

# A tropical approach to $m$ -dissimilarity maps

(partially joint work with C. Bocci)

Filip Cools

K.U.Leuven (Belgium)

September 2th, 2008

Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

## Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

## Some definitions

Let  $D$  be a square matrix, whose rows and columns are indexed by a finite set  $X$ . Usually, we take  $X = [n] := \{1, \dots, n\}$ .

- ▶  $D$  is a *dissimilarity matrix* if it is symmetric and all entries on the main diagonal are zero. We can see  $D$  as a map  $X^2 \rightarrow \mathbb{R}$  with  $D(i, j) = D(j, i)$  and  $D(i, i) = 0$  for all  $i, j \in X$ .
- ▶  $D$  has a *tree realization* if there exists a weighted tree  $T$  with leaf set  $X$  such that the distance between the leaves  $i, j \in X$  is exactly  $D(i, j)$ . In this case,  $D$  is called a *tree metric*.

## Some definitions

Let  $D$  be a square matrix, whose rows and columns are indexed by a finite set  $X$ . Usually, we take  $X = [n] := \{1, \dots, n\}$ .

- ▶  $D$  is a *dissimilarity matrix* if it is symmetric and all entries on the main diagonal are zero. We can see  $D$  as a map  $X^2 \rightarrow \mathbb{R}$  with  $D(i, j) = D(j, i)$  and  $D(i, i) = 0$  for all  $i, j \in X$ .
- ▶  $D$  has a *tree realization* if there exists a weighted tree  $T$  with leaf set  $X$  such that the distance between the leafs  $i, j \in X$  is exactly  $D(i, j)$ . In this case,  $D$  is called a *tree metric*.

## Some definitions

Let  $D$  be a square matrix, whose rows and columns are indexed by a finite set  $X$ . Usually, we take  $X = [n] := \{1, \dots, n\}$ .

- ▶  $D$  is a *dissimilarity matrix* if it is symmetric and all entries on the main diagonal are zero. We can see  $D$  as a map  $X^2 \rightarrow \mathbb{R}$  with  $D(i, j) = D(j, i)$  and  $D(i, i) = 0$  for all  $i, j \in X$ .
- ▶  $D$  has a *tree realization* if there exists a weighted tree  $T$  with leaf set  $X$  such that the distance between the leaves  $i, j \in X$  is exactly  $D(i, j)$ . In this case,  $D$  is called a *tree metric*.

## Tree Metric Theorem (P. Buneman, 1974)

Let  $D$  be a non-negative dissimilarity matrix on  $X$ . Then  $D$  is a tree metric on  $X$  if and only if, for every four (not necessarily distinct) elements  $i, j, k, l \in X$ , the maximum of the three numbers  $D(i, j) + D(k, l)$ ,  $D(i, k) + D(j, l)$  and  $D(i, l) + D(j, k)$  is attained at least twice. Moreover, the tree  $T$  that realizes  $D$  is unique.

### Remark

The condition above is called the *four-point condition*.

### Definition

The set of all tree metrics is the *space of trees*  $\mathcal{T}_n \subset \mathbb{R}^{\binom{n}{2}}$ .

## Generalization

Let  $m \leq n$  be an integer.

- ▶ A map  $D : X^m \rightarrow \mathbb{R}$  is called an  $m$ -dissimilarity map if  $D(i_1, \dots, i_m) = D(i_{\pi(1)}, \dots, i_{\pi(m)})$  for all permutations  $\pi \in \mathcal{S}_m$  and  $D(i_1, \dots, i_m) = 0$  if the elements  $i_1, \dots, i_m \in X$  are not pairwise distinct. We identify the map with  $D \in \mathbb{R}^{\binom{n}{m}}$ .
- ▶ An  $m$ -dissimilarity map  $D$  has a *tree realization* if there exists a tree  $T$  with leaf set  $X$  such that for every  $m$ -subset  $I = \{i_1, \dots, i_m\} \subset X$ , the weight of the smallest subtree of  $T$  containing  $I$  is equal to  $D(i_1, \dots, i_m)$ .

## Generalization

Let  $m \leq n$  be an integer.

- ▶ A map  $D : X^m \rightarrow \mathbb{R}$  is called an  $m$ -dissimilarity map if  $D(i_1, \dots, i_m) = D(i_{\pi(1)}, \dots, i_{\pi(m)})$  for all permutations  $\pi \in \mathcal{S}_m$  and  $D(i_1, \dots, i_m) = 0$  if the elements  $i_1, \dots, i_m \in X$  are not pairwise distinct. We identify the map with  $D \in \mathbb{R}^{\binom{n}{m}}$ .
- ▶ An  $m$ -dissimilarity map  $D$  has a *tree realization* if there exists a tree  $T$  with leaf set  $X$  such that for every  $m$ -subset  $I = \{i_1, \dots, i_m\} \subset X$ , the weight of the smallest subtree of  $T$  containing  $I$  is equal to  $D(i_1, \dots, i_m)$ .

## Generalization

Let  $m \leq n$  be an integer.

- ▶ A map  $D : X^m \rightarrow \mathbb{R}$  is called an  $m$ -dissimilarity map if  $D(i_1, \dots, i_m) = D(i_{\pi(1)}, \dots, i_{\pi(m)})$  for all permutations  $\pi \in \mathcal{S}_m$  and  $D(i_1, \dots, i_m) = 0$  if the elements  $i_1, \dots, i_m \in X$  are not pairwise distinct. We identify the map with  $D \in \mathbb{R}^{\binom{n}{m}}$ .
- ▶ An  $m$ -dissimilarity map  $D$  has a *tree realization* if there exists a tree  $T$  with leaf set  $X$  such that for every  $m$ -subset  $I = \{i_1, \dots, i_m\} \subset X$ , the weight of the smallest subtree of  $T$  containing  $I$  is equal to  $D(i_1, \dots, i_m)$ .

## Theorem (L. Pachter - D. Speyer)

Let  $T$  be a tree with  $n$  leaves and no vertices of degree 2. Let  $m \geq 3$  be an integer. If  $n \geq 2m - 1$ , then  $T$  is uniquely determined by its  $m$ -dissimilarity map  $D$ . If  $n = 2m - 2$ , this is not true.

### Some research problems

- ▶ description of a map  $\phi^{(m)} : \mathbb{R}^{\binom{n}{2}} \rightarrow \mathbb{R}^{\binom{n}{m}}$ , sending dissimilarity matrices of trees to the corresponding  $m$ -dissimilarity maps?
- ▶ description of  $\phi^{(m)}(\mathcal{T}_n) \subset \mathbb{R}^{\binom{n}{m}}$ ?

## Theorem (L. Pachter - D. Speyer)

Let  $T$  be a tree with  $n$  leaves and no vertices of degree 2. Let  $m \geq 3$  be an integer. If  $n \geq 2m - 1$ , then  $T$  is uniquely determined by its  $m$ -dissimilarity map  $D$ . If  $n = 2m - 2$ , this is not true.

## Some research problems

- ▶ description of a map  $\phi^{(m)} : \mathbb{R}^{\binom{n}{2}} \rightarrow \mathbb{R}^{\binom{n}{m}}$ , sending dissimilarity matrices of trees to the corresponding  $m$ -dissimilarity maps?
- ▶ description of  $\phi^{(m)}(\mathcal{T}_n) \subset \mathbb{R}^{\binom{n}{m}}$ ?

Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

An important problem in *computational biology* is to construct a *phylogenetic tree* from a sequence alignment of  $n$  species.

An example with  $n = 4$ :

Human	ACAATGTCATTAGCGAT...
Mouse	ACGTTGTCAATAGAGAT...
Rat	ACGTAGTCATTACACAT...
Chicken	GCACAGTCAGTAGAGCT...

The main statistical technique to select a tree model is computing the *maximum likelihood estimate* (MLE) for each of the

$$(2n - 5)!! = 1.3.5. \dots .(2n - 5)$$

combinatorial types of  $n$ -trees.

Unluckily, all the MLE computations are very difficult.

An important problem in *computational biology* is to construct a *phylogenetic tree* from a sequence alignment of  $n$  species.

An example with  $n = 4$ :

Human	ACAATGTCATTAGCGAT...
Mouse	ACGTTGTCAATAGAGAT...
Rat	ACGTAGTCATTACACAT...
Chicken	GCACAGTCAGTAGAGCT...

The main statistical technique to select a tree model is computing the *maximum likelihood estimate* (MLE) for each of the

$$(2n - 5)!! = 1.3.5. \dots .(2n - 5)$$

combinatorial types of  $n$ -trees.

Unluckily, all the MLE computations are very difficult.

## Distance based approach

This is a popular way to avoid this problem. It consists of two steps.

- ▶ First step : collaps the data into a dissimilarity matrix. Usually, the distance between any two strings is measured by some refined version of the Hamming distance.
- ▶ Second step : search for the weighted tree which represents this matrix, provided such a tree exists. Otherwise, search for the tree which closest to the matrix.

## Distance based approach

This is a popular way to avoid this problem. It consists of two steps.

- ▶ First step : collaps the data into a dissimilarity matrix.  
Usually, the distance between any two strings is measured by some refined version of the Hamming distance.
- ▶ Second step : search for the weighted tree which represents this matrix, provided such a tree exists. Otherwise, search for the tree which closest to the matrix.

## Distance based approach

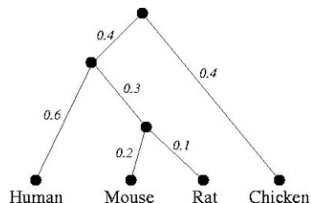
This is a popular way to avoid this problem. It consists of two steps.

- ▶ First step : collaps the data into a dissimilarity matrix.  
Usually, the distance between any two strings is measured by some refined version of the Hamming distance.
- ▶ Second step : search for the weighted tree which represents this matrix, provided such a tree exists. Otherwise, search for the tree which closest to the matrix.

## Example

$$D = \begin{matrix} & \begin{matrix} \text{H} & \text{M} & \text{R} & \text{C} \end{matrix} \\ \begin{matrix} \text{H} \\ \text{M} \\ \text{R} \\ \text{C} \end{matrix} & \begin{pmatrix} 0 & 1.1 & 1.0 & 1.4 \\ 1.1 & 0 & 0.3 & 1.3 \\ 1.0 & 0.3 & 0 & 1.2 \\ 1.4 & 1.3 & 1.2 & 0 \end{pmatrix} \end{matrix}$$

In this example,  $D \in \mathcal{T}_4$ .



For instance, the distance in this tree between “Human” and “Mouse” equals  $0.6 + 0.3 + 0.2 = 1.1$ .

## Why studying the case $m \geq 3$ ?

This problem has also some natural relevance in Phylogenetics.

Indeed, for example, it can be more reliable statistically to estimate the triple weights  $D(i, j, k)$  rather than the pairwise distances  $D(i, j)$ .

Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

We work in the *tropical semi-ring*

$$(\mathbb{R} \cup \{-\infty\}, \oplus, \otimes),$$

where

$$x \oplus y = \max\{x, y\} \quad \text{and} \quad x \otimes y = x + y.$$

Consider the polynomial ring  $\mathbb{R}[x_1, \dots, x_k]$ .

- ▶ Tropical monomials  $x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$  represent ordinary linear forms  $\sum_{i=1}^k a_i x_i$ , i.e. linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .
- ▶ Tropical polynomials  $\bigoplus_{a \in \mathcal{A}} C_a \otimes x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$ , with  $\mathcal{A} \subset \mathbb{N}^k$  finite and  $C_a \in \mathbb{R}$ , represent piecewise-linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .

We work in the *tropical semi-ring*

$$(\mathbb{R} \cup \{-\infty\}, \oplus, \otimes),$$

where

$$x \oplus y = \max\{x, y\} \quad \text{and} \quad x \otimes y = x + y.$$

Consider the polynomial ring  $\mathbb{R}[x_1, \dots, x_k]$ .

- ▶ Tropical monomials  $x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$  represent ordinary linear forms  $\sum_{i=1}^k a_i x_i$ , i.e. linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .
- ▶ Tropical polynomials  $\bigoplus_{a \in \mathcal{A}} C_a \otimes x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$ , with  $\mathcal{A} \subset \mathbb{N}^k$  finite and  $C_a \in \mathbb{R}$ , represent piecewise-linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .

We work in the *tropical semi-ring*

$$(\mathbb{R} \cup \{-\infty\}, \oplus, \otimes),$$

where

$$x \oplus y = \max\{x, y\} \quad \text{and} \quad x \otimes y = x + y.$$

Consider the polynomial ring  $\mathbb{R}[x_1, \dots, x_k]$ .

- ▶ Tropical monomials  $x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$  represent ordinary linear forms  $\sum_{i=1}^k a_i x_i$ , i.e. linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .
- ▶ Tropical polynomials  $\bigoplus_{a \in \mathcal{A}} C_a \otimes x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$ , with  $\mathcal{A} \subset \mathbb{N}^k$  finite and  $C_a \in \mathbb{R}$ , represent piecewise-linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .

We work in the *tropical semi-ring*

$$(\mathbb{R} \cup \{-\infty\}, \oplus, \otimes),$$

where

$$x \oplus y = \max\{x, y\} \quad \text{and} \quad x \otimes y = x + y.$$

Consider the polynomial ring  $\mathbb{R}[x_1, \dots, x_k]$ .

- ▶ Tropical monomials  $x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$  represent ordinary linear forms  $\sum_{i=1}^k a_i x_i$ , i.e. linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .
- ▶ Tropical polynomials  $\bigoplus_{a \in \mathcal{A}} C_a \otimes x_1^{a_1} \otimes \dots \otimes x_k^{a_k}$ , with  $\mathcal{A} \subset \mathbb{N}^k$  finite and  $C_a \in \mathbb{R}$ , represent piecewise-linear functions  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ .

Let  $K$  be the *field of Puiseux series*, i.e. the field of formal power series  $g = \sum_{q \in \mathbb{Q}} c_q t^q$  in the variable  $t$  such that the set  $Q_g = \{q \in \mathbb{Q} \mid c_q \neq 0\}$  is bounded below and has a finite set of denominators. For such an  $g \in K$ , the infimum of  $Q_g$  is equal to the minimum and we call it the *valuation*  $\text{val}(g)$  of  $g$ .

A polynomial

$$f = \sum_{a \in A} g_a \cdot x_1^{a_1} \cdots x_k^{a_k} \in K[x_1, \dots, x_k]$$

gives rise to the tropical polynomial, where  $C_a = -\text{val}(g_a)$ . We denote this tropical polynomial by  $\text{trop}(f)$ .

Let  $K$  be the *field of Puiseux series*, i.e. the field of formal power series  $g = \sum_{q \in \mathbb{Q}} c_q t^q$  in the variable  $t$  such that the set  $Q_g = \{q \in \mathbb{Q} \mid c_q \neq 0\}$  is bounded below and has a finite set of denominators. For such an  $g \in K$ , the infimum of  $Q_g$  is equal to the minimum and we call it the *valuation*  $\text{val}(g)$  of  $g$ .

A polynomial

$$f = \sum_{a \in A} g_a \cdot x_1^{a_1} \cdots x_k^{a_k} \in K[x_1, \dots, x_k]$$

gives rise to the tropical polynomial, where  $C_a = -\text{val}(g_a)$ . We denote this tropical polynomial by  $\text{trop}(f)$ .

## Definition Tropical Variety

- ▶ The *tropical hypersurface*  $\mathcal{T}(F) = \mathcal{T}(\text{trop}(f))$  is the corner locus of the piecewise-linear function  $F$ , i.e. the set of  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  such that the maximum of the set

$$\left\{ \sum_{i=1}^k a_i x_i + \lambda_a \right\}_{a \in A}$$

is attained at least twice.

- ▶ If  $I \subset K[x_1, \dots, x_k]$  is an ideal, we define the *tropical variety* of  $I$  as

$$\mathcal{T}(I) = \bigcap_{f \in I} \mathcal{T}(\text{trop}(f)).$$

## Definition Tropical Variety

- ▶ The *tropical hypersurface*  $\mathcal{T}(F) = \mathcal{T}(\text{trop}(f))$  is the corner locus of the piecewise-linear function  $F$ , i.e. the set of  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  such that the maximum of the set

$$\left\{ \sum_{i=1}^k a_i x_i + \lambda_a \right\}_{a \in \mathcal{A}}$$

is attained at least twice.

- ▶ If  $I \subset K[x_1, \dots, x_k]$  is an ideal, we define the *tropical variety* of  $I$  as

$$\mathcal{T}(I) = \bigcap_{f \in I} \mathcal{T}(\text{trop}(f)).$$

## Definition Tropical Variety

- ▶ The *tropical hypersurface*  $\mathcal{T}(F) = \mathcal{T}(\text{trop}(f))$  is the corner locus of the piecewise-linear function  $F$ , i.e. the set of  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  such that the maximum of the set

$$\left\{ \sum_{i=1}^k a_i x_i + \lambda_a \right\}_{a \in \mathcal{A}}$$

is attained at least twice.

- ▶ If  $I \subset K[x_1, \dots, x_k]$  is an ideal, we define the *tropical variety* of  $I$  as

$$\mathcal{T}(I) = \bigcap_{f \in I} \mathcal{T}(\text{trop}(f)).$$

## Theorem

The tropical variety  $\mathcal{T}(I)$  coincides with the closure in  $\mathbb{R}^k$  of the set

$$\{(-\text{val}(y_1), \dots, -\text{val}(y_k)) \mid (y_1, \dots, y_k) \in V(I)\} \subset \mathbb{Q}^k.$$

## Definition

If  $\mathcal{T}(I) \subset \mathbb{R}^k$  is a tropical variety, we say that  $\{f_1, \dots, f_r\}$  is a *tropical basis* of  $\mathcal{T}(I)$  if and only if  $I = \langle f_1, \dots, f_r \rangle$  and

$$\mathcal{T}(I) = \mathcal{T}(\text{trop}(f_1)) \cap \dots \cap \mathcal{T}(\text{trop}(f_r)).$$

## Remark

In general, a set of generators of  $I$  is not a tropical basis of  $\mathcal{T}(I)$ . Nevertheless, one can prove that every tropical variety has a tropical basis (software : GFan).

## Theorem

The tropical variety  $\mathcal{T}(I)$  coincides with the closure in  $\mathbb{R}^k$  of the set

$$\{(-\text{val}(y_1), \dots, -\text{val}(y_k)) \mid (y_1, \dots, y_k) \in V(I)\} \subset \mathbb{Q}^k.$$

## Definition

If  $\mathcal{T}(I) \subset \mathbb{R}^k$  is a tropical variety, we say that  $\{f_1, \dots, f_r\}$  is a *tropical basis* of  $\mathcal{T}(I)$  if and only if  $I = \langle f_1, \dots, f_r \rangle$  and

$$\mathcal{T}(I) = \mathcal{T}(\text{trop}(f_1)) \cap \dots \cap \mathcal{T}(\text{trop}(f_r)).$$

## Remark

In general, a set of generators of  $I$  is not a tropical basis of  $\mathcal{T}(I)$ . Nevertheless, one can prove that every tropical variety has a tropical basis (software : GFan).

## Theorem

The tropical variety  $\mathcal{T}(I)$  coincides with the closure in  $\mathbb{R}^k$  of the set

$$\{(-\text{val}(y_1), \dots, -\text{val}(y_k)) \mid (y_1, \dots, y_k) \in V(I)\} \subset \mathbb{Q}^k.$$

## Definition

If  $\mathcal{T}(I) \subset \mathbb{R}^k$  is a tropical variety, we say that  $\{f_1, \dots, f_r\}$  is a *tropical basis* of  $\mathcal{T}(I)$  if and only if  $I = \langle f_1, \dots, f_r \rangle$  and

$$\mathcal{T}(I) = \mathcal{T}(\text{trop}(f_1)) \cap \dots \cap \mathcal{T}(\text{trop}(f_r)).$$

## Remark

In general, a set of generators of  $I$  is not a tropical basis of  $\mathcal{T}(I)$ . Nevertheless, one can prove that every tropical variety has a tropical basis (software : GFan).

We are mainly interested in the tropical variety  $\mathcal{T}(I_{m,n})$ , where  $I_{m,n}$  is the ideal of the Grassmannian  $\mathbb{G}(m, n) \subset \mathbb{P}^{\binom{n}{m}-1}$ . Note that  $I_{m,n}$  consists of all algebraic relations among the determinants of the  $(m \times m)$ -minors of any  $(m \times n)$ -matrix with entries in a commutative ring.

### Definition

The *tropical Grassmannian*  $\mathcal{G}_{m,n}$  is the tropical variety  $\mathcal{T}(I_{m,n})$

### Theorem

The tropical Grassmannian  $\mathcal{G}_{m,n}$  is a polyhedral fan in  $\mathbb{R}^{\binom{n}{m}}$ . Each of its maximal cones has the same dimension, namely  $(n - m)m + 1$ .

We are mainly interested in the tropical variety  $\mathcal{T}(I_{m,n})$ , where  $I_{m,n}$  is the ideal of the Grassmannian  $\mathbb{G}(m, n) \subset \mathbb{P}^{\binom{n}{m}-1}$ . Note that  $I_{m,n}$  consists of all algebraic relations among the determinants of the  $(m \times m)$ -minors of any  $(m \times n)$ -matrix with entries in a commutative ring.

## Definition

The *tropical Grassmannian*  $\mathcal{G}_{m,n}$  is the tropical variety  $\mathcal{T}(I_{m,n})$

## Theorem

The tropical Grassmannian  $\mathcal{G}_{m,n}$  is a polyhedral fan in  $\mathbb{R}^{\binom{n}{m}}$ . Each of its maximal cones has the same dimension, namely  $(n - m)m + 1$ .

We are mainly interested in the tropical variety  $\mathcal{T}(I_{m,n})$ , where  $I_{m,n}$  is the ideal of the Grassmannian  $\mathbb{G}(m, n) \subset \mathbb{P}^{\binom{n}{m}-1}$ . Note that  $I_{m,n}$  consists of all algebraic relations among the determinants of the  $(m \times m)$ -minors of any  $(m \times n)$ -matrix with entries in a commutative ring.

### Definition

The *tropical Grassmannian*  $\mathcal{G}_{m,n}$  is the tropical variety  $\mathcal{T}(I_{m,n})$

### Theorem

The tropical Grassmannian  $\mathcal{G}_{m,n}$  is a polyhedral fan in  $\mathbb{R}^{\binom{n}{m}}$ . Each of its maximal cones has the same dimension, namely  $(n - m)m + 1$ .

## Special case : $m=2$

The ideal  $I_{2,n}$  is generated by the three-term Plücker relations

$$p_{ijkl} := x_{ik}x_{jl} - x_{ij}x_{kl} - x_{il}x_{jk} \quad (1 \leq i < j < k < l \leq n)$$

and one can prove that these quadrics form a tropical basis of  $\mathcal{G}_{2,n}$ .

Note that

$$\text{trop}(p_{ijkl}) = (x_{ij} \otimes x_{kl}) \oplus (x_{ik} \otimes x_{jl}) \oplus (x_{il} \otimes x_{jk})$$

hence  $D \in \mathcal{T}(\text{trop}(p_{ijkl}))$  if and only if the maximum of  $D(i, j) + D(k, l)$ ,  $D(i, k) + D(j, l)$  and  $D(i, l) + D(j, k)$  is attained at least twice.

Conclusion :  $\mathcal{T}_n = \mathcal{G}_{2,n}$  (D. Speyer - B. Sturmfels, 2004)

## Special case : $m=2$

The ideal  $I_{2,n}$  is generated by the three-term Plücker relations

$$p_{ijkl} := x_{ik}x_{jl} - x_{ij}x_{kl} - x_{il}x_{jk} \quad (1 \leq i < j < k < l \leq n)$$

and one can prove that these quadrics form a tropical basis of  $\mathcal{G}_{2,n}$ .

Note that

$$\text{trop}(p_{ijkl}) = (x_{ij} \otimes x_{kl}) \oplus (x_{ik} \otimes x_{jl}) \oplus (x_{il} \otimes x_{jk})$$

hence  $D \in \mathcal{T}(\text{trop}(p_{ijkl}))$  if and only if the maximum of  $D(i, j) + D(k, l)$ ,  $D(i, k) + D(j, l)$  and  $D(i, l) + D(j, k)$  is attained at least twice.

Conclusion :  $\mathcal{T}_n = \mathcal{G}_{2,n}$  (D. Speyer - B. Sturmfels, 2004)

## Special case : $m=2$

The ideal  $I_{2,n}$  is generated by the three-term Plücker relations

$$p_{ijkl} := x_{ik}x_{jl} - x_{ij}x_{kl} - x_{il}x_{jk} \quad (1 \leq i < j < k < l \leq n)$$

and one can prove that these quadrics form a tropical basis of  $\mathcal{G}_{2,n}$ .

Note that

$$\text{trop}(p_{ijkl}) = (x_{ij} \otimes x_{kl}) \oplus (x_{ik} \otimes x_{jl}) \oplus (x_{il} \otimes x_{jk})$$

hence  $D \in \mathcal{T}(\text{trop}(p_{ijkl}))$  if and only if the maximum of  $D(i, j) + D(k, l)$ ,  $D(i, k) + D(j, l)$  and  $D(i, l) + D(j, k)$  is attained at least twice.

Conclusion :  $\mathcal{T}_n = \mathcal{G}_{2,n}$  (D. Speyer - B. Sturmfels, 2004)

## General case : $m \leq n$ arbitrary

The ideal  $I_{m,n}$  is generated by quadrics. Among them are the (generalized) three-term Plücker relations

$$p_{R,ijkl} := x_{Rik}x_{Rjl} - x_{Rij}x_{Rkl} - x_{Ril}x_{Rjk}.$$

Hereby  $R$  is any  $(m-2)$ -subset of  $[n]$  and  $i, j, k, l \in [n] \setminus R$ .

The *three-term tropical Grassmannian*  $\mathcal{T}_{m,n}$  is the intersection

$$\bigcap_{R,i,j,k,l} \mathcal{T}(\text{trop}(p_{R,ijkl})) \subset \mathbb{R}^{\binom{n}{m}}.$$

In general, we have  $\mathcal{G}_{m,n} \subsetneq \mathcal{T}_{m,n}$ .

## General case : $m \leq n$ arbitrary

The ideal  $I_{m,n}$  is generated by quadrics. Among them are the (generalized) three-term Plücker relations

$$p_{R,ijkl} := x_{Rik}x_{Rjl} - x_{Rij}x_{Rkl} - x_{Ril}x_{Rjk}.$$

Hereby  $R$  is any  $(m-2)$ -subset of  $[n]$  and  $i, j, k, l \in [n] \setminus R$ .

The *three-term tropical Grassmannian*  $\mathcal{T}_{m,n}$  is the intersection

$$\bigcap_{R,i,j,k,l} \mathcal{T}(\text{trop}(p_{R,ijkl})) \subset \mathbb{R}^{\binom{n}{m}}.$$

In general, we have  $\mathcal{G}_{m,n} \subsetneq \mathcal{T}_{m,n}$ .

## General case : $m \leq n$ arbitrary

The ideal  $I_{m,n}$  is generated by quadrics. Among them are the (generalized) three-term Plücker relations

$$p_{R,ijkl} := x_{Rik}x_{Rjl} - x_{Rij}x_{Rkl} - x_{Ril}x_{Rjk}.$$

Hereby  $R$  is any  $(m-2)$ -subset of  $[n]$  and  $i, j, k, l \in [n] \setminus R$ .

The *three-term tropical Grassmannian*  $\mathcal{T}_{m,n}$  is the intersection

$$\bigcap_{R,i,j,k,l} \mathcal{T}(\text{trop}(p_{R,ijkl})) \subset \mathbb{R}^{\binom{n}{m}}.$$

In general, we have  $\mathcal{G}_{m,n} \subsetneq \mathcal{T}_{m,n}$ .

Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

## Description of $\phi^{(m)}$

Let  $n$  and  $m$  be integers such that  $n \geq m \geq 2$ . Let

$$\phi^{(m)} : \mathbb{R}^{\binom{n}{2}} \rightarrow \mathbb{R}^{\binom{n}{m}} : X = (X_{i,j}) \mapsto (X_{i_1, \dots, i_m})$$

be the map with

$$X_{i_1, \dots, i_m} = \frac{1}{2} \cdot \min_{\sigma \in \mathcal{C}_m} \{X_{i_1, i_{\sigma(1)}} + X_{i_{\sigma(1)}, i_{\sigma^2(1)}} + \dots + X_{i_{\sigma^{m-1}(1)}, i_{\sigma^m(1)}}\},$$

where  $\mathcal{C}_m \subset \mathcal{S}_m$  is the set of cyclic permutations.

If  $D \in \mathcal{G}_{2,n} = \mathcal{T}_n \subset \mathbb{R}^{\binom{n}{2}}$  is the dissimilarity matrix of an  $n$ -tree  $T$ , then the  $m$ -dissimilarity map of  $T$  is equal to  $\phi^{(m)}(D)$ . So the set of  $m$ -dissimilarity maps of  $n$ -trees is equal to  $\phi^{(m)}(\mathcal{G}_{2,n})$ .

Example :  $m = 4$ 

- ▶  $\mathcal{C}_4 = \{\sigma_1, \sigma_1^{-1}, \sigma_2, \sigma_2^{-1}, \sigma_3, \sigma_3^{-1}\}$  with

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}.$$

- ▶  $X_{i,j,k,l}$  is equal to the minimum of the three terms

$$X_{i,j} + X_{j,k} + X_{k,l} + X_{l,i},$$

$$X_{i,j} + X_{j,l} + X_{l,k} + X_{k,i},$$

$$X_{i,k} + X_{k,j} + X_{j,l} + X_{l,i},$$

divided by two.

Tropical description of  $\phi^{(m)}$ 

$$X_{i_1, \dots, i_m} = \left( \bigoplus_{\sigma \in \mathcal{C}_m} \left( X_{i_1, i_{\sigma(1)}} \otimes X_{i_{\sigma(1)}, i_{\sigma(2)}} \otimes \cdots \otimes X_{i_{\sigma(m-1)(1)}, i_1} \right)^{-1} \right)^{-\frac{1}{2}}$$

Inclusion for  $\phi^{(m)}(\mathcal{G}_{2,n})$ 

$$\phi^{(m)}(\mathcal{G}_{2,n}) \subset \mathcal{T}_{m,n} \cap \phi^{(m)}(\mathbb{R}^{\binom{n}{2}})$$

## Tropical description of $\phi^{(m)}$

$$X_{i_1, \dots, i_m} = \left( \bigoplus_{\sigma \in \mathcal{C}_m} \left( X_{i_1, i_{\sigma(1)}} \otimes X_{i_{\sigma(1)}, i_{\sigma(2)}} \otimes \cdots \otimes X_{i_{\sigma(m-1)(1)}, i_1} \right)^{-1} \right)^{-\frac{1}{2}}$$

## Inclusion for $\phi^{(m)}(\mathcal{G}_{2,n})$

$$\phi^{(m)}(\mathcal{G}_{2,n}) \subset \mathcal{T}_{m,n} \cap \phi^{(m)}(\mathbb{R}^{\binom{n}{2}})$$

## Remark

- ▶  $\mathcal{C}_3 = \{\sigma, \sigma^{-1}\}$  with  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$
- ▶  $X_{i,j,k} = \frac{1}{2} \cdot (X_{i,j} + X_{j,k} + X_{k,i})$

## Some results

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{T}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

and

$$\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$$

hence

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{G}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

## Remark

- ▶  $\mathcal{C}_3 = \{\sigma, \sigma^{-1}\}$  with  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$
- ▶  $X_{i,j,k} = \frac{1}{2} \cdot (X_{i,j} + X_{j,k} + X_{k,i})$

## Some results

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{T}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

and

$$\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$$

hence

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{G}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

## Remark

- ▶  $\mathcal{C}_3 = \{\sigma, \sigma^{-1}\}$  with  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$
- ▶  $X_{i,j,k} = \frac{1}{2} \cdot (X_{i,j} + X_{j,k} + X_{k,i})$

## Some results

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{T}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

and

$$\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$$

hence

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{G}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

## Remark

- ▶  $\mathcal{C}_3 = \{\sigma, \sigma^{-1}\}$  with  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$
- ▶  $X_{i,j,k} = \frac{1}{2} \cdot (X_{i,j} + X_{j,k} + X_{k,i})$

## Some results

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{T}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

and

$$\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$$

hence

$$\phi^{(3)}(\mathcal{G}_{2,n}) = \mathcal{G}_{3,n} \cap \phi^{(3)}(\mathbb{R}^{\binom{n}{2}})$$

Idea of proof of the inclusion  $\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$  :

Let  $T$  be a tree with 3-dissimilarity table  $\mathcal{D} = (D(i, j, k))_{i, j, k}$ .

*Remark* :  $\mathcal{G}_{3,n}$  is the closure in  $\mathbb{R}^{\binom{n}{3}}$  of the set

$$S := \{(-\text{val}(\det(M(i, j, k))))_{i, j, k} \mid M \in K^{3 \times n}\}.$$

*Claim* : If all the edges of  $T$  have rational lengths, we have that  $\mathcal{D} \in S$ .

Using the claim, we can finish the proof of the inclusion. Indeed, we can approximate  $T$  arbitrarily close by a tree  $\tilde{T}$  with rational edges. Hence the 3-dissimilarity map  $\tilde{\mathcal{D}}$  of  $\tilde{T}$  belongs to  $S$  and  $\mathcal{D} \in \bar{S} = \mathcal{G}_{3,n}$ .

Idea of proof of the inclusion  $\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$  :

Let  $T$  be a tree with 3-dissimilarity table  $\mathcal{D} = (D(i,j,k))_{i,j,k}$ .

*Remark* :  $\mathcal{G}_{3,n}$  is the closure in  $\mathbb{R}^{\binom{n}{3}}$  of the set

$$S := \{(-\text{val}(\det(M(i,j,k))))_{i,j,k} \mid M \in K^{3 \times n}\}.$$

*Claim* : If all the edges of  $T$  have rational lengths, we have that  $\mathcal{D} \in S$ .

Using the claim, we can finish the proof of the inclusion. Indeed, we can approximate  $T$  arbitrarily close by a tree  $\tilde{T}$  with rational edges. Hence the 3-dissimilarity map  $\tilde{\mathcal{D}}$  of  $\tilde{T}$  belongs to  $S$  and  $\mathcal{D} \in \bar{S} = \mathcal{G}_{3,n}$ .

Idea of proof of the inclusion  $\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$  :

Let  $T$  be a tree with 3-dissimilarity table  $\mathcal{D} = (D(i, j, k))_{i,j,k}$ .

*Remark* :  $\mathcal{G}_{3,n}$  is the closure in  $\mathbb{R}^{\binom{n}{3}}$  of the set

$$S := \{(-\text{val}(\det(M(i, j, k))))_{i,j,k} \mid M \in K^{3 \times n}\}.$$

*Claim* : If all the edges of  $T$  have rational lengths, we have that  $\mathcal{D} \in S$ .

Using the claim, we can finish the proof of the inclusion. Indeed, we can approximate  $T$  arbitrarily close by a tree  $\tilde{T}$  with rational edges. Hence the 3-dissimilarity map  $\tilde{\mathcal{D}}$  of  $\tilde{T}$  belongs to  $S$  and  $\mathcal{D} \in \bar{S} = \mathcal{G}_{3,n}$ .

Idea of proof of the inclusion  $\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$  :

Let  $T$  be a tree with 3-dissimilarity table  $\mathcal{D} = (D(i, j, k))_{i, j, k}$ .

*Remark* :  $\mathcal{G}_{3,n}$  is the closure in  $\mathbb{R}^{\binom{n}{3}}$  of the set

$$S := \{(-\text{val}(\det(M(i, j, k))))_{i, j, k} \mid M \in K^{3 \times n}\}.$$

*Claim* : If all the edges of  $T$  have rational lengths, we have that  $\mathcal{D} \in S$ .

Using the claim, we can finish the proof of the inclusion. Indeed, we can approximate  $T$  arbitrarily close by a tree  $\tilde{T}$  with rational edges. Hence the 3-dissimilarity map  $\tilde{\mathcal{D}}$  of  $\tilde{T}$  belongs to  $S$  and  $\mathcal{D} \in \bar{S} = \mathcal{G}_{3,n}$ .

Idea of proof of the inclusion  $\phi^{(3)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{3,n}$  :

Let  $T$  be a tree with 3-dissimilarity table  $\mathcal{D} = (D(i,j,k))_{i,j,k}$ .

*Remark* :  $\mathcal{G}_{3,n}$  is the closure in  $\mathbb{R}^{\binom{n}{3}}$  of the set

$$S := \{(-\text{val}(\det(M(i,j,k))))_{i,j,k} \mid M \in K^{3 \times n}\}.$$

*Claim* : If all the edges of  $T$  have rational lengths, we have that  $\mathcal{D} \in S$ .

Using the claim, we can finish the proof of the inclusion. Indeed, we can approximate  $T$  arbitrarily close by a tree  $\tilde{T}$  with rational edges. Hence the 3-dissimilarity map  $\tilde{\mathcal{D}}$  of  $\tilde{T}$  belongs to  $S$  and  $\mathcal{D} \in \bar{S} = \mathcal{G}_{3,n}$ .

## Proof of the claim for equidistant trees:

Pick a random number  $a(e)$  for each edge  $e$  of  $T$  and denote by  $h(e)$  the height of  $e$ , i.e. the distance between the top node of  $e$  and each leaf below  $e$ .

Define for each  $i \in [n]$  the polynomial  $x_i(t)$  as the sum of the monomials  $a(e)t^{2h(e)}$ , where  $e$  runs over all edges between the root  $r$  and the leaf  $i$ .

Then we have  $D(i, j) = \deg(x_j(t) - x_i(t))$ .

Proof of the claim for equidistant trees:

Pick a random number  $a(e)$  for each edge  $e$  of  $T$  and denote by  $h(e)$  the height of  $e$ , i.e. the distance between the top node of  $e$  and each leaf below  $e$ .

Define for each  $i \in [n]$  the polynomial  $x_i(t)$  as the sum of the monomials  $a(e)t^{2h(e)}$ , where  $e$  runs over all edges between the root  $r$  and the leaf  $i$ .

Then we have  $D(i, j) = \deg(x_j(t) - x_i(t))$ .

Proof of the claim for equidistant trees:

Pick a random number  $a(e)$  for each edge  $e$  of  $T$  and denote by  $h(e)$  the height of  $e$ , i.e. the distance between the top node of  $e$  and each leaf below  $e$ .

Define for each  $i \in [n]$  the polynomial  $x_i(t)$  as the sum of the monomials  $a(e)t^{2h(e)}$ , where  $e$  runs over all edges between the root  $r$  and the leaf  $i$ .

Then we have  $D(i, j) = \deg(x_j(t) - x_i(t))$ .

Proof of the claim for equidistant trees:

Pick a random number  $a(e)$  for each edge  $e$  of  $T$  and denote by  $h(e)$  the height of  $e$ , i.e. the distance between the top node of  $e$  and each leaf below  $e$ .

Define for each  $i \in [n]$  the polynomial  $x_i(t)$  as the sum of the monomials  $a(e)t^{2h(e)}$ , where  $e$  runs over all edges between the root  $r$  and the leaf  $i$ .

Then we have  $D(i, j) = \deg(x_j(t) - x_i(t))$ .

Consider the matrix  $M = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1(t) & x_2(t) & \dots & x_n(t) \\ x_1(t)^2 & x_2(t)^2 & \dots & x_n(t)^2 \end{pmatrix}$ .

Note that

$$\begin{aligned} \deg(\det(M(i, j, k))) &= \deg((x_j - x_i)(x_k - x_j)(x_k - x_i)) \\ &= D(i, j) + D(i, k) + D(j, k) \\ &= 2.D(i, j, k). \end{aligned}$$

Replacing each  $t$  in  $M$  by  $t^{-1/2}$ , we get that

$D(i, j, k) = -\text{val}(\det(M(i, j, k)))$ , hence  $\mathcal{D} \in S$ .

Consider the matrix  $M = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1(t) & x_2(t) & \dots & x_n(t) \\ x_1(t)^2 & x_2(t)^2 & \dots & x_n(t)^2 \end{pmatrix}$ .

Note that

$$\begin{aligned} \deg(\det(M(i, j, k))) &= \deg((x_j - x_i)(x_k - x_j)(x_k - x_i)) \\ &= D(i, j) + D(i, k) + D(j, k) \\ &= 2 \cdot D(i, j, k). \end{aligned}$$

Replacing each  $t$  in  $M$  by  $t^{-1/2}$ , we get that

$D(i, j, k) = -\text{val}(\det(M(i, j, k)))$ , hence  $\mathcal{D} \in S$ .

Consider the matrix  $M = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1(t) & x_2(t) & \dots & x_n(t) \\ x_1(t)^2 & x_2(t)^2 & \dots & x_n(t)^2 \end{pmatrix}$ .

Note that

$$\begin{aligned} \deg(\det(M(i, j, k))) &= \deg((x_j - x_i)(x_k - x_j)(x_k - x_i)) \\ &= D(i, j) + D(i, k) + D(j, k) \\ &= 2 \cdot D(i, j, k). \end{aligned}$$

Replacing each  $t$  in  $M$  by  $t^{-1/2}$ , we get that

$D(i, j, k) = -\text{val}(\det(M(i, j, k)))$ , hence  $\mathcal{D} \in S$ .

## Idea of proof of the claim for general trees:

- ▶ Consider a new metric  $D'$  such that  $D'$  is realizable by an equidistant tree  $T'$  and consider the matrix  $M' \in K^{3 \times n}$  such that

$$D'(i, j, k) = -\text{val}(\det(M'(i, j, k))).$$

- ▶ Replace  $M'$  by  $M$  so that

$$D(i, j, k) = -\text{val}(\det(M(i, j, k))).$$

Idea of proof of the claim for general trees:

- ▶ Consider a new metric  $D'$  such that  $D'$  is realizable by an equidistant tree  $T'$  and consider the matrix  $M' \in K^{3 \times n}$  such that

$$D'(i, j, k) = -\text{val}(\det(M'(i, j, k))).$$

- ▶ Replace  $M'$  by  $M$  so that

$$D(i, j, k) = -\text{val}(\det(M(i, j, k))).$$

Idea of proof of the claim for general trees:

- ▶ Consider a new metric  $D'$  such that  $D'$  is realizable by an equidistant tree  $T'$  and consider the matrix  $M' \in K^{3 \times n}$  such that

$$D'(i, j, k) = -\text{val}(\det(M'(i, j, k))).$$

- ▶ Replace  $M'$  by  $M$  so that

$$D(i, j, k) = -\text{val}(\det(M(i, j, k))).$$

## Theorem

$$\phi^{(4)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{4,n}$$

Introduction

Motivation : Computational biology

Tropical geometry

Some results

Open questions

- ▶  $\phi^{(4)}(\mathcal{G}_{2,n}) = \mathcal{G}_{4,n} \cap \phi^{(4)}(\mathbb{R}^{\binom{n}{2}})$ ?
- ▶  $\phi^{(m)}(\mathcal{G}_{2,n}) \subset \mathcal{G}_{m,n}$  for  $m \geq 5$  ?
- ▶  $\phi^{(m)}(\mathbb{R}^{\binom{n}{2}})$  tropical variety for  $m \geq 4$ ?
- ▶  $\mathcal{T}_{m,n}$  tropical variety for  $m \geq 3$ ?
- ▶ dimension of  $\mathcal{T}_{m,n}$  for  $m \geq 3$ ?