Normal and Unimodular Hierarchical Models

Daniel Irving Bernstein and Seth Sullivant

North Carolina State University

dibernst@ncsu.edu
`http://www4.ncsu.edu/~dibernst/`

`http://arxiv.org/abs/1502.06131`
`http://arxiv.org/abs/1508.05461`

October 3, 2015

Example

- Let T be the following $3 \times 2 \times 2$ table

$$\begin{array}{cc} \text{front} & \text{back} \\ \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{pmatrix} & \begin{pmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 2 \end{pmatrix} \end{array}$$

- If we sum entries going down, we get the 2-way margin below. If we sum entries going left and back, we get the 1-way margin below.

$$\begin{pmatrix} 3 & 6 \\ 6 & 2 \end{pmatrix} \qquad \begin{pmatrix} 5 \\ 6 \\ 6 \end{pmatrix}$$

- We are interested in the matrix that maps tables to margins

Main Definition

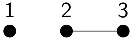
- $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ is an integer vector, $d_i \geq 2$
- $\mathcal{C}$ denotes a simplicial complex on $[n]$
- facet($\mathcal{C}$) denotes the inclusion-maximal faces of $\mathcal{C}$

Definition

Let $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ be the matrix defined as follows:

- Columns are indexed by elements of $\bigoplus_{i=1}^{n}[d_i]$
- Rows are indexed by $\bigoplus_{F\in\text{facet}(\mathcal{C})}\bigoplus_{j\in F}[d_j]$
- Entry in row $(F, (j_1, \ldots, j_k))$ and column $(i_1, \ldots, i_n)$ is 1 if $i|_F = (j_1, \ldots, j_k)$
- All other entries are 0

Example

- Let $n = 3$ with $d_1 = 3, d_2 = 2, d_3 = 2$
- Let $\mathcal{C}$ be the complex 

- Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is the following matrix:

$$\begin{array}{l|cccccccccccc}
 & \substack{1\\1\\1} & \substack{1\\1\\2} & \substack{1\\2\\1} & \substack{1\\2\\2} & \substack{2\\1\\1} & \substack{2\\1\\2} & \substack{2\\2\\1} & \substack{2\\2\\2} & \substack{3\\1\\1} & \substack{3\\1\\2} & \substack{3\\2\\1} & \substack{3\\2\\2} \\
\hline
\{1\},1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\{1\},2 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
\{1\},3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\hline
\{2,3\},11 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\{2,3\},12 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
\{2,3\},21 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
\{2,3\},22 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}$$

Motivating Question

Definition (Unimodularity)

Assume $A \in \mathbb{Z}^{d \times n}$ has full row rank. We say that A is **unimodular** if all nonsingular $d \times d$ submatrices have determinant ± 1.

Definition (Normality)

We say that $A \in \mathbb{Z}^{d \times n}$ is **normal** if $\mathbb{Z}A \cap \mathbb{R}_{\geq 0}A = \mathbb{N}A$. This is a weaker condition than unimodularity.

Question

When is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular? When is it normal?

Observation

If $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is unimodular/normal, then so is $\mathcal{A}_{\mathcal{C},(2,\ldots,2)}$.

Our Results

Our results include:

- Necessary and sufficient conditions on $\mathcal{C}$ guaranteeing unimodularity of $\mathcal{A}_{\mathcal{C},\mathbf{2}}$
- Progress towards a similar classification for normal $\mathcal{A}_{\mathcal{C},\mathbf{2}}$

Note

We abuse language and say that a simplicial complex $\mathcal{C}$ is unimodular/normal to mean that $\mathcal{A}_{\mathcal{C},(2,\ldots,2)}$ is unimodular/normal.

Applications include:

- Integer programming
- Disclosure limitation
- Compute Markov basis via toric fiber product (Rauh-Sullivant 2014)

Unimodularity-Preserving Operations

Definition (Adding a cone vertex)

If $\mathcal{C}$ is a simplicial complex on $[n]$, define $\mathrm{cone}(\mathcal{C})$ to be the complex on $[n+1]$ with facets

$$\mathrm{facet}(\mathrm{cone}(\mathcal{C})) = \{F \cup \{n+1\} : F \in \mathrm{facet}(\mathcal{C})\}.$$

Definition (Adding a ghost vertex)

If $\mathcal{C}$ is a simplicial complex on $[n]$, define $G(\mathcal{C})$ to be the simplicial complex on $[n+1]$ that has exactly the same faces as $\mathcal{C}$.

Definition (Alexander Duality)

If $\mathcal{C}$ is a simplicial complex on $[n]$, then the *Alexander dual* complex $\mathcal{C}^*$ is the simplicial complex on $[n]$ with facets

$$\mathrm{facet}(\mathcal{C}^*) = \{[n] \setminus S : S \text{ is a minimal non-face of } \mathcal{C}\}.$$

Unimodularity: Constructive Classification

Definition

We say that a simplicial complex $\mathcal{C}$ is *nuclear* if it satisfies one of the following:

1. $\mathcal{C} = \Delta_k$ for some $k \geq -2$ (i.e. a simplex)
2. $\mathcal{C} = \Delta_m \sqcup \Delta_n$ (i.e. a disjoint union of simplices)
3. $\mathcal{C} = \text{cone}(\mathcal{D})$ where $\mathcal{D}$ is nuclear
4. $\mathcal{C} = G(\mathcal{D})$ where $\mathcal{D}$ is nuclear
5. $\mathcal{C}$ is the Alexander dual of a nuclear complex.

Theorem (B-Sullivant 2015)

The matrix $\mathcal{A}_\mathcal{C}$ is unimodular if and only if $\mathcal{C}$ is nuclear.

Simplicial Complex Minors

Definition (Deletion)

Let $\mathcal{C}$ be a simplicial complex on $[n]$. Let $v \in [n]$ be a vertex of $\mathcal{C}$. Then $\mathcal{C} \setminus v$ denotes the induced simplicial complex on $[n] \setminus \{v\}$.

Definition (Link)

Let $\mathcal{C}$ be a simplicial complex on $[n]$. Let $v \in [n]$ be a vertex of $\mathcal{C}$. Then $\text{link}_v(\mathcal{C})$ denotes the simplicial complex on $[n] \setminus \{v\}$ with facets

$$\text{facet}(\text{link}_v(\mathcal{C})) = \{F \setminus \{v\} : F \text{ is a facet of } \mathcal{C} \text{ with } v \in F\}.$$

Definition (Simplicial Complex Minor)

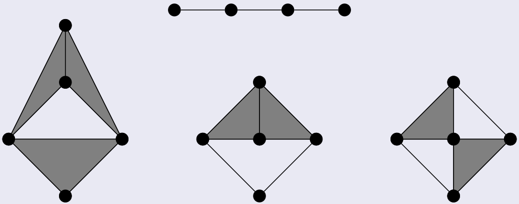
Let $\mathcal{C}, \mathcal{D}$ be simplicial complexes. If $\mathcal{D}$ can be obtained from $\mathcal{C}$ by taking links of vertices and deleting vertices, then we say that $\mathcal{D}$ is a *minor* of $\mathcal{C}$.

Unimodularity: Excluded Minor Classification

Theorem (B-Sullivant 2015)

The matrix $\mathcal{A}_{\mathcal{C}}$ is unimodular if and only if $\mathcal{C}$ has no simplicial complex minors isomorphic to any of the following

- $\partial\Delta_k \sqcup \{v\}$, *the disjoint union of the boundary of a simplex and an isolated vertex*
- O_6, *the boundary complex of an octahedron, or its Alexander dual O_6^**
- *The four simplicial complexes shown below*



Sketch of Proof

- $\mathcal{C}$ nuclear $\implies$ $\mathcal{C}$ unimodular
  - Simplices are unimodular
  - A disjoint union of two simplices is unimodular
  - Adding cone and ghost vertices and taking duals preserves unimodularity
- $\mathcal{C}$ unimodular $\implies$ $\mathcal{C}$ avoids forbidden minors
  - The forbidden minors are not unimodular
  - Taking minors preserves unimodularity
- $\mathcal{C}$ avoids forbidden minors $\implies$ $\mathcal{C}$ nuclear
  - If $\mathcal{C}$ avoids the forbidden minors but has a 4-cycle, then it must be an iterated cone over the 4-cycle. This is nuclear.
  - So focus on 4-cycle-free complexes. Then the 1-skeleton is either a complete graph, or two complete graphs glued along a clique.
  - Complex induction argument based on the link of a vertex of $\mathcal{C}$.

Next Steps - Unimodularity

Question

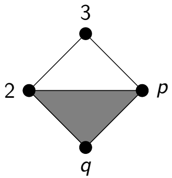
Given a simplicial complex $\mathcal{C}$ on $[n]$ and an integer vector $\mathbf{d} = (d_1, \ldots, d_n)$ with $d_i \geq 2$, is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular?

Corollary (B-Sullivant 2015)

If $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is unimodular, then $\mathcal{C}$ is nuclear.

Question

Let $\mathcal{C}$ and $\mathbf{d}$ be specified by the figure below. For which values of p and q is $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ unimodular?



Known Classification Results - Normality

Theorem (Sullivant 2010)

If $\mathcal{C}$ is a graph, then $\mathcal{A}_{\mathcal{C},\mathbf{2}}$ is normal if and only if $\mathcal{C}$ is free of K_4-minors.

Theorem (Bruns, Hemmecke, Hibi, Ichim, Ohsugi, Köppe, Söger 2007-2011)

Let $\mathcal{C}$ be a complex whose facets are all $m-1$ element subsets of $[m]$. Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is normal in precisely the following situations up to symmetry:

1. *At most two of the d_v are greater than two*
2. $m = 3$ *and* $\mathbf{d} = (3, 3, a)$ *for any* $a \in \mathbb{N}$
3. $m = 3$ *and* $\mathbf{d} = (3, 4, 4), (3, 4, 5)$ *or* $(3, 5, 5)$.

Theorem (Rauh-Sullivant 2014)

Let $\mathcal{C}$ be the four-cycle graph. Then $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is normal if $\mathbf{d} = (2, a, 2, b)$ or $\mathbf{d} = (2, a, 3, b)$ with $a, b, \in \mathbb{N}$.

Corollary of Unimodular Classification

Definition

Let $\mathcal{C}$ be a simplicial complex on $[n]$. We say a facet of $\mathcal{C}$ that has $n-1$ vertices is called a **big facet**.

Proposition

If $\mathcal{C}$ is a complex with a big facet, then $\mathcal{C}$ is normal if and only if unimodular.

So our classification result on unimodular $\mathcal{C}$ immediately gives a classification of the normal $\mathcal{C}$ when $\mathcal{C}$ has a big facet.

Normality Preserving Operations

Theorem (Sullivant 2010)

Normality of $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is preserved under the following operations on the simplicial complex

1. *Deleting vertices*
2. *Contracting edges*
3. *Gluing two simplicial complexes along a common face*
4. *Adding or removing a cone or ghost vertex.*

Theorem (B-Sullivant 2015)

Normality of $\mathcal{A}_{\mathcal{C},\mathbf{d}}$ is preserved when taking links of vertices of $\mathcal{C}$.

Minimally Non-Normal Simplicial Complexes

Question

Which simplicial complexes are minimally non-normal with respect to the operations of deleting vertices, contracting edges, gluing two complexes along a facet, removing cone and ghost vertices, and taking links of vertices?

Computational method:

- All simplicial complexes on 3 or fewer vertices are normal
- Choose two normal simplicial complexes $\mathcal{C}, \mathcal{D}$ on $n-1$ vertices. Create simplicial complex $\mathcal{C}'$ on n vertices by attaching a new vertex v to $\mathcal{C}$ such that $\text{link}_v(\mathcal{C}') = \mathcal{D}$
- See if (non)normality of $\mathcal{C}'$ can be certified by reducing to a smaller complex via our normality-preserving operations
- If not, check normality of $\mathcal{C}'$ using Normaliz. If non-normal, then minimally non-normal

Minimally Non-Normal Simplicial Complexes

We were able to use the computational method to determine normality of all but 6 of the complexes on up to 6 vertices.

So far, we know that the set of minimally non-normal simplicial complexes consists of:

- 20 sporadic complexes, obtained by computational method
- Two infinite families, obtained by theoretical means

References

Daniel Irving Bernstein and Seth Sullivant.
Unimodular Binary Hierarchical Models.
ArXiv:1502.06131, 2015.

Daniel Irving Bernstein and Seth Sullivant.
Normal Binary Hierarchical Models
ArXiv:1508.05461, 2015

Winfried Bruns, Raymond Hemmecke, Bogdan Ichimc, Matthias Kpped and Christof Sgera.
Challenging Computations of Hilbert Bases of Cones Associated with Algebraic Statistics
Exp. Math. 20, 25 - 33 (2011)

Takayuki HIbi and Hidefumi Ohsugi
Toric Ideals Arising from Contingency Tables
Commutative Algebra and Combinatorics
Ramanujan Mathematical Society Lecture Notes Series, Number 4, Ramanujan Math. Soc., Mysore, 2007, pp. 91-115.

Johannes Rauh and Seth Sullivant
Lifting markov bases and higher codimension toric fiber products
ArXiv:1404.6392 2014

Bernd Sturmfels.
Gröbner Bases and Convex Polytopes, volume 8 of *University Lecture Series*.
American Mathematical Society, Providence, RI, 1996.

Seth Sullivant.
Normal binary graph models.
Ann. Inst. Statist. Math., 62(4):717–726, 2010.