Bounds on the number of inference functions of a graphical model

Sergi Elizalde

Dartmouth College

Kevin Woods

UC Berkeley

FPSAC 2006

Overview

- Background
  - Sequence alignment
  - Hidden Markov model
  - Graphical models
  - Inference functions
- Upper bound on the number of inference functions

  Sketch of the proof:
  - From inference functions to vertices of a polytope
  - The number of vertices of a Minkowski sum of polytopes
- Lower bound on the number of inference functions
- Application to sequence alignment

Sequence alignment

Consider the sequences $S = \texttt{AGGAATTCG}$ and $T = \texttt{GGACGGTAT}$.

Consider the sequences $S = \texttt{AGGAATTCG}$ and $T = \texttt{GGACGGTAT}$.

Which of the following two alignments is better?

$S_1 =$ AGGA-ATTCG	$S_2 =$ --A-GGAATTCG
$T_1 =$ -GGACGGTAT	$T_2 =$ GGACGGTAT---

Consider the sequences $S = \texttt{AGGAATTCG}$ and $T = \texttt{GGACGGTAT}$.

Which of the following two alignments is better?

$S_1 = \texttt{AGGA-ATTCG}$	$S_2 = \texttt{--A-GGAATTCG}$
$T_1 = \texttt{-GGACGGTAT}$	$T_2 = \texttt{GGACGGTAT---}$

4	$\longleftarrow$	$z = \#$matches	$\longrightarrow$	5
4	$\longleftarrow$	$x = \#$mismatches	$\longrightarrow$	1
1	$\longleftarrow$	$y = \#$insertions	$\longrightarrow$	3

Consider the sequences $S = \texttt{AGGAATTCG}$ and $T = \texttt{GGACGGTAT}$.

Which of the following two alignments is better?

$S_1 = \texttt{AGGA-ATTCG}$	$S_2 = \texttt{--A-GGAATTCG}$
$T_1 = \texttt{-GGACGGTAT}$	$T_2 = \texttt{GGACGGTAT---}$

$$4 \longleftarrow z = \#\text{matches} \longrightarrow 5$$
$$4 \longleftarrow x = \#\text{mismatches} \longrightarrow 1$$
$$1 \longleftarrow y = \#\text{insertions} \longrightarrow 3$$

One possible model: matches are rewarded by 1, mismatches are penalized by α, and insertions are penalized by β.

$$4 - 4\alpha - \beta \longleftarrow \text{score} := z - x\alpha - y\beta \longrightarrow 5 - \alpha - 3\beta$$

Inference functions for sequence alignment

- An *optimal alignment* is one with highest score.

  Optimality depends on the values of the parameters α and β.

Inference functions for sequence alignment

- An *optimal alignment* is one with highest score.

  Optimality depends on the values of the parameters α and β.

- For fixed parameters, we get a function that maps each pair of sequences to their optimal alignment.

  Such a map is called an *inference function*.

Inference functions for sequence alignment

- An *optimal alignment* is one with highest score.

  Optimality depends on the values of the parameters α and β.

- For fixed parameters, we get a function that maps each pair of sequences to their optimal alignment.

  Such a map is called an *inference function*.

- The total number of maps from pairs of sequences of length n to alignments is doubly exponential in n.

  However, most of them are not inference functions for any parameter values.

Inference functions for sequence alignment

- An *optimal alignment* is one with highest score.

  Optimality depends on the values of the parameters α and β.

- For fixed parameters, we get a function that maps each pair of sequences to their optimal alignment.

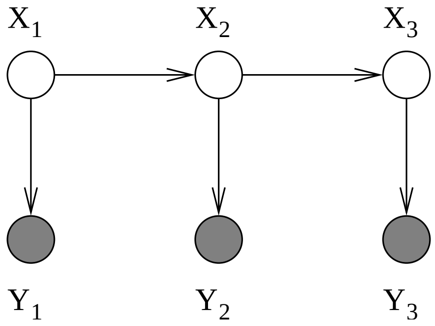
  Such a map is called an *inference function*.

- The total number of maps from pairs of sequences of length n to alignments is doubly exponential in n.

  However, most of them are not inference functions for any parameter values.

Problem: How many of them are inference functions?

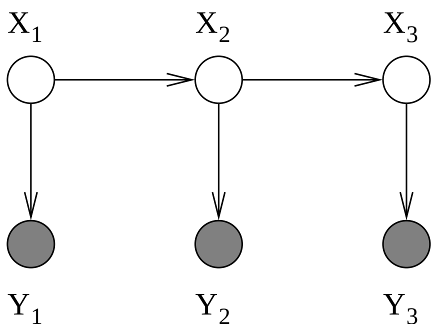
Hidden Markov model



} hidden variables $X_i \in \Sigma$

} observed variables $Y_i \in \Sigma'$

Hidden Markov model



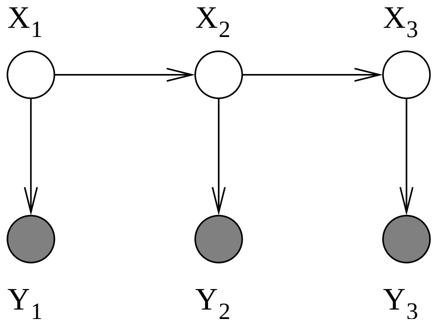
} hidden variables $X_i \in \Sigma$

} observed variables $Y_i \in \Sigma'$

For $i, j \in \Sigma$, $\ell \in \Sigma'$,

$$\theta_{i\ell} \;=\; \text{transition probability from } X_k = i \text{ to } Y_k = \ell$$

$$\theta_{ij} \;=\; \text{transition probability from } X_k = i \text{ to } X_{k+1} = j$$



} hidden variables $X_i \in \Sigma = \{\mathrm{in}, \mathrm{ex}\}$

} observed variables $Y_i \in \Sigma' = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$

For $i, j \in \Sigma$, $\ell \in \Sigma'$,

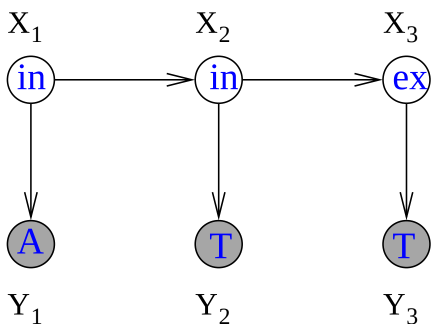
$\theta_{i\ell}$ = transition probability from $X_k = i$ to $Y_k = \ell$

θ_{ij} = transition probability from $X_k = i$ to $X_{k+1} = j$

} 12 parameters

This model can be used to determine what parts of the genome are introns and what parts are exons.

Hidden Markov model



} hidden variables $X_i \in \Sigma = \{\mathrm{in}, \mathrm{ex}\}$

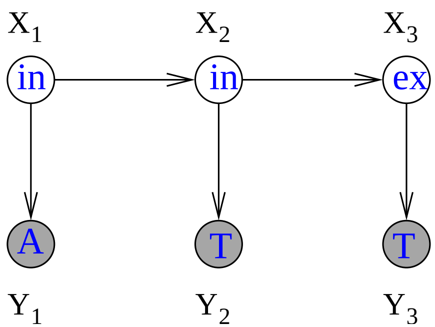
} observed variables $Y_i \in \Sigma' = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$

The model is represented by a polynomial map

$$\mathbf{f} : \quad \mathbb{R}^{12} \longrightarrow \mathbb{R}^{64}$$

$$(\theta_{\mathrm{in},\mathrm{in}}, \theta_{\mathrm{in},\mathrm{ex}}, \theta_{\mathrm{in},\mathtt{A}}, \ldots) \mapsto (f_{\mathtt{AAA}}, f_{\mathtt{AAC}}, f_{\mathtt{AAG}}, f_{\mathtt{AAT}}, f_{\mathtt{ACA}}, \ldots)$$

Hidden Markov model



} hidden variables $X_i \in \Sigma = \{\mathrm{in}, \mathrm{ex}\}$

} observed variables $Y_i \in \Sigma' = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$

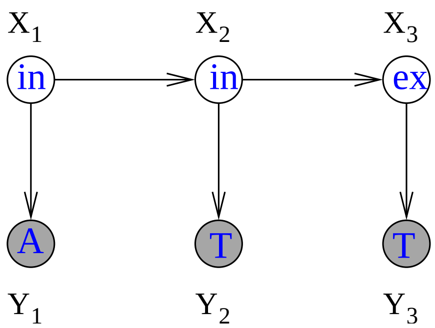
The model is represented by a polynomial map

$$\mathbf{f} : \quad \mathbb{R}^{12} \longrightarrow \mathbb{R}^{64}$$

$$(\theta_{\mathrm{in},\mathrm{in}}, \theta_{\mathrm{in},\mathrm{ex}}, \theta_{\mathrm{in},\mathtt{A}}, \ldots) \mapsto (f_{\mathtt{AAA}}, f_{\mathtt{AAC}}, f_{\mathtt{AAG}}, f_{\mathtt{AAT}}, f_{\mathtt{ACA}}, \ldots)$$

$$f_{\mathtt{ATT}} = \underbrace{\frac{1}{2}\theta_{\mathrm{in},\mathtt{A}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}}_{X_1=X_2=X_3=\mathrm{in}} + \underbrace{\frac{1}{2}\theta_{\mathrm{in},\mathtt{A}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}\theta_{\mathrm{in},\mathrm{ex}}\theta_{\mathrm{ex},\mathtt{T}}}_{X_1=X_2=\mathrm{in},\ X_3=\mathrm{ex}} + \cdots$$

(each coordinate gives the probability of an outcome)



} hidden variables $X_i \in \Sigma = \{\mathrm{in}, \mathrm{ex}\}$

} observed variables $Y_i \in \Sigma' = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$

The model is represented by a polynomial map

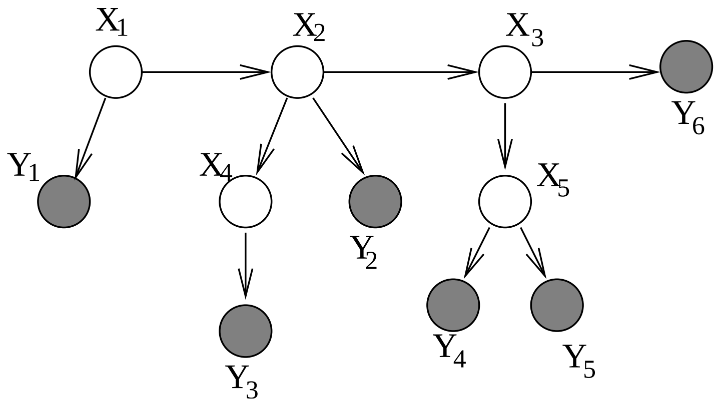
$$\mathbf{f}: \quad \mathbb{R}^{12} \longrightarrow \mathbb{R}^{64}$$
$$(\theta_{\mathrm{in},\mathrm{in}}, \theta_{\mathrm{in},\mathrm{ex}}, \theta_{\mathrm{in},\mathtt{A}}, \ldots) \mapsto (f_{\mathtt{AAA}}, f_{\mathtt{AAC}}, f_{\mathtt{AAG}}, f_{\mathtt{AAT}}, f_{\mathtt{ACA}}, \ldots)$$

$$f_{\mathtt{ATT}} = \underbrace{\frac{1}{2}\theta_{\mathrm{in},\mathtt{A}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}}_{X_1=X_2=X_3=\mathrm{in}} + \underbrace{\frac{1}{2}\theta_{\mathrm{in},\mathtt{A}}\theta_{\mathrm{in},\mathrm{in}}\theta_{\mathrm{in},\mathtt{T}}\theta_{\mathrm{in},\mathrm{ex}}\theta_{\mathrm{ex},\mathtt{T}}}_{X_1=X_2=\mathrm{in},\ X_3=\mathrm{ex}} + \cdots$$

(each coordinate gives the probability of an outcome)

Given an observation (e.g. ATT), one wants to find the most likely values of X_1, X_2, X_3.

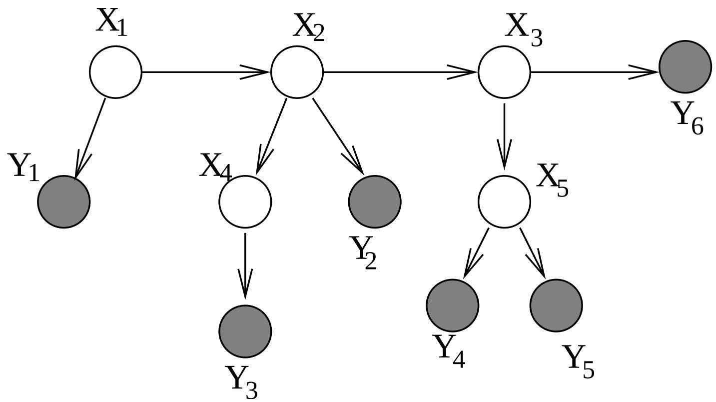
More generally:



Hidden: $X_1, X_2, \ldots, X_q \in \Sigma$

Observed: $Y_1, Y_2, \ldots, Y_n \in \Sigma'$

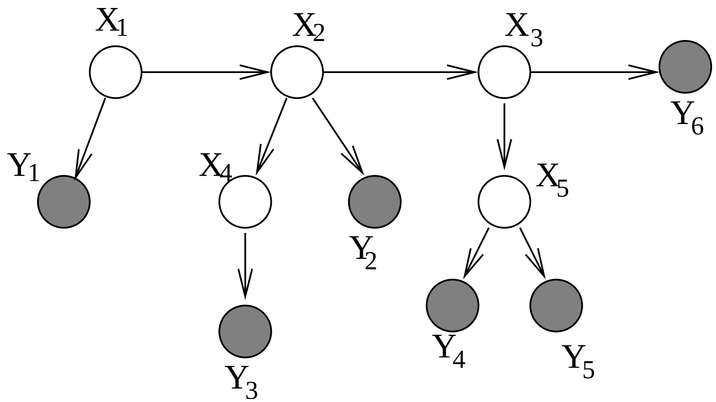
More generally:



Hidden: $X_1, X_2, \ldots, X_q \in \Sigma$
Observed: $Y_1, Y_2, \ldots, Y_n \in \Sigma'$

$$\mathbf{f} : \quad \mathbb{R}^d \longrightarrow \mathbb{R}^{(l')^n}$$
$$(\theta_1, \theta_2, \ldots, \theta_d) \mapsto (f_\tau, \ldots)$$

More generally:



Hidden: $X_1, X_2, \ldots, X_q \in \Sigma$

Observed: $Y_1, Y_2, \ldots, Y_n \in \Sigma'$

$$\mathbf{f} : \quad \mathbb{R}^d \longrightarrow \mathbb{R}^{(l')^n}$$
$$(\theta_1, \theta_2, \ldots, \theta_d) \mapsto (f_\tau, \ldots)$$

For each observation $\tau \in (\Sigma')^n$,

$$f_\tau = \mathrm{Prob}(\mathbf{Y} = \tau) = \sum_{\mathbf{h} \in \Sigma^q} \underbrace{\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau)}_{\text{monomial in } \theta_1, \theta_2, \ldots, \theta_d}$$

Basic inference problem: Given an observation τ, determine the value of the hidden data $\mathbf{h}$ that maximizes

$$\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau).$$

Basic inference problem: Given an observation τ, determine the value of the hidden data $\mathbf{h}$ that maximizes

$$\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau).$$

- The solution $\widehat{\mathbf{h}}$ is called an *explanation* of τ, and it depends on the values of the parameters $\theta_1, \theta_2, \ldots, \theta_d$.

Basic inference problem: Given an observation τ, determine the value of the hidden data $\mathbf{h}$ that maximizes

$$\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau).$$

- The solution $\widehat{\mathbf{h}}$ is called an *explanation* of τ, and it depends on the values of the parameters $\theta_1, \theta_2, \ldots, \theta_d$.

Definition. An *inference function* is a map

$$\begin{array}{ccc} (\Sigma')^n & \longrightarrow & \Sigma^q \\ \tau & \mapsto & \widehat{\mathbf{h}} \end{array}$$

that sends each observation to its explanation.

- Each choice of parameter values defines an inference function.

Basic inference problem: Given an observation τ, determine the value of the hidden data $\mathbf{h}$ that maximizes

$$\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau).$$

- The solution $\widehat{\mathbf{h}}$ is called an *explanation* of τ, and it depends on the values of the parameters $\theta_1, \theta_2, \ldots, \theta_d$.

Definition. An *inference function* is a map

$$\begin{array}{ccc} (\Sigma')^n & \longrightarrow & \Sigma^q \\ \tau & \mapsto & \widehat{\mathbf{h}} \end{array}$$

that sends each observation to its explanation.

- Each choice of parameter values defines an inference function.

In the previous example, an inference function is a map $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}^n \longrightarrow \{\mathrm{in}, \mathrm{ex}\}^n$. These are called *gene finding functions*.

The number of inference functions

In general, if $l = |\Sigma|$, $l' = |\Sigma'|$, there are in total $l^{q(l')^n}$ functions

$$(\Sigma')^n \longrightarrow \Sigma^q.$$

The number of inference functions

In general, if $l = |\Sigma|$, $l' = |\Sigma'|$, there are in total $l^{q(l')^n}$ functions

$$(\Sigma')^n \longrightarrow \Sigma^q.$$

However, most of these maps are not inference functions for any value of the parameters.

In general, if $l = |\Sigma|$, $l' = |\Sigma'|$, there are in total $l^{q(l')^n}$ functions

$$(\Sigma')^n \longrightarrow \Sigma^q.$$

However, most of these maps are not inference functions for any value of the parameters.

Problem: How many of them are inference functions?

The number of inference functions

In general, if $l = |\Sigma|$, $l' = |\Sigma'|$, there are in total $l^{q(l')^n}$ functions

$$(\Sigma')^n \longrightarrow \Sigma^q.$$

However, most of these maps are not inference functions for any value of the parameters.

Problem: How many of them are inference functions?

Main Theorem. Fix $d > 0$. Consider a graphical model with d parameters, and let E be the number of edges of the underlying graph. Then, the number of inference functions of the model is at most

$$O(E^{d(d-1)}).$$

In general, if $l = |\Sigma|$, $l' = |\Sigma'|$, there are in total $l^{q(l')^n}$ functions

$$(\Sigma')^n \longrightarrow \Sigma^q.$$

However, most of these maps are not inference functions for any value of the parameters.

Problem: How many of them are inference functions?

Main Theorem. Fix $d > 0$. Consider a graphical model with d parameters, and let E be the number of edges of the underlying graph. Then, the number of inference functions of the model is at most

$$O(E^{d(d-1)}).$$

Usually, E is a linear in n, so the number of inference functions is in fact $O(n^{d(d-1)})$.

Geometric interpretation of the inference problem

Definition. The *Newton polytope* of

$$f_{\tau}(\theta_1, \theta_2, \ldots, \theta_d) = \sum_i \theta_1^{a_{1,i}} \theta_2^{a_{2,i}} \cdots \theta_d^{a_{d,i}},$$

$\mathrm{NP}(f_{\tau})$, is the convex hull of the vectors $(a_{1,i}, a_{2,i}, \ldots, a_{d,i}) \in \mathbb{R}^d$.

Geometric interpretation of the inference problem

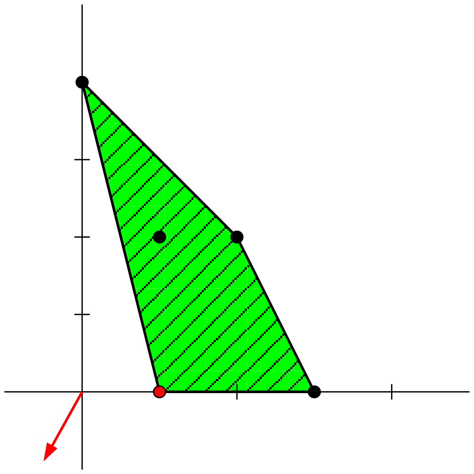
Definition. The *Newton polytope* of

$$f_{\tau}(\theta_1, \theta_2, \ldots, \theta_d) = \sum_i \theta_1^{a_{1,i}} \theta_2^{a_{2,i}} \cdots \theta_d^{a_{d,i}},$$

$\mathrm{NP}(f_{\tau})$, is the convex hull of the vectors $(a_{1,i}, a_{2,i}, \ldots, a_{d,i}) \in \mathbb{R}^d$.

Example.

$$f(\theta_1, \theta_2) = \theta_1^3 + \theta_1^2\theta_2^2 + \theta_1\theta_2^2 + \theta_1 + \theta_2^4$$



Geometric interpretation of the inference problem

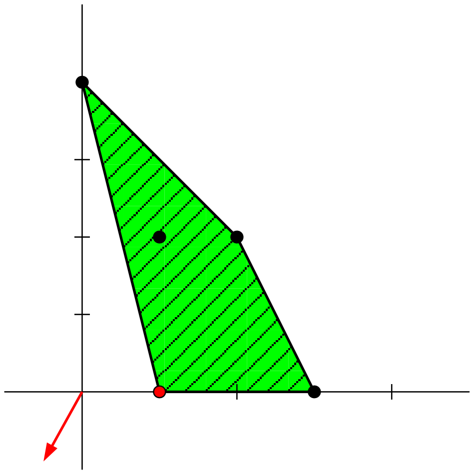
Definition. The *Newton polytope* of

$$f_{\tau}(\theta_1, \theta_2, \ldots, \theta_d) = \sum_i \theta_1^{a_{1,i}} \theta_2^{a_{2,i}} \cdots \theta_d^{a_{d,i}},$$

$\mathrm{NP}(f_{\tau})$, is the convex hull of the vectors $(a_{1,i}, a_{2,i}, \ldots, a_{d,i}) \in \mathbb{R}^d$.

Example.

$$f(\theta_1, \theta_2) = \theta_1^3 + \theta_1^2\theta_2^2 + \theta_1\theta_2^2 + \theta_1 + \theta_2^4$$



For fixed parameters θ_i, if we let $v_i = \log \theta_i$, then the explanation $\widehat{\mathbf{h}}$ maximizing $\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau)$ is given by the vertex of $\mathrm{NP}(f_{\tau})$ that maximizes $v_1 x_1 + \ldots + v_d x_d$.

Geometric interpretation of the inference problem

Definition. The *Newton polytope* of

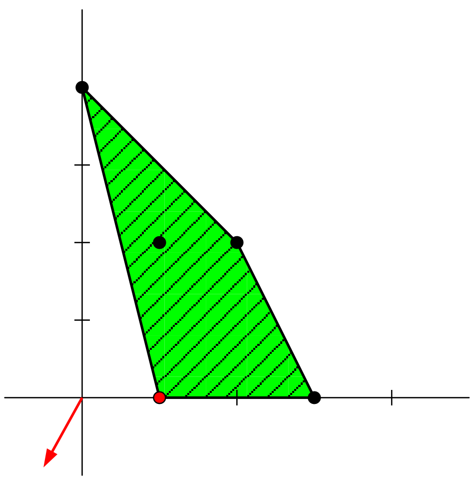
$$f_{\tau}(\theta_1, \theta_2, \ldots, \theta_d) = \sum_i \theta_1^{a_{1,i}} \theta_2^{a_{2,i}} \cdots \theta_d^{a_{d,i}},$$

$\mathrm{NP}(f_{\tau})$, is the convex hull of the vectors $(a_{1,i}, a_{2,i}, \ldots, a_{d,i}) \in \mathbb{R}^d$.

Example.

$$f(\theta_1, \theta_2) = \theta_1^3 + \theta_1^2\theta_2^2 + \theta_1\theta_2^2 + \theta_1 + \theta_2^4$$

$$\theta_1 = 0.6, \quad \theta_2 = 0.4,$$
$$\mathbf{v} = (\log\theta_1, \log\theta_2) = (-0.51, -0.92)$$

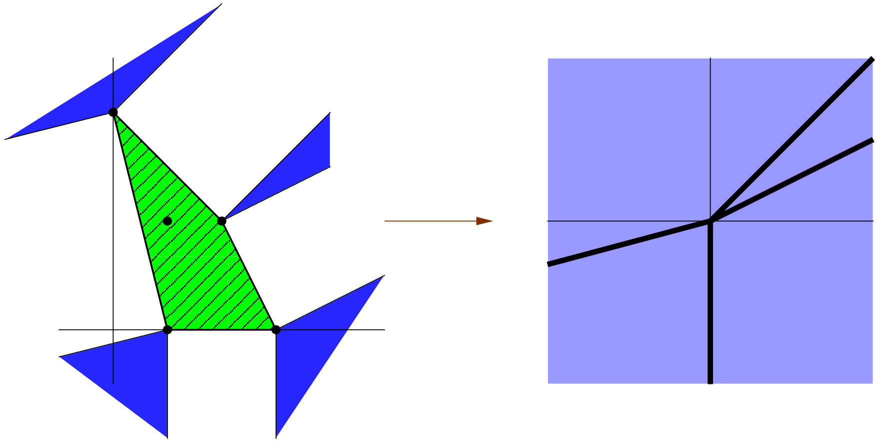


For fixed parameters θ_i, if we let $v_i = \log\theta_i$, then the explanation $\widehat{\mathbf{h}}$ maximizing $\mathrm{Prob}(\mathbf{X} = \mathbf{h}, \mathbf{Y} = \tau)$ is given by the vertex of $\mathrm{NP}(f_{\tau})$ that maximizes $v_1 x_1 + \ldots + v_d x_d$.

Let
$$\Phi_{\mathbf{v}} : (\Sigma')^n \longrightarrow \Sigma^q$$
denote the inference function where $\mathbf{v} = (\log(\theta_1), \ldots, \log(\theta_d)) \in \mathbb{R}^d$.

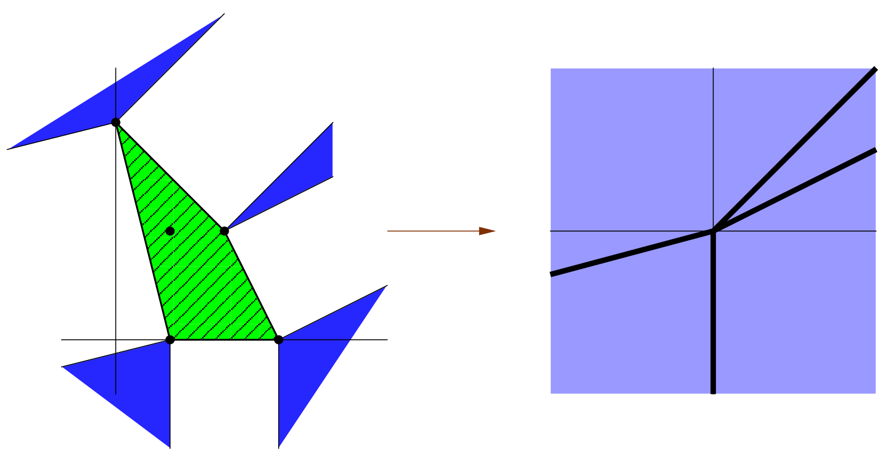
Let
$$\Phi_{\mathbf{v}} : (\Sigma')^n \longrightarrow \Sigma^q$$
denote the inference function where $\mathbf{v} = (\log(\theta_1), \ldots, \log(\theta_d)) \in \mathbb{R}^d$.

Definition. The *normal fan* $\mathcal{F}(P)$ of a polytope P is the decomposition of the space according to the directions in which each face is maximal.



Let
$$\Phi_{\mathbf{v}} : (\Sigma')^n \longrightarrow \Sigma^q$$
denote the inference function where $\mathbf{v} = (\log(\theta_1), \ldots, \log(\theta_d)) \in \mathbb{R}^d$.

Definition. The *normal fan* $\mathcal{F}(P)$ of a polytope P is the decomposition of the space according to the directions in which each face is maximal.



$\Phi_{\mathbf{v}}(\tau) = \Phi_{\mathbf{v}'}(\tau)$ iff $\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$.

Common refinement of fans

$$\Phi_{\mathbf{v}}(\tau) = \Phi_{\mathbf{v}'}(\tau) \quad \Longleftrightarrow \quad \text{$\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of } \mathcal{F}(\mathrm{NP}(f_\tau))$$

Common refinement of fans

$\Phi_{\mathbf{v}}(\tau) = \Phi_{\mathbf{v}'}(\tau) \iff$ $\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$

$\Phi_{\mathbf{v}} \equiv \Phi_{\mathbf{v}'}$ as inference functions $\iff$ $\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$ for all observations $\tau \in (\Sigma')^n$

Common refinement of fans

$$\Phi_{\mathbf{v}}(\tau) = \Phi_{\mathbf{v}'}(\tau) \quad \Longleftrightarrow$$

$\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$

$\Phi_{\mathbf{v}} \equiv \Phi_{\mathbf{v}'}$ as inference functions $\Longleftrightarrow$

$\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$ for all observations $\tau \in (\Sigma')^n$

$$\Updownarrow$$

$\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of

$$\underbrace{\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau))}_{\text{common refinement of fans}}$$

Common refinement of fans

$\Phi_{\mathbf{v}}(\tau) = \Phi_{\mathbf{v}'}(\tau) \iff$ $\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$

$\Phi_{\mathbf{v}} \equiv \Phi_{\mathbf{v}'}$ as inference functions $\iff$ $\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\mathcal{F}(\mathrm{NP}(f_\tau))$ for all observations $\tau \in (\Sigma')^n$

$\Updownarrow$

$\mathbf{v}$ and $\mathbf{v}'$ belong to the same cone of $\underbrace{\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau))}_{\text{common refinement of fans}}$

The number of inference functions equals the **number of cones in the common refinement of fans**.

Minkowski sum of polytopes

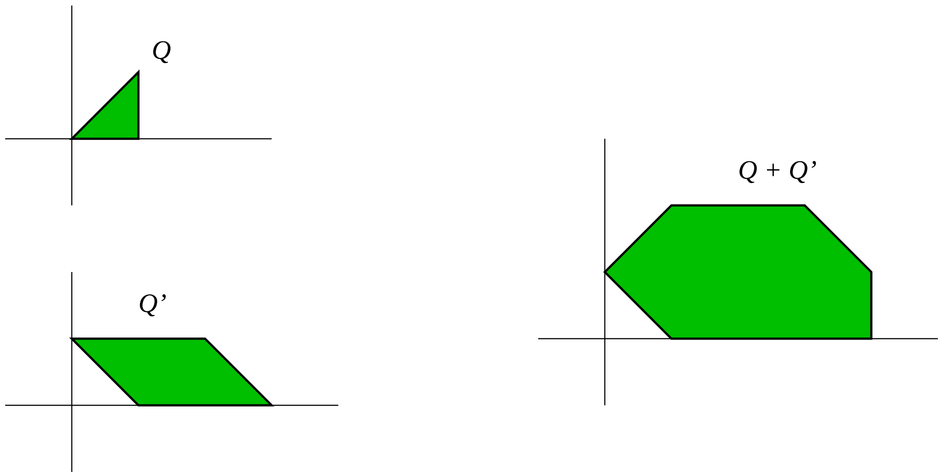
The common refinement of the normal fans is the normal fan of the Minkowski sum of polytopes:

$$\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau)) = \mathcal{F}\Big(\bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)\Big)$$

The common refinement of the normal fans is the normal fan of the Minkowski sum of polytopes:

$$\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau)) = \mathcal{F}\Big(\bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)\Big)$$

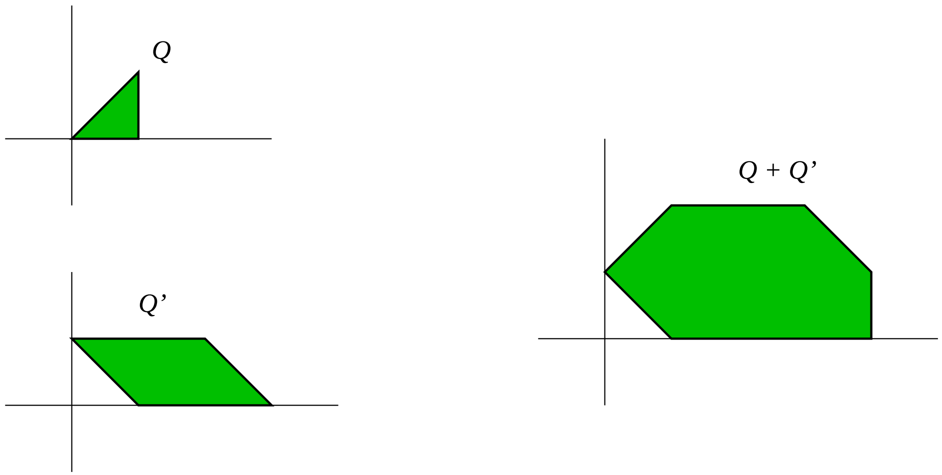
Definition. *Minkowski sum*: $Q \oplus Q' := \{\mathbf{x} + \mathbf{x}' : \mathbf{x} \in Q,\ \mathbf{x}' \in Q'\}$.



The common refinement of the normal fans is the normal fan of the Minkowski sum of polytopes:

$$\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau)) = \mathcal{F}\Big(\bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)\Big)$$

Definition. *Minkowski sum*: $Q \oplus Q' := \{\mathbf{x} + \mathbf{x}' : \mathbf{x} \in Q,\, \mathbf{x}' \in Q'\}$.



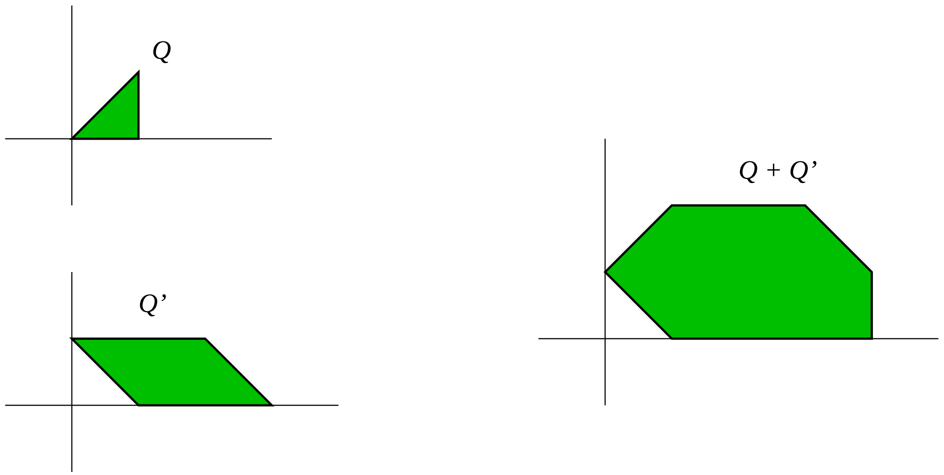
Let $P := \bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)$.

Minkowski sum of polytopes

The common refinement of the normal fans is the normal fan of the Minkowski sum of polytopes:

$$\bigwedge_{\tau \in (\Sigma')^n} \mathcal{F}(\mathrm{NP}(f_\tau)) = \mathcal{F}\Big(\bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)\Big)$$

Definition. *Minkowski sum*: $Q \oplus Q' := \{\mathbf{x} + \mathbf{x}' : \mathbf{x} \in Q,\ \mathbf{x}' \in Q'\}$.



Let $P := \bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)$.

The number of inference functions equals the **number of vertices of** P.

Edges of $\mathrm{NP}(f_\tau)$

- For each τ,

$$\mathrm{NP}(f_\tau) \subseteq [0, E]^d,$$

since each θ_i can appear as a factor of a monomial of f_τ at most E times.

Edges of $\mathrm{NP}(f_\tau)$

- For each τ,

$$\mathrm{NP}(f_\tau) \subseteq [0, E]^d,$$

  since each θ_i can appear as a factor of a monomial of f_τ at most E times.

- Besides, $\mathrm{NP}(f_\tau)$ is a lattice polytope (i.e., its vertices have integral coordinates).

Edges of $\mathrm{NP}(f_\tau)$

- For each τ,
$$\mathrm{NP}(f_\tau) \subseteq [0, E]^d,$$
since each θ_i can appear as a factor of a monomial of f_τ at most E times.

- Besides, $\mathrm{NP}(f_\tau)$ is a lattice polytope (i.e., its vertices have integral coordinates).

$$\Downarrow$$

Edges of $\mathrm{NP}(f_\tau)$ are vectors where each coordinate is an integer between $-E$ and E.

Edges of $\mathrm{NP}(f_\tau)$

- For each τ,

  $$\mathrm{NP}(f_\tau) \subseteq [0, E]^d,$$

  since each θ_i can appear as a factor of a monomial of f_τ at most E times.

- Besides, $\mathrm{NP}(f_\tau)$ is a lattice polytope (i.e., its vertices have integral coordinates).

$$\Downarrow$$

Edges of $\mathrm{NP}(f_\tau)$ are vectors where each coordinate is an integer between $-E$ and E.

$$\Downarrow$$

The polytopes $\mathrm{NP}(f_\tau)$ have at most $(2E+1)^d$ nonparallel edges in total.

Vertices of P

The polytopes $\mathrm{NP}(f_\tau)$ have at most $(2E+1)^d$ nonparallel edges in total.

$$\Big\Downarrow \text{ [Gritzmann, Sturmfels]}$$

The number of vertices of $P = \bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)$ is at most

$$2 \sum_{j=0}^{d-1} \binom{(2E+1)^d - 1}{j}.$$

Vertices of P

The polytopes $\mathrm{NP}(f_\tau)$ have at most $(2E+1)^d$ nonparallel edges in total.

$$\Big\Downarrow \text{ [Gritzmann, Sturmfels]}$$

The number of vertices of $P = \bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)$ is at most

$$2 \sum_{j=0}^{d-1} \binom{(2E+1)^d - 1}{j}.$$

As $E \to \infty$, the dominant term is $\frac{2^{d^2-d+1}}{(d-1)!} E^{d(d-1)}$.

The polytopes $\mathrm{NP}(f_\tau)$ have at most $(2E+1)^d$ nonparallel edges in total.

$$\Downarrow \text{[Gritzmann, Sturmfels]}$$

The number of vertices of $P = \bigoplus_{\tau \in (\Sigma')^n} \mathrm{NP}(f_\tau)$ is at most

$$2 \sum_{j=0}^{d-1} \binom{(2E+1)^d - 1}{j}.$$

As $E \to \infty$, the dominant term is $\frac{2^{d^2-d+1}}{(d-1)!} E^{d(d-1)}$.

$$\Downarrow$$

The number of inference functions is $O(E^{d(d-1)})$. $\square$

A lower bound

Theorem. *The above upper bound* $O(E^{d(d-1)})$ *on the number of inference functions is tight.*

A lower bound

Theorem. *The above upper bound* $O(E^{d(d-1)})$ *on the number of inference functions is tight.*

Idea of the proof:

- Construct an HMM $\mathcal{M}_n$ of length n with d parameters s.t. for any $a = (a_1, \ldots, a_d) \in \mathbb{Z}_+^d$ with $\sum_i a_i < n$, there is an observed sequence which has one explanation if
  $a_1 v_1 + \cdots + a_d v_d > 0$ and another explanation if
  $a_1 v_1 + \cdots + a_d v_d < 0$, where $v_i = \log(\theta_i)$.

A lower bound

Theorem. *The above upper bound* $O(E^{d(d-1)})$ *on the number of inference functions is tight.*

Idea of the proof:

- Construct an HMM $\mathcal{M}_n$ of length n with d parameters s.t. for any $a = (a_1, \ldots, a_d) \in \mathbb{Z}_+^d$ with $\sum_i a_i < n$, there is an observed sequence which has one explanation if
  $a_1 v_1 + \cdots + a_d v_d > 0$ and another explanation if
  $a_1 v_1 + \cdots + a_d v_d < 0$, where $v_i = \log(\theta_i)$.

- Show that the hyperplane arrangement consisting of the hyperplanes of the form $\{x : \langle a, x \rangle = 0\}$ with $a \in \mathbb{Z}_+^d$ and $\sum_i a_i < n$ has at least $\Omega(n^{d(d-1)})$ chambers.

- So, $\mathcal{M}_n$ has $\Omega(n^{d(d-1)}) = \Omega(E^{d(d-1)})$ distinct inference functions.

Application to sequence alignment

- The 2-parameter model for sequence alignment is a particular case of a *pair hidden Markov model*.

Application to sequence alignment

- The 2-parameter model for sequence alignment is a particular case of a *pair hidden Markov model*.

General graphical model		**Sequence alignment**
observation τ	$\longleftrightarrow$	pair of sequences of length n
explanation $\widehat{\mathbf{h}}$	$\longleftrightarrow$	optimal alignment

Application to sequence alignment

- The 2-parameter model for sequence alignment is a particular case of a *pair hidden Markov model*.

General graphical model		**Sequence alignment**
observation τ	$\longleftrightarrow$	pair of sequences of length n
explanation $\widehat{\mathbf{h}}$	$\longleftrightarrow$	optimal alignment
Newton polytope $\mathrm{NP}(f_\tau)$	$\longleftrightarrow$	convex hull of points (x, y, z) where $x = \#$mismatches, $y = \#$insertions, $z = \#$matches for each alignment (this polytope lies on the plane $x + y + z = n$)

- The 2-parameter model for sequence alignment is a particular case of a *pair hidden Markov model*.

General graphical model		**Sequence alignment**
observation τ	$\longleftrightarrow$	pair of sequences of length n
explanation $\widehat{\mathbf{h}}$	$\longleftrightarrow$	optimal alignment
Newton polytope $\mathrm{NP}(f_\tau)$	$\longleftrightarrow$	convex hull of points (x, y, z) where $x = \#$mismatches, $y = \#$insertions, $z = \#$matches for each alignment (this polytope lies on the plane $x + y + z = n$)

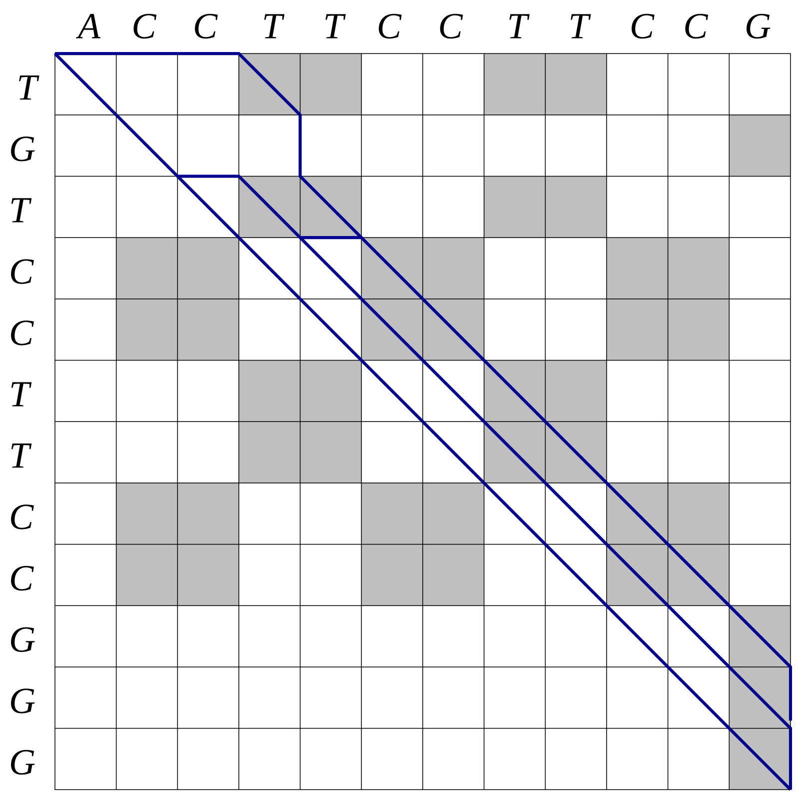
By the Main Theorem, the number of inference functions of this model is

$$O(n^{d(d-1)}) = O(n^2).$$

In fact, it is $\Theta(n^2)$.

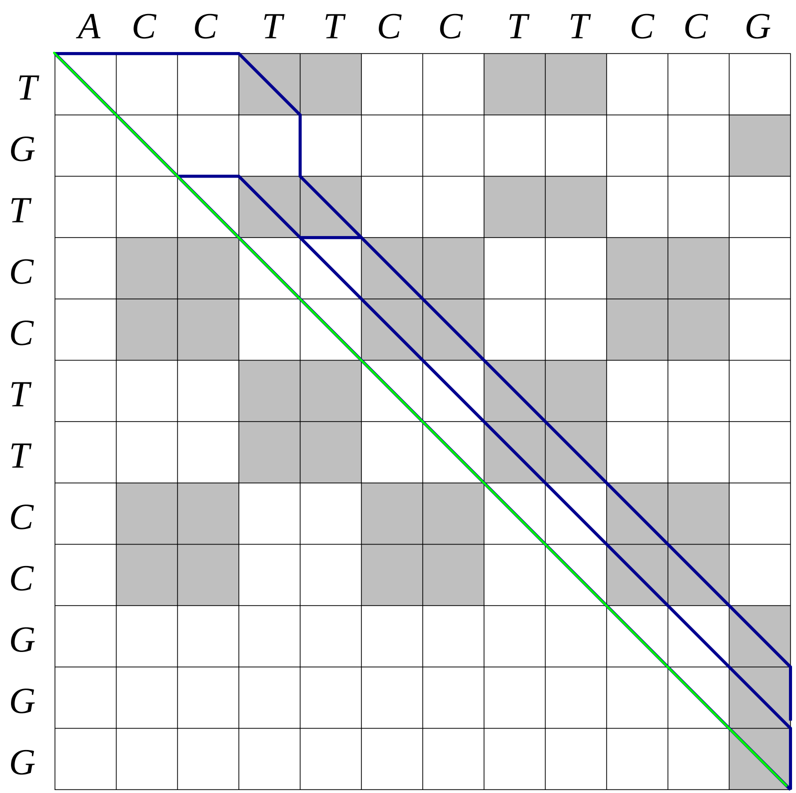
Alignment graphs

- Alignments can be represented as paths from the upper-left to the lower-right corner in the alignment graph:



x	y	z

- Alignments can be represented as paths from the upper-left to the lower-right corner in the alignment graph:

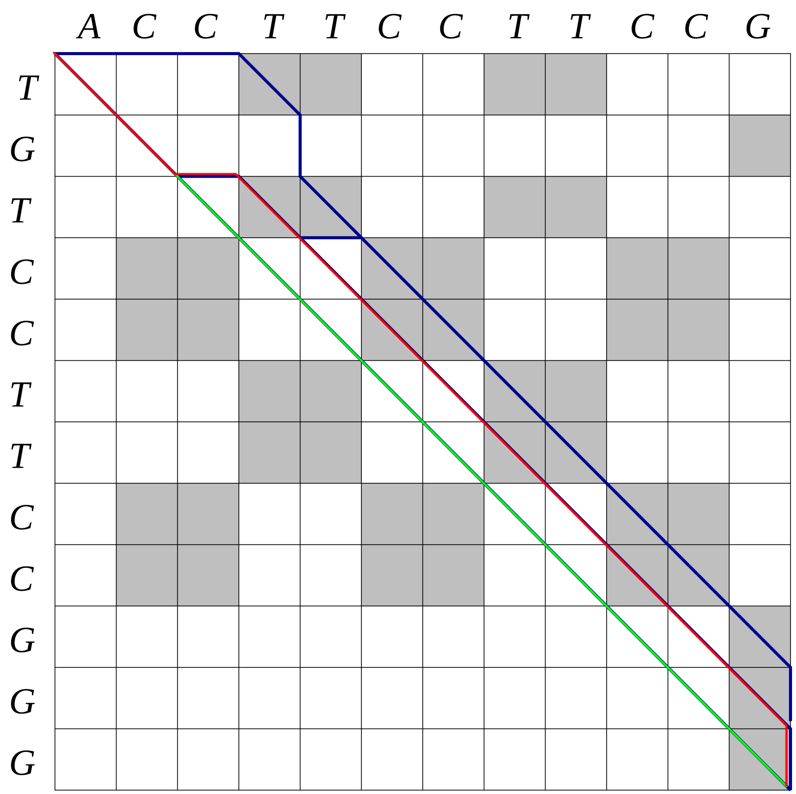


x	y	z
11	0	1

```
TGTCCTTCCGGG
ACCTTCCTTCCG
```

Alignment graphs

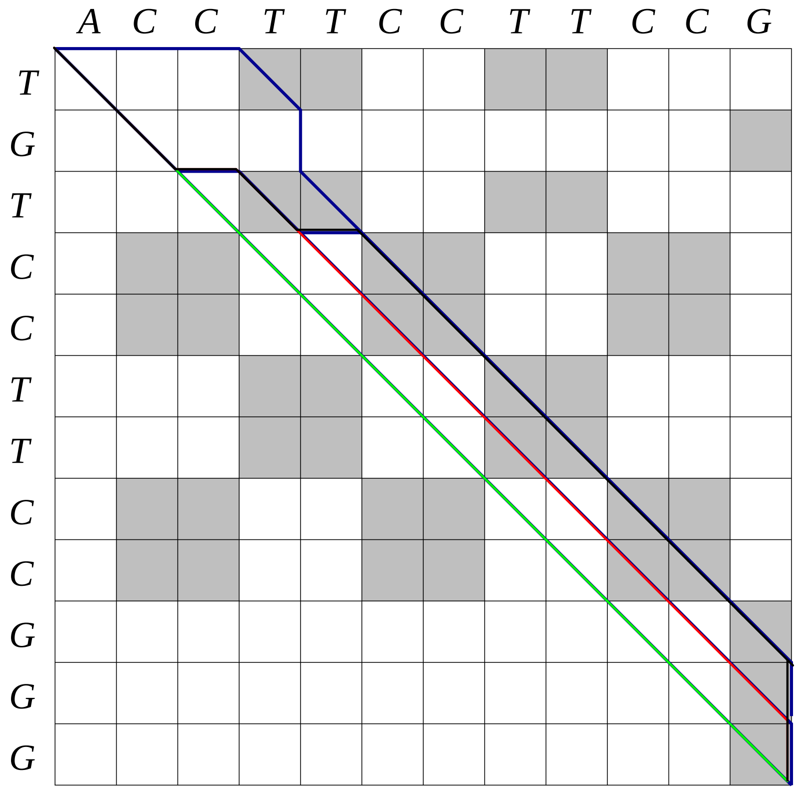
- Alignments can be represented as paths from the upper-left to the lower-right corner in the alignment graph:



x	y	z
11	0	1
6	1	5

```
TGTCCTTCCGGG   TG-TCCTTCCGGG
ACCTTCCTTCCG   ACCTTCCTTCCG-
```

- Alignments can be represented as paths from the upper-left to the lower-right corner in the alignment graph:

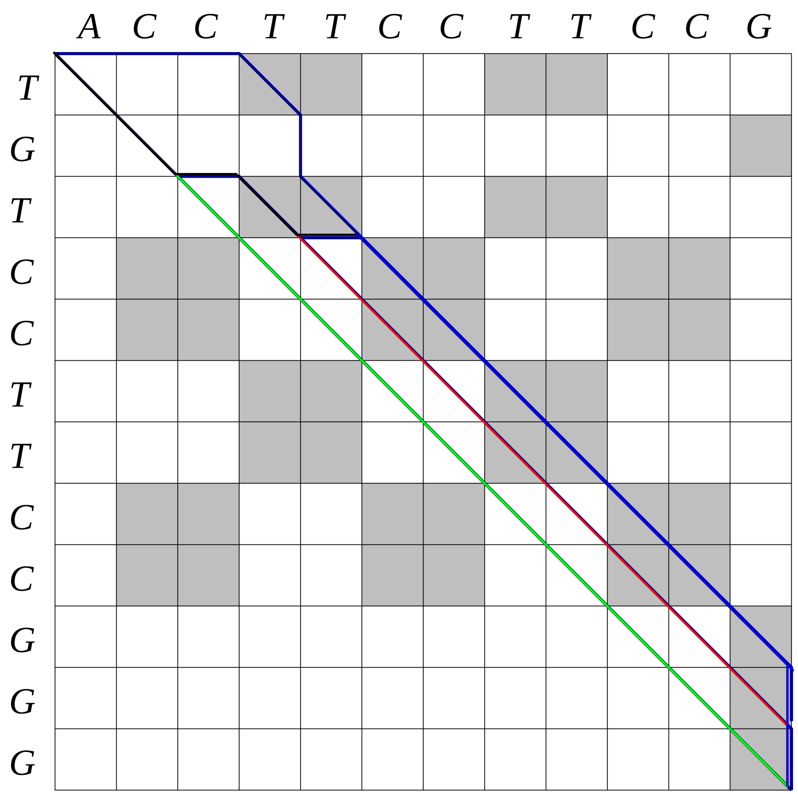


x	y	z
11	0	1
6	1	5
2	2	8

```
TGTCCTTCCGGG    TG-TCCTTCCGGG    TG-T-CCTTCCGGG
ACCTTCCTTCCG    ACCTTCCTTCCG-    ACCTTCCTTCCG--
```

Alignment graphs

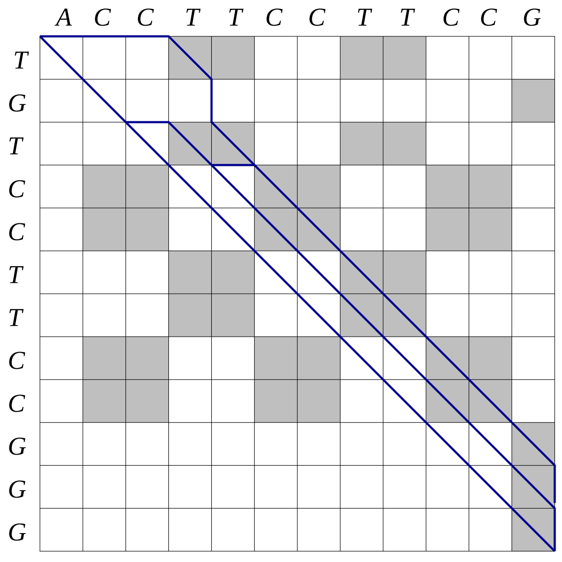
- Alignments can be represented as paths from the upper-left to the lower-right corner in the alignment graph:



x	y	z
11	0	1
6	1	5
2	2	8
0	3	9
⋮	⋮	⋮

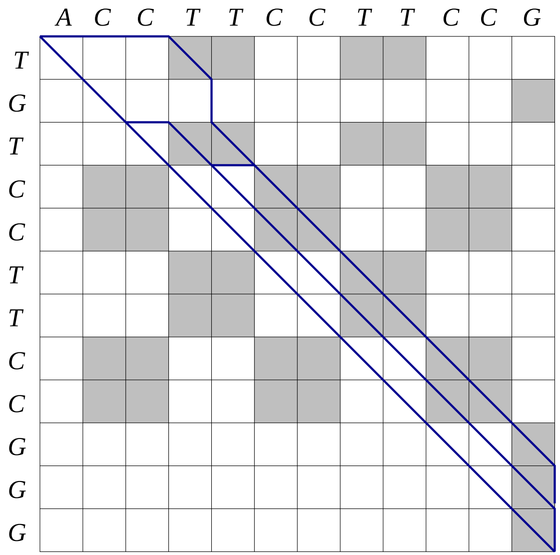
```
TGTCCTTCCGGG     TG-TCCTTCCGGG     TG-T-CCTTCCGGG     ---TGTCCTTCCGGG
ACCTTCCTTCCG     ACCTTCCTTCCG-     ACCTTCCTTCCG--     ACCT-TCCTTCCG--
```

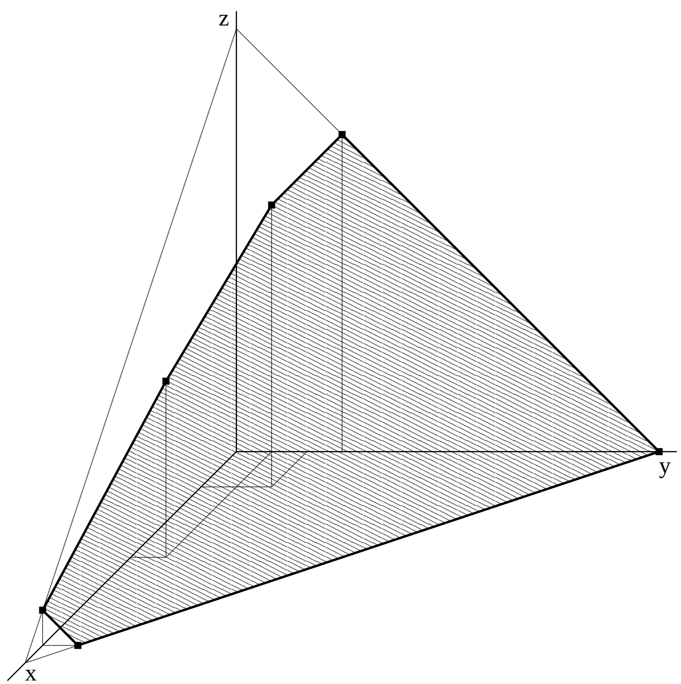
The Newton polytope of a pair of sequences



x	y	z
11	0	1
6	1	5
2	2	8
0	3	9
11	1	0
0	12	0

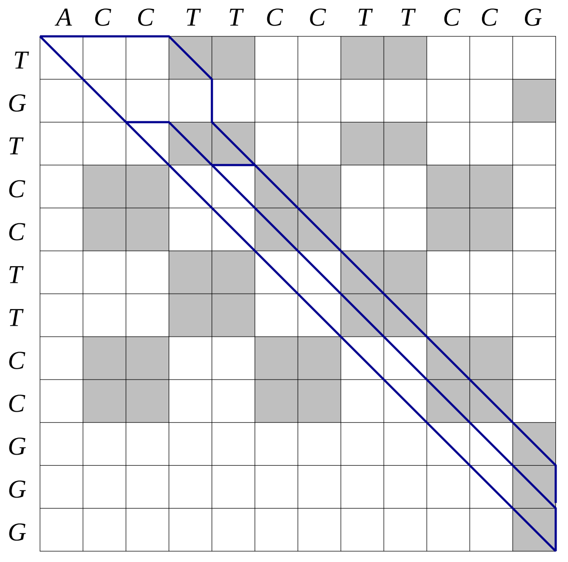
The Newton polytope of a pair of sequences



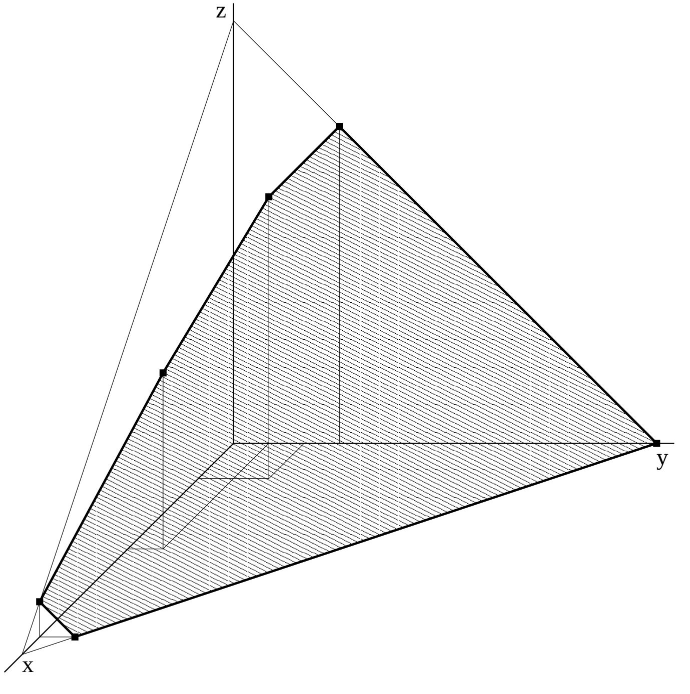


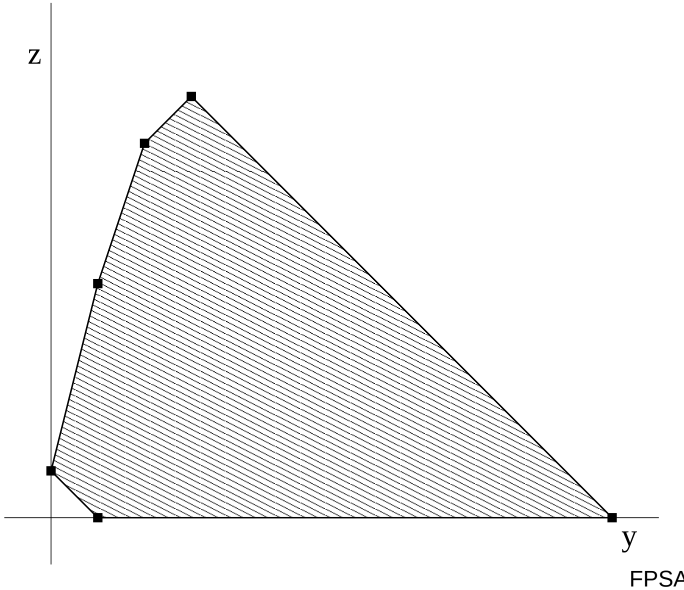
x	y	z
11	0	1
6	1	5
2	2	8
0	3	9
11	1	0
0	12	0

The Newton polytope of a pair of sequences



x	y	z
11	0	1
6	1	5
2	2	8
0	3	9
11	1	0
0	12	0





Open problem

In the 2-parameter model for sequence alignment,

$$P = \bigoplus_{\tau} \mathrm{NP}(f_\tau) \text{ has } \Theta(n^2) \text{ vertices } (= \# \text{ inference functions}).$$

Open: For fixed $\tau = (S, T)$, how many vertices can $\mathrm{NP}(f_\tau)$ have?

Open problem

In the 2-parameter model for sequence alignment,

$$P = \bigoplus_{\tau} \mathrm{NP}(f_\tau) \text{ has } \Theta(n^2) \text{ vertices } (= \# \text{ inference functions}).$$

Open: For fixed $\tau = (S, T)$, how many vertices can $\mathrm{NP}(f_\tau)$ have?

This is the number of different alignments of S and T that can be optimal for some values of the parameters α and β.

In the 2-parameter model for sequence alignment,

$$P = \bigoplus_{\tau} \mathrm{NP}(f_\tau) \text{ has } \Theta(n^2) \text{ vertices } (= \# \text{ inference functions}).$$

Open: For fixed $\tau = (S, T)$, how many vertices can $\mathrm{NP}(f_\tau)$ have?

This is the number of different alignments of S and T that can be optimal for some values of the parameters α and β.

Known to be $O(n^{2/3})$. [Gusfield et al '94, Fernández-Baca et al '02]

In the 2-parameter model for sequence alignment,

$$P = \bigoplus_{\tau} \mathrm{NP}(f_\tau) \text{ has } \Theta(n^2) \text{ vertices } (= \# \text{ inference functions}).$$

Open: For fixed $\tau = (S, T)$, how many vertices can $\mathrm{NP}(f_\tau)$ have?

This is the number of different alignments of S and T that can be optimal for some values of the parameters α and β.

Known to be $O(n^{2/3})$. [Gusfield et al '94, Fernández-Baca et al '02]

Conjecture: It is $O(\sqrt{n})$ for binary sequences.

In the 2-parameter model for sequence alignment,

$$P = \bigoplus_{\tau} \mathrm{NP}(f_\tau) \text{ has } \Theta(n^2) \text{ vertices } (= \# \text{ inference functions}).$$

Open: For fixed $\tau = (S, T)$, how many vertices can $\mathrm{NP}(f_\tau)$ have?

This is the number of different alignments of S and T that can be optimal for some values of the parameters α and β.

Known to be $O(n^{2/3})$. [Gusfield et al ’94, Fernández-Baca et al ’02]

Conjecture: It is $O(\sqrt{n})$ for binary sequences.

Conjecture (?): It is $O(\sqrt{n})$ for sequences on any finite alphabet.