. Mulmuley and O. Schwarzkopf, Randomized algorithms, in: *Handbook of Discrete and Computational Geometry* (J.E. Goodman and J. O'Rourke, eds.), CRC Press, 1997, 633–652.

[10] J. Pach and P.K. Agarwal, *Combinatorial Geometry*, Wiley-Interscience, New York 1995.

[11] M. Sharir and P.K. Agarwal, *Davenport Schinzel Sequences and Their Geometric Applications*, Cambridge University Press, New York 1995.

[12] M. Sharir and S. Smorodinsky, Polytopes in arrangements, in preparation.

[13] M. Sharir, S. Smorodinsky and G. Tardos, An improved bound for k-sets in three dimensions, *Discrete Comput. Geom.*, to appear.

[14] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry, *Combinatorics, Probability and Computing* 6 (1997), 353–358.