Geometry of Hidden-Visible Products of Statistical Models

Max Planck Institute for Mathematics in the Sciences

Guido Montúfar
montufar@mis.mpg.de

Jason Morton
morton@math.psu.edu

PENNSTATE University Park 1855

Abstract

Dimension of marginals of exponential families

- naïve Bayes models / mixtures of products / Segre secants
- restricted Boltzmann machines (RBMs) / Hadamard products of Segre secants
- mixtures of interaction models
- exponential family harmoniums

Solution following Draisma [4] and Cueto, Morton, and Sturmfels [3] by

- tropicalization of hidden-visible Kronecker product models
- inference functions (slicings of polytopes by normal-fans)

Hidden-Visible Product Models

$$\Delta_{\mathcal{X}} := \left\{ (p_v)_{v\in\mathcal{X}} \in \mathbb{R}^{\mathcal{X}} \colon p_v > 0, \sum_{v\in\mathcal{X}} p_v = 1 \right\}$$

$$\mathcal{E}_{A\otimes B} := \left\{ \exp\left(\theta^\top [A\otimes B] - \psi(\theta)\right) : \theta\in\mathbb{R}^{ab} \right\} \subseteq \Delta_{\mathcal{X}\times\mathcal{Y}}$$

$$\mathcal{V}_{A\otimes B} := \left\{ \sum_{h\in\mathcal{Y}} \exp\left(\theta^\top [A\otimes B_h] - \psi(\theta)\right) : \theta\in\mathbb{R}^{ab} \right\} \subseteq \Delta_{\mathcal{X}}$$

Jacobian Rank and Tropical Model

Jacobian for large parameters

$$\max_\theta \operatorname{rank}\left(J_{\mathcal{V}_{A\otimes B}}(\theta)\right) + 1 \geq \max_\theta \operatorname{rank}\left(A_x \otimes B_{h_\theta(x)}\right)_x$$

Inference function

$$h_\theta(x) := \operatorname{argmax}_y p_\theta(y|x) = \operatorname{argmax}_y \langle \Theta A_x, B_y\rangle$$

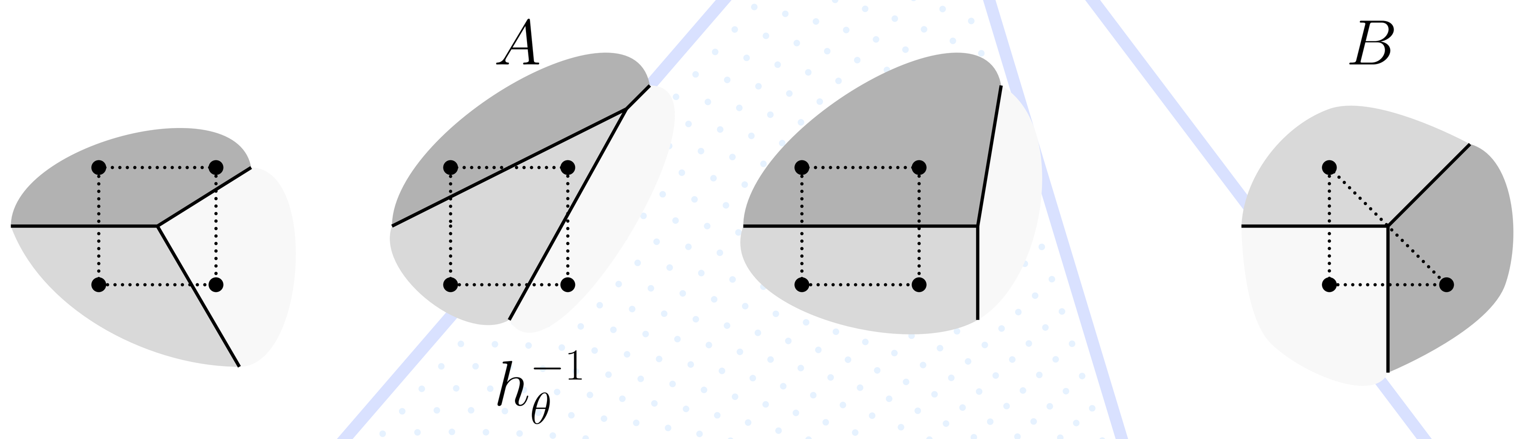
Tropical morphism

$$\Phi(x;\theta) \;=\; \theta^\top \mathcal{A}_\theta \;=\; \theta^\top \left[A_x \otimes B_{h_\theta(x)}\right]$$

By Bieri-Groves

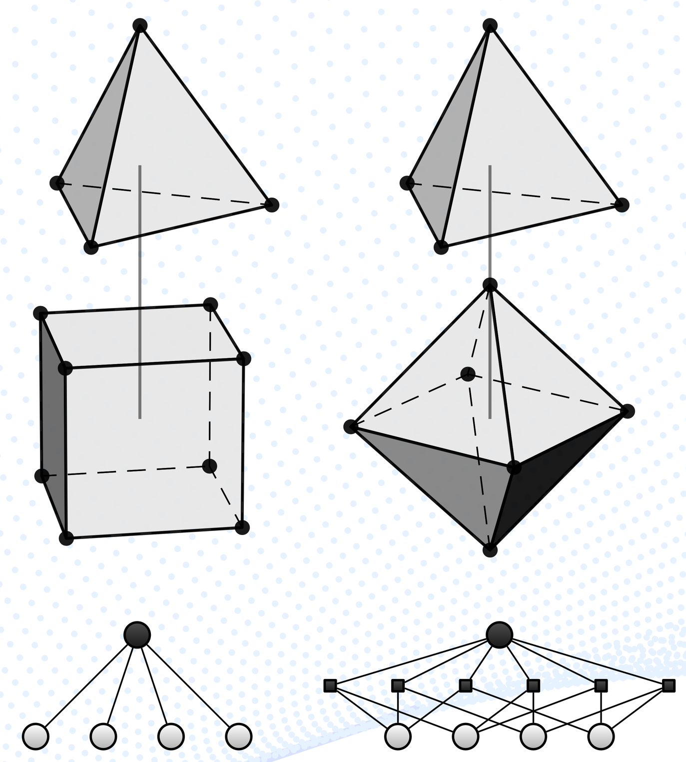
$$\dim(\mathcal{V}_{A\otimes B}) + 1 \geq \max_\theta \operatorname{rank} \mathcal{A}_\theta$$

Slicings



Mixtures

(a) (b)

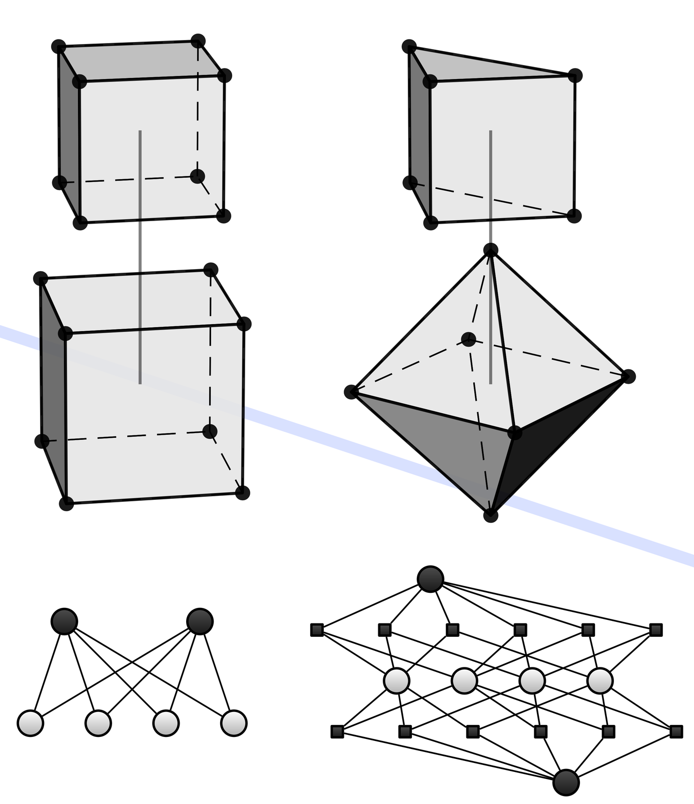


(a) Segre secants [1, 4]
(b) Mixtures of hierarchical models

$$\mathcal{A}_\theta \;=\; \oplus_y A_{h_\theta^{-1}(y)}$$

Products of Mixtures

(a) (b)

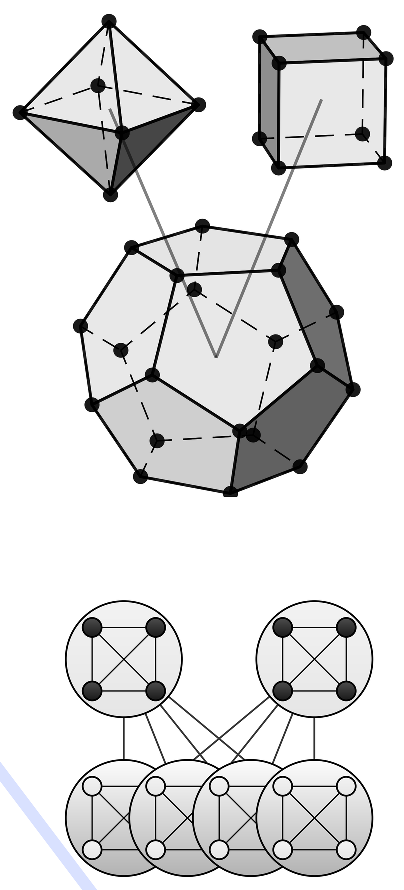


(a) RBMs [3, 8]
(b) Products of mixtures of hierarchical models

$$\mathcal{A}_\theta \;=\; \left(\mathcal{A}_{\theta^1}\middle|\cdots\middle|\mathcal{A}_{\theta^m}\right)$$

$$\mathcal{A}_{\theta^j} \;=\; \oplus_{y_j} A_{h_{\theta^j}^{-1}(y_j)}$$

Products of Restricted Mixtures



- Exponential family harmonium
- Kronecker products of exponential families / toric varieties

$$\mathcal{A}_\theta \;=\; \left(\mathcal{A}_{\theta^1}\middle|\cdots\middle|\mathcal{A}_{\theta^m}\right)$$

$$\mathcal{A}_{\theta^j} \;=\; \left[A_x \otimes B^j_{h_{\theta^j}(x)}\right]_x$$

Results

- Combinatorial classification of non-defective cases (extensive but still incomplete)
- Examples of defective cases
- Preprint will be available soon!

Open Problems

- Combinatorics of Kronecker product polytopes
  Markov bases of bipartite graphs (J. Rauh)
- Ranks of Khatri-Rao products (block-wise Kronecker products)
- Sub-model dimensions
  non-defective mixture $\overset{?}{\Rightarrow}$ non-defective restricted mixture
  Orthogonal mixed parametrizations
- Deep Models

[1] M. V. Catalisano, A. V. Geramita, and A. Gimigliano. Secant varieties of $\mathbb{P}^1\times\cdots\times\mathbb{P}^1$ (n-times) are not defective for $n\geq 5$. J. Algebraic Geometry, 20:295–327, 2011.

[2] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein. Covering Codes. North-Holland Mathematical Library. Elsevier Science, 2005.

[3] M. A. Cueto, J. Morton, and B. Sturmfels. Geometry of the restricted Boltzmann machine. In M. A. G. Viana and H. P. Wynn, editors, Algebraic methods in statistics and probability II, AMS Special Session, volume 2. American Mathematical Society, 2010.

[4] J. Draisma. A tropical approach to secant dimensions. J. Pure Appl. Algebra, 212(2):349–363, 2008.

[5] E. Gilbert. A comparison of signalling alphabets. Bell System Technical Journal, 31:504–522, 1952.

[6] S. Hoşten and S. Sullivant. Gröbner bases and polyhedral geometry of reducible and cyclic models. J. Combin. Theory Ser. A, 100:277–301, 2002.

[7] G. Montúfar. Mixture decompositions of exponential families using a decomposition of their sample spaces. Kybernetika, 49(1), 2013.

[8] G. Montúfar and J. Morton. Discrete restricted Boltzmann machines. 2013. Preprint available at http://arxiv.org/abs/1301.3529.

[9] R. Varshamov. Estimate of the number of signals in error correcting codes. Doklady Akad. Nauk SSSR, 117:739–741, 1957.

Algebraic Statistics 2014 at Illinois Institute of Technology (as2014 at iit), May 19-22, 2014 Chicago, IL, USA.