Chapter 4

Random sampling

The *randomization* method has proved useful in computational geometry. This usefulness can be ascribed in large part to a few probabilistic theorems which rely on combinatorial properties of certain geometric problems. The probabilities involved in those theorems concern random samples from the set of data, and do not involve statistical assumptions about the distribution of these data.

The goal of this chapter is to present those probabilistic theorems on which the analysis of the randomized incremental method is based. This method is described in chapters 5 and 6. We express these theorems in a framework general enough to be adaptable to different geometric settings. All the randomized algorithms presented below fit into the same framework as the one that we define here.

The first part of this chapter recalls the necessary definitions and notation. The second part proves the basic two theorems: the sampling theorem, and the moment theorem. These theorems provide the main tools to analyze the average performance of randomized algorithms.

4.1 Definitions

4.1.1 Objects, regions, and conflicts

In the framework presented here, any geometric problem can be formulated in terms of objects, regions, and conflicts between these objects and regions.

Objects are elements of a universe $\mathcal{O}$, usually infinite. The input to some problem will be a set $\mathcal{S}$ of objects of $\mathcal{O}$. The objects under consideration are typically subsets of the Euclidean space $\mathbb{E}^d$ such as points, line segments, lines, half-planes, hyperplanes, half-spaces, etc.

A *region* is a member of a set $\mathcal{F}$ of regions. Each region is associated with two sets of objects: those that *determine* it, and those that *conflict* with it.

The set of objects that determine a region is a finite subset of $\mathcal{O}$, of cardinality bounded by some constant b. The constant b depends on the nature of the problem, but not on the actual instance nor on its size. This restriction is required for all the probabilistic theorems to be expressed within the framework.

The set of objects that conflict with a given region is usually infinite and is called the *domain of influence* of the region.

Let $\mathcal{S}$ be a set of objects. A region F of $\mathcal{F}$ is *defined over* $\mathcal{S}$ if the set of objects that determines it is contained in $\mathcal{S}$. A region F is said to be *without conflict over* $\mathcal{S}$ if its domain of influence contains no member of $\mathcal{S}$, and otherwise is said to *have j conflicts over* $\mathcal{S}$ if its domain of influence contains j objects of $\mathcal{S}$.

For each geometric application, the notions of objects, regions, and conflicts are defined in such a way that the problem is equivalent to finding all the regions defined and without conflict over $\mathcal{S}$.

Let us immediately discuss a concrete example. Let $\mathcal{S}$ be a set of n points in the d-dimensional Euclidean space $\mathbb{E}^d$. The convex hull of $\mathcal{S}$ is the smallest convex set containing $\mathcal{S}$; suppose we wish to compute it. Assume the points are in general position[1]. The convex hull $conv(\mathcal{S})$ is a polytope whose special properties will be studied further in chapter 7. For now, it suffices to notice that, in order to compute the convex hull, we have to find all the subsets of d points in $\mathcal{S}$ such that one of the half-spaces bounded by the hyperplane passing through these d points contains no other point that belong to $\mathcal{S}$ (see figure 4.1). In this example, the objects are points, and the regions are open half-spaces in $\mathbb{E}^d$. Every set of d points determines two regions: the open half-spaces whose boundaries are the hyperplane passing through these points. A point is in conflict with a half-space if it lies inside it. To find the convex hull, one must find all the regions determined by points of $\mathcal{S}$ and without conflict over $\mathcal{S}$.

The preceding definitions call for a few comments.

Remark 1. A region is determined by a finite and bounded number of objects and this restriction is the only fundamental condition that objects, regions, and conflicts must satisfy. Nevertheless, we do not demand that all the regions be determined by exactly the same number of objects. In the case of the convex hull of n points in $\mathbb{E}^d$, all the regions are determined by exactly d points. One may envision other settings (as in the case of the vertical decomposition of a set a line segments in the plane, discussed in subsection 5.2.2), where the regions can be determined by a variable number i of objects, provided that $1 \le i \le b$ for some constant b.

Remark 2. A region does not conflict with the objects that determine it. This

[1]A set of points is in general position if every subset of $k+1 \le d+1$ points is affinely independent, or in other words if it generates an affine subspace of dimension k.

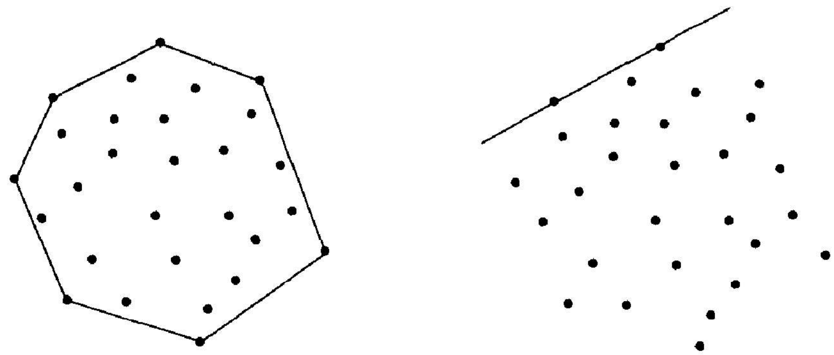

Figure 4.1. Convex hull and empty half-spaces.

simple convention greatly simplifies the statements and proofs of the theorems below, and does not modify their meaning. In the case of the convex hull, this can be easily achieved by defining the domain of influence of a region as an *open* half-space.

Remark 3. A region is characterized by two sets of objects: the set of objects that determine it, and the set of objects that conflict with it. Regions determined by different objects will be considered as different, even if they share the same domain of influence. In this context, a set $\mathcal{S}$ of objects is in general position precisely if any two regions determined by different subsets of $\mathcal{S}$ have distinct domains of influence.

Remark 4. A set of b or fewer objects may determine one, or more, or zero regions. Usually, the number of regions determined by a given set of (less than b) objects is bounded by a constant. For instance, in the case of convex hulls, every subset of d points determines exactly two regions. In this case, the total number of regions defined over a set of cardinality n is $O(n^b)$.

If $\mathcal{S}$ is a finite set of objects, say with n elements, we denote by $\mathcal{F}(\mathcal{S})$ the set of regions defined over $\mathcal{S}$ and, for each integer j in $[0, n]$, we denote by $\mathcal{F}_j(\mathcal{S})$ the set of all regions defined over $\mathcal{S}$ that have j conflicts over $\mathcal{S}$. In particular, $\mathcal{F}_0(\mathcal{S})$ is the set of those regions that are defined over $\mathcal{S}$ and without conflict over $\mathcal{S}$. Furthermore, we denote by $\mathcal{F}_{\leq k}(\mathcal{S})$ the subset of regions defined over $\mathcal{S}$ that have at most k conflicts over $\mathcal{S}$.

When the regions are determined by a variable number i of objects ($1 \leq i \leq b$), the preceding notation may be refined to denote by $\mathcal{F}^i_j(\mathcal{S})$, $\mathcal{F}^i_{\leq k}(\mathcal{S})$, $\mathcal{F}^i_{\geq k}(\mathcal{S})$, the subsets of those regions defined by exactly i objects of $\mathcal{S}$, with (respectively) exactly, at most, at least, k conflicts with the objects of $\mathcal{S}$.

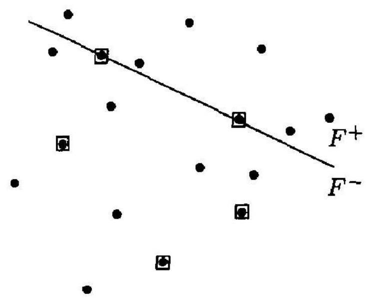

Figure 4.2. Instances of regions.

4.1.2 Random sampling

Let $\mathcal{R}$ be a subset of $\mathcal{S}$ with cardinality r. This subset $\mathcal{R}$ is a random sample of $\mathcal{S}$ if its elements are randomly chosen among all those of $\mathcal{S}$, such that each subset is equally likely to be chosen with probability $1/\binom{n}{r}$. In what follows, we shall call such a subset a *random r-sample* of the set $\mathcal{S}$.

The notation defined in the previous subsection is valid over any subset $\mathcal{R}$ of $\mathcal{S}$. In particular, $\mathcal{F}(\mathcal{R})$ is the set of regions defined over $\mathcal{R}$, $\mathcal{F}_j(\mathcal{R})$ is the set of regions defined over $\mathcal{R}$ that have j conflicts over $\mathcal{R}$, and $\mathcal{F}_j^i(\mathcal{R})$ is the set of of regions defined by exactly i objects of $\mathcal{R}$ that have j conflicts over $\mathcal{R}$. Since we may also be interested in the conflicts over $\mathcal{S}$ of a region defined over $\mathcal{R}$, or the converse, we will avoid any ambiguities by setting up a special terminology. Henceforth, by a *region defined and without conflict over* $\mathcal{R}$, we shall mean a region defined over $\mathcal{R}$ and without conflict over $\mathcal{R}$; these are the regions of $\mathcal{F}_0(\mathcal{R})$. Likewise, a *region defined and with j conflicts over* $\mathcal{R}$ is a region defined over $\mathcal{R}$ and that has j conflicts over $\mathcal{R}$; these are the regions of $\mathcal{F}_j(\mathcal{R})$.

In figure 4.2, the points of the subset $\mathcal{R}$ are enclosed by squares, the half-space F^+ belongs to $\mathcal{F}_6(\mathcal{S})$ and to $\mathcal{F}_0(\mathcal{R})$, while F^- belongs to $\mathcal{F}_{10}(\mathcal{S})$ and to $\mathcal{F}_3(\mathcal{R})$.

From now on, we are primarily interested in the regions defined over a random sample $\mathcal{R}$ from $\mathcal{S}$. Generally speaking, if $g(\mathcal{R})$ is a function of the sample $\mathcal{R}$, we denote by $g(r, \mathcal{S})$ the expected value of $g(\mathcal{R})$ for a random r-sample of $\mathcal{S}$. In particular, the following functions are defined: We denote by $f_j(\mathcal{R})$ the number of regions defined and with j conflicts over a subset $\mathcal{R}$ of $\mathcal{S}$ (in mathematical notation, $f_j(\mathcal{R}) = |\mathcal{F}_j(\mathcal{R})|$). Following our convention, $f_j(r, \mathcal{S})$ denotes the expected number of regions defined and with j conflicts over a random r-sample of $\mathcal{S}$. Likewise, $f_j^i(\mathcal{R})$ stands for the number of regions defined by i objects of $\mathcal{R}$ and with j conflicts over $\mathcal{R}$ (in mathematical notation, $f_j^i(\mathcal{R}) = |\mathcal{F}_j^i(\mathcal{R})|$). Then $f_j^i(r, \mathcal{S})$ is the expected number of such regions for a random r-sample of $\mathcal{S}$.

4.2 Probabilistic theorems

In this section, we prove two probabilistic theorems, the sampling theorem and the moment theorem. These two theorems lay the foundations for our analysis of randomized algorithms as described in chapters 5 and 6. The reader mostly interested in the algorithmic applications of these theorems may skip this section in a first reading. In order to understand the results, it would be enough to memorize the definition of a moment, to look up lemma 4.2.5, and to admit corollary 4.2.7.

The probabilistic theorems below are based on certain combinatorial properties of the geometric objects. The probabilities involved concern mainly random samples from the input data. In particular, these theorems do not make any assumptions on the statistical distribution of the input data. The theorems are stated in the formal framework introduced in the preceding section. Nevertheless, to shape the intuition of the reader, we start by stating them explicitly for the specific problem of computing the convex hull of a set of points in the plane.

Let $\mathcal{S}$ be a set of n points in the plane, assumed to be in general position, let k be an integer smaller than n and let $\mathcal{R}$ be a random sample of $\mathcal{S}$ of size $r = \lfloor n/k \rfloor$. The sampling theorem links the number of half-spaces defined over $\mathcal{S}$ and containing at most k points of $\mathcal{S}$, with the expected number of half-spaces defined and without conflict over $\mathcal{R}$, which is precisely the number of edges of the convex hull $conv(\mathcal{R})$. Let A and B be points of $\mathcal{S}$. Segment AB is an edge of the convex hull $conv(\mathcal{R})$ if and only if A and B are points of $\mathcal{R}$ and also one the half-planes H^+_{AB} and H^-_{AB} bounded by the line AB does not contain any points of $\mathcal{R}$. The *sampling theorem* relies on the fact that the segment AB joining two points of $\mathcal{S}$ is an edge of the convex hull $conv(\mathcal{R})$ with a probability that increases as the smallest number of points in either H^+_{AB} or H^-_{AB} decreases.

The *moment theorem* concerns the number of points in $\mathcal{S}$ and in its sample $\mathcal{R}$ that belong to some half-plane. If the size of $\mathcal{R}$ is large enough, the sample is representative of the whole set, and the number of points of $\mathcal{R}$ in a half-plane is roughly the number of points of $\mathcal{S}$ in this half-plane scaled by the appropriate factor r/n.

In fact, the moment theorem is a little more restrictive and concerns only those half-planes defined and without conflict over the sample. Any edge E of $conv(\mathcal{R})$ corresponds to a region defined and without conflict over $\mathcal{R}$: the half-plane $H^-(E)$ bounded by the line supporting E that contains no point of $\mathcal{R}$. The first moment of $\mathcal{R}$ relative to $\mathcal{S}$, or moment of order 1, is defined to be the sum, over all edges E of the convex hull $conv(\mathcal{R})$, of the number of points of $\mathcal{S}$ lying inside $H^-(E)$. In other words, the moment of order 1 of $\mathcal{R}$ with respect to $\mathcal{S}$ counts each point of $\mathcal{S} \setminus \mathcal{R}$ with a multiplicity equal to the number of edges of $conv(\mathcal{R})$ whose supporting lines separate it from $conv(\mathcal{R})$ itself. Figure 4.3

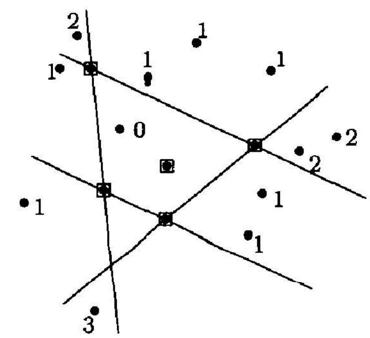

Figure 4.3. Moment of order 1.

indicates the multiplicity of each point, and the first-order moment of the sample is 16.

The moment theorem shows that, if the size of the sample is big enough, the expected moment of order 1 is at most $n - r$.

4.2.1 The sampling theorem

The sampling theorem yields an upper bound on the number of regions defined and with at most k conflicts over a set $\mathcal{S}$ of n elements. This bound depends on the expected number of regions defined and without conflict over a random $\lfloor n/k \rfloor$-sample of $\mathcal{S}$. The proof of this theorem relies on the simple idea that, the fewer objects in conflict with a region, the more likely this region is to have no conflict with a random sample $\mathcal{R}$ of $\mathcal{S}$. The proof uses the two fundamental lemmas below.

Lemma 4.2.1 *Let $\mathcal{S}$ be a set of n objects and F a region in conflict with j objects of $\mathcal{S}$ and determined by i objects of $\mathcal{S}$. If $\mathcal{R}$ is a r-sample of $\mathcal{S}$, the probability $p^i_{j,k}(r)$ that F be a region defined and with k conflicts over $\mathcal{R}$ is*

$$p^i_{j,k}(r) = \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}.$$

Proof. Let $\mathcal{R}$ be a random r-sample of $\mathcal{S}$. The region F of $\mathcal{F}^i_j(\mathcal{S})$ belongs to $\mathcal{F}^i_k(\mathcal{R})$ if it is determined by i objects in $\mathcal{R}$ and conflicts with k objects in $\mathcal{R}$. For this to be the case, the i objects determining F must be part of $\mathcal{R}$. The k objects of $\mathcal{R}$ conflicting with F must be chosen among the j objects of $\mathcal{S}$ that conflict with F. Finally, the $r - i - k$ remaining objects in $\mathcal{R}$ not in conflict with F must be chosen among the $n-i-j$ objects in $\mathcal{S}$ that do not determine or conflict with F. □

We denote by $p_j^i(r)$ the probability $p_{j,0}^i(r)$ that a region F of $\mathcal{F}_j^i(\mathcal{S})$ be defined and without conflict over a random r-sample of $\mathcal{S}$:

$$p_j^i(r) = \frac{\binom{n-i-j}{r-i}}{\binom{n}{r}}.$$

Lemma 4.2.2 *Let $\mathcal{S}$ be a set of n objects and $\mathcal{R}$ a random r-sample of $\mathcal{S}$. The expected number $f_k^i(r,\mathcal{S})$ of regions determined by i objects of $\mathcal{R}$ that conflict with k objects of $\mathcal{R}$ is given by the formula*

$$f_k^i(r,\mathcal{S}) = \sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}.$$

Proof. The expected number of regions in the set $\mathcal{F}_k^i(\mathcal{R})$ is the sum, over all the regions determined by i objects of $\mathcal{S}$, of the probability that this region belongs to the set $\mathcal{F}_k^i(\mathcal{R})$. This probability is given by the lemma 4.2.1 above. □

Theorem 4.2.3 (Sampling theorem) *Let $\mathcal{S}$ be a set of n objects and k an integer such that $2 \le k \le \frac{n}{b+1}$. Then*

$$|\mathcal{F}_{\le k}(\mathcal{S})| \le 4(b+1)^b k^b f_0(\lfloor n/k \rfloor, \mathcal{S}).$$

where b is an upper bound on the number of objects that determine a region, $|\mathcal{F}_{\le k}(\mathcal{S})|$ is the number of regions defined and with at most k conflicts over $\mathcal{S}$, and $f_0(\lfloor n/k \rfloor, \mathcal{S})$ is the expected number of regions defined and without conflict over a random $\lfloor n/k \rfloor$-sample of $\mathcal{S}$.

Proof. For each i, $1 \le i \le b$, we shall prove the following inequality bounding the number of regions determined by i objects:

$$|\mathcal{F}_{\le k}^i(\mathcal{S})| \le 4(b+1)^i k^i f_0^i(\lfloor n/k \rfloor, \mathcal{S}).$$

Then the theorem can be easily proved by summing over all the values of i between 1 and b.

Let k be an integer such that $2 \leq k \leq \frac{n}{b+1}$ and $\mathcal{R}$ a random sample of $\mathcal{S}$, of size $r = \lfloor n/k \rfloor$. From lemma 4.2.2, we recall that the expected number $f_0^i(r, \mathcal{S})$ of regions defined and without conflict over $\mathcal{R}$ is

$$f_0^i(r, \mathcal{S}) = \sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \frac{\binom{n-i-j}{r-i}}{\binom{n}{r}} \geq |\mathcal{F}_{\leq k}^i(\mathcal{S})| \frac{\binom{n-i-k}{r-i}}{\binom{n}{r}}.$$

The remainder of this proof is a mere computation on factorials, which shows that for each k such that $2 \leq k \leq \frac{n}{b+1}$, and $r = \lfloor n/k \rfloor$,

$$\frac{\binom{n-i-k}{r-i}}{\binom{n}{r}} \geq \frac{1}{4(b+1)^i k^i}.$$

Indeed,

$$\frac{\binom{n-i-k}{r-i}}{\binom{n}{r}} = \frac{r!}{(r-i)!} \frac{(n-i)!}{n!} \frac{(n-r)!}{(n-r-k)!} \frac{(n-i-k)!}{(n-i)!}.$$

We compute

$$\begin{aligned}
\frac{(n-r)!}{(n-r-k)!} \frac{(n-i-k)!}{(n-i)!} &\geq \left(\frac{n-r-k+1}{n-i-k+1}\right)^k \\
&\geq \left(\frac{n-n/k-k+1}{n-k}\right)^k \\
&\geq (1-1/k)^k \\
&\geq 1/4 \;\; (\text{if } 2 \leq k),
\end{aligned}$$

and

$$\begin{aligned}
\frac{r!}{(r-i)!} \frac{(n-i)!}{n!} &= \prod_{l=0}^{i-1} \frac{r-l}{n-l} \geq \prod_{l=1}^{i} \frac{r+1-l}{n} \\
&\geq \prod_{l=1}^{i} \frac{n/k-l}{n} \geq \frac{1}{k^i}\left(1-\frac{bk}{n}\right)^i \\
&\geq \frac{1}{k^i(b+1)^i} \;\; \left(\text{if } k \leq \frac{n}{b+1}\right),
\end{aligned}$$

proving the inequality stated by the theorem. □

Remark 1. The sampling theorem deals with the numbers $|\mathcal{F}_{\leq k}(\mathcal{S})|$ of regions with at most k conflicts, for values of k between 2 and $\frac{n}{b+1}$.

For the case of regions without or with at most one conflict, however, it is possible to prove the following bound

$$|\mathcal{F}_0(\mathcal{S})| \leq |\mathcal{F}_{\leq 1}(\mathcal{S})| \leq |\mathcal{F}_{\leq 2}(\mathcal{S})| \leq 4(b+1)^b 2^b f_0(\lfloor n/2 \rfloor, \mathcal{S}),$$

valid whenever $n \geq 2(b+1)$.

Moreover, for values of k close to n, there is always the trivial bound

$$|\mathcal{F}_{\leq k}(\mathcal{S})| \leq |\mathcal{F}(\mathcal{S})| = O(n^b)$$

if, as in remark 4 of subsection 4.1.1, we suppose that each subset of size at most b determines at most q regions, for a constant number q that depends on the interpretation of objects and regions.

Remark 2. The sampling theorem yields a deterministic combinatorial result when an upper bound on $f_0(\lfloor n/k \rfloor, \mathcal{S})$ can be derived. For instance, in chapter 14, we will use an upper bound on the number of faces of a d-dimensional polytope to yield, via the sampling theorem, an upper bound on the number of faces at level at most k in an arrangement of hyperplanes.

The following corollary is very useful for analyzing the average performance of randomized algorithms. It shows that the expected number of regions defined and with one or two conflicts over a random r-sample of a set $\mathcal{S}$ is of the same order of magnitude as the expected number of regions defined and without conflict over such a sample.

Corollary 4.2.4 *Let $\mathcal{S}$ be a set of n objects, with $n \geq 2(b+1)$. For each integer r such that $n \geq r \geq 2(b+1)$, we have*

$$\begin{aligned} f_1(r, \mathcal{S}) &\leq \beta f_0(\lfloor r/2 \rfloor, \mathcal{S}) \\ f_2(r, \mathcal{S}) &\leq \beta f_0(\lfloor r/2 \rfloor, \mathcal{S}) \end{aligned}$$

where $f_j(r, \mathcal{S})$ is the expected number of regions defined and with j conflicts over a random r-sample of $\mathcal{S}$, and β is the real constant

$$\beta = 4(b+1)^b 2^b.$$

Proof. Let $\mathcal{R}$ be a subset of $\mathcal{S}$ of size r, such that $2(b+1) \leq r$. Applied to $\mathcal{R}$, remark 1 following theorem 4.2.3 yields

$$|\mathcal{F}_1(\mathcal{R})| \leq 4(b+1)^b 2^b f_0(\lfloor r/2 \rfloor, \mathcal{R}).$$

The first inequality is obtained by taking expectations on the two members of this equation. Indeed, $f_0(\lfloor r/2 \rfloor, \mathcal{R})$ is the expected number of regions defined and without conflict over a random $\lfloor r/2 \rfloor$-sample of $\mathcal{R}$, and the expectation of this expected number when $\mathcal{R}$ itself is a random r-sample of $\mathcal{S}$ is simply $f_0(\lfloor r/2 \rfloor, \mathcal{S})$. The second inequality can be proved in much the same way. □

4.2.2 The moment theorem

Let $\mathcal{S}$ be a set of n objects and $\mathcal{R}$ be a subset of $\mathcal{S}$. The moment theorem bounds the total number of conflicts between the objects of $\mathcal{S}$ and the regions defined and without conflict over $\mathcal{R}$.

Let k be a integer less than or equal to n. The *moment of order* k of $\mathcal{R}$ with respect to $\mathcal{S}$, denoted by $m_k(\mathcal{R},\mathcal{S})$, is the sum

$$m_k(\mathcal{R},\mathcal{S}) = \sum_{F\in\mathcal{F}_0(\mathcal{R})} \binom{|\mathcal{S}(F)|}{k},$$

where $\mathcal{F}_0(\mathcal{R})$ stands for the set of regions defined and without conflict over $\mathcal{R}$, and $|\mathcal{S}(F)|$ is the cardinality of the set $\mathcal{S}(F)$ of objects in $\mathcal{S}$ that conflict with a region F.

The moment of order 0, $m_0(\mathcal{R},\mathcal{S})$, is simply the number of regions defined and without conflict over $\mathcal{R}$:

$$m_0(\mathcal{R},\mathcal{S}) = |\mathcal{F}_0(\mathcal{R})|.$$

The moment of order 1, $m_1(\mathcal{R},\mathcal{S})$, is the total number of conflicts between the elements of $\mathcal{S}$ and the regions defined and without conflict over $\mathcal{R}$:

$$m_1(\mathcal{R},\mathcal{S}) = \sum_{F\in\mathcal{F}_0(\mathcal{R})} |\mathcal{S}(F)|.$$

The expectation of $m_k(\mathcal{R},\mathcal{S})$ for a random r-sample $\mathcal{R}$ of $\mathcal{S}$ is denoted by $m_k(r,\mathcal{S})$. In particular, $m_0(r,\mathcal{S}) = f_0(r,\mathcal{S})$.

Lemma 4.2.5

$$m_k(r,\mathcal{S}) = \sum_{i=1}^{b}\sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \binom{j}{k} p_j^i(r).$$

Proof. Recall that $p_j^i(r)$ stands for the probability that a given region F of $\mathcal{F}_j^i(\mathcal{S})$ be defined and without conflict over a random r-sample of $\mathcal{S}$, whence

$$m_k(r,\mathcal{S}) = \sum_{i=1}^{b}\sum_{j=0}^{n-i} \sum_{F\in\mathcal{F}_i^j(\mathcal{S})} \binom{j}{k} p_j^i(r).$$ □

Theorem 4.2.6 (Moment theorem) *Let $\mathcal{S}$ be a set of n objects. The expectation $m_k(r,\mathcal{S})$ of the moment of order k of a random r-sample of $\mathcal{S}$ is related to the expected number $f_k(r,\mathcal{S})$ of regions defined and with k conflicts over a random r-sample of $\mathcal{S}$ by the relation*

$$m_k(r,\mathcal{S}) \le f_k(r,\mathcal{S}) \frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!},$$

where each region is determined by at most b objects.

Proof. According to the previous lemma 4.2.5, and to lemma 4.2.1 which gives the expression for the probability $p_j^i(r)$, we have

$$\begin{aligned} m_k(r,\mathcal{S}) &= \sum_{i=1}^{b}\sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \binom{j}{k} \frac{\binom{n-i-j}{r-i}}{\binom{n}{r}} \\ &= \sum_{i=1}^{b}\sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}} \frac{(n-j-r+k)!}{(n-j-r)!}\frac{(r-i-k)!}{(r-i)!} \\ &\le \frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!} \sum_{i=1}^{b}\sum_{j=0}^{n-i} |\mathcal{F}_j^i(\mathcal{S})| \frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}. \end{aligned}$$

As proved by the same lemma 4.2.1, however, the factor

$$\frac{\binom{j}{k}\binom{n-i-j}{r-i-k}}{\binom{n}{r}}$$

is nothing else but the probability $p_{j,k}^i(r)$ that a region F of $\mathcal{F}_j^i(\mathcal{S})$ belong to $\mathcal{F}_k^i(\mathcal{R})$, whence

$$m_k(r,\mathcal{S}) \le f_k(r,\mathcal{S})\frac{(n-r+k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!}.$$ □

Corollary 4.2.7 *Let $\mathcal{S}$ be a set of n objects. There exists a real constant γ and an integer r_0, both independent of n, such that for each $n \ge r \ge r_0$,*

$$\begin{aligned} m_1(r,\mathcal{S}) &\le \gamma\frac{n-r}{r} f_0(\lfloor r/2 \rfloor, \mathcal{S}) \\ m_2(r,\mathcal{S}) &\le \gamma\frac{(n-r)^2}{r^2} f_0(\lfloor r/2 \rfloor, \mathcal{S}), \end{aligned}$$

where $m_k(r,\mathcal{S})$ is the expected number of the k-th moment of a random r-sample of $\mathcal{S}$, and $f_0(r,\mathcal{S})$ is the expected number of regions defined and without conflict over a random r-sample of $\mathcal{S}$.

Proof. For $k = 1$, the moment theorem yields

$$m_1(r,\mathcal{S}) \le f_1(r,\mathcal{S})\frac{n-r+1}{r-b}$$

and the upper bound is a consequence of corollary 4.2.4. The second inequality can be proved very much the same way. □

4.3 Exercises

Exercise 4.1 (Backward analysis) In this exercise, regions are determined by at most b objects of a set $\mathcal{S}$. Let $f_j(r, \mathcal{S})$ be the expected number of regions defined and without conflict over a random r-sample of $\mathcal{S}$. Corollary 4.2.4 to the sampling theorem proves that $f_1(r, \mathcal{S}) = O(f_0(r, \mathcal{S}))$. *Backward analysis* can be used to prove this without invoking the sampling theorem.

Let $\mathcal{R}$ be a subset of $\mathcal{S}$ of cardinality r, and $f_0(r-1, \mathcal{R})$ the expected number of regions defined and without conflict over a random sample of $\mathcal{R}$ of size $r-1$. Show that

$$f_0(r-1, \mathcal{R}) \leq \frac{1}{r}|\mathcal{F}_1(\mathcal{R})| + \frac{r-1}{r}|\mathcal{F}_0(\mathcal{R})| \quad (4.1)$$

$$f_0(r-1, \mathcal{R}) \geq \frac{1}{r}|\mathcal{F}_1(\mathcal{R})| + \frac{r-b}{r}|\mathcal{F}_0(\mathcal{R})|. \quad (4.2)$$

From this, show that $f_1(r, \mathcal{S}) = O(f_0(r, \mathcal{S}))$. Similarly, show that $f_2(r, \mathcal{S}) = O(f_0(r, \mathcal{S})$.

Hint: Backward analysis consists in observing that a random $(r-1)$-sample $\mathcal{R}'$ of $\mathcal{R}$ can be obtained by removing one random object from $\mathcal{R}$. Any region in $\mathcal{F}_0(\mathcal{R}')$ is defined over $\mathcal{R}$ and belongs either to $\mathcal{F}_0(\mathcal{R})$ or to $\mathcal{F}_1(\mathcal{R})$. A region F that belongs to $\mathcal{F}_0(\mathcal{R})$ determined by i objects is a region of $\mathcal{F}_0(\mathcal{R}')$ if the removed object is not one of the i objects that determine F; this happens with probability $\frac{r-i}{r}$. A region F that belongs to $\mathcal{F}_1(\mathcal{R})$ is a region of $\mathcal{F}_0(\mathcal{R}')$ if the removed object is precisely the one that was removed from $\mathcal{R}$, which happens with probability $\frac{1}{r}$. To show that $f_1(r, \mathcal{S}) = O(f_0(r, \mathcal{S}))$, it suffices to take expectations in equation 4.2 over all r-samples of $\mathcal{S}$ and to assume that $f_0(r, \mathcal{S})$ is a non-decreasing function of r.

Exercise 4.2 (The moment theorem, using backward analysis) Let $\mathcal{R}$ be a random r-sample of a set $\mathcal{S}$ of n objects, and O a random object of $\mathcal{S} \setminus \mathcal{R}$. Show that the expected number of regions defined and without conflict over $\mathcal{R}$ but conflicting with O is $O(\frac{1}{r+1} f_1(r+1, \mathcal{S}))$. From this, show that the expected value $m_1(r, \mathcal{S})$ of the moment of order 1 with respect to $\mathcal{S}$ of a random r-sample is $O(\frac{n-r}{r+1} f_1(r+1, \mathcal{S}))$. From this, deduce an alternative proof of the moment theorem by using the result of the previous exercise or corollary 4.2.4 to the sampling theorem.

Hint: Note that $\mathcal{R} \cup \{O\}$ is a random $(r+1)$-sample of $\mathcal{S}$ and that a region of $\mathcal{F}_0(\mathcal{R})$ that conflicts with O is a region of $\mathcal{F}_1(\mathcal{R} \cup \{O\})$ that conflicts with O.

Exercise 4.3 (An extension of the moment theorem) A function w is called *convex* if it satisfies, for all x, y in $\mathbb{R}$ and all α in $[0, 1]$,

$$w(\alpha x + (1-\alpha)y) \geq \alpha w(x) + (1-\alpha)w(y).$$

We are interested in regions determined by at most b objects of a set $\mathcal{S}$ of n objects. For each subset $\mathcal{R}$ of $\mathcal{S}$ and any convex function w, we define:

$$w_k(\mathcal{R}) = \sum_{F\in\mathcal{F}_0(\mathcal{R})} \binom{w(|\mathcal{S}(F)|)}{k},$$

where $\mathcal{F}_0(\mathcal{R})$ is the set of regions defined and without conflict over $\mathcal{R}$ and $|\mathcal{S}(F)|$ is the number of objects in $\mathcal{S}$ that conflict with F. Let $w_k(r,\mathcal{S})$ stand for the expected value of $w_k(\mathcal{R})$ for a random r-sample of $\mathcal{S}$. Show that

$$w_k(r,\mathcal{S}) \leq f_0(r,\mathcal{S})\, w\left(\frac{(n-r-k)!}{(n-r)!}\frac{(r-b-k)!}{(r-b)!}\frac{f_k(r,\mathcal{S})}{f_0(r,\mathcal{S})}\right).$$

Exercise 4.4 (Non-local subset of regions) We still work with the framework of objects, regions, and conflicts, each region being determined by at most b objects. In this exercise, we are mostly interested, for a subset $\mathcal{R}$ of objects in $\mathcal{S}$, in a subset $\mathcal{G}_0(\mathcal{R})$ of regions defined and without conflict over $\mathcal{R}$. The definition of $\mathcal{G}_0(\mathcal{R})$ is not necessarily local, however: a region F of $\mathcal{F}_0(\mathcal{R})$ belongs to $\mathcal{G}_0(\mathcal{R})$ depending on *all* the elements of $\mathcal{R}$, not only those in conflict with F or that determine F. Nevertheless, suppose that the subsets of the form $\mathcal{G}_0(\mathcal{R})$ satisfy the following property: *If F is a region of $\mathcal{G}_0(\mathcal{R})$, $\mathcal{R}'$ a subset of $\mathcal{R}$, and if $\mathcal{R}'$ contains the elements that determine F, then F is a region of $\mathcal{G}_0(\mathcal{R}')$.*

Let $w_k(r,\mathcal{S})$ be the expected value of the sum

$$\sum_{F\in\mathcal{G}_0(\mathcal{R})} |\mathcal{S}(F)|^k$$

where $|\mathcal{S}(F)|$ is the number of objects of $\mathcal{S}$ in conflict with F. We are interested in showing the moment theorem for the regions in $\mathcal{G}_0(\mathcal{R})$, in other words that

$$w_k(r,\mathcal{S}) = O\left(\frac{n^k}{r^k} g_0(r,\mathcal{S})\right),$$

where $g_0(r,\mathcal{S})$ is the expected number of regions in $\mathcal{G}_0(\mathcal{R})$ for a random r-sample of $\mathcal{S}$.

Hint: 1. Let $p(r,F)$ be the probability that F be a region of $\mathcal{G}_0(\mathcal{R})$ for a random r-sample $\mathcal{R}$ of $\mathcal{S}$. Show that, for all $t < r \leq n$,

$$p(r,F) \leq \frac{r!}{(r-b)!}\frac{(t-b)!}{t!}p(t,F).$$

2. Let us propose an incremental algorithm to compute $\mathcal{G}_0(\mathcal{S})$. The probability that a region F appear in $\mathcal{G}_0(\mathcal{R})$ precisely at step r is

$$\frac{b}{r}p(r,F).$$

The probability that it disappear from $\mathcal{G}_0(\mathcal{R})$ at the next step $r+1$ is at least

$$\frac{|\mathcal{S}(F)|}{n-r}p(r,F).$$

Show that, for all $r_1 < r_2$,

$$p(r_1, F) + \sum_{r_1+1}^{r_2} \frac{b}{r} p(r, F) \geq \sum_{r_1}^{r_2} \frac{|\mathcal{S}(F)|}{n-r} p(r, F).$$

3. Using the previous inequality, show by induction on k that

$$\forall r, r/2 \leq r_0 \leq r \quad \sum_{t=r_0}^{r} w_k(t, \mathcal{S}) \leq \gamma'_k r \frac{n^k}{r^k} g_0(r, \mathcal{S})$$

and that

$$w_k(r, \mathcal{S}) \leq \gamma_k \frac{n^k}{r^k} g_0(r, \mathcal{S}),$$

where γ'_k and γ_k are constants depending only on k.

Exercise 4.5 (Tail estimates) Let b be the maximum number of objects that determine a single region. Suppose again that a set of at most b objects determine at most q regions, q being a constant, or that the number of regions determined by a set $\mathcal{S}$ of n objects is $O\left(n^b\right)$.

1. Let $\mathcal{S}$ be a set of n objects and $\mathcal{R}$ a random r-sample of $\mathcal{S}$. Let α be a real constant in $]0,1[$. Let $\pi_0(\alpha, r)$ denote the probability over all samples $\mathcal{R}$ that some region defined and without conflict over $\mathcal{R}$ have at least $\lceil \alpha n \rceil$ conflicts with $\mathcal{S}$. Show that, for r big enough,

$$\pi_0(\alpha, r) = O\left(r^b (1-\alpha)^r\right).$$

2. Show that for any constant $\lambda > b$, the probability $\pi_0(\lambda \log r/r, r)$ that some region F, defined and without conflict over $\mathcal{R}$, have at least $\lambda n \log r/r$ conflicts with $\mathcal{S}$ decreases to 0 as r increases.

Exercise 4.6 (Extension of the previous tail estimates) We propose to generalize the tail estimates given in exercise 4.5. Again, let b be the maximum number of objects that determine a single region, and suppose that the number of regions determined by a set $\mathcal{S}$ of n objects is $O\left(n^b\right)$.

Let $\mathcal{S}$ be a set of n objects, $\mathcal{R}$ a random r-sample of $\mathcal{S}$, α a real constant in $]0,1[$, and m a positive integer. Denote by $\pi_m^-(\alpha, r)$ the probability that there exists a region F defined over $\mathcal{R}$ with at most m conflicts over $\mathcal{R}$, and at least $\lceil \alpha n \rceil$ conflicts over $\mathcal{S}$. Likewise, denote by $\pi_m^+(\alpha, r)$ the probability that there exists a region F defined over $\mathcal{R}$ with at least m conflicts over $\mathcal{R}$, and at most $\lceil \alpha n \rceil$ conflicts over $\mathcal{S}$.

Show that if the size r of the sample is big enough while still smaller than $\sqrt{n}/2$, then

$$\text{if } m \leq \alpha(r-b), \qquad \pi_m^-(\alpha, r) = \frac{O(r^b)}{(1-\alpha)^b} \left[\sum_{j=0}^{m} \binom{r}{j} \alpha^j (1-\alpha)^{r-j}\right],$$

$$\text{if } m \geq \alpha(r-b), \qquad \pi_m^+(\alpha, r) = \frac{O(r^b)}{(1-\alpha)^b} \left[\sum_{j\geq m} \binom{r}{j} \alpha^j (1-\alpha)^{r-j}\right].$$

Then show that, if $\alpha(r) = \lambda \log r/r$ and $m(r) = \log r/\log\log r$,

$$\lim_{r\to\infty} \pi^-_{m(r)}(\alpha(r), r) = 0.$$

Exercise 4.7 (An upper bound on $f_0(\mathcal{S})$) Consider the set $\mathcal{F}(\mathcal{S})$ of regions defined over a set $\mathcal{S}$, each region being determined by at most b objects. Let $f_j(\mathcal{S})$ be the number of regions defined and having j conflicts with $\mathcal{S}$, and $f_0(n)$ be the maximum of $f_0(\mathcal{S})$ over all sets $\mathcal{S}$ of n objects. Suppose that there is a relation between the number of regions defined and without conflict over $\mathcal{S}$ on one hand, and the number of regions defined over $\mathcal{S}$ and conflicting with one element of $\mathcal{S}$ on the other. Suppose further that this relation is of the type

$$cf_0(\mathcal{S}) \le f_1(\mathcal{S}) + d(n) \tag{4.3}$$

where c is an integer constant and $d(n)$ a known function of n. Let $t = b - c$. Show then that

$$f_0(n) = O\left(n^t\left(1 + \sum_{j=t+1}^{n} \frac{d(j)}{j^{t+1}}\right)\right).$$

In particular,

$$\begin{aligned}
f_0(n) &= O(n^t) && \text{if } d(n) = O(n^{t'}) \text{ for } t' < t,\\
f_0(n) &= O(n^t \log n) && \text{if } d(n) = O(n^t),\\
f_0(n) &= O(n^{t'}) && \text{if } d(n) = O(n^{t'}) \text{ for } t' > t.
\end{aligned}$$

Hint: Combining equation 4.2, written for a random $(n-1)$-sample of $\mathcal{S}$, and equation 4.3 yields

$$\begin{aligned}
\frac{n-b+c}{n} f_0(\mathcal{S}) &= \frac{n-b}{n} f_0(\mathcal{S}) + \frac{c}{n} f_0(\mathcal{S})\\
&\le \frac{n-b}{n} f_0(\mathcal{S}) + \frac{1}{n}\left(f_1(\mathcal{S}) + d(n)\right)\\
&\le f_0(n-1, \mathcal{S}) + \frac{1}{n} d(n).
\end{aligned}$$

Exercise 4.8 (Union of parallel hypercubes) Consider a set of *parallel* hypercubes in $\mathbb{E}^d$, that is, hypercubes whose sides are parallel to the axes.

Show that the union of n hypercubes has at most $O(n^{\lceil d/2 \rceil})$ faces for each $d \ge 1$.

Furthermore, show that the complexity of the union of n hypercubes of equal size is $O(n^{\lfloor d/2 \rfloor})$ when $d \ge 2$ and remains $O(n)$ in dimension 1.

Hint: Each vertex of the union belongs to a bounded number of faces of the union. Hence it suffices to bound the number of vertices of the union to bound the total complexity.

The proof works by induction on d. The proof is trivial in dimension 1, and easy in dimension 2.

In dimension d, each cube has $2d$ pairwise parallel facets. Let us denote by $F_j^+(C)$ the facet of the cube C that is perpendicular to the x_j-axis with maximal j-coordinate, and by $F_j^-(C)$ the facet of the cube C that is perpendicular to the x_j-axis with minimal

j-coordinate. Let $\mathcal{C}$ be a set of axis-parallel cubes in $\mathbb{E}^d$, and denote by $\mathcal{U}(\mathcal{C})$ the union of these cubes and $\mathcal{A}(\mathcal{C})$ their arrangement, that is, the decomposition of $\mathbb{E}^d$ induced by the cubes (see part IV for an introduction to arrangements). Each vertex of $\mathcal{U}(\mathcal{C})$ or of $\mathcal{A}(\mathcal{C})$ is at the intersection of d facets of cubes, one perpendicular to each axis direction. Such a vertex P is denoted by $(C_1^{\epsilon_1}, C_2^{\epsilon_2}, \ldots, C_d^{\epsilon_d})$ if at the intersection of facets $F_j^{\epsilon_j}(C_j)$, for $j = 1, \ldots, d$ and $\epsilon_j = +$ or $-$. The vertex P is called *outer* if it belongs to a $(d-2)$-face of one of the cubes (then not all the cubes C_j are distinct). It is called an *inner* vertex if it is at the intersection of d facets of pairwise distinct cubes. A vertex of $\mathcal{A}(\mathcal{C})$ is at *level* k if it belongs to the interior of k cubes of $\mathcal{C}$. The vertices of the union are precisely the vertices at level 0 in the arrangement $\mathcal{A}(\mathcal{C})$. Let $w_k(\mathcal{C})$ be the number of inner vertices of $\mathcal{A}(\mathcal{C})$ at level k, and $v_k(\mathcal{C})$ be the number of outer vertices at level k, and $v_k(n,d)$ (resp. $w_k(n,d)$) the maximum of $v_k(\mathcal{C})$ (resp. of $w_k(\mathcal{C})$) over all possible sets $\mathcal{C}$ of n axis-parallel hypercubes in $\mathbb{E}^d$.

1. The maximum number $v_0(n,d)$ of outer vertices of the union is $O(n^{\lceil d/2 \rceil})$ (and $O(n^{\lfloor d/2 \rfloor})$ when the cubes have same size). Indeed, any outer vertex of $\mathcal{U}(\mathcal{C})$ belongs to a $(d-2)$-face H of one of the cubes in $\mathcal{C}$ and is a vertex (either outer or inner) of the union of all $(d-2)$-cubes $\mathcal{C} \cap \mathit{aff}(H)$, where $\mathit{aff}(H)$ is the affine hull of H. Consequently,

$$v_0(n,d) \leq 2nd(d-1)\big(\hat{v}_0(n-1, d-2) + \hat{w}_0(n-1, d-2)\big),$$

where $\hat{v}_0(n-1, d-2)$ and $\hat{w}_0(n-1, d-2)$ respectively stand for the maximum numbers of outer or inner vertices in the union of $n-1$ cubes in a $(d-2)$-dimensional space lying inside a given $(d-2)$-cube.

2. Applying the sampling theorem (theorem 4.2.3) and its corollary 4.2.4, we derive a similar bound on the maximum number $v_1(n,d)$ of outer vertices at level 1.

3. To count the number of inner vertices, we use the following charging scheme. For each vertex $P = (C_1^{\epsilon_1}, C_2^{\epsilon_2}, \ldots, C_d^{\epsilon_d})$ of $\mathcal{U}(\mathcal{C})$, and each direction $j = 1, \ldots d$, slide along the edge of $\mathcal{A}(\mathcal{C})$ that lies inside the cube C_j (this edge is $\bigcap_{i \neq j} F_i^{\epsilon_i}(C_i)$) until the other vertex P' of this edge is reached.

If P' belongs to the facet $F_j^{-\epsilon_j}(C_j)$ of cube C_j, we do not charge anything. This case cannot happen unless the cubes have different side lengths and C_j is the smallest of the cubes intersecting at P.

If P' belongs to a $(d-2)$-face of one of the cubes C_i $(i \neq j)$ intersecting at P, P' is an outer vertex at level 1, and is charged one unit for P. Note that P' cannot be charged more than twice for this situation.

If P' belongs to another cube C' distinct from all the C_i intersecting at P, then P' is an inner vertex at level 1, and is charged one unit for P. Any inner vertex P' of this type may be charged up to d times for this situation. However, when it is charged more than once, say m times, we may redistribute the extra $m-1$ charges on the outer vertices at level 0 or 1, and these vertices will only be charged once in this fashion.

In the case of cubes with different sizes, the induction is

$$(d-1)w_0(\mathcal{C}) \leq w_1(\mathcal{C}) + 3v_1(\mathcal{C}) + v_0(\mathcal{C}).$$

In the case of cubes with identical sizes, we obtain

$$dw_0(\mathcal{C}) \leq w_1(\mathcal{C}) + 3v_1(\mathcal{C}) + v_0(\mathcal{C}).$$

It suffices to apply exercise 4.7 to conclude.

4.4 Bibliographical notes

Randomized methods revolutionized computational geometry. Most of the material in this chapter is taken from the ground-breaking work of Clarkson and Shor [71]. The randomized incremental algorithms in the next two chapters are concrete applications of the formalism developed in this chapter, and we invite the reader to consult the bibliographical notes of these chapters for more references. Clarkson and Shor proved the tail estimates and their extension as stated in exercises 4.5 and 4.6, which are the corner stone on which all analyses of randomized divide-and-conquer algorithms rely. In their article, they also prove the extension to the moment theorem proposed in exercise 4.3. This extension will prove useful in exercise 5.8 for the analysis of an algorithm that triangulates a simple polygon, due to Clarkson, Cole, and Tarjan [69].

The extension of the moment theorem to a non-local set of regions defined and without conflict over a random sample (exercise 4.4) is due to de Berg, Dobrindt, and Schwarzkopf [76]. The result stated in exercise 4.4 will be used in chapter 15 to analyze the randomized incremental algorithm that builds a single cell in an arrangement of line segments.

The *backward analysis* method proposed in exercises 4.1 and 4.2 was used by Chew [59] to analyze an algorithm that builds the Voronoi diagram of a convex polygon (see exercise 19.4). The method was used later in a systematic fashion by Seidel [203] and Devillers [80].

The method used in exercise 4.7 to obtain an upper bound on the expected number of regions defined and without conflict over a set of objects is due to Tagansky [212]. The analysis of the complexity of the union of parallel hypercubes in d dimensions (see exercise 4.8) given by Boissonnat, Sharir, Tagansky, and Yvinec [34] illustrates the power of this method.