Maximum likelihood estimation for log-linear models in dimension two

Eliana Duarte

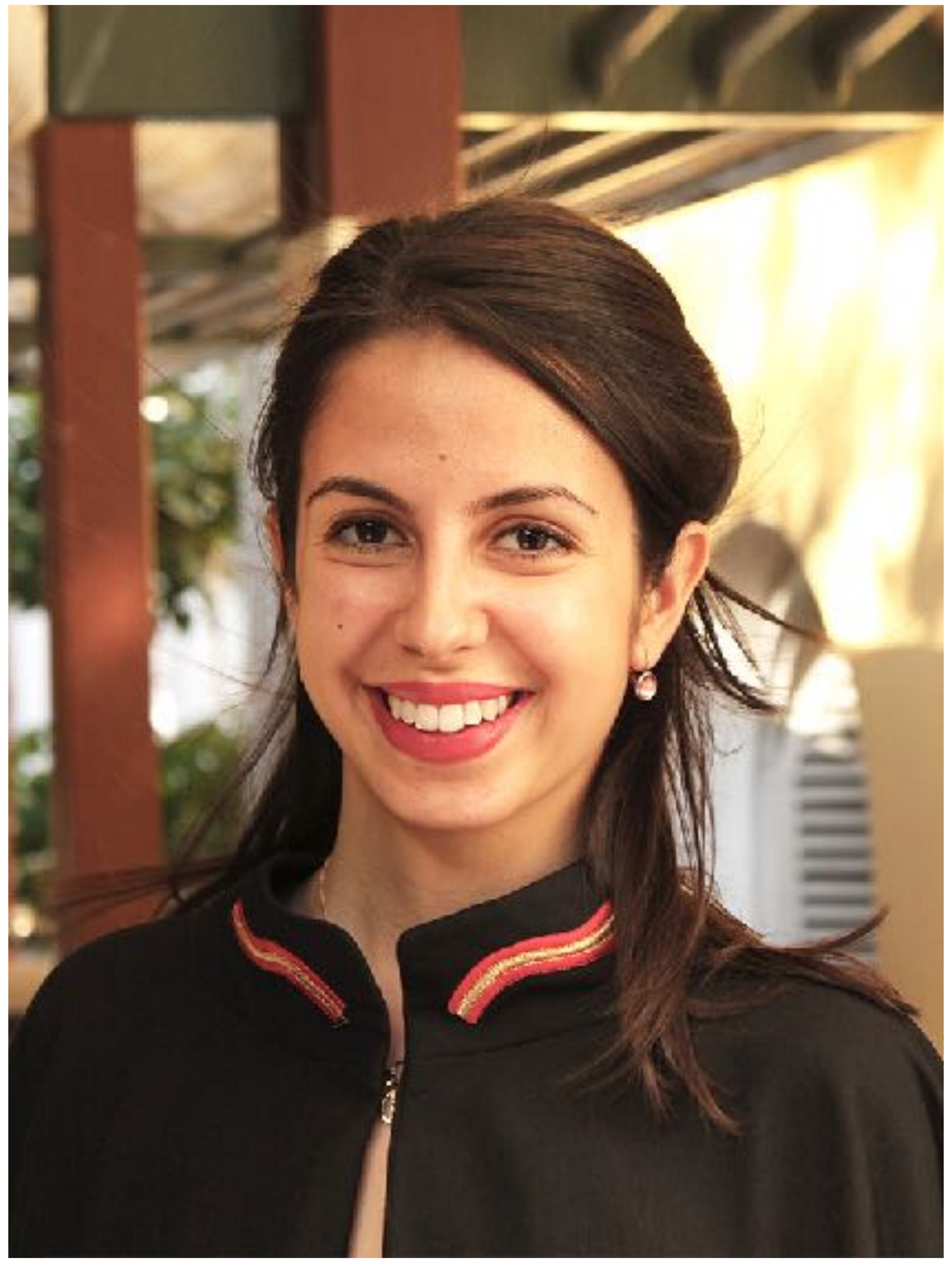
May 11th, 2021

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

DFG-Graduiertenkolleg
MATHEMATISCHE
KOMPLEXITÄTSREDUKTION

Joint work with:

Isobel Davies
OVGU



Irem Portakal
MPI MIS, Leipzig

Miruna-Stefana Sorea
Sissa, Trieste, Italy

$$A \in \mathbb{Z}^{d\times r}, \quad \mathbf{1} \in \text{rowspan}(A), \quad w \in \mathbb{R}^r_{\geq 0}$$

Log-linear models: The log-linear model $M_{A,w}$ is the set of probability distributions

$$M_{A,w} := \{p \in \Delta^\circ_{r-1} : \log p \in \log w + \text{rowspan}(A)\}$$

The model is parametrized by monomials

$$\varphi^{A,w}\colon \mathbb{R}^d \to \mathbb{R}^r$$

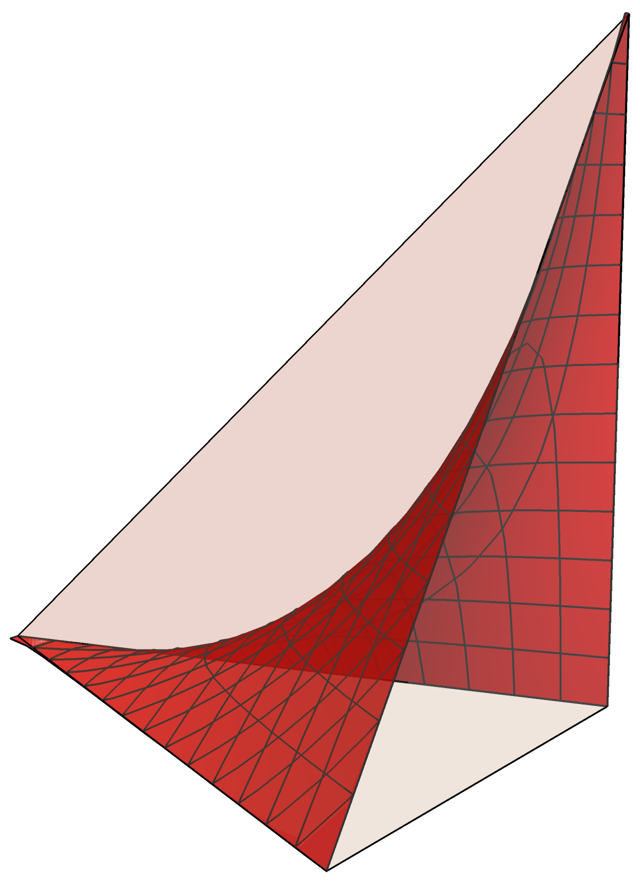
$$(t_1, \ldots, t_d) \mapsto \left(w_1 \prod_{i=1}^{d} t_i^{a_{i1}}, \ldots, w_r \prod_{i=1}^{d} t_i^{a_{ir}} \right)$$

$$M_{A,w} = \varphi^{A,w}(\mathbb{R}^d) \cap \Delta^\circ_{r-1}$$

$$A = \begin{matrix} t_0 \\ t_1 \\ s_0 \\ s_1 \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad w = (1,\ 1,\ 1,\ 1)$$

$$(t_0, t_1, s_0, s_1) \mapsto (t_0 s_0, t_0 s_1, t_1 s_0, t_1 s_1)$$

$$\begin{aligned} M_{A,w} &= \varphi^{A,w}(\mathbb{R}^4) \cap \Delta^\circ_3 \\ &= \{(p_{00}, p_{01}, p_{10}, p_{11}) \in \Delta^\circ_3 : \\ &\quad p_{00}p_{11} - p_{01}p_{10} = 0\} \end{aligned}$$



$$A \in \mathbb{Z}^{d\times r}, \quad \mathbf{1} \in \text{rowspan}(A), \quad w \in \mathbb{R}^r_{\geq 0}$$

Log-linear models: The log-linear model $M_{A,w}$ is the set of probability distributions

$$M_{A,w} := \{p \in \Delta^\circ_{r-1} : \log p \in \log w + \text{rowspan}(A)\}$$

The model is parametrized by monomials

$$\varphi^{A,w}\colon \mathbb{R}^d \to \mathbb{R}^r$$

$$(t_1, \ldots, t_d) \mapsto \left(w_1 \prod_{i=1}^{d} t_i^{a_{i1}}, \ldots, w_r \prod_{i=1}^{d} t_i^{a_{ir}} \right)$$

$$M_{A,w} = \varphi^{A,w}(\mathbb{R}^d) \cap \Delta^\circ_{r-1}$$

Why log-liner models?

- Useful and popular for analysis of categorical data.
- Hierarchical models
- undirected graphical models
- Social sciences, biology, medicine, data mining, language processinr

Let $M \subset \Delta_{r-1}$ be a discrete statistical model and $(u_1, \ldots, u_n) \in \mathbb{N}^r$ a i.i.d data vector

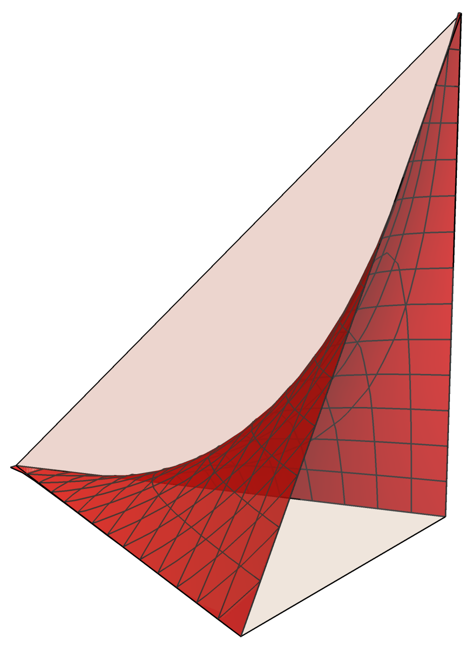
The **likelihood function** is $L(p \mid u) = \prod_{i=1}^{r} p_i^{u_i}$

The maximum likelihood estimate (MLE) for $(u_1, \ldots, u_r)$ is

$$\hat{p} = \text{argmax}_{p \in M} L(p \mid u)$$

The function $\Phi : \mathbb{N}^r \to M, \; u \mapsto \hat{p}$ is the maximum likelihood estimator.

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \quad (u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2} \right)$$



$$u_{i+} = u_{i0} + u_{i1}, \quad u_{j+} = u_{0j} + u_{1j}$$

$$A \in \mathbb{Z}^{d\times r}, \quad \mathbf{1} \in \text{rowspan}(A), \quad w \in \mathbb{R}^r_{\geq 0}$$

Birch's Theorem: Let $u = (u_1, \ldots, u_r)$ be the vector of counts of n i.i.d samples. Then the MLE in $M_{A,w}$ given u is the unique solution if it exists, to the equations

$$Au = nAp, \;\; p \in M_{A,w}$$

$$Au = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} u_{00} \\ u_{01} \\ u_{10} \\ u_{11} \end{pmatrix} = u_{++} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{u_{0+}u_{+0}}{u_{++}^2} \\ \frac{u_{0+}u_{+1}}{u_{++}^2} \\ \frac{u_{1+}u_{+0}}{u_{++}^2} \\ \frac{u_{1+}u_{+1}}{u_{++}^2} \end{pmatrix}$$

A model has rational MLE if $\Phi : \mathbb{N}^r \to M, \;\; u \mapsto \hat{p}$ is a rational function of u.

$A \in \mathbb{Z}^{d\times r}, \quad \mathbf{1} \in \text{rowspan}(A), \quad w \in \mathbb{R}^r_{\geq 0}$

<u>Birch's Theorem:</u> Let $u = (u_1, \ldots, u_r)$ be the vector of counts of n i.i.d samples. Then the MLE in $M_{A,w}$ given u is the unique solution if it exists, to the equations

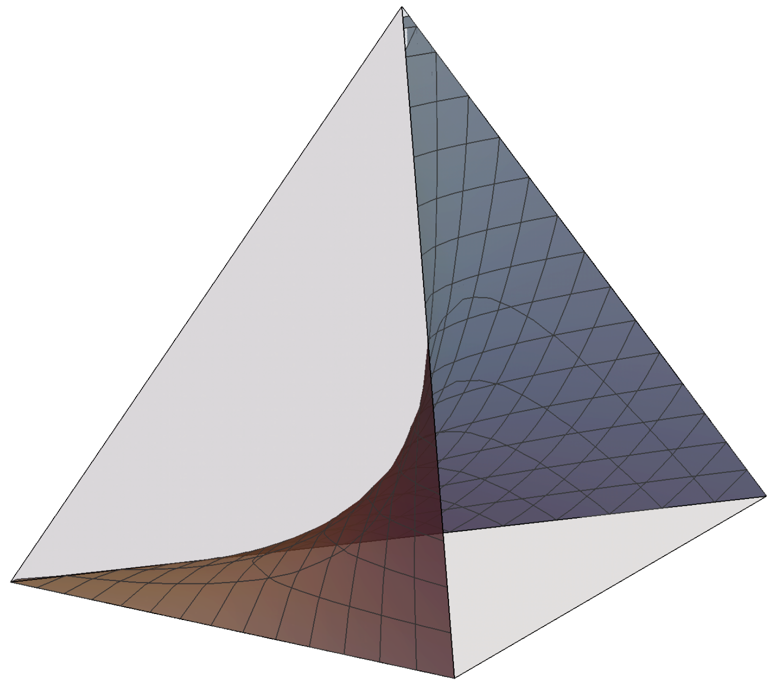
$$Au = nAp, \;\; p \in M_{A,w}$$

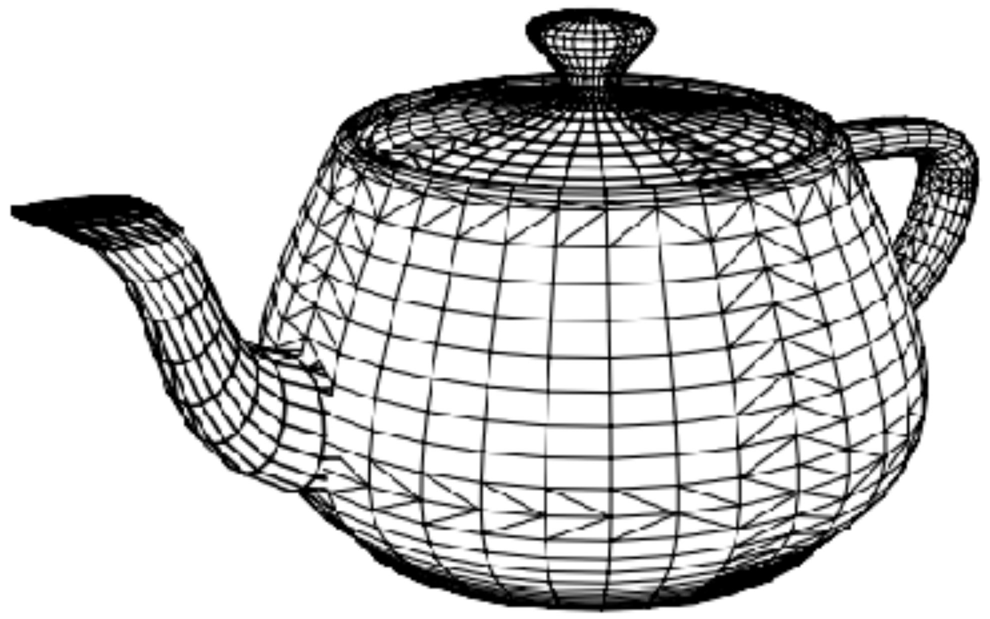
$$Au = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} u_{00} \\ u_{01} \\ u_{10} \\ u_{11} \end{pmatrix} = u_{++} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{u_{0+}u_{+0}}{u_{++}^2} \\ \frac{u_{0+}u_{+1}}{u_{++}^2} \\ \frac{u_{1+}u_{+0}}{u_{++}^2} \\ \frac{u_{1+}u_{+1}}{u_{++}^2} \end{pmatrix}$$

A model has rational MLE if $\Phi : \mathbb{N}^r \rightarrow M, \;\; u \mapsto \hat{p}$ is a rational function of u.

GOAL: Classify all log-linear models with rational MLE

Algebraic Statistics



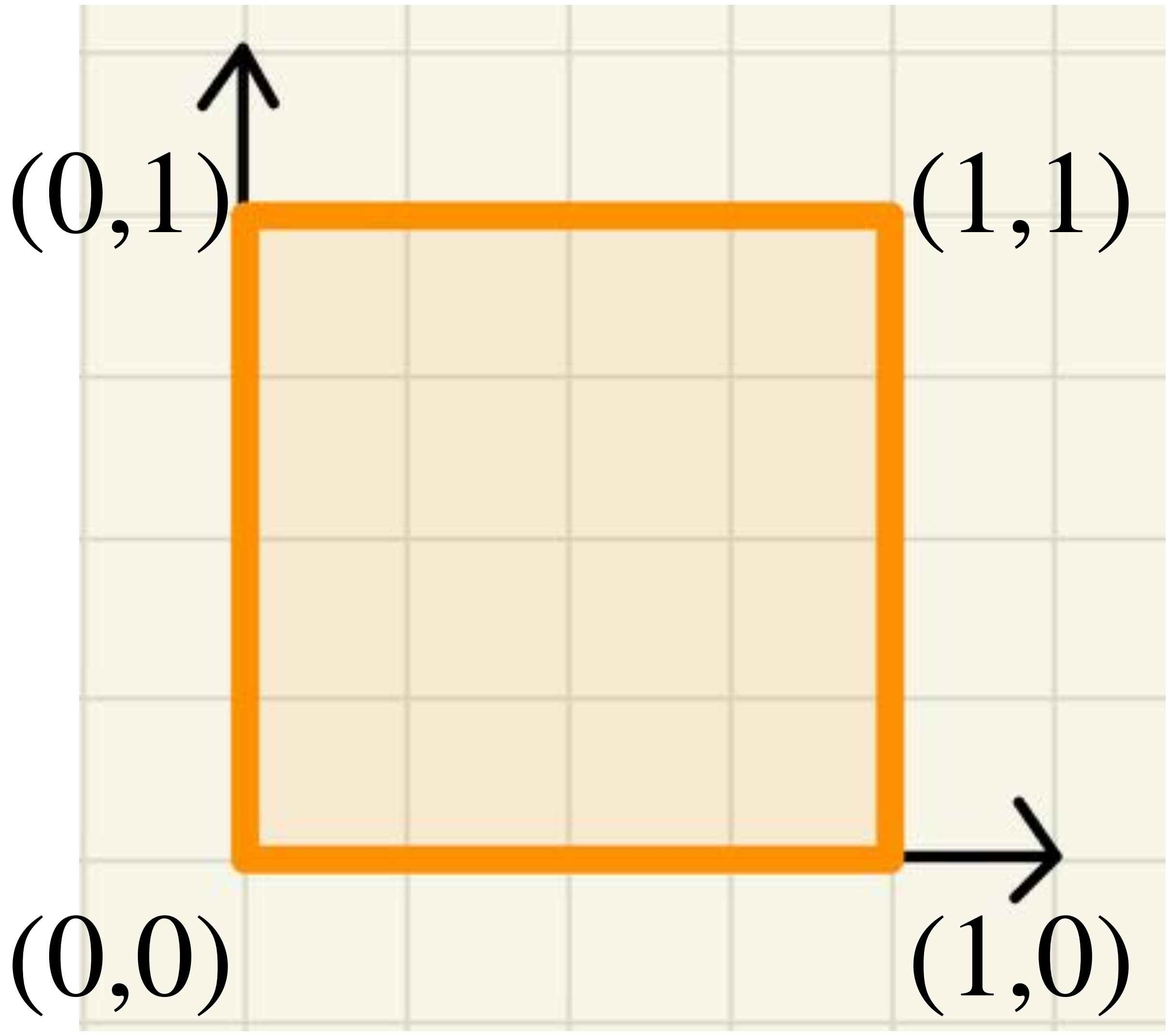


Geometric Modeling

$A \in \mathbb{Z}^{d \times r}, \quad \mathbf{1} \in \text{rowspan}(A), \quad w \in \mathbb{R}^r_{\geq 0},$

$P = \text{conv}(a_1, \ldots a_r)$ where a_j is the j-th column of A

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

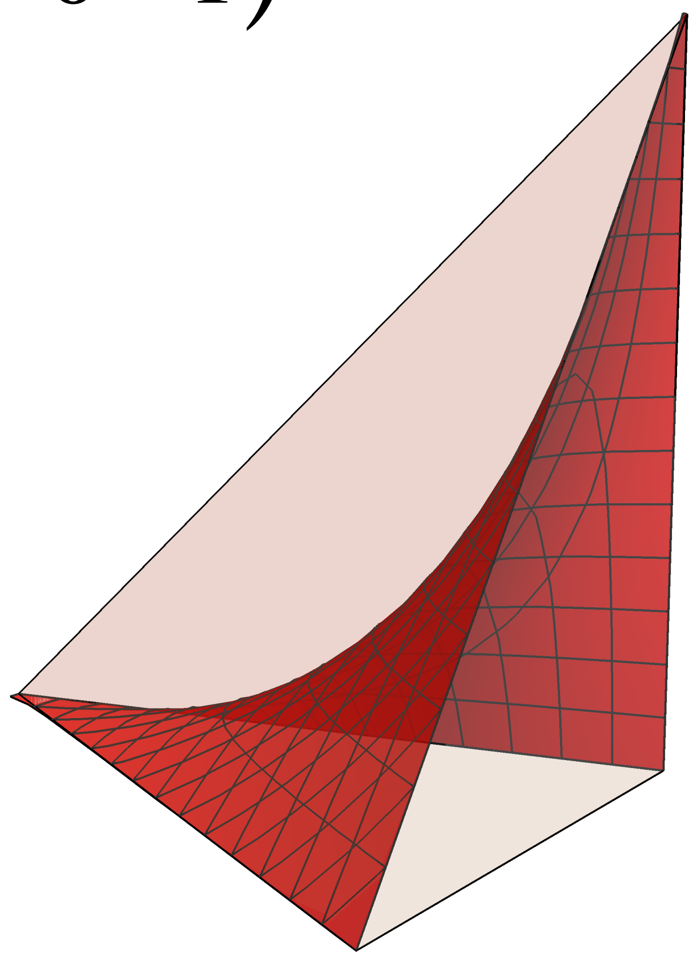


$$(s_0, s_1, t_0, t_1) \mapsto (s_0 t_0, s_0 t_1, s_1 t_0, s_1 t_1)$$

Maximum Likelihood Estimator (MLE)

$$(u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2} \right)$$

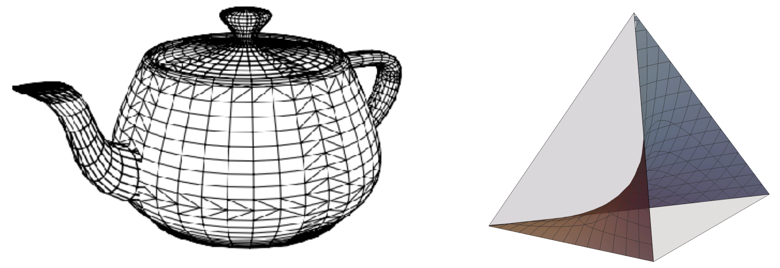
$$u_{i+} = u_{i0} + u_{i1}, \quad u_{j+} = u_{0j} + u_{1j}$$



<u>**Theorem:**</u> The statistical model $M_{A,w}$ has rational MLE if and only if the pair (P, w) has <u>rational linear precision.</u>

Garcia-Puente and Sottile (Adv. Comput. Math 2009)

Algebraic Statistics



Geometric Modeling

D. Cox and P. Clarke (2020): Give charaterization of pairs (P, w) that have strict linear precision.

Method of proof:
Horn matrices associated to statistical models with rational MLE

Questions:

- How does the Horn matrix relate to the geometry of the polytope?
- Does the normal fan of the polytope relate to the Horn matrix via primitive collections?
- Classify Horn matrices of polytopes with rational linear precision.

Theorem: A discrete statistical model has rational MLE Φ if and only if there exists a Horn pair (H, λ) such that $\Phi(u) = \left(\lambda_1 \cdot (Hu)^{h_0}, \ldots, \lambda_n \cdot (Hu)^{h_n}\right)$.

$$H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \quad \lambda = (1, 1, 1, 1), \;\; u = \begin{pmatrix} u_{00} \\ u_{01} \\ u_{10} \\ u_{11} \end{pmatrix}, \;\; Hu = \begin{pmatrix} u_{0+} \\ u_{1+} \\ u_{+0} \\ u_{+1} \\ -u_{++} \\ -u_{++} \end{pmatrix}$$

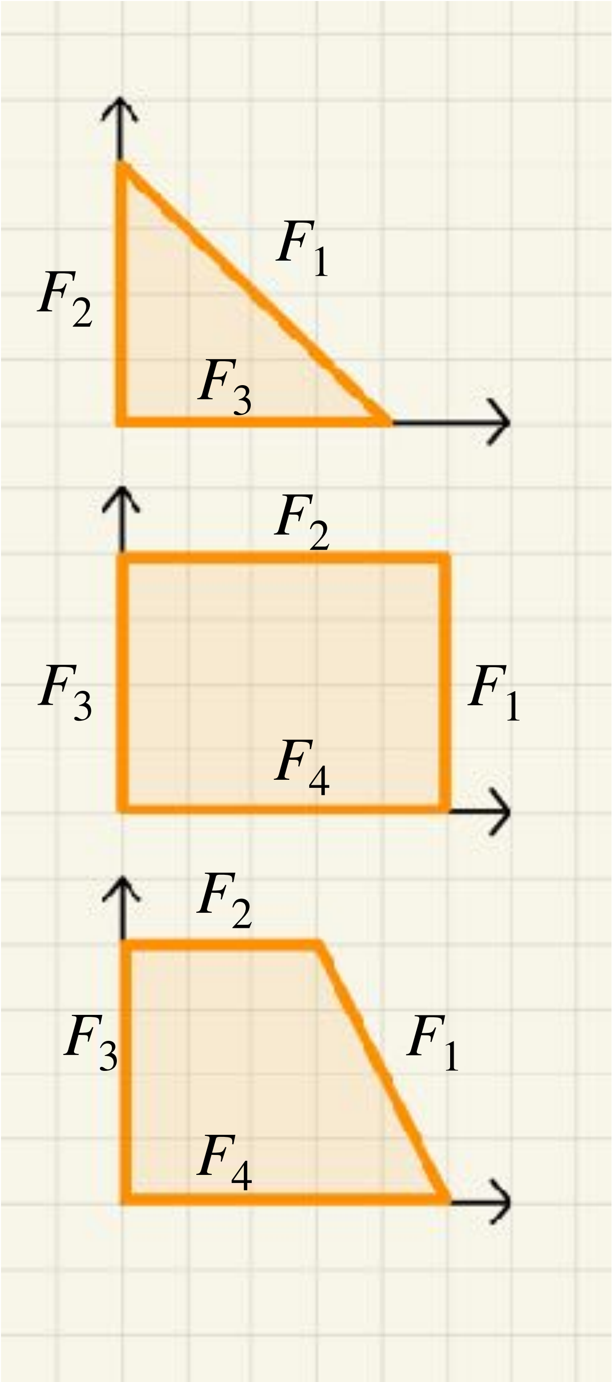
$$u_{i+} = u_{i0} + u_{i1}, \qquad u_{j+} = u_{0j} + u_{1j}$$

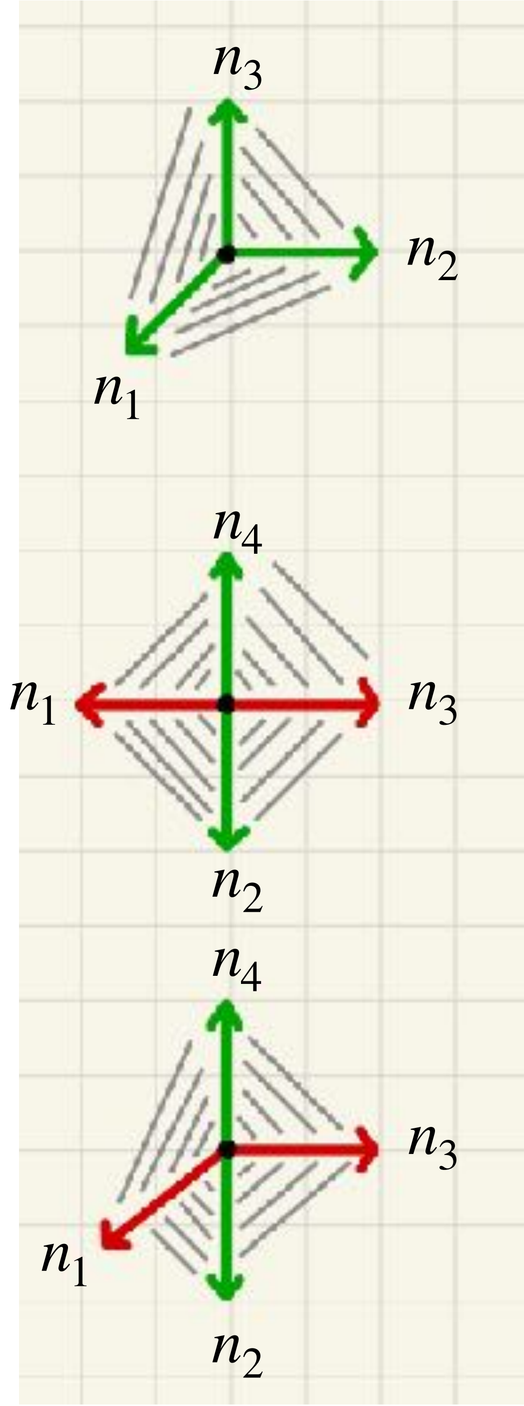
$$(u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2} \right)$$

Duarte, Marigliano, Sturmfels (Bernoulli 2020)

Breakout rooms

- How to construct a Horn pair (H, λ) ?
- Write down some Horn matrix that comes to mind. Is it the MLE of a statistical model for a suitable λ ?
- How to obtain the Horn matrices from the picture?





$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}, \quad (a\Delta_2, w), \quad a = 1, \quad \lambda = (-1, \ -1, \ -1)$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \quad (a\Delta_1 \times b\Delta_1, w), \quad a = b = 1, \quad \lambda = (1, \ 1, \ 1, \ 1)$$

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -1 & -1 \end{pmatrix}, \quad (T_{a,b,d}, w), a = b = d = 1, \quad \lambda = (-1, -2, -1, 1, 1)$$

Theorem: A discrete statistical model has rational MLE Φ if and only if there exists a Horn pair (H, λ) such that $\Phi(u) = \left(\lambda_1 \cdot (Hu)^{h_0}, \ldots, \lambda_n \cdot (Hu)^{h_n}\right)$.

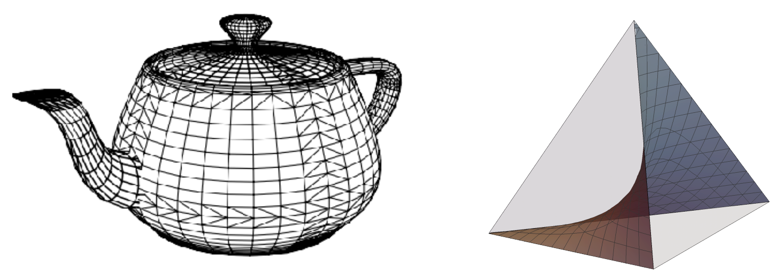
$$H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \quad \lambda = (1, 1, 1, 1), \;\; u = \begin{pmatrix} u_{00} \\ u_{01} \\ u_{10} \\ u_{11} \end{pmatrix}, \;\; Hu = \begin{pmatrix} u_{0+} \\ u_{1+} \\ u_{+0} \\ u_{+1} \\ -u_{++} \\ -u_{++} \end{pmatrix}$$

$$u_{i+} = u_{i0} + u_{i1}, \quad u_{j+} = u_{0j} + u_{1j}$$

$$(u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2} \right)$$

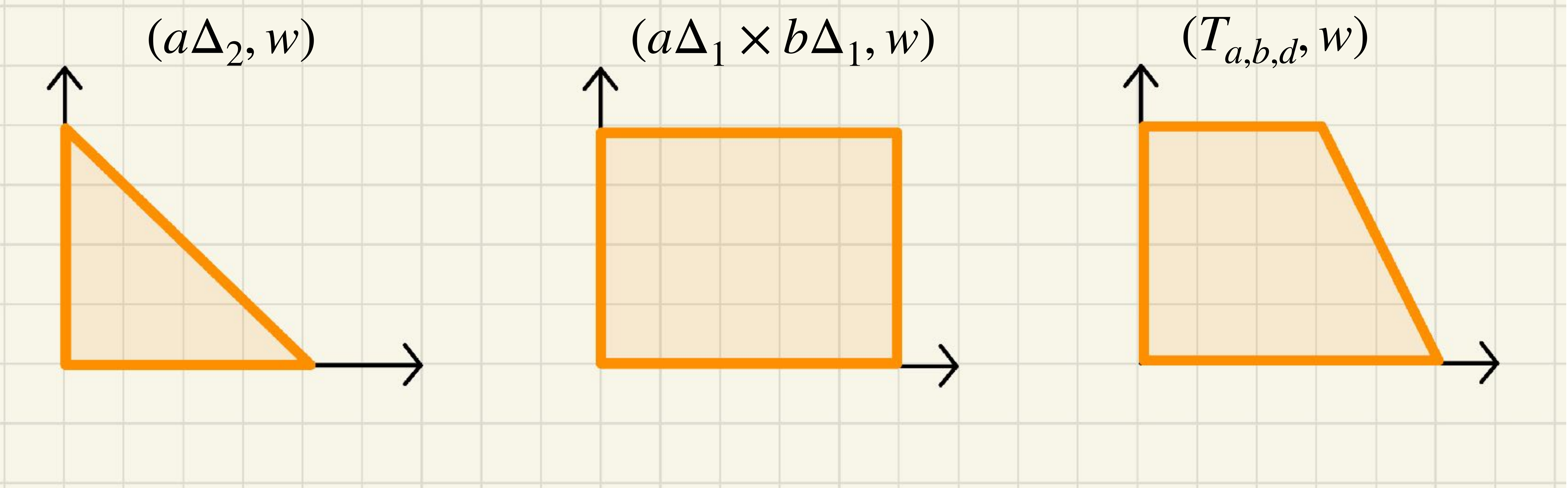
Duarte, Marigliano, Sturmfels (Bernoulli 2020)

Algebraic Statistics



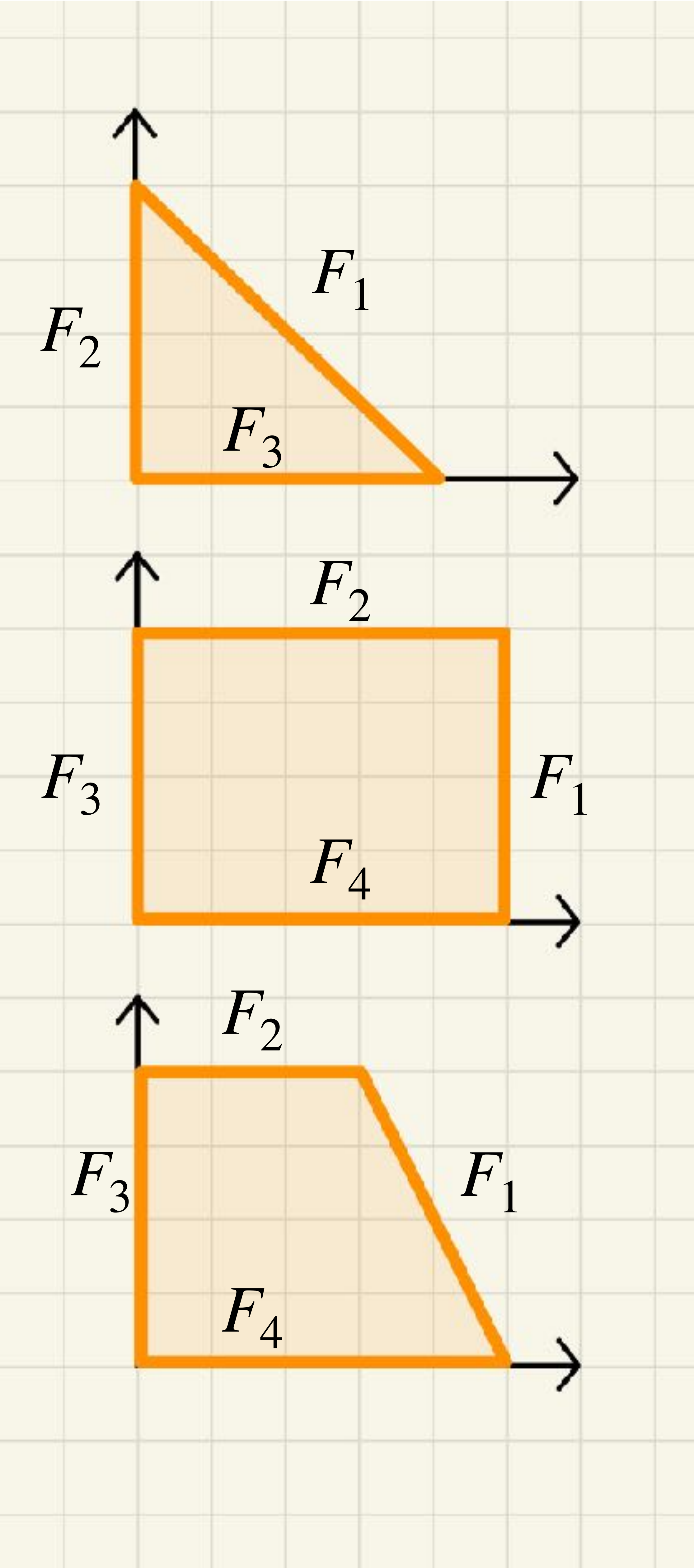
Geometric Modeling

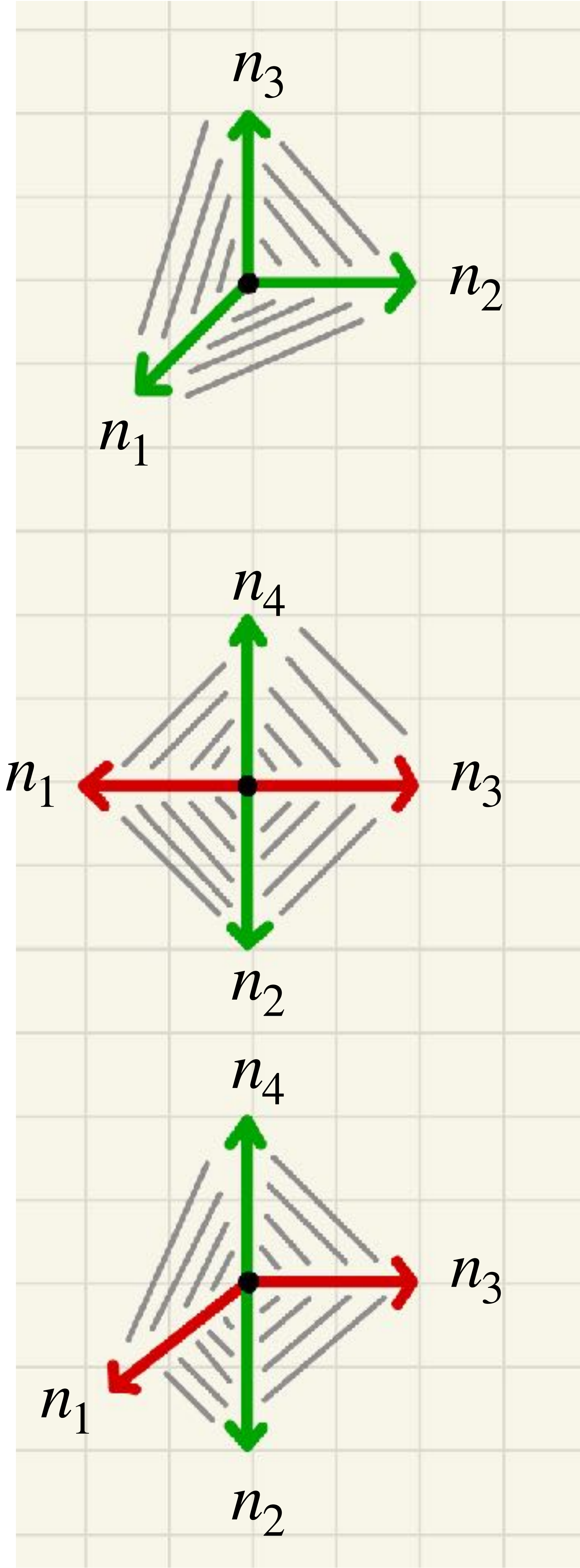
Polytopes with rational linear precision in dimension two



Bezier triangle patch **Tensor product patch** **Trapezoidal patch**

Von Bothmer, Ranestad and Sottile (FoCM 2010)





Normal Fans and primitive collections

A fan Σ in $N_{\mathbb{R}}$ is a finite collection of cones such that:

(1) Every $\sigma \in \Sigma$ is a strongly convex rational polyhedral cone.

(2) For each $\sigma \in \Sigma$, each face of σ is also in Σ.

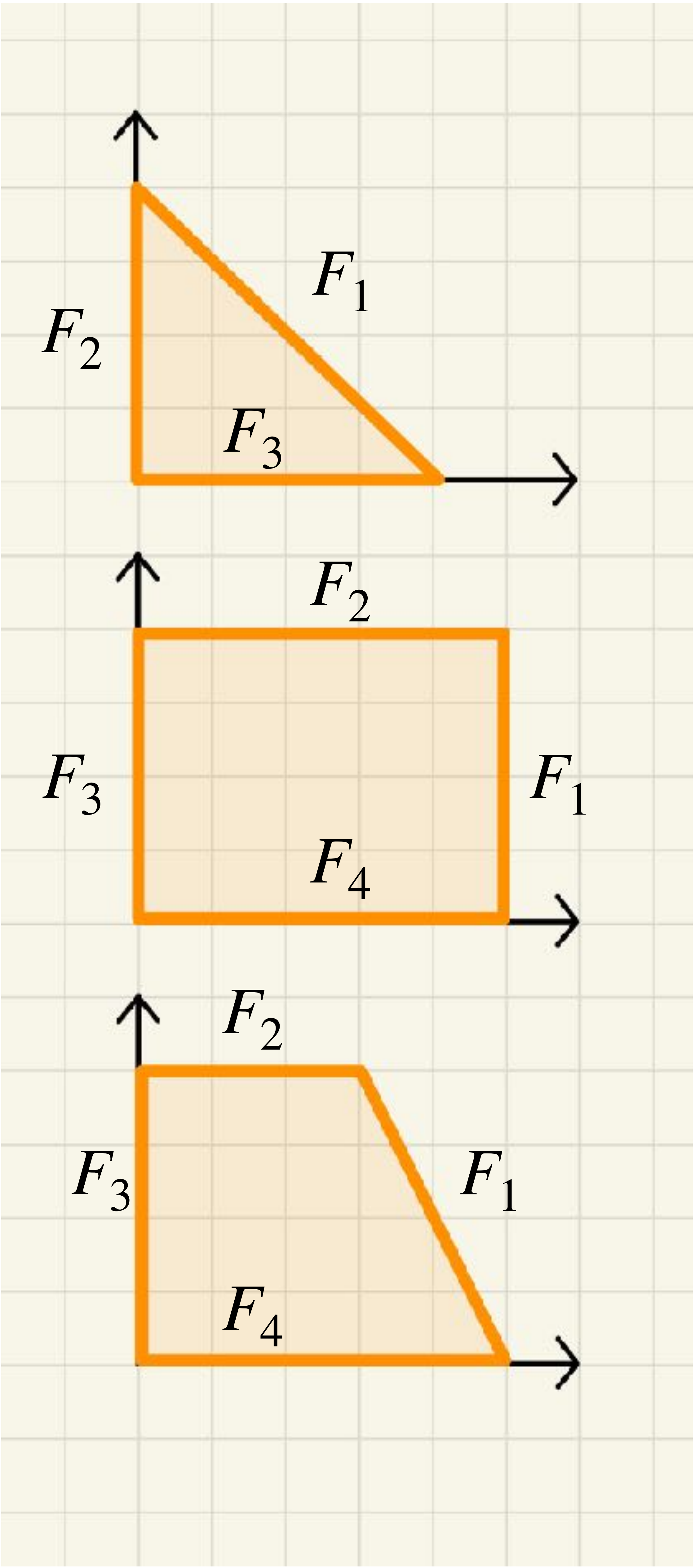
(3) For all $\sigma_1, \sigma_2 \in \Sigma$, the intersection $\sigma_1 \cap \sigma_2$ is a face of each.

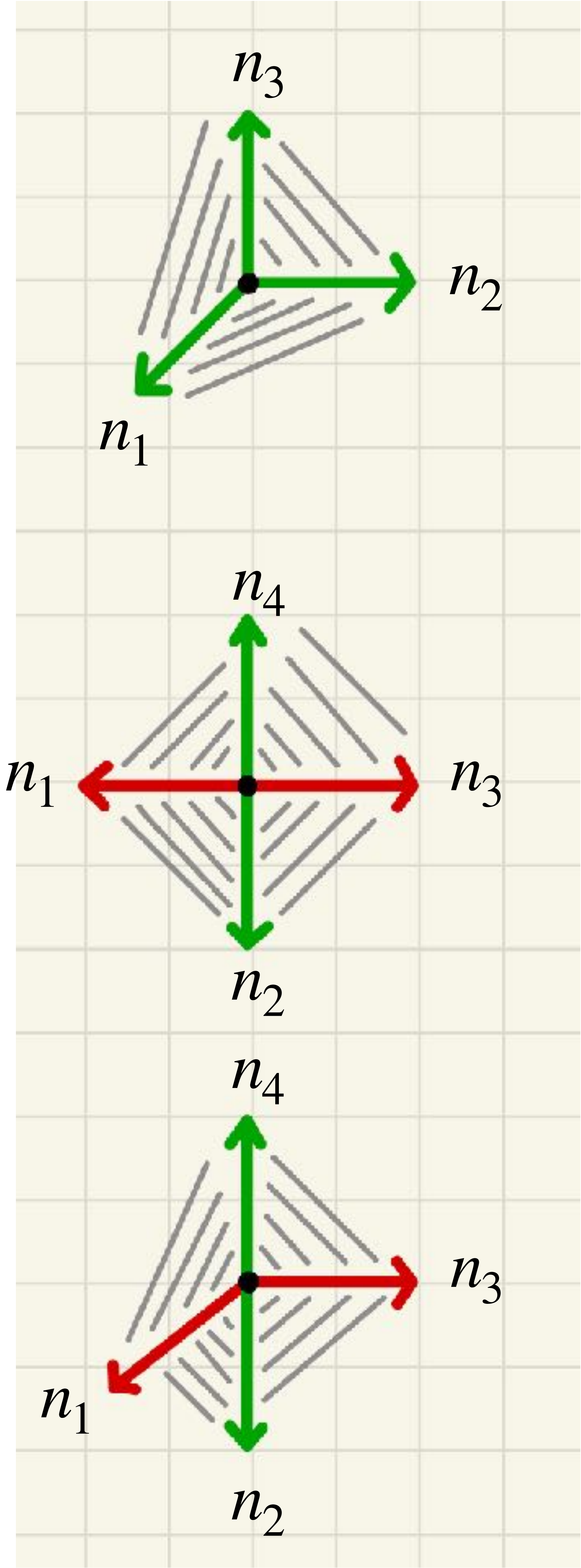
$\Sigma(1) :=$ one-dimensional faces

A subset $C \subset \Sigma(1)$ is a **primitive collection** if:

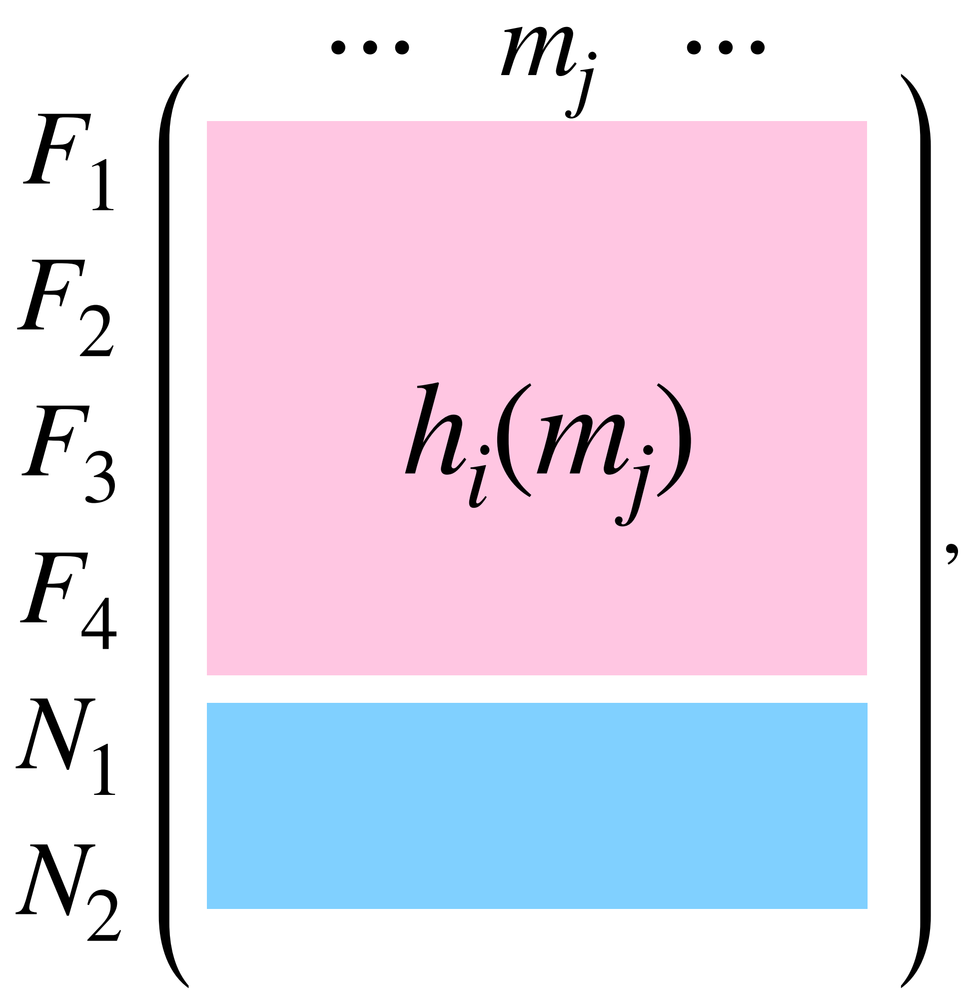
(a) $C \subsetneq \sigma(1)$ for all $\sigma \in \Sigma$,

(b) For every proper subset $C' \subsetneq C$ there is $\sigma \in \Sigma$ with $C' \subset \sigma(1)$





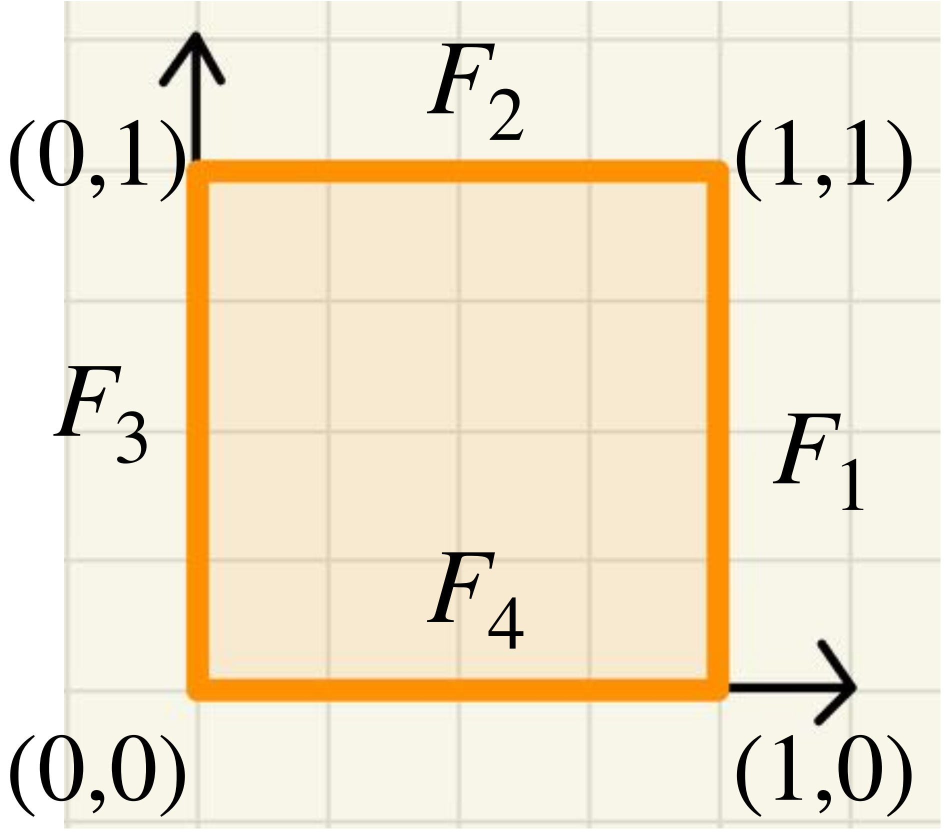
<u>**Theorem:**</u> For all pairs (P, w) in dimension two with rational linear precision, the Horn matrix for the statistical model associated to (P, w) is:

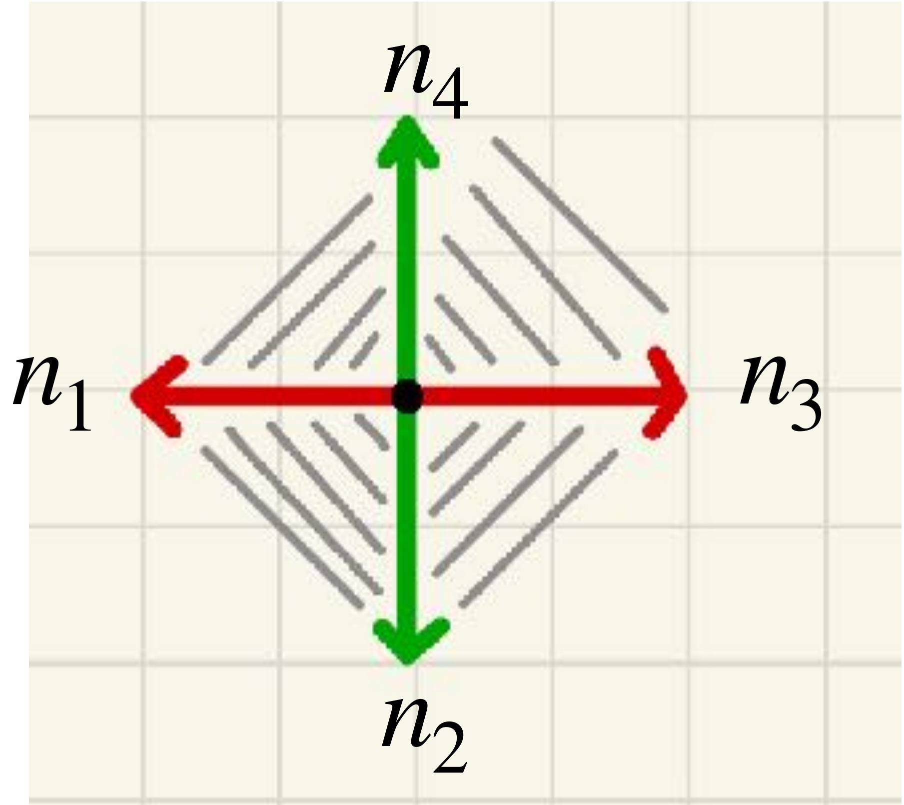


where $h_i(m_j)$ is the lattice distance from F_i to the lattice point m_j . The negative rows are obtained by adding the faces in the same primitive collection.

Davies, **Duarte**, Portakal and Sorea (in preparation)

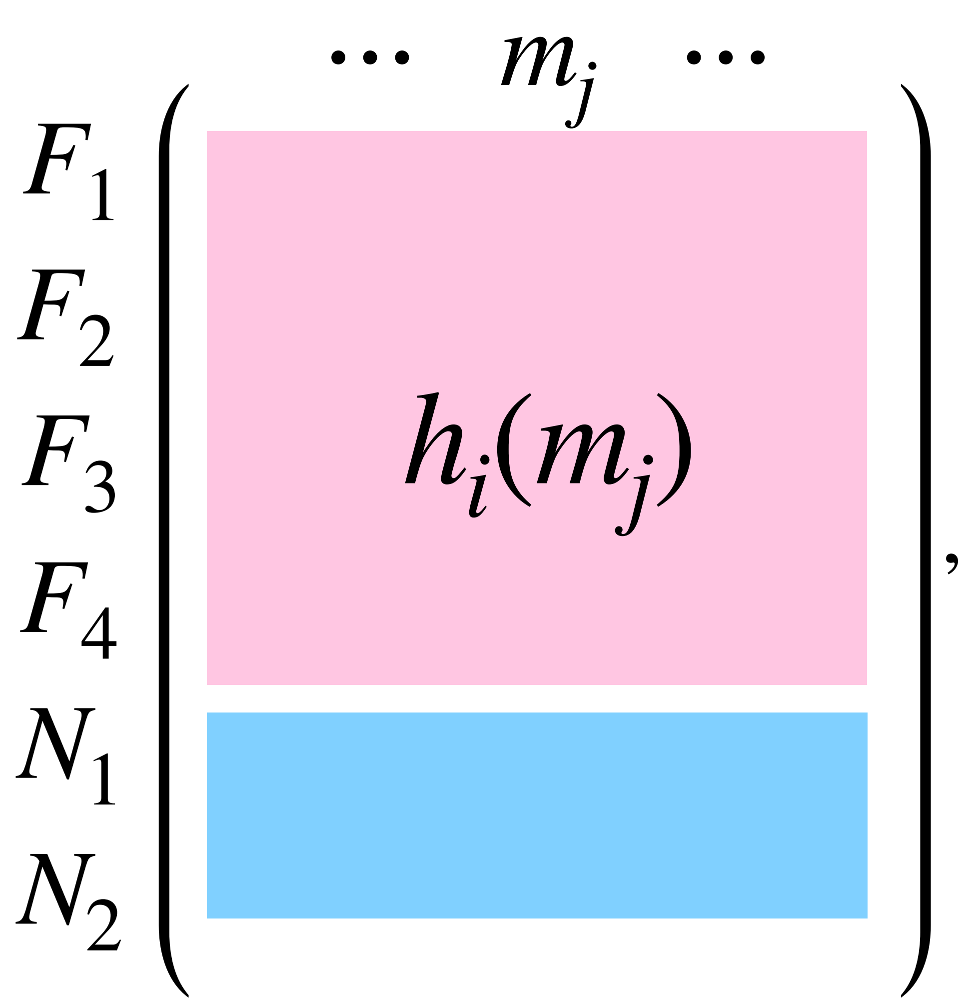
$(a\Delta_1 \times b\Delta_1, w),\ \ a = b = 1$





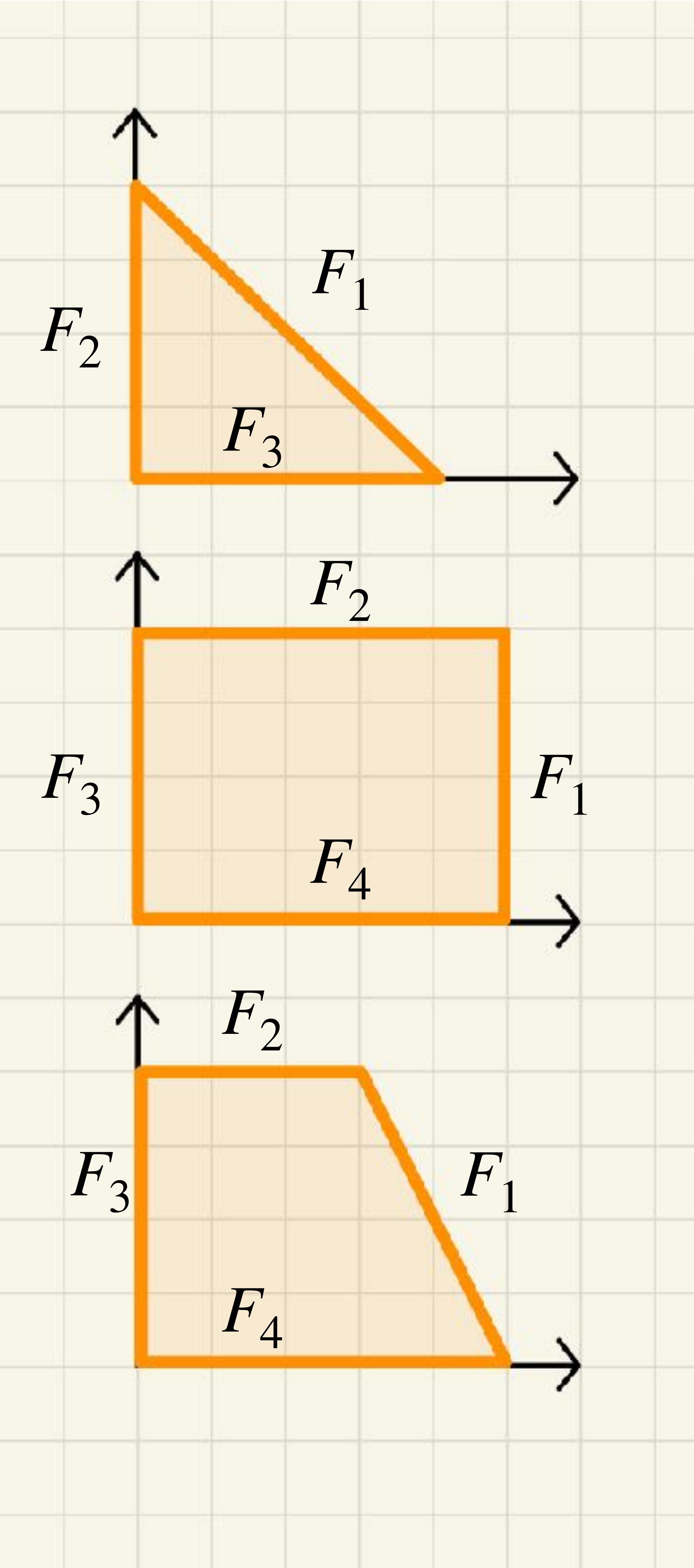
$$
\begin{array}{c} \\ F_2 \\ F_4 \\ F_3 \\ F_1 \\ -(F_2+F_4) \\ -(F_1+F_3) \end{array}
\begin{array}{c} \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \end{array} \\
\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix} \end{array},
\quad \lambda = (1,\ \ 1,\ \ 1,\ \ 1)
$$

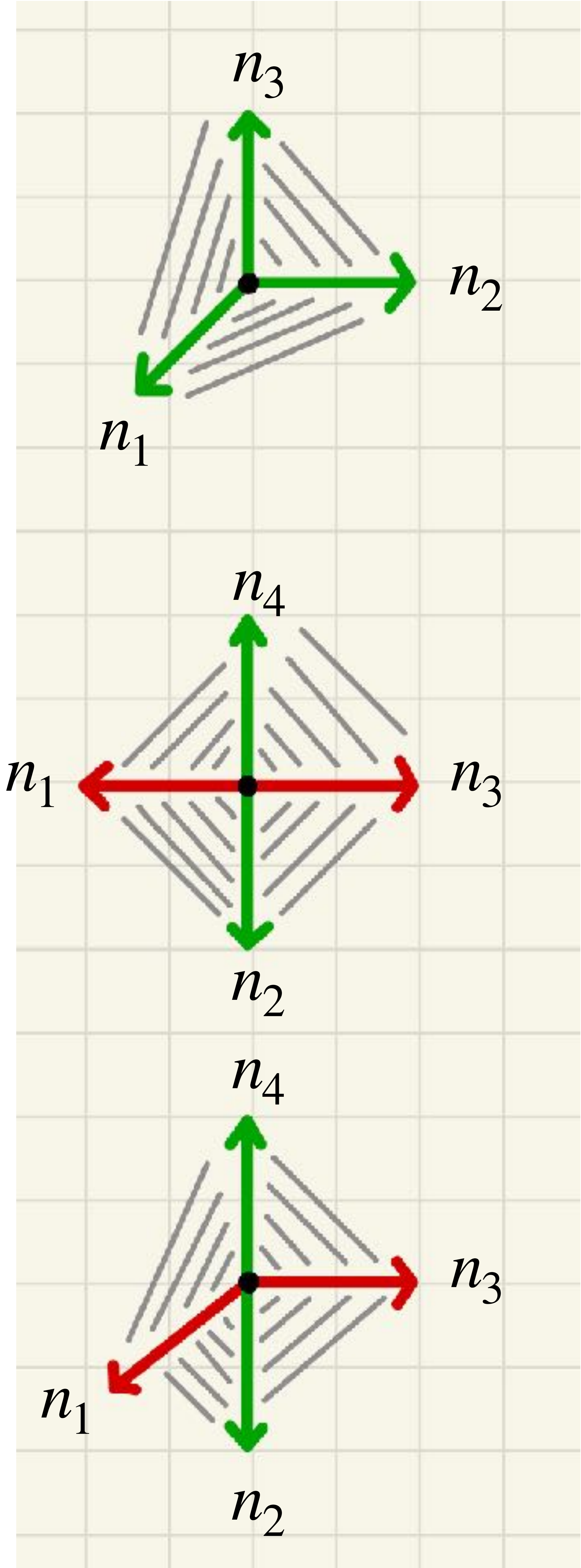
Theorem: For all pairs (P, w) in dimension two with rational linear precision, the Horn matrix for the statistical model associated to (P, w) is:



where $h_i(m_j)$ is the lattice distance from F_i to the lattice point m_j. The negative rows are obtained by adding the faces in the same primitive collection.

Davies, **Duarte**, Portakal and Sorea (in preparation)





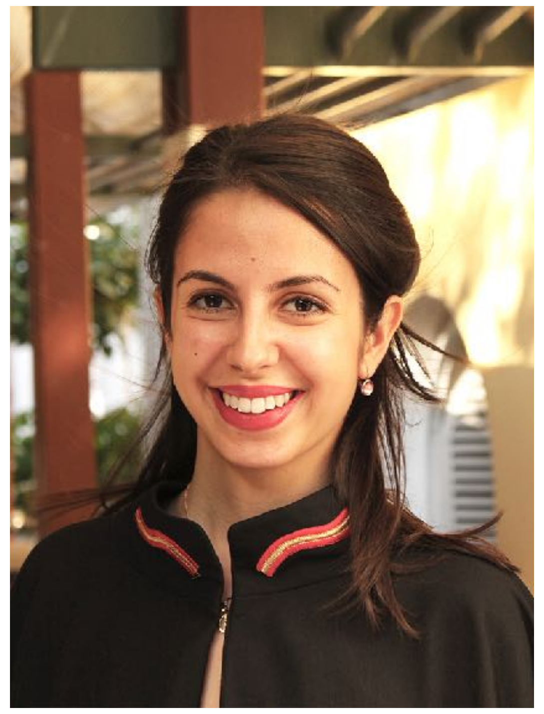
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}, \quad (a\Delta_2, w), \quad a = 1 \quad \lambda = (-1, \; -1, \; -1)$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \quad (a\Delta_1 \times b\Delta_1, w), \; a = b = 1 \quad \lambda = (1, \; 1, \; 1, \; 1)$$

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -1 & -1 \end{pmatrix}, \quad (T_{a,b,d}, w), a = b = d = 1 \quad \lambda = (-1, -2, -1, 1, 1)$$

Thank you

Isobel Davies
OVGU



Irem Portakal
MPI MIS, Leipzig

Miruna-Stefana Sorea
Sissa, Trieste, Italy

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

DFG-Graduiertenkolleg
MATHEMATISCHE
KOMPLEXITÄTSREDUKTION

Thank you

Tensor Voices

Eliana Duarte and Thomas Kahle

SCIENCE · 2021

Tensor Voices is a short podcast series about tensors.

Episodes

29. MARCH

Kaie Kubjas

Kaie Kubjas - Tensors are a natural way to encode multivariate data completion Tensor network complexity of multilinear maps Tensor c Dimension of tensor network varieties