Chapter 1

Sampling and the Moments Technique

By Sariel Har-Peled, February 16, 2023[1]

Sun and rain and bush had made the site look old, like the site of a dead civilization. The ruins, spreading over so many acres, seemed to speak of a final catastrophe. But the civilization wasn't dead. It was the civilization I existed in and in fact was still working towards. And that could make for an odd feeling: to be among the ruins was to have your time-sense unsettled. You felt like a ghost, not from the past, but from the future. You felt that your life and ambition had already been lived out for you and you were looking at the relics of that life. You were in a place where the future had come and gone.

A bend in the river, V. S. Naipaul

1.1. General settings

Let S be a set of objects. For a subset $\mathsf{R} \subseteq \mathsf{S}$, we define a collection of 'regions' called $\mathcal{F}(\mathsf{R})$. For the case of vertical decomposition of segments (i.e., Theorem **??**), the objects are segments, the regions are trapezoids, and $\mathcal{F}(\mathsf{R})$ is the set of vertical trapezoids in $\mathcal{A}^{|}(\mathsf{R})$. Let

$$\mathcal{T} = \mathcal{T}(\mathsf{S}) = \bigcup_{\mathsf{R} \subseteq \mathsf{S}} \mathcal{F}(\mathsf{R})$$

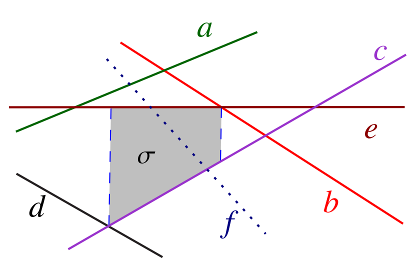

Figure 1.1.1: $D(\sigma) = \{b, c, d, e\}$ and $K(\sigma) = \{f\}$.

denote the set of ***all possible regions*** defined by subsets of S.

In the vertical trapezoids case, the set $\mathcal{T}$ is the set of all vertical trapezoids that can be defined by any subset of the given input segments.

We associate two subsets $D(\sigma), K(\sigma) \subseteq \mathsf{S}$ with each region $\sigma \in \mathcal{T}$.

The ***defining set*** $D(\sigma)$ of σ is the subset of S defining the region σ (the precise requirements from this set are specified in the axioms below). We assume that for every $\sigma \in \mathcal{T}$, $|D(\sigma)| \leq d$ for a (small) constant d. The constant d is sometime referred to as the ***combinatorial dimension***. In the case of Theorem **??**, each trapezoid σ is defined by at most four segments (or lines) of S that define the region covered by the trapezoid σ, and this set of segments is $D(\sigma)$. See Figure 1.1.1.

The ***stopping set*** $K(\sigma)$ of σ is the set of objects of S such that including any object of $K(\sigma)$ in R prevents σ from appearing in $\mathcal{F}(\mathsf{R})$. In many applications $K(\sigma)$ is just the set of objects intersecting the cell σ; this is also the case in Theorem **??**, where $K(\sigma)$ is the set of segments of S intersecting the interior of the trapezoid σ (see Figure 1.1.1). Thus, the stopping set of a region σ, in many cases, is just the conflict list of this region, when it is being created by an RIC algorithm. The ***weight*** of σ is $\omega(\sigma) = |K(\sigma)|$.

Axioms. Let $\mathsf{S}, \mathcal{F}(\mathsf{R}), D(\sigma)$, and $K(\sigma)$ be such that for any subset $\mathsf{R} \subseteq \mathsf{S}$, the set $\mathcal{F}(\mathsf{R})$ satisfies the following axioms:

(i) For any $\sigma \in \mathcal{F}(\mathsf{R})$, we have $D(\sigma) \subseteq \mathsf{R}$ and $\mathsf{R} \cap K(\sigma) = \emptyset$.

(ii) If $D(\sigma) \subseteq \mathsf{R}$ and $K(\sigma) \cap \mathsf{R} = \emptyset$, then $\sigma \in \mathcal{F}(\mathsf{R})$.

1.1.1. Examples of the general framework

(A) **Vertical decomposition.** Discussed above.

(B) **Points on a line.** Let S be a set of n points on the real line. For a set $\mathsf{R} \subseteq \mathsf{S}$, let $\mathcal{F}(\mathsf{R})$ be the set of atomic intervals of the real lines formed by R; that is, the partition of the real line into maximal connected sets (i.e., intervals and rays) that do not contain a point of R in their interior.

Clearly, in this case, an interval $\mathcal{I} \in \mathcal{F}(\mathsf{R})$ the defining set of $\mathcal{I}$ (i.e., $D(\mathcal{I})$) is the set containing the (one or two) endpoints of $\mathcal{I}$ in R. The stopping set of an $\mathcal{I}$ is the set $K(\mathcal{I})$, which is the set of all points of S contained in $\mathcal{I}$.

(C) **Vertices of the convex-hull in 2d.** Consider a set S of n points in the plane. A vertex on the convex hull is defined by the point defining the vertex, and the two edges before and after it on the convex hull. To this end, a ***certified vertex*** of the convex hull (say this vertex is q) is a triplet (p, q, s), such that p, q and s are consecutive vertices of $\mathcal{CH}(\mathsf{S})$ (say, in clockwise order). Observe, that computing the convex-hull of S is equivalent to computing the set of certified vertices of S.

For a set $\mathsf{R} \subseteq \mathsf{S}$, let $\mathcal{F}(\mathsf{R})$ denote the set of certified vertices of R (i.e., this is equivalent to the set of vertices of the convex-hull of R. For a certified vertex $\sigma \in \mathcal{F}(\mathsf{R})$, its defining set is the set of three vertices p, q, s that (surprise, surprise) define it. Its stopping set, is the set of all points in S, that either on the "wrong" side of the line spanning pq, or on the "wrong" side of the line spanning $q\mathsf{s}$. Equivalently, $K(\sigma)$ is the set of all points $t \in \mathsf{S} \setminus \mathsf{R}$, such that the convex-hull of p, q, s, and t does not form a convex quadrilateral.

(D) **Edges of the convex-hull in 3d.**

Let S be a set of points in three dimensions. An edge e of the convex-hull of a set $\mathsf{R} \subseteq ObjSet$ of points in $\mathbb{R}^3$ is defined by two vertices of S, and it can be certified as being on the convex hull $\mathcal{CH}(\mathsf{R})$, by the two faces f, f' adjacent to e. If all the points of R are on the "right" side of both these two faces then e is an edge of the convex hull of R. Computing all the certified edges of S is equivalent to computing the convex-hull of S.

In the following, assume that each face of any convex-hull of a subset of points of S is a triangle. As such, a face of the convex-hull would be defined by three points. Formally, the ***butterfly*** of an edge e of $\mathcal{CH}(\mathsf{R})$ is (e, p, q), where $p, q \in \mathsf{R}$, and such that all the points of R are on the same side as q of the plane spanned by e and p (we have symmetric condition requiring that all the points of S are on the same as p of the plane spanned by e and q).

For a set $\mathsf{R} \subseteq \mathsf{P}$, let $\mathcal{F}(\mathsf{R})$ be its set of butterflies. Clearly, computing all the butterflies of S (i.e., $\mathcal{F}(\mathsf{S})$) is equivalent to computing the convex-hull of S.

For a butterfly $\sigma = (\mathsf{e}, p, q) \in \mathcal{F}(\mathsf{R})$ its defining set (i.e., $D(\sigma)$) is a set of four points (i.e., the two points defining its edge e, and the to additional vertices defining the two faces $Face$ and f' adjacent to it). Its stopping set $K(\sigma)$, is the set of all the points of $\mathsf{S} \setminus \mathsf{R}$ that of different sides of the plane spanned by e and p (resp. e and q) than q (resp. p) [here, the stopping set is the union of these two sets].

(E) **Delaunay triangles in 2d.**

For a set of S of n points in the plane. Consider a subset $\mathsf{R} \subseteq \mathsf{S}$. A ***Delaunay circle*** of R is a disc D that has three points p_1, p_2, p_3 of R on its boundary, and no points of R in its interior. Naturally, these three points define a ***Delaunay triangle*** $\triangle = \triangle p_1 p_2 p_3$. The defining set is $D(\triangle) = \{p_1, p_2, p_3\}$, and the stopping set $K(\triangle)$ is the set of all points in S that are contained in the interior of the disk D.

1.2. Analysis: Bounding the moments

Consider a different randomized algorithm that in a first round samples r objects, $\mathsf{R} \subseteq \mathsf{S}$ (say, segments), computes the arrangement induced by these r objects (i.e., $\mathcal{A}^{|}(\mathsf{R})$), and then inside each region σ it computes the arrangement of the $\omega(\sigma)$ objects intersecting the interior of this region, using an algorithm that takes $O\Big((\omega(\sigma))^c\Big)$ time, where $c > 0$ is some fixed constant. The overall expected running time of this algorithm is

$$\mathbb{E}\Big[\sum_{\sigma \in \mathcal{F}(\mathsf{R})} (\omega(\sigma))^c\Big].$$

We need the following theorem – we do not provide a proof of it, and instead just state it.

Theorem 1.2.1 (Bounded moments theorem). *Let $\mathsf{R} \subseteq \mathsf{S}$ be a random subset of size r. Let $\mathbb{E}f(r) = \mathbb{E}[|\mathcal{F}(\mathsf{R})|]$ and let $c \geq 1$ be an arbitrary constant. Then,*

$$\mathbb{E}\Big[\sum_{\sigma \in \mathcal{F}(\mathsf{R})} (\omega(\sigma))^c\Big] = O\Big(\mathbb{E}f(r)\, \Big(\frac{n}{r}\Big)^c\Big).$$

1.3. Algorithms using Randomized Incremental Construction

1.3.1. Convex hull in two dimensions

Let P be a set of n points in the plane. Let $P = \langle p_1, \ldots, p_n \rangle$ be a random permutation of P. Let $P_i = \langle p_1, \ldots, p_i \rangle$ be the prefix of the random permutation. Let $C_i = \mathcal{CH}(P_i)$.

Let $V_i = \mathsf{V}(C_i)$ be the set of vertices of C_i. For each vertex, we remember with it the two edges adjacent to it in C_i. We refer to this entity as a ***butterfly***. See Figure 1.3.1 for an example.

Thus, the convex-hull of a set of points is a collection of butterflies (a group of butterflies is called *kaleidoscope* according to the web). Let $\mathcal{F}_i = \mathcal{F}(C_i)$ be the butterflies of C_i. A point $p \in P \setminus P_i$ is in ***conflict*** with a butterfly β, if β is no longer present in $\mathcal{CH}(\mathsf{V}(\beta) + p)$. In particular, let $\mathrm{cl}(\beta)$ be the set of all the points of P that are in conflict with β. The list $\mathrm{cl}(\beta)$ is the ***conflict-list*** of β.

A useful concept is the ***conflict graph***. It is a graph having the butterflies of $\mathcal{F}_i$ on one side, and the points of P on the other side, with an edge between a point p and a butterfly β it they are in conflict.

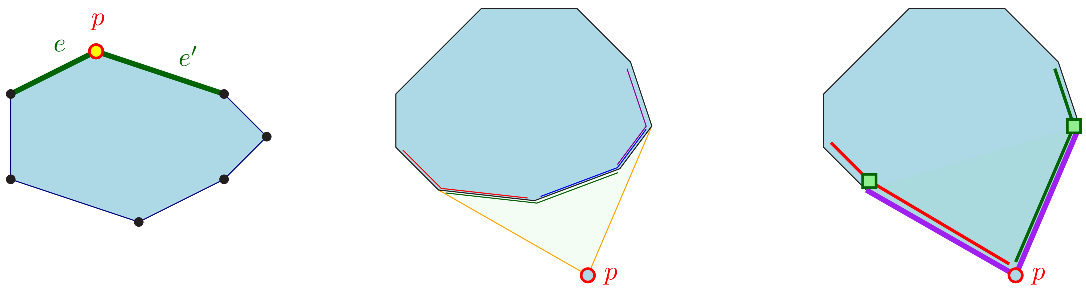

Figure 1.3.1: Left: A 2d butterfly is made out of a vertex and its two adjacent edges. Here the butterfly is (e, p, e'). Middle: The butterflies deleted by the insertions of the new point. Right: The three new butterflies created.

Imagine that we maintain this graph using a data-structure where all the natural basic operations takes constant time.

When inserting the point p_{i+1} is not in conflict with any butterfly of $\mathcal{F}_i$, then it is inside the current convex-hull and can be ignored. The more interesting case is when it is in conflict with some butterflies. The butterflies it is in conflict with form a path on the convex-hull. The two extreme butterflies in this chain are of interest, and they give birth to two new butterflies after p_{i+1} is inserted (i.e., they "flap" their wings). All the other butterflies in the middle are deleted. There is also a new butterfly in the new vertex.

Clearly, we can compute the new conflict graph for C_{i+1} in linear time in the size of the conflict lists of the two extreme butterflies being deleted.

Recap of convex-hull algorithm in 2d. We randomly permute the points of P. At each stage we compute $C_i = \mathcal{CH}(P_i)$ from C_{i-1}. To this end, we maintain the set of butterflies $\mathcal{F}_i = \mathcal{F}(C_i)$, their conflict lists, and their conflict graph. When inserting the point p_i, we delete the butterflies in conflict with the newly inserted point, and create the new three butterflies. We also compute the conflict lists of the new butterflies, from the conflict-lists of the two old butterflies. In the end of the execution of this RIC algorithm (i.e., randomized incremental construction algorithm) we have the $\mathcal{CH}(P)$).

Analysis.

Lemma 1.3.1. *Let P be a set of n points in the plane. The above algorithm computes $\mathcal{CH}(P)$ in $O(n \log n)$ expected time.*

Proof: Observe that $\mathbb{E}f(i) = |\mathcal{F}_i| = O(i)$. By Theorem 1.2.1, we have that

$$T_i = \mathbb{E}\Big[\sum_{\beta \in \mathcal{F}_i} |\mathrm{cl}(\beta)|\Big] = O\bigg(\mathbb{E}f(i) \left(\frac{n}{i}\right)^1 \bigg) = O\Big(i \cdot \frac{n}{i}\Big) = O(n).$$

Using backward analysis, and taking into account only the newly constructed butterflies (i.e., we charge the deletion work, to the creation work), we have that the work in the ith iteration is proportional to the size of the newly created conflict lists. Formally, if $X_\beta = 1 \iff \beta$ was created in the ith iteration, then we have that the "creation" work in the ith iteration is proportional to

$$W_i = \sum_{\beta \in \mathcal{F}_i} X_i \left|\mathrm{cl}(\beta)\right|.$$

Since the probability of a butterfly $\beta \in \mathcal{F}_i$ to be created in the ith iteration is at most $3/i$, we have that $\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] \leq 3/i$. As such, fix the points appearing in P_i (i.e., this is the set $\widehat{P_i} = \{p_1, \ldots, p_i\}$), and consider the different random ordering of them in P_i. Conditions on $\widehat{P_i}$, we have

$$\mathbb{E}\Big[W_i \,\Big|\, \widehat{P_i}\Big] = \mathbb{E}\Bigg[\sum_{\beta \in \mathcal{F}_i} X_i \left|\mathrm{cl}(\beta)\right| \,\Bigg|\, \widehat{P_i}\Bigg] \leq \sum_{\beta \in \mathcal{F}_i} \mathbb{E}\Big[X_i \,\Big|\, \widehat{P_i}\Big] \left|\mathrm{cl}(\beta)\right| \leq \frac{3}{i} \sum_{\beta \in \mathcal{F}_i} \left|\mathrm{cl}(\beta)\right|.$$

Thus, we have

$$\mathbb{E}[W_i] = \mathbb{E}\Big[\mathbb{E}\Big[W_i \,\Big|\, \widehat{P_i}\Big]\Big] \leq \mathbb{E}\Bigg[\frac{3}{i} \sum_{\beta \in \mathcal{F}_i} \left|\mathrm{cl}(\beta)\right|\Bigg] \leq \frac{3}{i}\, \mathbb{E}\Bigg[\sum_{\beta \in \mathcal{F}_i} \left|\mathrm{cl}(\beta)\right|\Bigg] \leq \frac{3}{i} T_i = O\Big(\frac{n}{i}\Big).$$

We thus have that the expected running time of the algorithm is

$$\mathbb{E}\Bigg[\sum_i W_i\Bigg] = \sum_{i=1}^{n} \mathbb{E}[W_i] = \sum_{i=1}^{n} O\Big(\frac{n}{i}\Big) = O(n \log n). \qquad \blacksquare$$

1.3.2. Convex hull in three dimensions

The above algorithm extends verbatim to 3d. The convex-hull of the first i points still have complexity $O(i)$. The butterflies are now two triangles hanging on a common edge. Going through the details, we get the following:

Lemma 1.3.2. *Let P be a set of n points in 3d. The above algorithm computes $\mathcal{CH}(P)$ in $O(n \log n)$ expected time.*