The Geometry of the Birkhoff Polytope

Ludovick Bouthat

Université Laval

CMS Winter Meeting; December 2023

Acknowledgement

This research is a collaborate effort with Pr. Javad Mashreghi and Pr. Frédéric Morneau-Guérin.

It was done with the financial help of the FRQNT, NSERC, and the Vanier Scholarship.

Doubly stochastic matrices

Definition

A square matrix is *doubly stochastic* if :

- nonnegative coefficients ;
- row sums $= 1$;
- column sums $= 1$.

The set of $n \times n$ doubly stochastic matrices is denoted by Ω_n.

Example

$$D = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

Important cases

Example

The *uniform matrix* and the *identity matrix*

$$J_n = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad \& \quad I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Important properties (Algebraic structure)

- Ω_n is a <u>*semigroup*</u> : $D_1 D_2 \in \Omega_n$ if $D_1, D_2 \in \Omega_n$.

Important properties (Algebraic structure)

- Ω_n is a *semigroup* : $D_1 D_2 \in \Omega_n$ if $D_1, D_2 \in \Omega_n$.
- Ω_n is a *monoid* : $DI_n = I_n D = D$ for every $D \in I_n$.

Important properties (Algebraic structure)

- Ω_n is a *semigroup* : $D_1 D_2 \in \Omega_n$ if $D_1, D_2 \in \Omega_n$.
- Ω_n is a *monoid* : $DI_n = I_n D = D$ for every $D \in I_n$.
- Ω_n has an *absorbing element* : $DJ_n = J_n D = J_n$ for every $D \in \Omega_n$.

Important properties (Geometric structure)

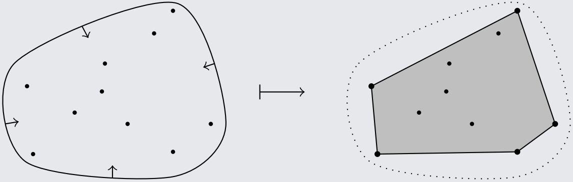
- Ω_n is a *convex polytope*.

Important properties (Geometric structure)

- Ω_n is a *convex polytope*.

Definition

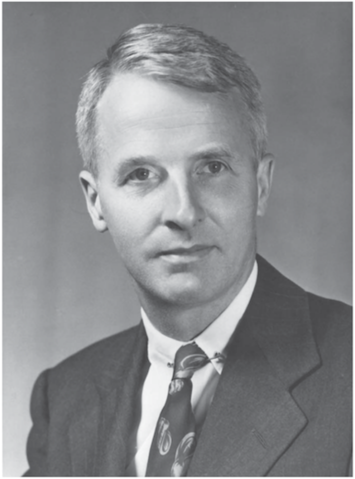
A *convex polytope* is a convex hull of a finite and nonempty set of points in $\mathbb{R}^n$.



Birkhoff's theorem

Theorem (Birkhoff ; 1946)

The set Ω_n of $n \times n$ doubly stochastic matrices is the convex hull of the $n \times n$ permutation matrices. Furthermore, the permutation matrices are precisely the extreme points of Ω_n.



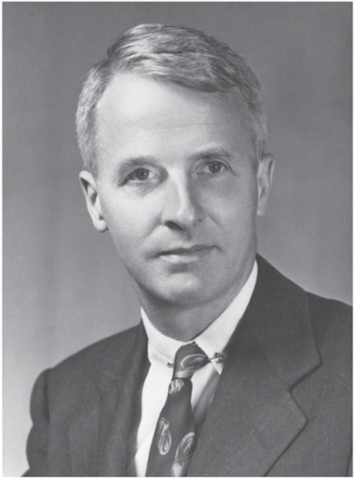
Garrett Birkhoff

Birkhoff's theorem

Theorem (Birkhoff ; 1946)

The set Ω_n of $n \times n$ doubly stochastic matrices is the convex hull of the $n \times n$ permutation matrices. Furthermore, the permutation matrices are precisely the extreme points of Ω_n.

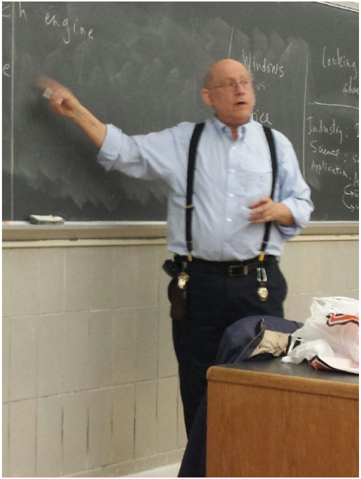
- $D = \sum_i \alpha_i P_i$, where P_i are permutation matrices and $\alpha_i > 0$, $\sum_i \alpha_i = 1$.



Garrett Birkhoff

Past results (Balinski and Russakoff; 1974)

Michel Balinski



Andrew Russakoff

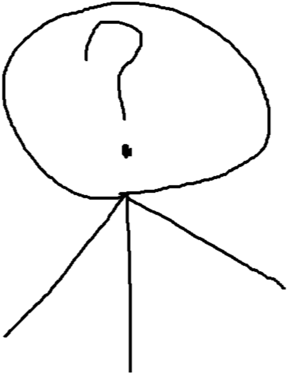
Past results (Balinski and Russakoff ; 1974)

Theorem (Balinski and Russakoff ; 1974)

1. *The graph* $G(\Omega_n)$ *is vertex-symmetric ;*
2. *The degree of each vertex of* $G(\Omega_n)$ *is equal to* $N(n) = \sum_{k=2}^{n} \binom{n}{k}(k-1)!$ *;*
3. *The number of edges of* Ω_n *is equal to* $\frac{n!}{2} N(n)$.

Past results (Brualdi and Gibson ; 1975–1977)

Richard A. Brualdi



Peter M. Gibson

Past results (Brualdi and Gibson ; 1975–1977)

Theorem (Brualdi and Gibson ; 1975–1977)

Let B be a $(0,1)$-matrices of order n and denote by $\mathscr{F}(B)$ the set of all $D \in \Omega_n$ satisfying $d_{ij} \leq b_{ij}$ for all $1 \leq i, j \leq n$. The faces of Ω_n are given by $\mathscr{F}(B)$. Moreover, the $n \times n$ permutation matrices P satisfying $P \leq B$ are the vertices of the face $\mathscr{F}(B)$.

Theorem (Brualdi and Gibson ; 1975–1977)

If $n > 2$, then Ω_n has n^2 facets. If $n = 2$, then Ω_2 only has 2 facets corresponding to the vertices of the polytope.

The discrete volume of Ω_n

- Values of the volume of Ω_n were given for
  - $n \leq 7$ by Sturmfels (1997) ;
  - $n = 8$ by Chan and Robbins (1999) ;
  - $n = 9, 10$ by Beck and Pixton (2003).

The discrete volume of Ω_n

n	$\mathrm{Vol}(\Omega_n)$
1	1
2	2
3	$\frac{9}{8}$
4	$\frac{176}{2835}$
5	$\frac{23590375}{167382319104}$
6	$\frac{9700106723}{1319281996032000000}$
7	$\frac{77436678274508929033}{137302963682235238399868928000000000}$
8	$\frac{5562533838576105333259507434329}{125890362600954779500814809426933398033089280000000000}$
9	$\frac{559498129702796022246895686372766052475496691}{2153302766311808894781211017508325066061406891577233480945238016000000000000000}$
10	$\frac{727291284016786420977508457990121862548823260052557333386607889}{828160860106766855125676318796872729344622463533089422677980721388055739956270293750883504892820848640000000}$

Volume of the Birkhoff polytope for $n = 1, 2, \ldots, 10$

An asymptotic formula

Theorem (Canfield and Mckay ; 2009)

For any $\varepsilon > 0$,

$$\mathrm{Vol}(\Omega_n) = \frac{1}{(2\pi)^{n-\frac{1}{2}} n^{(n-1)^2}} \exp\left(n^2 + \frac{1}{3} + O\left(n^{-\frac{1}{2}+\varepsilon}\right)\right)$$

as $n \to \infty$.

My supervisors and I

Javad Mashreghi

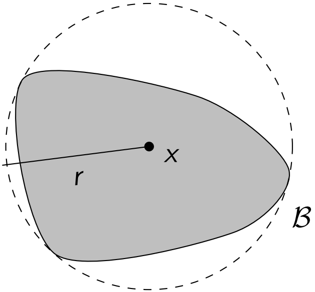
Frédéric Morneau-Guérin

My cat and I

The Smallest Enclosing Ball Problem

Definition

Given a metric space $(\mathcal{U}, d)$ and a nonempty closed, bounded set $\mathcal{B} \subseteq \mathcal{U}$, a *minimal bounding ball* of $\mathcal{B}$ relative to the metric space $(\mathcal{U}, d)$ is a ball $B(x, r) \subseteq \mathcal{U}$ containing the set $\mathcal{B}$ and such that r is smallest possible.



The Chebyshev radius and centers

Definition

If a minimal bounding ball $B(x, r)$ of $\mathcal{B}$ relative to the metric space $(\mathcal{U}, d)$ exists, then

1. r is called the *Chebyshev radius* of $\mathcal{B}$ relative to $(\mathcal{U}, d)$ and is noted $R_d(\mathcal{B})$;
2. x is a *Chebyshev center* of $\mathcal{B}$ relative $(\mathcal{U}, d)$.

Observe that we have

$$R_d(\mathcal{B}) = \inf_{x\in\mathcal{U}} \sup_{y\in\mathcal{B}} d(x, y).$$

Permutation invariant norms

Definition

A matrix norm $\|\cdot\|$ is *permutation-invariant* if

$$\|PAQ\| = \|A\|, \qquad \forall A \in M_n,$$

for every permutation matrix P and Q.

Example

1. The *entrywise p-norms* $\|A\|_p := \left(\sum_{i,j=1}^n |a_{ij}|^p\right)^{\frac{1}{p}}$;
2. The *operator p-norms* $\|A\|_{\ell^p \to \ell^p} := \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$;
3. The *Schatten p-norms* $\|A\|_{S_p} := \left(\sum_{i=1}^n |\sigma_i(A)|^p\right)^{\frac{1}{p}}$.

Properties of the Chebyshev centers of Ω_n

Proposition (B., Mashreghi, Morneau-Guérin ; 2023)

Let $\mathcal{U} \subseteq M_n(\mathbb{R})$ be a convex permutation-invariant set and let $\|\cdot\|$ be a permutation-invariant norm. If $A_1, A_2 \in \mathcal{U}$ are Chebyshev centers of Ω_n relative to the metric space $(\mathcal{U}, \|\cdot\|)$, then

1. *PA_iQ $(i = 1, 2)$ is a Chebyshev center of Ω_n for any permutation matrices P and Q ;*
2. *Any convex combination of A_1 and A_2 is a Chebyshev center of Ω_n ;*
3. *$\|A_i - P\| = R_{\|\cdot\|}(\Omega_n)$ $(i = 1, 2)$ for any permutation matrix P.*

Main theorem

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

Let $\mathcal{U} \subseteq M_n(\mathbb{R})$ be a convex permutation-invariant set and let $\|\cdot\|$ be a permutation-invariant norm. If there exist a Chebyshev center A of Ω_n relative to the metric space $(\mathcal{U}, \|\cdot\|)$, then the matrix $J_nAJ_n = \left(\frac{1}{n}\sum_{i,j=1}^{n} a_{ij}\right)J_n$ is also a Chebyshev center of Ω_n relative to the same metric space. Moreover, the Chebyshev radius of Ω_n in this setting is given by

$$R_{\|\cdot\|}(\Omega_n) = \|J_nAJ_n - I_n\| = \inf_{\substack{\alpha\in\mathbb{R} \\ \alpha J_n \in \mathcal{U}}} \|\alpha J_n - I_n\|$$

and the infimum is attained by $\alpha = \frac{1}{n}\sum_{i,j=1}^{n} a_{ij}$.

A corollary

Corollary (B., Mashreghi, Morneau-Guérin ; 2023)

If $\|\cdot\|$ is a permutation-invariant norm, then the uniform matrix J_n is a Chebyshev center of Ω_n relative to $(\Omega_n, \|\cdot\|)$. Moreover, the associated Chebyshev radius is given by $R_{\|\cdot\|}(\Omega_n) = \|J_n - I_n\|$.

About unicity

About unicity

. . .

The Schatten p-norms

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

For $1 \leq p \leq \infty$, the uniform matrix J_n is the unique Chebyshev center of Ω_n relative to the metric space $(M_n(\mathbb{R}), \|\cdot\|_{S_p})$, and the associated Chebyshev radius is equal to $(n-1)^{1/p}$.

The operator p-norms

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

For $1 \leq p \leq \infty$, the uniform matrix J_n is the unique Chebyshev center of Ω_n relative to the metric space $(\Omega_n, \|\cdot\|_{\ell^p \to \ell^p})$. Moreover, for $p = 1, \infty$, the Chebyshev radius of Ω_n is equal to $2\left(1 - \frac{1}{n}\right)$ and for $p = 2$, it is equal to 1.

The operator p-norms

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

For $1 \leq p \leq \infty$, the uniform matrix J_n is the unique Chebyshev center of Ω_n relative to the metric space $(\Omega_n, \|\cdot\|_{\ell^p \to \ell^p})$. Moreover, for $p = 1, \infty$, the Chebyshev radius of Ω_n is equal to $2\left(1 - \frac{1}{n}\right)$ and for $p = 2$, it is equal to 1.

- For a general $p \in (1, \infty)$, the value of the Chebyshev radius $R_p(\Omega_n) = \|I_n - J_n\|_{\ell^p \to \ell^p}$ is unknown.

A conjecture

Conjecture (B., Mashreghi, Morneau-Guérin ; 2023)

For $1 < p < \infty$ *and* $p \neq 2$, *define* $\rho := p - 1$. *Let* x_p *be the unique root of the function*

$$x \longmapsto \rho\left(1 + x^{\frac{1}{\rho}}\right)\left(1 - x^{\rho-1}\right) + \left(1 + x^{\rho}\right)\left(1 - x^{\frac{1}{\rho}-1}\right)$$

in the closed interval $[0,1]$. *If* $m_1 := \left\lfloor \frac{n}{x_p+1} \right\rfloor$ *and* $m_2 := \left\lceil \frac{n}{x_p+1} \right\rceil$, *then*

$$R_p(\Omega_n) = \max_{m \in \{m_1, m_2\}} \frac{\left(\left(\frac{n}{m} - 1\right)^{p-1} + 1\right)^{\frac{1}{p}} \left(\left(\frac{n}{m} - 1\right)^{\frac{1}{p-1}} + 1\right)^{1-\frac{1}{p}}}{\frac{n}{m}}.$$

A problem in stochastic processes

- In some cases, the transition matrices in a Markov chain are not only stochastic, but doubly stochastic.

A problem in stochastic processes

- In some cases, the transition matrices in a Markov chain are not only stochastic, but doubly stochastic.

- Let $(D_k) \subseteq \Omega_n$ be such that D_k is the transition matrix of a doubly stochastic Markov chain at step k.

A problem in stochastic processes

• In some cases, the transition matrices in a Markov chain are not only stochastic, but doubly stochastic.

• Let $(D_k) \subseteq \Omega_n$ be such that D_k is the transition matrix of a doubly stochastic Markov chain at step k.

Question : What is the long-term behavior of a Markov chain formed from doubly stochastic matrices, i.e., what can we say about $D_1 D_2 D_3 \cdots$?

A problem in stochastic processes

• In some cases, the transition matrices in a Markov chain are not only stochastic, but doubly stochastic.

• Let $(D_k) \subseteq \Omega_n$ be such that D_k is the transition matrix of a doubly stochastic Markov chain at step k.

Question : What is the long-term behavior of a Markov chain formed from doubly stochastic matrices, i.e., what can we say about $D_1 D_2 D_3 \cdots$?

Proposition (B., Mashreghi, Morneau-Guérin ; 2023)

For almost all doubly stochastic matrices $D \in \Omega_n$, $D^k \to J_n$ as $k \to \infty$.

A sufficient result

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

Let $D_1, D_2, \dots$ be a sequence of $n \times n$ doubly stochastic matrices and let $\sigma_2(A)$ be the second largest singular value of A. If $\sum_{k=1}^{\infty}(1 - \sigma_2(D_k)) = \infty$, then $\lim_{m\to\infty} D_1 D_2 \cdots D_m = J_n$.

A sufficient result

Theorem (B., Mashreghi, Morneau-Guérin ; 2023)

Let $D_1, D_2, \ldots$ *be a sequence of* $n \times n$ *doubly stochastic matrices and let* $\sigma_2(A)$ *be the second largest singular value of* A*. If* $\sum_{k=1}^{\infty}(1-\sigma_2(D_k)) = \infty$*, then* $\lim_{m\to\infty} D_1 D_2 \cdots D_m = J_n$*.*

Theorem (Schwarz ; 1980)

Let $D_1, D_2, \ldots$ *be a sequence of* $n \times n$ *doubly stochastic matrices and let* $\nu(A) := \min_{i,j} a_{ij}$*. If* $\sum_{k=1}^{\infty} \nu(D_k) = \infty$*, then* $\lim_{m\to\infty} D_1 D_2 \cdots D_m = J_n$*.*

Improving a result of Schwarz

- $n\nu(D) \leq 1 - \sigma_2(D)$ for any $D \in \Omega_n$.

Improving a result of Schwarz

- $n\nu(D) \leq 1 - \sigma_2(D)$ for any $D \in \Omega_n$.

$$\sum_{k=1}^{\infty} \nu(A_k) = \infty \quad \Longrightarrow \quad \sum_{k=1}^{\infty} (1 - \sigma_2(A_k)) = \infty.$$

Improving a result of Schwarz

- $n\nu(D) \leq 1 - \sigma_2(D)$ for any $D \in \Omega_n$.

$$\sum_{k=1}^{\infty} \nu(A_k) = \infty \quad \Longrightarrow \quad \sum_{k=1}^{\infty} (1 - \sigma_2(A_k)) = \infty.$$

Example

Consider the case $A_k = A$ for each k, where

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

Then $A \in \Omega_n$, $\nu(A) = 0$, and the singular values of A are $1, 1/2$ and 0 so that $1 - \sigma_2(A) = 1/2$.

References

- Ludovick Bouthat, Javad Mashreghi, and Frédéric Morneau-Guérin. On the norm of normal matrices. *RIMS Kôkyûroku Bessatsu*, 93 :183–222, 2023.
- Ludovick Bouthat, Javad Mashreghi, Frédéric Morneau-Guérin. On the Geometry of the Birkhoff Polytope. I. The operator ℓ_n^p-norms, *Linear Algebra Appl.*, Submitted, 2023.
- Ludovick Bouthat, Javad Mashreghi, Frédéric Morneau-Guérin. On the Geometry of the Birkhoff Polytope. II. The Schatten p-norms, *Linear Algebra Appl.*, Submitted, 2023.
- Ludovick Bouthat, Nicolas Doyon, Javad Mashreghi, Frédéric Morneau-Guérin. On the convergence of doubly stochastic Markov Chains, Preprint, 2023.