12.1 Alignment Polytopes

Recall that the alignment polytope is the convex hull of all alignment summaries. In last lecture, we saw that the alignment polytope for two sequences is the Newton polytope of the statistical model for the two sequences. Here are some results that give bounds on the number of vertices.

**Theorem 12.1.** The number of vertices of the alignment polytope in the case of two free parameters is $O((n + m)^{2/3})$, where $n$ and $m$ are the lengths of the sequences.

**Remark.** It is conjectured that the number of vertices is $\Theta((n + m)^{2/3})$ for any fixed alphabet. In fact, it is not even known if this is true for the case of a binary alphabet $\{0, 1\}$.

**Theorem 12.2.** (Upper-bound for number of vertices, G. Andrews, 1967) For any fixed number $d$, there is a constant $C_d$ such that the number of vertices of any lattice polytope $P$ in $\mathbb{R}^d$ is bounded from above by $C_d \left( \operatorname{vol}(P)^{\frac{d-1}{d+1}} \right)$.

Note that Theorem 12.2 implies Theorem 12.1, because in the case of two free parameters, $\operatorname{vol}(P) = O((n + m)^{d-1/(d+1)})$ so for $d = 2$, the number of vertices of $P$ is $O((n + m)^{2/3})$. Moreover, in the case $d = 2$, this theorem gives a good running time bound for the polytope propagation algorithm. Indeed, the run times for the Minkowski sum and the convex hull of unions are both linear in the number of vertices and the polytope propagation algorithm executes these two operations over a grid of size $nm$, so the algorithm has a time complexity of $O(nm(n + m)^{2/3})$.

12.2 Normal Fan

We would like to decompose the parameter space of the alignment model into regions which correspond with faces of the alignment polytope. The construction is as follows:

1. Fix a polytope $P \subset \mathbb{R}^d$. Given a vector $w \in \mathbb{R}^d$, let

$$\text{face}_w(P) = \{x \in P \mid x \cdot w \leq y \cdot w, \forall y \in P\}.$$ 

Note that $\text{face}_w(P)$ is a polytope because it is the intersection of $P$ with the hyperplane $x \cdot w = \min\{x \cdot w : x \in P\}$. We call such a polytope a face of $P$. 
2. For $F$ a face of polytope $P$, define the normal cone of $P$ at $F$ to be

$$N_P(F) = \{ w \in \mathbb{R}^d \mid \text{face}_w(P) = F \}.$$ 

3. Finally, the normal fan of $P$ is the collection of all the normal cones of $P$.

$$N(P) = \{ N_P(F) \mid F \text{ a face of } P \}.$$ 

Note that $N(P)$ partitions $\mathbb{R}^d$ into cones which are in bijection with the faces of $P$. As an example, suppose we have the polytope

For each face, we have a normal cone, displayed here:

The collection of these cones gives the normal fan of the polytope.
12.3 Graham Scan

The Graham scan is an algorithm that computes the convex hull of a set of points in $\mathbb{R}^2$. In particular, we can use it to compute the convex hull of the union of two polytopes in $\mathbb{R}^2$ by taking the vertices of the two polytopes and the points where their boundaries intersect as input for the algorithm. The following implementation of the Graham scan takes a set $S$ of $n$ distinct points and returns a sequence $T$ of points on the boundary of the convex hull.

1. Pick a point $p_0 \in S$ lying on the boundary of the convex hull of $S$. This can be done by finding a point which minimizes a linear functional, e.g. the $y$-coordinate.

2. Sort the remaining points, arranging them clockwise according to their radial angle from $p_0$. Label them $p_1, \ldots, p_{n-1}, p_n = p_0$. This sorting has complexity $O(n \log n)$.

3. If $S$ has only one or two points, then return $T = (p_0)$ or $T = (p_0, p_1)$ respectively. Otherwise, initialize $T = (p_0, p_1)$ and $i = 2$.

4. Let $x$ and $y$ be the last and second-to-last points in $T$ respectively. If we make a right turn going from $y$ to $x$ to $p_i$, append $p_i$ to the end of $T$ and increase $i$ by 1. If we make a left turn, then delete $x$ from $T$. Repeat this step if $i < n$.

5. Output $T$.

12.4 Inference Functions

In earlier lectures, we studied some graphical models with hidden variables, such as the hidden Markov model. In those situations, given an observation, it would be useful to find the most likely hidden data that caused this observation. Given fixed parameters, an inference function is a map that assigns a most likely hidden data point to each observation. We call this most likely hidden data an explanation for the observation.

For example, in the case of sequence alignments, fixing the parameters $M$, $X$ and $G$ gives an inference function

$$\Phi_{M,X,G} : \{A, C, G, T\}^{n+m} \rightarrow \mathcal{A}_{n,m}$$

which maps each pair of sequences of length $n$ and $m$ to an alignment maximizing the score $mM + xX + gG$. Here, $\mathcal{A}_{n,m}$ is the set of global alignments for the two sequences. We order the elements of $\mathcal{A}_{n,m}$ to break ties when more than one alignment maximizes the above score. In general, consider a graphical model $f$ which has the $k$ observable variables with the alphabet $\Sigma'$ and $l$ hidden variables with the alphabet $\Sigma$. Then, the inference function corresponding to given parameters $\theta$ is of the form

$$\Phi_{\theta} : (\Sigma')^k \rightarrow \Sigma^l.$$
An interesting question is the following: how many different maps $\Phi_{\theta}$ are there as the parameters $\theta$ are varied? For sequence alignments, a trivial upper bound would be

$$\left(\frac{n + m}{n}\right)^{4n+m}.$$ 

This bound is doubly exponential in $n+m$. However, as the theorem below shows, most of the functions counted in the bound are not inference functions for any values of the parameters.

**Theorem 12.3.** The number of inference functions is $O((n + m)^2)$ for the 2-parameter alignment model. In fact, it is $\Theta((n + m)^2)$. More generally, for a graphical model of complexity $N$, the bound is $O(N^{d(d-1)})$.

**Remark.** The complexity of a graphical model will not be defined here, but it is usually bounded linearly in the number $k$ of observable variables.

**Example.** Consider the hidden Markov model for binary states of length 5 represented by the graph below, where the dark nodes are observed and the white nodes are hidden.

![Graph](image)

There are 32 possible observations and 32 possible explanations. The trivial upper bound is thus $32^{32}$ but the actual number of inference functions is only 5266.

To see why this theorem is true, we first need to introduce some new concepts and results. Let $N(P)$ denote the normal fan of a polytope $P$. For polytopes $P_1, \ldots, P_k$, the common refinement of their normal fans is the collection of cones obtained as the intersection of a cone from each of the individual fans, and is denoted by $N(P_1) \land \cdots \land N(P_k)$.

**Lemma 12.4.**

$$N(P_1) \land \cdots \land N(P_k) = N(\circ_{i=1}^k P_i)$$

**Proof:** Assigned as a homework. □

**Theorem 12.5.** (Gritzmann-Sturmfels) Let $P_1, \ldots, P_k$ be polytopes in $\mathbb{R}^d$, and let $m$ denote the number of nonparallel edges. Then the number of vertices of $P_1 \circ \cdots \circ P_k$ is bounded above by

$$2 \sum_{j=0}^{d-1} \binom{m-1}{j}$$
**Remark.** Note that this bound does not depend on $k$.

Recall that for every sequence pair $\sigma$, the alignment model with parameters $\theta$ is given by a polynomial $f_\sigma(\theta)$ where each monomial of $f_\sigma$ represents the probability of some alignment of the two sequences. Then, the alignment polytope $P_\sigma$ is the Newton polytope of $f_\sigma$ and its normal fan is $N(P_\sigma)$. Each vertex of $P_\sigma$ represents some alignment summary, so more than one alignment may be mapped to the same vertex.

An optimal alignment for given parameters $\theta \in \mathbb{R}^d$ is an alignment mapped to a vertex on $P_\sigma$ that is maximal in the direction of $\tilde{\theta} = \log \theta$. Ties between optimal alignments are broken by a predetermined order on the set of alignments. In the language of normal cones, each cone in the normal fan $N(P_\sigma)$ represents a set of parameter vectors $\tilde{\theta}$ with the same optimal alignment. As we vary over all sequence pairs $\sigma$, for two parameter vectors $\tilde{\theta}$ and $\tilde{\theta}'$, we see that $\Phi_{\tilde{\theta}}(\sigma) = \Phi_{\tilde{\theta}'}(\sigma)$ if and only if $\tilde{\theta}$ and $\tilde{\theta}'$ lie in the same of cone of the common refinement $N(P_1) \wedge \cdots \wedge N(P_k) = N(\circ_{i=1}^k P_i)$. Hence, we have the following inequality:

$$\text{number of inference functions} \leq \text{number of vertices of } \circ_{\sigma} P_\sigma.$$ 

Now, since each $P_\sigma$ lies in $[0, n+m]^d$, every edge is a vector with coordinates in $[-(n+m), n+m]$. Therefore, the number of non-parallel edges is at most $(2(n+m)+1)^d$, and

$$\text{number of vertices of } \circ_{\sigma} P_\sigma \leq 2 \sum_{j=0}^{d-1} \binom{(2(n+m)+1)^d - 1}{j} \leq \frac{2^{d^2-d+1}}{(d-1)!} (n+m)^{d(d-1)}.$$

### 12.5 Homework

Show that

$$N(P_1) \wedge \cdots \wedge N(P_k) = N(\circ_{i=1}^k P_i).$$