Computational Geometry: A Retrospective

BERNARD CHAZELLE
Department of Computer Science
Princeton University
Princeton, NJ 08544, USA

Abstract

Computational geometry is at a crossroads. New challenges and opportunities are likely to reshape the field rather drastically in the years ahead. I will survey some of its principal accomplishments, and in light of recent developments, I will discuss the profound transformations the field has begun to undergo. There are reasons to believe that computational geometry will emerge from this transition far richer and stronger but barely recognizable from what it was ten years ago.

Over the last two decades the field has enjoyed tremendous successes. Some of them might be dismissed as the cheap payoffs to be expected from any field lacking maturity. But others are the products of indisputable creativity and should be held as genuine scientific achievements. More important, the field is now able to claim a broad, solid foundation upon which its future can be securely built.

To mature fully as an original subfield of computer science, however, computational geometry must broaden its connections to applied mathematics while at the same time pay more than lip service to the applications areas that it purports to serve. Happily, active efforts to meet these challenges are underway.

Three recent developments are particular encouraging: one is the building of a theory of geometric sampling and its revolutionary impact on the design of geometric algorithms. Another is the maturing of computational real-algebraic geometry and computational topology: both subjects are being revitalized by the introduction of geometric (as opposed to purely algebraic) methods. On the practical end of the spectrum, the emergence of a sub-area concerned specifically with issues of finite precision and degeneracy in geometric computing is a most welcome development.

1 Introduction

Nearly 20 years ago Shamos and Hoey [250] introduced Voronoi diagrams to the world of theoretical computer science. Computational geometry was born. To be sure, Voronoi diagrams had been known long before (surprisingly even Voronoi seemed to have been cognizant of them), but to compute them efficiently was a new concern. In short order, researchers identified a list of top-priority geometric problems to be solved. The oft-cited reason for working on these problems were their relevance to computer graphics, chemistry, statistical analysis, pattern recognition, geographic databases, and other equally vital applications areas.

I forget how convincing these claims might have seemed at the time, but with the benefit of hindsight they appear to have been fairly legitimate. Of course this might have been a happy coincidence. Taking their cue from Sir Hillary who, to the question, “Why did you climb Mount Everest?” replied, “Because it’s there!” researchers felt that certain problems were central and compelling enough to merit their attention, with scant regard for practical fallout. Classical examples are (i) computing the convex hull or the Voronoi diagram of a set of points; (ii) computing the intersection of several objects; (iii) subdividing a complex geometric object into simpler components, e.g., triangulating a simple polygon.

These are a sample of geometric primitives which we would expect any geometric software library to have. Two other important classes of problems emerged: geometric searching and fixed-dimensional geometric optimization. The first category includes problems such as point location: store a planar subdivision in some data structure so that given a query point, the face that contains it can be found quickly. It also includes range searching, e.g., set up a database to support queries such as: how many free) STOC proceedings are there within x miles of city y? (Of course, we have here the added difficulty that reading the question might change...
the answer...) **Geometric optimization** includes such problems as: given a set of points, compute its two nearest neighbors, its diameter, its smallest enclosing ellipsoid, etc.

The good news is that most of the theoretical challenges laid down over a decade ago have been met. The field is thus ready to forge ahead and broaden its scope without leaving behind massive areas of darkness. To be sure, trouble spots remain (e.g., k-set theory, hidden-surface removal), but most of the classical theoretical questions have been answered. Most satisfying is the fact that a whole body of mathematical tools and data structuring techniques have been created in the process.

The less-than-exhilarating news is that computational geometry has not yet had the practical impact it was expected to have. Issues such as robustness, stability, and degeneracies were long treated as annoying distractions from loftier intellectual pursuits. Computers have trouble with real numbers; theoreticians don't. As a result, the transfer of algorithmic ideas to the "real" world of geometric computing has lagged behind. Fortunately, this is being remedied through several promising research efforts, and one is entitled to a certain degree of optimism.

It is crucial for computational geometry to keep its eyes on the practical aspects of computing. Granted, good research should produce general, not ad hoc, results and this often necessitates a certain distance from specific, narrowly focused problems. Granted, practical problems have a knack for bringing up a hybrid mess of distinct, often contradictory, issues, which no mind of good taste would ever think of putting together. Trying to tackle all these issues at once is likely to result in the algorithmic analog of non-alcoholic diet champagne: pleasing no one by trying to please everyone. And yet one must pause and wonder: What is the point of designing convex hull algorithms if they can never be used or at least influence the way actual code is written? Practical applications are the lifeline of computational geometry. If it ever ceases to be so, chances are computational geometry will cease to be.

One final introductory remark: This little essay does not represent any views besides mine. Feel free to disagree. Of course, if my points don't come across, there won't be much to disagree about, and you might want to look at this piece as a harmless survey that the program committee chairman gently coerced me into writing. Naturally, as a survey it does suffer from its limited scope and my personal biases... All right, having thus pretty much covered all my bases in the disclaiming department, I can now begin.

In the first three sections of this paper I discuss the current status of the "classical problems" (geometric primitives, searching, optimization) and mention some of the most important open problems in these areas. Section 5 is devoted to lower bounds, a particularly sore topic (though no sorer than elsewhere in computer science). Section 6 discusses the mini-revolution of sorts caused by the newly developed theory of geometric sampling. Sections 7 and 8 give a brief account of the emerging areas of computational real-algebraic geometry and computational topology. Section 9 rapidly surveys the current efforts to cope with finite precision and degeneracies. Section 10 mentions some important developments in combinatorial geometry, which have direct bearing on the complexity of certain geometric algorithms. Finally, a few closing words are given in Section 11.

## 2 Geometric Primitives

Historically, the three main categories are: partition, intersection, visibility. The latter is more aptly classified as geometric searching and will be discussed in the next section. Convex hulls and Voronoi diagrams occupy a category of their own, mostly for historical reasons. I will not mention the practical motivations behind these problems, but I will include appropriate references where they can be found.

**Polygon Triangulation.** Triangulating a simple polygon with \( n \) vertices has attracted a great deal of attention over the years. After an early \( O(n \log n) \) solution by Garey et al. [138] dating back to 1978, little tangible progress was reported for many years, besides efficient schemes for restricted classes of polygons. The first major breakthrough came in 1986 with Tarjan and Van Wyk's discovery of an \( O(n \log \log n) \) deterministic algorithm [257]. This was followed in 1989 by a simpler solution with the same complexity due to Kirkpatrick, Klawe, and Tarjan [176]. In the meantime, an \( O(n \log^* n) \) expected-time Las Vegas algorithm was proposed by Clarkson, Tarjan, and Van Wyk [89], later simplified by Seidel [247]. Finally, in 1990, an \( O(n) \) deterministic solution was discovered by the author [55]. (This is a rare case where determinism beats randomness.) The algorithm was parallelized optimally by Goodrich [144].

I suspect that the key to linear complexity is not that a polygon is simple (i.e., free of self-intersections) but that its boundary is a connected curve. Perhaps the most interesting problem left in this general area is to find a "grand unified" solution to polygonal triangula-
tion. Given \( p \) (nonsimple) polygons with a total of \( n \) vertices, compute all \( k \) pairwise intersections and triangulate the induced planar map in optimal \( O(n+k+p \log p) \) time. Such an algorithm would unify, among others, linear-time triangulation \( (p = 1, k = 0) \) with the optimal line segment intersection algorithm of [60] \((n = 2p)\). Another interesting question is to find a truly simple optimal polygon triangulation algorithm, possibly using randomization.

**Partition Problems.** Polygon triangulation is but one example of a large class of problems whose objective is to partition a geometric object into a "few good" pieces. The subject has been heavily researched because of its relevance to the finite-element method, mesh generation, topology of 3-manifolds, interpolation theory, tool design, etc. [19, 21, 20, 24, 34, 35, 51, 59, 72, 76, 104, 219, 222, 227]. For example, the Delaunay triangulation is commonly used in the finite-element method because of its desirable morphology and overall simplicity.

Two goals are usually sought. One is to have partitions with relatively few pieces and the other is to avoid long, skinny ones. For example, we might want to decompose a three-dimensional polytope into a small number of convex pieces with bounded aspect ratios. Avoiding thin angles, large aspect ratios, or other kinds of ill behavior is the main focus of the work in [35, 207].

There is also a tradeoff involved in determining how many Steiner points (i.e., new vertices) can or should be added to the object. Steiner points are a nuisance in practice because they make representations more complicated and cause round-off errors. Unfortunately, they can rarely be avoided. To make matters worse, determining whether they are needed is NP-hard, as was shown by Ruppert and Seidel [238].

To subdivide a polygon into as few convex pieces as possible can be done in polynomial time, but most variants or extensions of the problem (e.g., polygons with holes, covering instead of partitioning) are NP-hard — see [222] for many references on this subject. The author showed that an \( n \)-vertex 3-polytope can always be subdivided into \( O(n^2) \) convex pieces and that this bound is optimal in the worst case [51]. Refinements of this result can be found in [19, 20, 24, 72, 76]. An interesting open problem is to find the complexity of approximating a minimum-size decomposition within, say, a constant factor. Similar questions can be asked involving the other standard parameters (e.g., aspect ratio, angles).

**Intersection Problems.** In two dimensions the archetypical problem is this: given \( n \) line segments, compute all its \( k \) pairwise intersections. In 1979 Bentley and Ottmann [31] used the by-now classical plane-sweep paradigm to solve the problem in \( O(n \log n + k \log k) \) time.\(^2\) To remove the \( \log k \) factor took some effort, and it was not until 1988 that an optimal \( O(n \log n + k) \) algorithm was discovered (Chazelle and Edelsbrunner [60]). Simpler probabilistic algorithms with optimal expected complexity were given by Clarkson and Shor [88] and Mulmuley [210, 213, 219].

In three dimensions, a natural operation is to intersect two convex polyhedra. Muller and Preparata [209] gave an \( O(n \log n) \) algorithm in 1978, but an optimal \( O(n) \) algorithm was found only much later, in 1989, by the author [56]. The method makes extensive use of a versatile polyhedral hierarchy developed by Dobkin and Kirkpatrick [106].

If all convexity assumptions are removed, only slightly subquadratic overhead can be achieved at this point, i.e., \( O(n^{5/4} + k \log k) \), where \( k \) is the number of intersected edges and \( \varepsilon > 0 \) is any fixed constant (Pellegrini [226]), and there is probably much room for progress. The case where one of the polyhedra is convex is treated by Mehlhorn and Simon [199] and requires \( O((n + k) \log n) \) time. It is likely that further improvement will require an increased understanding of lines in 3-space and ray shooting (see next section). An intriguing open question is whether a convex \( n \)-vertex 3-polytope can be preprocessed in linear space, so that its intersection with a plane can be computed in time \( O(k + \log n) \), where \( k \) is the size of the cross-section.

**Convex Hulls and Voronoi Diagrams.** I lump these two notions together because, as was shown by Edelsbrunner and Seidel [121], the Voronoi diagram of a set of points in \( E^d \) is isomorphic to (part of) the convex hull of an appropriate lift of the points into \( E^{d+1} \). In other words, Voronoi diagrams are special cases of convex hulls. Computing the convex hull of \( n \) points in two and three dimensions in \( O(n \log n) \) time can be done by using standard algorithmic tools (Graham [149], Preparata and Hong [228]). For Voronoi diagrams in two dimensions a simple plane-sweep method was developed by Fortune [127]; see also Fortune’s survey [130].

Things get considerably more difficult in higher dimensions, i.e., \( d > 3 \). The full facial complexity of the hull is \( \Theta(n^{d/2}) \) in the worst case. Seidel [244] gave an optimal \( O(n^{d/2}) \)-time algorithm for the case of even dimension \( d \). An \( O(n^{d/2} \log n) \) solution that fell slightly short of optimal was also discovered by Seidel [245]. A remarkable Las Vegas algorithm with optimal expected complexity was found by Clarkson and Shor [88], and a simpler variant was proposed by Sei-

\(^2\)Space sweeping is the algorithmic version of Morse theory; it is one of the most popular techniques in computational geometry.
Finally, an optimal $O(n^{4/3})$ deterministic algorithm was found by the author [58]. Interestingly, the method is based on derandomizing the Clarkson-Shor algorithm. The tools used for the derandomization are fairly involved, however (see Section 6). The technique was recently simplified somewhat and generalized by Brönnimann, Chazelle, and Matoušek [42].

Optimal algorithms now exist, but only in the worst-case sense. An interesting open problem is to find an optimal output-sensitive algorithm. The best known general solution is due to Seidel [245]: its running time is $O(n^2 + h \log h)$, where $h$ is the facial complexity of the hull. As was shown by Matoušek [189], the quadratic overhead can be slightly reduced to $O(n^{2-\epsilon})$, for any fixed $\epsilon > 0$. Whether the complexity can be reduced to the lower bound of $\Omega(h + n \log h)$ is an outstanding open problem. Currently, the only optimal output-sensitive algorithms are due to Kirkpatrick and Seidel [177] (deterministic) in two dimensions and Clarkson and Shor [88] (Las Vegas) in three dimensions. The latter algorithm was derandomized optimally by Chazelle and Matoušek [71].

There is a vast amount of literature on Voronoi diagrams. Far-reaching generalizations of the basic notion have been developed over the years: for example, abstract Voronoi diagrams and power diagrams are treated thoroughly by Klein [178] and Aurenhammer [22], respectively. I highly recommend the comprehensive survey of Voronoi diagrams and their applications by Aurenhammer [23].

Visibility and Form Factors. Ray-shooting in the plane is now well understood [2, 6, 61, 68, 164]: given a simple polygon with $n$ vertices, it is possible to shoot a ray in any direction and find the first point that it hits in $O(n \log n)$ time, using only $O(n)$ space. The dynamic case was tackled recently by Goodrich and Tamassia [148]. The three-dimensional case is still open, however, despite some progress reported in [19, 35, 154, 192]. Because the problem is akin to ray tracing in computer graphics [139] its practical importance is considerable.

Agarwal and Matoušek [5] have shown how to answer a query in time $O((n + m)/m)$, using $O(m^{1+\epsilon})$ space, for any $n \leq m \leq n^3$, which improved on earlier work [6]. Further improvement is likely to require a better computational understanding of the Plücker surface for lines in 3-space. The work in [63, 226] contains preliminary steps in that direction. Generally speaking, however, it seems fair to say that computing with Grassmannian coordinates has been grossly under-researched. Finally, I should mention the interesting use of parametric search for ray shooting discussed by Agarwal and Matoušek [4].

Visibility is a central issue in the rendering problem. Current state-of-the-art rendering techniques use radiosity methods to trace light in order to provide the most realistic rendering of a scene with lighting [91]. At the inner loop of these methods is the computation of form factors. Informally, a form factor is defined for a pair of atomic objects (typically, triangles) as the fraction of one object visible to the other through the obstacles in the scene. Fast, practical methods for computing form factors would have a great impact in rendering.

3 Geometric Searching

Hit detection and ray tracing in computer graphics, range searching in databases, point location in geographical maps are all classical examples of geometric searching for which efficient, preferably dynamic, solutions are sought. We quickly review the state of the art and mention possible directions for future research.

Point Location. In its static version, the problem of locating a query point in a planar map is completely solved [92, 119, 175, 180, 240]: it is possible to store a planar subdivision with linear storage, so that given any point, finding the region that encloses it can be done in logarithmic time.

If insertions and deletions of facial features are allowed, however, an optimal $O(\log n)$ query/update time has been elusive. Many solutions exist, but all suffer from extra logarithmic factors in one or several of the relevant resource measures (query, update time, storage) [78, 79, 147, 231]. As was shown in [232] an efficient solution in the plane would have impact on the static three-dimensional case. An interesting line of attack is to dynimize randomized incremental methods. Very little work has been done without stochastic assumptions on the sequencing of dynamic operations. A notable exception is Mulmuley’s recent work on randomized multidimensional search trees [215, 216, 217], which might contain useful ideas to solve the original problem.

In higher dimensions, searching among $n$ hyperplanes in $\mathbb{R}^d$ means locating a query point in the arrangement formed by the hyperplanes. A solution due to Clarkson [81] requires $O(n^{d+\epsilon})$ space and $O(\log n)$ query time. The storage was reduced to $O(n^d)$ by the author [57]. Probabilistic solutions are described in Mulmuley’s book [219].
problem is to store the points in some data structure so
that given any range, the points that it contains can be
reported or counted efficiently. By and large the clas-
sical instances of that problem have been solved sat-
sfactorily. For example, simplex range searching on $n$
points in $\mathbb{R}^d$ can be solved in space $m$ and query time
$O(n/m^{1/4})$ (up to logarithmic factors), where $m$ is an
adjustable parameter. This means that $n$ points in $\mathbb{R}^d$
can be stored by using $m$ memory units, so that given any
simplex the number of points that it encloses can be
determined within the indicated query time. As it
turns out, this result is nearly optimal (see Section 5).

An earlier, quasi-optimal solution by Chazelle, Sharir,
and Welzl [75] was improved by Matoušek [187], who
introduced the powerful concept of simplicial partition-
s by extending the earlier notions of partition trees [267]
and low-stabbing spanning trees [77, 264]. See also Ma-
toušek [186]. A nice survey by Welzl on the subject of
spanning trees with low stabbing numbers appears in
[265].

Halfspace range searching (which is the special case
of simplex range searching where the query range is a
halfspace) arises in many problems, in some cases quite
unexpectedly. A general linearization technique of Yao
and Yao [272] shows that many range searching prob-
lems that are expressed by high-degree polynomials can
be reduced to halfspace range searching by lifting them
into higher-dimensional space. An alternative treat-
ment of generalized range searching (where the ranges
are arbitrary semi-algebraic sets) was provided by Agar-
wal and Matoušek [5].

Efficient solutions for the reporting case of halfspace
range searching (where the points must be explicitly
enumerated as opposed to just counted) were proposed
by Clarkson [82] and Matoušek [188]: the first solution
provides $O(\log n + k)$ query time for $O(n^{d/2} + k)$
storage, where $n$ is the number of points and $k$ is the
size of the output; the second method has a query time
of roughly $O(n^{1-1/(d/2)})$ for linear storage. Two gen-
eral techniques, filtering search and fractional cas-
cading, have proven quite useful in speeding up range reporting
algorithms. The first one, developed by the author [52],
exploits the simple idea that if the output is large then
the algorithm can slow down proportionately, since the
output size can be used to amortize the search cost; this
in turn leads to savings in storage. The second tech-
nique, developed jointly with Guibas [69], deals with
the problem of searching for the same key in several
dictionaries at once: auxiliary dictionaries are created
that allow instant transitions from one dictionary to the
next.

In closing this section, I will mention the comprehen-
sive survey on range searching by Matoušek [190], which
I highly recommend.

4 Geometric Optimization

I now briefly discuss recent developments that have had
an important impact in geometric optimization. One is
parametric search; the other concerns fixed-dimensional
linear programming.

Parametric Search. Almost a decade ago, Megiddo
devised a beautiful algorithmic paradigm known as
parametric search [195, 196]. A clever improvement
which under some conditions speeds up the method was
proposed by Cole [93]. The main feature of parametric
search is to turn a checking algorithm for an optimiza-
tion problem into one that actually searches for a solu-
tion. Certain conditions must be satisfied, of course, but
the setting is general enough to have wide applicability.
For example, it can be used to solve the 2-center prob-
lem in the plane, selecting slopes or distances, comput-
ing the $k$th leftmost vertex in a line arrangement, per-
form ray-shooting in 3-space, etc. See, e.g., [4, 8, 65, 94].

One of the main conditions of Megiddo’s technique is
that the checking algorithm be parallelizable. This
condition can be somewhat weakened, but actually it
would appear that in a geometric setting the require-
ment should be reducible to nothing but a mild form of
multidimensional searching.

An interesting open problem is computing the diam-
ter of $n$ points in $E^3$ (i.e., the distance between its
two furthest points). Clarkson and Shor [88] gave a
randomized algorithm with an optimal expected complex-
ity of $O(n \log n)$. Attempts to derandomize it have
been unsuccessful so far. Building on previous work
[58, 65, 191, 235], Brönnimann, Chazelle, and Matoušek
[42] gave an $O(n \log^2 n)$-time deterministic solution. It
is currently the best known but it is unlikely to be opti-
mal. The algorithm uses parametric search as well as
the entire derandomization machinery of [58]: one won-
ders whether the diameter problem really requires all
these hybrid complicated techniques.

We close our discussion of parametric searching by
mentioning an important generalization of the method,
where the parameter to optimize is not a single value but
a point in higher-dimensional space [90, 96, 189, 221].

Linear Programming. In 1984 Megiddo [197] gave
the first linear-time algorithm for linear programming
with a fixed number of variables. The running time of
his algorithm is $O(c_d n)$, with $c_d = 2^d$, where $n$ is
the number of constraints and $d$ is the number of variables
(the dimension of the ambient space). The constant $c_d$
was improved to $3^d$ by Clarkson [80] and Dyer [111].
Still lower dependency on $d$ was achieved by randomized
methods (Dyer and Frieze [113], Clarkson [83], Seidel
[246]), the current best being Clarkson’s algorithm, with
an expected complexity of $O(d^2 n + d^{1/2 + O(1)} \log n)$. Kalai [171] and then independently Matoušek, Sharir, and Welzl [193] developed algorithms with a subexponential dependency on both $n$ and $d$. These (dual) simplex-like methods rekindle the hope that a strongly polynomial-time algorithm for general linear programming might exist. Combining these methods with Clarkson's yields a randomized algorithm for linear programming with expected running time $O(d^2 n + e^{O(\sqrt{d \ln d})})$. To derandomize these algorithms appears far from obvious. The best current deterministic solutions for linear programming in fixed dimension have complexity $d^{O(d)} n$ [9, 70].

An abstract framework for optimization problems similar to linear programming but much more general was introduced by Sharir and Welzl [253] and Matoušek, Sharir, and Welzl [193]. Similar techniques to those used in the solutions we mentioned above can be made to bear on these $LP$-type problems. Gártner [137] was able to derive a subexponential method for a subclass of these problems. Similarly, the method of Chazelle and Matoušek [70] can be used to solve in linear deterministic time problems such as computing the minimum-volume ellipsoid enclosing a set of points or finding the maximum volume ellipsoid within the feasible set of linear constraints. Solutions with higher dependency on the dimension were found earlier by Dyer [112].

5 Lower Bounds

Sadly enough, surveying the status of lower bounds in computational geometry is a fairly easy task. As long as the information-theoretic bound is good enough, there are nice tools at one's disposal. A proof technique by Yao [289] for proving that any algebraic decision tree for computing the convex hull of $n$ points in the plane must have depth $\Omega(n \log n)$ was generalized by Ben-Or [29] and can be applied to many problems. The case of integer inputs was treated by Yao [270].

If the information-theoretic bound is too weak, however, we are seriously out of luck. Range searching is one of the few exceptions to this rule, because it has a nice combinatorial characterization. In a range searching problem, think of each point as a variable over a semigroup. The answer to a query is the semigroup sum of the variables associated with the points falling in the query range.

In that model, originally introduced by Fredman and Yao, remarkably strong lower bounds can be proven. Fredman established optimal lower bounds for dynamic orthogonal range searching [134] and quasi-optimal ones for dynamic halfplane range searching [135]. Yao [288] established a lower bound for orthogonal range searching in two dimensions, while for arbitrary dimensions Vaidya [259] provided nontrivial but suboptimal lower bounds. The author obtained quasi-optimal lower bounds for orthogonal range reporting, orthogonal range searching, and simplex range searching [54, 53]. Other lower bounds were established by the author jointly with Rosenberg [73] (simplex range reporting) and Brönnimann and Pach [43] (halfspace range searching).

Unfortunately, if we allow subtractions, i.e., if we operate over a group as opposed to a semigroup, all these lower bounds collapse miserably. This is very similar to the difficulty of proving lower bounds for general circuits as opposed to monotone circuits. New ideas seem to be required. At this point, any nontrivial lower bound for any nontrivial (constructive) instance of range searching in the group model would be a major result.

The second challenge is to handle computational as opposed to combinatorial models. For example, none of the current methods can be used to derive nontrivial lower bounds for Hopcroft's problem: this is the problem of deciding whether, among $n$ lines and $n$ points, all the points are disjoint from all the lines. Several solutions exist that require roughly $n^{4/3}$ time, the best of which is due to Matoušek [186], but $\Omega(n \log n)$ is the only existing lower bound.

Problems of that sort can be formalized as follows: Given a $k$-variate polynomial $P$ of fixed degree with rational coefficients, and given $n$ numbers $x_1, \ldots, x_n$, is there a $k$-tuple among the $x_i$'s that is in the zero set of $P$? There exists a trivial $O(n^k)$ solution, to which slight improvements can be made by using fairly complicated batching techniques. But again the best known lower bound is only on the order of $n \log n$.

Björner, Lovász, and Yao [38] have derived lower bounds (for nongeometric problems) that make novel use of topological invariants and might provide useful tools. However, these tools in themselves are still inherently information-theoretic and therefore unlikely to break the $n \log n$ barrier. Erickson and Seidel [125] have proved elegant lower bounds for problems such as testing if a set of points is in general position, but their model of computation is far too restrictive to give much hope that a satisfactory solution should be anywhere in sight.

6 Randomization and Derandomization

Randomization has had a tremendously positive effect in computational geometry. Not only has it led to a host of fairly simple, efficient probabilistic algorithms, but it has led to a complete rethinking in the design of deterministic algorithms through the process of derandomization. Perhaps this phenomenon is best illustrated by
the convex hull problem discussed in Section 2.

After several years without noticeable progress on the problem, Clarkson and Shor [88] produced an optimal randomized algorithm. Around the same time through the efforts of several researchers, in particular Matoušek [183, 184, 185, 186], a beautiful theory of deterministic sampling in range spaces of finite VC-dimension was being developed. Also, general derandomization techniques began to emerge, such as the method of conditional probabilities of Raghavan and Spencer [234, 254]. It was only with the help of all these powerful tools that the author was able to derive an optimal convex hull algorithm [58]. It is one of a growing number of deterministic algorithms that literally cannot be understood without referring to their probabilistic counterparts. Unfortunately, in this case, derandomization exacts a hefty price, as the algorithm (and especially, its analysis) are quite complicated. If history is a guide, however, one should expect much simpler versions to emerge in the near future.

One of the main ingredients of the algorithm is a method of deterministic Monte Carlo integration, using ε-nets and ε-approximations (see definitions below) as tools for estimating functions that are too hard to evaluate exactly. It is a fascinating open question whether this can be extended to other problems, such as higher-order Voronoi diagrams or even line segment intersection.

The papers of Clarkson [81], Clarkson and Shor [88], and Haussler and Welzl [160], all concerned with probabilistic geometric algorithms, proved to be enormously influential. Two main classes of algorithms emerged: some use divide-and-conquer via random sampling; others, called randomized incremental constructions, use insertions in random order. For a small random sample of points relevant to the first class, see [81, 82, 87, 88, 90, 116, 117, 160, 230]. I also recommend the surveys by Clarkson [89] and Seidel [248], as well as Mulmuley's textbook [219] for excellent introductions to the use of random sampling in computational geometry.

A key notion in the analysis of probabilistic geometric algorithms is that of range spaces of finite VC-dimension. Let \( \Sigma = (X, \mathcal{R}) \) be a set system. Given \( Y \subseteq X \), let \( \mathcal{R}|_Y \) denote the set system \( \{ R \cap Y; R \in \mathcal{R} \} \).

We say that \( Y \) is shattered (by \( \mathcal{R} \)) if \( \mathcal{R}|_Y = 2^Y \). The maximum size of a shattered subset of \( X \) is called the \textit{Vapnik-Chervonenkis dimension}, or VC-dimension, for short, of the set system \( \Sigma \). Given \( \varepsilon \in (0, 1) \), a subset \( A \subseteq X \) is an \( \varepsilon \)-approximation for \( \Sigma \), if

\[
\left| \frac{|A \cap R|}{|A|} - \frac{|R|}{|X|} \right| \leq \varepsilon
\]

for every set \( R \in \mathcal{R} \). A subset \( S \subseteq X \) is called an \( \varepsilon \)-net for \( (X, \mathcal{R}) \) provided that \( S \cap R \neq \emptyset \) for every set \( R \in \mathcal{R} \) with \( |R|/|X| > \varepsilon \). Remarkably, as long as it has finite VC-dimension, \( \Sigma \) admits \( \varepsilon \)-approximations and \( \varepsilon \)-nets whose sizes are independent of \( |X| \) (Vapnik and Chervonenkis [260], Haussler and Welzl [160]).

The question of constructing \( \varepsilon \)-nets deterministically was investigated by Chazelle and Friedman [67], who gave polynomial-time algorithms for computing an \( \varepsilon \)-net of size \( O(\varepsilon^{-1} \log \varepsilon^{-1}) \). A major advance was subsequently made by Matoušek [185], who showed that under some natural computational assumptions about the set system \( \Sigma \), it is possible to compute \( \varepsilon \)-approximations and \( \varepsilon \)-nets of size \( O(\varepsilon^{-2} \log \varepsilon^{-1}) \) and \( O(\varepsilon^{-1} \log \varepsilon^{-1}) \), respectively, in time \( O(\varepsilon^{-2d} \log^d \varepsilon^{-1}) |X| \). A simpler proof was given by Chazelle and Matoušek [70] and slightly better running times were achieved by Brönnimann, Chazelle, and Matoušek [42] by introducing the notion of sensitive \( \varepsilon \)-approximations.

With the use of these tools (among others) derandomizing divide-and-conquer type algorithms has been very successful. But lately, the attention has shifted towards randomized incremental constructions [40, 41, 66, 87, 88, 90, 102, 103, 153, 210, 211, 212, 213, 214, 246, 247]. These algorithms examine the input elements in random order and at each step maintain a partial solution based on the portion of input examined so far. This approach has produced surprisingly simple Las Vegas algorithms of low expected complexity. Unfortunately, their derandomization has proven elusive so far. Actually, the author's convex hull algorithm [58] is currently the only successful derandomization of a randomized incremental construction.

To understand the computational power of randomness has been a central preoccupation in theoretical computer science. While only few general derandomization techniques have been discovered [12, 15, 182, 234, 254], several clever schemes have been found to reduce the amount of randomness needed by probabilistic algorithms [14, 28, 32, 33, 170, 208, 220].

As was shown by Mulmuley [218], such techniques can be used to prove that most randomized geometric algorithms require few (truly) random bits; but still too many, however, to allow efficient deterministic simulations. Interesting geometric use of expanders (which act as quasi-random graphs) has been made by Ajtai and Megiddo [11] and Katz and Sharir [174].

Matoušek, Welzl, and Wernisch [194] established that the effectiveness of \( \varepsilon \)-nets and \( \varepsilon \)-approximations for divide-and-conquer in computational geometry is closely tied to the discrepancy of certain set systems. A better understanding of the relationship between discrepancy and sampling might hold the key to understanding what makes derandomization feasible or not.
7 Computational Real-Algebraic Geometry

Planning the motion of a robot amidst obstacles usually boils down to asking connectivity questions about semi-algebraic sets. Although reasonably efficient methods are already known for answering such questions, the larger picture is still blurred. For example, one would often like to navigate in configuration space in real-time, perform point location and ray-shooting, modify the constraints on-line, etc. Also, it might be desirable to obtain more information about the cycle structure besides path-connectivity: for example, one might want to compute the full homology of a cell decomposition (see next section). Although much progress has been done in that general area since Collins' pioneering work [97], e.g., Schwartz and Sharir [242], Canny [44], two central questions remain open: (1) triangulating real-algebraic varieties in singly-exponential time, and (2) keeping polynomial coefficients below doubly exponential. Here is a brief explanation:

Given $n$ polynomials in $\mathbb{Q}[x_1, \ldots, x_d]$, we wish to partition $\mathbb{R}^d$ into simple cells over which each polynomial remains sign-invariant. A simple cell could be, for example, a smooth manifold (i) topologically equivalent to a $k$-ball ($k \leq d$) and (ii) specified by a constant-size sentence in the theory of the reals. This is the motivation behind a particular type of decomposition known as a stratification [44, 64, 74, 97, 233, 242, 256]. Property (ii) is used to ensure that the cells of the decomposition are computationally “tractable.”

Informally a triangulation is a special case of a stratification where adjacent faces glue “nicely” to each other, i.e., in a lattice-like fashion. This added property is crucial for, say, navigating in configuration space, e.g., moving a robot amidst obstacles: in configuration space the robot is a point in a space of dimension equal to its degree of freedom and the obstacles are modeled by (usually) algebraic constraints.

As shown by Chazelle and Sharir [74], an economical sign-invariant stratification yields more efficient algorithms for problems as diverse as computing the longest segment fitting inside a polygon, finding the minimum vertical separation between two sets of segments in $\mathbb{R}^2$, computing the first time the convex hull of a set of moving objects reaches its stationary configuration, etc.

Another motivation for studying this problem is that its language is powerful enough to express any multi-dimensional searching problem expressed as a first-order predicate in the theory of reals.

An effective algorithm for constructing a sign-invariant stratification of size roughly $O(n^{2d-5})$ was given by Chazelle et al. [64]. Although the algorithm represents an attractive alternative to Collins' decomposition (whose size is doubly exponential in $d$), it raises three major open questions: One is to reduce the size to $O(n^d)$, which by the Thom-Milnor bound, would then be optimal. (We omit to mention the dependency on the size and degree of the polynomials.) Another is to reduce the complexity of the algebraic numbers used to describe the cells from doubly to singly exponential (in the dimension). Although it is known that eliminating quantifiers is inherently doubly exponential (Davenport and Heintz [100], Weispfenning [263]), many restricted problems related to the theory of the reals can be solved in singly-exponential time.

The mathematical landscape in real-algebraic geometry has changed dramatically in the last few years. An eminently readable survey by Heintz, Recio and Roy is given in [163]. A good introduction to the (non-computational part of the) field is the text (in French) by Bochnak, Coste, and Roy [39]. Let us mention a few examples of the recent progress. A very useful lemma by Grigor’ev and Vorobjov [152] gives an upper bound on the size of a ball about the origin that misses any component of a semi-algebraic set (as a function of the number of polynomials, the maximum degree, the coefficient sizes, and the number of variables). This was an important tool used by Goodman, Pollack, and Sturmfels [142] to give sharp estimates on how large a grid must be if it is to accommodate all point configurations of a given size (up to order type).

Pollack and Roy [228] showed that the number of connected components of all non-empty realizations of sign conditions of $n$ polynomials in $d$ variables, each of degree at most $b$, is bounded by $O(nb/d)^d$, thus improving on an earlier bound based on a result of Heintz [161]. (Note that the coefficients need not be in $\mathbb{Q}$ but in any real-closed field.) See also Basu, Pollack, and Roy [27] for a more general result concerning the cross-section of a semi-algebraic set with a given variety.

In [26] Basu, Pollack, and Roy give an algorithm that returns a small set containing an algebraic point in each connected component of each non-empty sign condition in time $n(n/d)^dO(d)$—see also Canny [47, 48, 49]; this improves on the $(nb)^O(d)$ algorithms of Canny [45] and Renegar [237].

Since Collins' work [97], important results in the general (i.e., fully quantified) first-order theory of the reals have been obtained. The works of Ben-Or, Kozen, and Reif [30] and Fitchas, Galligo, and Morgenstern [126] led to an interesting arithmetic-operation counterpart to Collins' bit-operation construction. Grigor’ev [151] discovered a method that is doubly exponential only in the number of quantifier alternations. The best complexity in both the arithmetic and bit models was achieved by Renegar [237]. Another quantifier elimination method was also given by Heintz, Roy, and Solerno [162]. Further improvements over both methods are reported by
Basu, Pollack, and Roy [25].

Somewhat outside the scope of this discussion, the theory of Groebner bases ought to be at least mentioned, because it is a very active area in computational algebraic geometry with many applications in geometric modeling. A good introduction can be found in the book by Cox, Little, and O'Shea [98]. A more in-depth treatment is provided by Mishra [206].

8 Computational Topology

The purpose of the subject is to study computational questions regarding the classical topological invariants, e.g., fundamental group, homology groups. Some of these problems, like the word problem for automatic groups, are very interesting from an algorithmic and data structuring viewpoint, but they are not germane to computational geometry. It is a different matter with homology theory, however. Computing homology groups can be done by algebraic methods, but it is often highly inefficient. Exploiting the specific geometry of the manifolds can lead to much faster algorithms. Similarly, checking homotopy equivalence between curves or homeomorphisms between surfaces can gain a lot from a computational-geometric approach.

Homology Groups. The connectivity properties of a geometric structure can reveal a lot of information about it, but they are usually much more subtle and difficult to compute than for combinatorial structures. The homology groups describe the cycle structure of a manifold. Their main advantage over the fundamental group or higher-dimensional homotopy groups is that they lend themselves to effective computations. It would be extremely useful to be able to compute homology groups efficiently for low-dimensional geometric cell complexes. There are many important applications in pattern recognition and classification in biology and chemistry as well as in robotics and scene analysis [99]. Comparing homologies is a good (if not foolproof) test to rule out topological equivalence: two manifolds with different homologies cannot be homeomorphic; unfortunately, the converse is not true.

There are also fascinating applications to the classification of time series and the analysis of dynamical systems (Freedman [136]). Spectral analysis has long been the standard tool for classifying time series. In the absence of sharp harmonics, the signal is dismissed as noise. Analysis for nonlinear dynamics can be done, however, by computing topological invariants of the attractor set for the system being measured. This can be done by mapping the time series into \( \mathbb{R}^d \) by scanning a “comb” of length \( d \) across the series and thinking of each \( d \)-tuple of entries at the teeth as a point in \( \mathbb{R}^d \). (This is the time-lag imbedding of the series; in practice the imbedding dimension \( d \) is often less than 10.) By grouping all \((k + 1)\)-tuples of points at a distance less than \( \varepsilon \) from one another, we can define \( k \)-simplices. By introducing the obvious boundary operators, this yields a complex whose \( \varepsilon \)-homology is called the \( \varepsilon \)-homology of the set. Because \( \varepsilon \)-homology commutes with inverse limits, this is an appropriate methods for computing the homology of the attractor. Moreover the use of \( \varepsilon \)-homology (where simplices are defined only in terms of pairwise distances) should have computational advantages. A similar notion, called alpha-shape, has been developed by Edelsbrunner. This is a very interesting area to explore where computational-geometric methods can be brought to bear.

Schwartz and Sharir [242] gave a general method for computing the homology of a real-algebraic variety, but unfortunately it is doubly exponential in the dimension of the ambient space. More recently, Donald and Chang [109] proposed a randomized algorithm for computing the homology groups of a cell complex, which takes advantage of the sparsity of its incidence structure. By and large, however, the problem has been treated in full generality, with little focus on the specifically geometric nature of the complex. A notable exception is the work of Delfinado and Edelsbrunner [101], where a nearly optimal method for computing the betti numbers of a subcomplex of a triangulation of \( S^3 \) is given. (Recall that these particular homology groups are free, i.e., without torsion, so they are specified entirely by the betti numbers). This result has interesting applications to the so-called alpha-shapes mentioned earlier, which are generalizations of convex hulls used in geometric modeling.

Mayer-Vietoris sequences should be useful in designing efficient divide-and-conquer schemes for computing the homology of higher-dimensional complexes [101, 168, 239]. This is an exciting open area of research. The goal is, in a somewhat narrow sense, to do computational linear algebra by geometric means.

Edelsbrunner [115] has given efficient algorithms for computing topological, combinatorial, and metric properties of the union of finitely many balls in \( \mathbb{R}^d \). These algorithms are based on short inclusion-exclusion formulas derived from a complex dual to a decomposition of the balls. These methods have applications to biology and chemistry, where unions of balls are commonly used as models of molecules. Further work on triangulating topological spaces has been reported by Edelsbrunner and Shah [123].

Topology of Curves. Interesting questions in two-dimensional computational topology have been investigated recently. Mehlhorn and Yap [200] and Vegter [261] considered an algorithmic discrete version of
the Whitney-Graustein theorem, which states that up to kink-free deformations a regular closed curve in the plane is completely determined by its winding number. In particular, Vegter gives a linear algorithm for deforming a regular polygon into any other regular polygon with the same winding number.

Motivated by the desire to be able to decide quickly whether two surfaces are homeomorphic or whether curves on them are homotopic, Vegter and Yap [262] looked at the classical question of putting in canonical form the polygonal schema of a closed triangulated surface. They showed how to implement the Brahana procedure (of the Dehn-Reeborg classification theorem) in $O(n \log n)$ time, where $n$ is the size of the triangulation.

In [241] Schipper presented an algorithm for deciding whether a curve on a triangulated surface is null-homotopic, i.e., contractible to a point. A similar question regarding the detection of null-homologous cycles on surfaces was investigated by Dey [105]. See also [258].

9 Robustness and Degeneracies

There are two kinds of numerical difficulties in implementing geometric algorithms: finite precision causes round-off errors or inconsistent results, while often more brutal, degeneracies or near-degeneracies cause programs to crash. A program is usually termed robust if it is immune to these ills (as much as possible). A growing research effort is underway to provide the programmer and algorithm designer with tools, guidelines, and options to deal with these difficulties. Fortune has an excellent survey [131] on the subject to which the reader is referred. Hoffmann’s book [165] is also a good source. I will briefly sketch some of the issues and lines of research in that area.

Finite Precision. A typical atomic geometric operation is to determine whether three points make a left turn or if a fourth one lies inside the circle passing through the first three. In general, there is a certain determinant whose sign must be determined. An interesting approach for doing that is given by Clarkson [84].

If we must use exact representation, one option is to use modular arithmetic. As Fortune notes in [131], fairly little seems to be known in a specifically geometric context. This is an interesting subject to explore.

Rational arithmetic is another option. Again, the absence of error is the major drawing card. An obvious problem, however, is that the representation length grows rather quickly. Karasick, Lieber, and Nackman [172] report some cases where appropriate optimization and careful use of interval arithmetic result in running times that are competitive with floating-point computation.

How general these techniques are, however, remains unclear.

Yap [275] has argued that exact geometric computation need not be hopelessly costly, and that the answer might be to enhance traditional BigNumber packages and customize them to geometric situations. See also Fortune and Van Wyk [133].

Away from exact computation, another route, followed by Greene and Yao [150], is to round newly computed points to the nearest point on a fine enough lattice. The usual incidence laws no longer apply, however, and to preserve some measure of consistency is no easy task.

A variant on the Greene-Yao rounding idea that increases its flexibility involves “fattening” the lines. Building on this idea, Milenkovic [204, 205] was able to develop low bit-complexity algorithms for fundamental geometric problems in higher dimensions. See also Milenkovic [203].

Sugihara and Iri [256] considered the problem of maintaining low-bit complexity in the face of solid-modeling operations: these are often nested computations which tend to blow up representation lengths.

In practice the user sometimes has no choice but to use a floating point representation. A standard approach is to add a third truth value, “don’t-know”, to the logic. Variants of interval analysis techniques (so-called epsilon-geometry) were used by Guibas, Salesin, and Stolfi [155]. Dobkin and Silver [107] suggested adding the use of randomization and sampling.

Backward analysis is another standard error analysis technique which has been adapted with success to some geometric situations. Fortune [128] redesigned the Graham scan method for computing planar convex Hulls, so that the output is the true convex hull of an appropriate perturbation of the input points. A variant is described by Li and Milenkovic [179]. The complexity of computing a line arrangement robustly (by stretching lines into pseudolines when necessary) was analyzed by Milenkovic [201, 202] and Fortune and Milenkovic [132].

Other interesting work on maintaining consistency and small error using floating-point arithmetic appears in (Sugihara and Iri [255], Hoffmann, Hopcroft, and Karasick [166, 167], Karasick [173], Fortune [129]). The problem of performing rotations was addressed by Canny, Donald, and Ressler in [50].

Degeneracies. Perturbing the input data ever so slightly can sometimes ensure that it is in general position. As has long been known in linear programming, the perturbation can be made symbolic. This approach was adapted to a geometric context by Edelsbrunner and Mücke [120] and Yap [273, 274].

The overhead in using these techniques can be quite high. Emiris and Canny have proposed schemes to alle-
viate the cost of computing long sequences of determinants [124]. Two problems inherent to these methods are the lack of control over the perturbation and the lack of discrimination between coincidental and mathematically implied degeneracies; e.g., three input points might or might not be collinear, but they always are if one is the midpoint of the other two. We close this brief overview by mentioning an interesting investigation of perturbations techniques by Seidel [249].

10 Combinatorial Geometry

I will briefly mention a few developments in combinatorial geometry that have had a considerable impact in the analysis of geometric algorithms.

Thanks to the effort of several researchers, (notably Sharir), the theory of Davenport-Schinzel sequences is now fairly advanced [7, 159]. We omit the discussion of this subject and instead refer the reader to the forthcoming book by Sharir and Agarwal [252]. There have been numerous applications to motion planning, triangulation, optimization, etc.

Zone theorems have also played a prominent role lately. Given an arrangement of \( n \) hyperplanes in \( \mathbb{R}^d \), the zone complexity of any hyperplane \( h \) is the total size of all the cells that \( h \) intersects. Edelsbrunner, Seidel, and Sharir [122] showed that the maximum zone complexity is \( O(n^{d-1}) \). Aronov, Pellegrini, and Sharir [18] generalized the proof technique to the case of the zone of an algebraic surface, while Aronov, Matoušek, and Sharir [17] bounded the mean square cell complexity.

A related question is to bound the sum of the complexities of \( m \) cells in an arrangement of \( n \) hyperplanes in \( \mathbb{R}^d \). In the two-dimensional case, Clarkson et al. [89] prove that this complexity is at most \( O(n + m + m^{2/3}n^{2/3}) \), which is tight. An efficient algorithm for constructing the corresponding cells is given by Edelsbrunner, Guibas, and Sharir [117]. The higher-dimensional case is treated by Edelsbrunner, Guibas, and Sharir [118], Aronov, Matoušek, and Sharir [17], and Agarwal and Aronov [3].

In problems of placement or Voronoi diagrams-related questions, it is important to have estimates on the combinatorial complexity of the lower envelope of an arrangement of hyperplanes or surfaces: this is the locus of points visible from a point at infinity. This problem is easier than the zone problem, but unfortunately the cases of interest are those for which no zone theorems are available. However, Sharir [251] recently provided a quasi-optimal upper bound of \( O(n^{d-1+\varepsilon}) \), for any fixed \( \varepsilon > 0 \), on the size of the lower envelope of \( n \) surface patches in \( \mathbb{R}^d \) bounded by algebraic surfaces of constant maximum degree. Specializations to three dimensions were investigated earlier by Halperin and Sharir [157, 158]. Many applications are discussed in [252].

We conclude by returning to geometric sampling and mentioning an intriguing extension of \( \varepsilon \)-nets that was proposed by Alon et al. [13]. Given a finite point set \( P \), they define a weak \( \varepsilon \)-net (for convex sets) to be a collection \( S \) of points such that any convex set that contains a fraction \( \varepsilon \) of the points in \( P \) must intersect \( S \). Alon et al. proved that although the underlying range space does not have finite VC-dimension, such a set \( S \) exists (often outside of \( P \)) such that the size \( |S| \) is polynomial in \( 1/\varepsilon \) and, most important, independent of \( |P| \). The bound was slightly improved by Chazelle et al. [62] for the general case, and optimal bounds were found for special cases by using tilings of the hyperbolic plane. Nevertheless, the problem is still by and large wide open.

We close this section by mentioning the nice survey by Guibas and Sharir [156] and the forthcoming text on combinatorial geometry by Pach and Agarwal [224].

11 Concluding Remarks

Computational geometry is more active and diverse than ever. I have not touched on vast areas of research, e.g., motion planning [243], assembly, kinodynamics [110, 108], Groebner bases [98], transversal theory [143], parallel computational geometry [10, 145], hidden surface removal [146, 271], art-gallery theorems [222], Borsuk-Ulam type search (ham-sandwich cuts, centerpoints) [181], arrangements and configurations [140, 141], \( k \)-set theory [225], mesh generation [36, 37], geometric problems in computer vision [16, 169].

There exist several introductory textbooks in computational geometry: Edelsbrunner [114], Mehlinrn [198], Mulmuley [219], O'Rourke [223], and Shamos and Preparata [230].

Today, the field covers a pretty wide front and is still rapidly expanding. There is no way to tell what it will look like in the year 2000. Whatever the new millenium has in store for computational geometry, however, I am sure that it will continue to be vibrant and fun!

Acknowledgments

I wish to thank Pankaj Agarwal, Hervé Brönnimann, Herbert Edelsbrunner, Jiřka Matoušek, Ricky Pollack, Marie-Françoise Roy, Micha Sharir, and Emo Welzl for reading a draft of this survey and giving me many useful comments.
References


[67] Chazelle, B., Friedman, J. A deterministic view of random sampling and its use in geometry, Combinatorica, 10 (1990), 229-249.


[197] Megiddo, N. Linear programming in linear time when the dimension is fixed, J. ACM, 31 (1984), 114–127.


[227] Pellegrini, M. On point location and motion planning among simplices, these proceedings.


[235] Ramos, E., Intersection of unit-balls and diameter of a set of points in \( \mathbb{R}^3 \), manuscript, 1994.


