Convergence and Orchestrated Divergence of Polygons

Eric Hintikka, Siddharth Patel, and Rachel Robinson

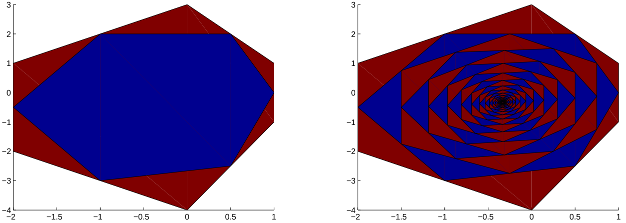
Missouri State University Math REU

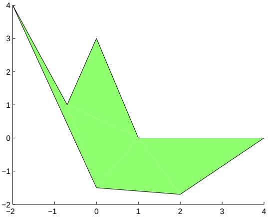
August 1, 2013

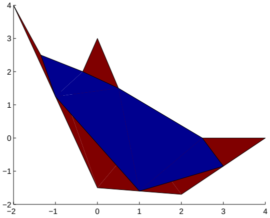
Overview

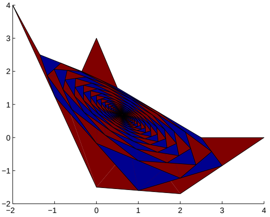
M. Rosenman's Midpoint Problem

Let Π be a closed polygon in the plane with vertices $z_0, z_1, \ldots, z_{n-1}$. Denote by $z_0^{(1)}, z_1^{(1)}, \ldots, z_{n-1}^{(1)}$ the midpoints of the sides $z_0 z_1, z_1 z_2, \ldots, z_{n-1} z_0$, respectively. Using $z_0^{(1)}, z_1^{(1)}, \ldots, z_{n-1}^{(1)}$ as vertices, we derive a new polygon, denoted by $\Pi^{(1)}$. Apply the same procedure to derive the polygon $\Pi^{(2)}$. After k constructions, we obtain polygon $\Pi^{(k)}$. Show that $\Pi^{(k)}$ converges, as $k \to \infty$, to the centroid of the original points $z_0, z_1, \ldots, z_{n-1}$.









$$z_0^{(1)} = \frac{1}{2}(z_0 + z_1)$$

$$z_1^{(1)} = \frac{1}{2}(z_1 + z_2)$$

$$\vdots$$

$$z_{n-1}^{(1)} = \frac{1}{2}(z_{n-1} + z_0)$$

Matrix Expression

The vertices of polygon Π can be represented as the column vector $\Pi = (z_0, z_1, z_2, \cdots, z_{n-1})^T$. If we define A to be a circulant $n \times n$ (row) stochastic matrix with first row $(\frac{1}{2}\ \frac{1}{2}\ 0\ \cdots\ 0)$, we have $\Pi^{(1)} = Az$.

$$\begin{bmatrix} z_0^{(1)} \\ z_1^{(1)} \\ z_2^{(1)} \\ \vdots \\ z_{n-1}^{(1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 \\ 0 & 0 & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & 0 & 0 & \cdots & \frac{1}{2} \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(z_0 + z_1) \\ \frac{1}{2}(z_1 + z_2) \\ \frac{1}{2}(z_2 + z_3) \\ \vdots \\ \frac{1}{2}(z_{n-1} + z_0) \end{bmatrix}$$

Note: We begin our matrix indexing at 0, so the top-left entry of a matrix A is denoted by $(A)_{00}$.

Some Solutions

- use complex coordinates
- each polygon transformation interpreted as a weighted average of vertices
- Huston's geometric solution *(1933)*
- Schoenberg's solution: Fourier analysis *(1950)*
- Charles' solution: Markov chains *(2012)*

Deterministic Problem

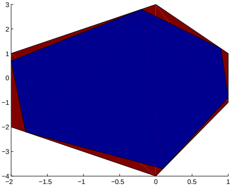
Let Π be a closed polygon in the plane with vertices $z_0, z_1, ..., z_{n-1}$, and let $0 < \delta \leq 1/2$ be a given constant. Select $z_{i-1}^{(1)}$ on the edge $z_{i-1}z_i$ such that

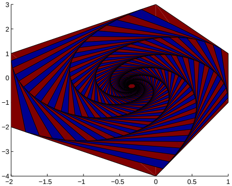
$$\min(\text{dist}(z_{i-1}^{(1)}, z_{i-1}), \text{dist}(z_{i-1}^{(1)}, z_i)) \geq \delta \, \text{dist}(z_{i-1}, z_i), \quad i = 0, 1, \ldots, n-1.$$

In this fashion, we derive a new polygon $\Pi^{(1)} = z_0^{(1)}, z_1^{(1)}, ..., z_{n-1}^{(1)}$. Apply the same procedure to derive the polygon $\Pi^{(2)}$. After k constructions, we obtain polygon $\Pi^{(k)}$. Show that $\Pi^{(k)}$ converges to a point as $k \to \infty$.

If we choose exactly $z_i^{(k)} = \delta z_i^{(k-1)} + (1-\delta) z_{i+1}^{(k-1)}$ at every iteration for every $i = 0, 1, \ldots, n-1$, we get:

$$\begin{bmatrix} z_0^{(1)} \\ z_1^{(1)} \\ z_2^{(1)} \\ \vdots \\ z_{n-1}^{(1)} \end{bmatrix} = \begin{bmatrix} \delta & 1-\delta & 0 & \cdots & 0 \\ 0 & \delta & 1-\delta & \cdots & 0 \\ 0 & 0 & \delta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-\delta & 0 & 0 & \cdots & \delta \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} \delta z_0 + (1-\delta) z_1 \\ \delta z_1 + (1-\delta) z_2 \\ \delta z_2 + (1-\delta) z_3 \\ \vdots \\ \delta z_{n-1} + (1-\delta) z_0 \end{bmatrix}$$





Some Definitions

- A matrix Q is *ergodic* (or *primitive*) if $Q^s > 0$ for some $s \in \mathbb{N}$.
- A matrix A is *(row) stochastic* if $\sum_{j=1}^{n} (A)_{ij} = 1$ for each $i \in \{1, \ldots, n\}$.
- A matrix A is *circulant* if it is of the form

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ & \vdots & & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}$$

- A *root of unity* is any complex number that gives 1 when raised to some integer power k. The *n-th roots of unity* are given by $\omega_\nu = e^{2\pi i \frac{\nu}{n}}$ for $\nu = 0, 1, \ldots, n-1$.

Schoenberg's Solution

- Fourier analysis: a bridge

$$\begin{array}{ccc} B & \xleftarrow{\mathcal{F}^{-1}} & F(B) \\ & & \uparrow \\ A & \xrightarrow[\mathcal{F}]{} & F(A) \end{array}$$

- takes an input (usually a time-based function) and decomposes it into its frequencies
- solves a generalization of the midpoint problem
- establishes an exponential convergence rate

Finite (Discrete) Fourier transform

Let $z_\nu = \zeta_0 + \zeta_1\omega_\nu + \zeta_2\omega_\nu^2 + \cdots + \zeta_{n-1}\omega_\nu^{n-1}$ for $\nu = 0, 1, \ldots, n-1$ with $\omega_\nu = e^{2\pi i\nu/n}$ (the nth roots of unity). We call this representation of z_ν the finite Fourier (f.F.) expansion of the sequence $z_0, z_1, \ldots, z_\nu$ and we call $\zeta_0, \zeta_1, \ldots, \zeta_{n-1}$ the f.F. coefficients of the sequence (z_ν). This too has a matrix expression:

$$
\begin{aligned}
[z_\nu]^T &= F[\zeta_\nu]^T \\
\begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{n-1} \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_1 & \omega_1^2 & \cdots & \omega_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & \omega_{n-1} & \omega_{n-1}^2 & \cdots & \omega_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} \zeta_0 \\ \zeta_1 \\ \vdots \\ \zeta_{n-1} \end{bmatrix} \\
&= \begin{bmatrix} \zeta_0 + \zeta_1 + \zeta_2 + \cdots + \zeta_{n-1} \\ \zeta_0 + \zeta_1\omega_1 + \zeta_2\omega_1^2 + \cdots + \zeta_{n-1}\omega_1^{n-1} \\ \vdots \\ \zeta_0 + \zeta_1\omega_\nu + \zeta_2\omega_\nu^2 + \cdots + \zeta_{n-1}\omega_\nu^{n-1} \end{bmatrix}.
\end{aligned}
$$

In fact, this approach is a generalization of the midpoint problem: choose any stochastic circulant matrix A and let the vertices of our n-gon be the vector $\Pi = (z_0, \cdots, z_{n-1})^T \in \mathbb{C}^n$. Let

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_0 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_1 & \cdots & \omega_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_{n-1} & \cdots & \omega_{n-1}^{n-1} \end{bmatrix}.$$

If our first-iteration polygon $\Pi^{(1)} = (z_0^{(1)}, \dots, z_{n-1}^{(1)})^T$ is given by $\Pi^{(1)} = A\Pi$, then we have $F\Pi^{(1)} = FA\Pi$, a column vector where each entry $(FA\Pi)_\nu$ is equal to $f(\omega_\nu)$.

Theorem (1, Schoenberg)

Now if we subject (z_ν) *to the cyclic transformation*

$$\begin{aligned} z_0' &= a_0 z_0 + a_1 z_1 + \cdots + a_{n-1} z_{n-1} \\ z_1' &= a_{n-1} z_0 + a_0 z_1 + \cdots + a_{n-2} z_{n-1} \\ &\vdots \\ z_{n-1}' &= a_1 z_0 + a_2 z_1 + \cdots + a_0 z_{n-1} \end{aligned}$$

then the f.F. coefficients of the sequence (z_ν') *are* $\zeta_\nu' = \zeta_\nu f(\omega_\nu)$ *where* $f(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1}$. *We call* $f(z)$ *the representative polynomial of this cyclic transformation.*

This representative polynomial has other nice applications: for a circulant matrix A, we have $f(\omega_\nu) = \lambda_\nu$ where λ_ν are the eigenvalues of A and the associated eigenvectors are $x_\nu = (1, \omega_\nu, \omega_\nu^2, \ldots, \omega_\nu^{n-1})^T$ for $\nu = 0, 1, \ldots, n-1$. For example, take the midpoint matrix $A = circ[\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0]$ for a 6-gon:

ν	$f(\omega_\nu) = \lambda_\nu$	x_ν
0	1	$(1,1,1,1,1,1)^T$
1	$\frac{1}{2} + \frac{1}{2}e^{2\pi i \frac{1}{6}}$	$(1, \omega_1, \omega_1^2, \omega_1^3, \omega_1^4, \omega_1^5)^T$
2	$\frac{1}{2} + \frac{1}{2}e^{2\pi i \frac{2}{6}}$	$(1, \omega_2, \omega_2^2, \omega_2^3, \omega_2^4, \omega_2^5)^T$
3	0	$(1, \omega_3, \omega_3^2, \omega_3^3, \omega_3^4, \omega_3^5)^T$
4	$\frac{1}{2} + \frac{1}{2}e^{2\pi i \frac{4}{6}}$	$(1, \omega_4, \omega_4^2, \omega_4^3, \omega_4^4, \omega_4^5)^T$
5	$\frac{1}{2} + \frac{1}{2}e^{2\pi i \frac{5}{6}}$	$(1, \omega_5, \omega_5^2, \omega_5^3, \omega_5^4, \omega_5^5)^T$

Room for Generalizations

While elegant, Schoenberg's technique is only applicable to a very particular problem, where the sequence of polygons $(\Pi^{(i)})_{i\geq 1}$ is generated by repeatedly applying a single circulant matrix A to an initial polygon $\Pi^{(0)}$, so that $\Pi^{(k)} = A^k \Pi^{(0)}$.

- What if we do not force A to be circulant?
- What if we allow A to vary with each iteration? That is, what if we define a sequence of matrices $(A_i)_{i\geq 1}$ and say instead that $\Pi^{(k)} = A_k \Pi^{(0)}$?

These questions motivate three generalizations.

The First Generalization

We follow the first line of inquiry, and allow A to be chosen from a slightly broader class of matrices, so that

$$A = \begin{bmatrix} \alpha_1 & 1-\alpha_1 & 0 & \cdots & & 0 \\ 0 & \alpha_2 & 1-\alpha_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_n & 0 & \cdots & & 0 & \alpha_n \end{bmatrix},$$

where $0 < \alpha_i < 1$. Since in general $\alpha_i \neq \alpha_j$, A is no longer circulant, but note that it is still stochastic.

We still let our sequence of polygons $(\Pi^{(i)})_{i\geq 1}$ be given by

$$\Pi^{(k)} = A^k \Pi^{(0)}.$$

Note that the question of whether $\Pi^{(k)}$ converges to a polygon with identical vertices for any choice of $\Pi^{(0)}$ is equivalent to the question of whether A^k converges to a rank one matrix.

This is because the product of stochastic matrices is itself stochastic, since matrices A, B are stochastic if and only if $Ae = e$ and $Be = e$, where $e = (1, \ldots, 1)^T$, so if A and B are stochastic, then $ABe = Ae = e$, and thus AB is stochastic.

What is more, if a matrix A is rank one, then each of its rows must be a scalar multiple of its first row. If A is stochastic, then each of those scalars must be 1, since otherwise the matrix would have some row sums not equal to 1. Thus, a stochastic rank one matrix has all rows equal.

So, if $\lim_{k\to\infty} A^k = L$ for some rank one matrix L, then we have $\lim_{k\to\infty} \Pi^{(k)} = L\Pi^{(0)} = \Pi$, for some polygon Π with all components equal. That is, the sequence $(\Pi^i)_{i\geq 1}$ converges to a single point, in the sense that there exists $p \in \mathbb{C}$ such that $\lim_{k\to\infty} z_\nu^{(k)} = p$ for all ν, where $z_\nu^{(k)}$ is the ν-th vertex of $\Pi^{(k)}$.

Thus, it seems reasonable to approach our problem from a purely matrix-analytical point of view, rather than a geometrical one.

One useful result in matrix analysis is called Perron's Theorem, after the German mathematician Oskar Perron. Though the complete formulation is somewhat longer, we excerpt the relevant portion from the text *Matrix Analysis* by Roger A. Horn and Charles R. Johnson.

Theorem (Perron)

If A is an $n \times n$ matrix over $\mathbb{C}$ and $(A)_{ij} > 0$ for all i, j, then

$$[\rho(A)^{-1}A]^m \to L \text{ as } m \to \infty,$$

where $\rho(A) = \max\{|\lambda| : \lambda$ is an eigenvalue of $A\}$, $L \equiv xy^T$, $Ax = \rho(A)x$, $A^T y = \rho(A)y$, $x > 0$, $y > 0$, and $x^T y = 1$.

In the case where A is stochastic, it is well known that $\rho(A) = 1$. This follows immediately from Lemma 8.1.21 in Horn and Johnson's book, which states that if A is an $n \times n$ complex-valued matrix with all entries non-negative and if the row sums of A are constant, then $\rho(A) = \|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |(A)_{ij}|$.

So, in this case, the result of Perron's theorem takes the nice form $A^m \to L$ as $m \to \infty$. However, one of the assumptions of the theorem is that the matrix A has strictly positive entries. Certainly, if as before we have

$$A = \begin{bmatrix} \alpha_1 & 1-\alpha_1 & 0 & \cdots & & 0 \\ 0 & \alpha_2 & 1-\alpha_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_n & 0 & \cdots & & 0 & \alpha_n \end{bmatrix},$$

then this is not the case.

However, we note that A^{n-1} *does* have strictly positive entries. This can be proven via mathematical induction, where the key observation is that, for any N, the $(i, (i + N \mod n) + 1)$-th entry of A^{N+1} is greater than or equal to $1 - \alpha_i$ times the $((i \mod n) + 1, (i + N \mod n) + 1)$-th entry of A^N. For the sake of time, we suppress the details of the proof.

Thus, we apply Perron's theorem to $B = A^{n-1}$.

Let $y = \begin{pmatrix} \frac{1}{1-\alpha_1} \\ \frac{1}{1-\alpha_2} \\ \vdots \\ \frac{1}{1-\alpha_n} \end{pmatrix}$, and let $x = \begin{pmatrix} \frac{1}{\sum_{i=1}^n \frac{1}{1-\alpha_i}} \\ \vdots \\ \frac{1}{\sum_{i=1}^n \frac{1}{1-\alpha_i}} \end{pmatrix}$.

Then it is just a matter of calculation to verify that $A^T y = y$, and it follows easily that $B^T y = y$. It is also not difficult to see that $Bx = x$, since B is stochastic and x has equal components. Finally, a quick check verifies that $x^T y = 1$, and clearly $x, y > 0$.

So, Perron's theorem guarantees that $\lim_{k\to\infty} B^k = L$, where

$$L = xy^T = \begin{bmatrix} \frac{1}{(1-\alpha_1)\sum_{i=1}^{n}\frac{1}{1-\alpha_i}} & \cdots & \frac{1}{(1-\alpha_n)\sum_{i=1}^{n}\frac{1}{1-\alpha_i}} \\ \vdots & & \vdots \\ \frac{1}{(1-\alpha_1)\sum_{i=1}^{n}\frac{1}{1-\alpha_i}} & \cdots & \frac{1}{(1-\alpha_n)\sum_{i=1}^{n}\frac{1}{1-\alpha_i}} \end{bmatrix}$$

But, we're really interested in $\lim_{k\to\infty} A^k$, not $\lim_{k\to\infty} B^k$. Making this transition is not as easy as it may seem. However, since A^i is stochastic for any $i \in \mathbb{N}$ and L is a rank one matrix, it can be shown that $L = A^i L$ for any i. We can use this fact to our advantage!

For any $i \in \{0, \ldots, n-2\}$, we can write

$$L = A^i L = A^i \lim_{k\to\infty} A^{k(n-1)} = \lim_{k\to\infty} A^{k(n-1)+i}$$

Now, choose $\epsilon > 0$, and for each $i \in \{0, \ldots, n-2\}$ let N_i be such that $\|A^{m(n-1)+i} - L\| < \epsilon$ for all $m \geq N_i$. Let $N = \max\{N_i\}$, and suppose that $j \geq N(n-1) + (n-2)$. Then by the division theorem, we can write $j = m(n-1) + i$ for some integer m and some $i \in \{0, \ldots, n-2\}$. So, $j = m(n-1) + i \geq N(n-1) + (n-2) \geq N(n-1) + i \geq N_i(n-1) + i$. So, $m \geq N_i$. Now $\|A^j - L\| < \epsilon$, and hence $\lim_{k\to\infty} A^k = L$.

This is the result we wanted, and it allows us to immediately reach the following theorem, which concludes our first generalization of the polygon problem.

Theorem

Let $\Pi^{(0)} = (z_1^{(0)}, z_2^{(0)}, \ldots, z_n^{(0)})^T$ be an n-gon in the complex plane. Choose $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $0 < \alpha_i < 1$, and write

$$A = \begin{bmatrix} \alpha_1 & 1-\alpha_1 & 0 & \cdots & & 0 \\ 0 & \alpha_2 & 1-\alpha_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_n & 0 & \cdots & & 0 & \alpha_n \end{bmatrix}.$$

Let $\Pi^{(1)} = A\Pi^{(0)}$ be a new polygon inscribed in $\Pi^{(0)}$. Repeat this process to obtain $\Pi^{(2)}$ inscribed in $\Pi^{(1)}$, etc., so that in general $\Pi^{(k)} = A^k\Pi^{(0)}$. Then $\lim_{k\to\infty} \Pi^{(k)} = P$, where $P = \sum_{j=1}^{n}((1-\alpha_j)\sum_{i=1}^{n}\frac{1}{1-\alpha_i})^{-1} z_j^{(0)} \cdot (1, \ldots, 1)^T$. Note that all components of P are identical.

The Second Generalization

We now consider the case in which the matrix used to derive descendant polygons varies with each iteration. Speaking roughly, this gives us more freedom in choosing our sequence of polygons. More precisely, the situation we are now interested can be stated as follows.

Let $\Pi^{(0)}$ be an n-gon as before, and choose $\delta \in (0, 1/2)$. For each natural number i, choose $\alpha_0^{(i)}, \alpha_1^{(i)}, \dots, \alpha_n^{(i)}$ from the open interval $(\delta, 1-\delta)$, let

$$A_k = \begin{bmatrix} \alpha_0^{(k)} & 1-\alpha_0^{(k)} & 0 & \cdots & & 0 \\ 0 & \alpha_1^{(k)} & 1-\alpha_1^{(k)} & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1-\alpha_{n-1}^{(k)} & 0 & \cdots & & 0 & \alpha_{n-1}^{(k)} \end{bmatrix},$$

let $\mathcal{A}_k = A_k A_{k-1} \cdots A_1$, and define $\Pi^{(k)} = \mathcal{A}_k \Pi^{(0)}$ for $k \geq 1$.

As before, we wish to determine whether the resulting sequence of polygons $(\Pi^{(k)})_{k\geq 1}$ necessarily converges to a point.

Unfortunately, our primary tool from the previous section, Perron's theorem, is no longer applicable. However, we utilize *coefficients of ergodicity* to help handle this more general case. The relevant definition, taken from a paper by Ipsen and Selee, is as follows.

The 1-norm ergodicity coefficient $\tau_1(S)$ for an $n \times n$ stochastic matrix S is given by

$$\tau_1(S) = \max_{\substack{\|z\|_1=1 \\ z^T e=0}} \|S^T z\|_1$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^n$ and the maximum ranges over $z \in \mathbb{R}^n$. If $n = 1$, we say $\tau_1(S) = 0$.

Equivalently, we can write

$$\tau_1(S) = \frac{1}{2} \max_{ij} \sum_{k=1}^{n} |(S)_{ik} - (S)_{jk}|.$$

This is the expression that we will be using, because it makes calculations easier. (For a proof that the two expressions are identical, see Ipsen and Selee's paper.)

These coefficients will prove useful to us for two reasons: they can distinguish rank 1 matrices, and they are submultiplicative. That is, for any $n \times n$ stochastic matrices S and T, we have

- $\tau_1(S) = 0 \iff \text{rank}(S) = 1$
- $\tau_1(ST) \leq \tau_1(S)\tau_1(T)$

However, before we continue, we make a few preliminary observations.

First, if $\mathcal{A}_k = A_k A_{k-1} \dots A_1$, where each matrix A_i is as described at the beginning of this section, then

$$\mathcal{A}_{n-1} > \delta^{n-1}$$

The proof is very similar to our earlier proof that $A^{n-1} > 0$, with just a slight modification. Again, for the sake of time we suppress the details.

Second, we observe that if S is a positive $n \times n$ stochastic matrix and $\epsilon > 0$ is such that $(S)_{ij} > \epsilon$ for all $i, j \in \{0, 1, \ldots, n-1\}$, then

$$\tau_1(S) \leq 1 - n\epsilon.$$

The proof is slightly more involved, but ultimately, it boils down to a few observations:

- $\tau_1(S) = \tau_1(S - \epsilon)$
- The row sums of $S - \epsilon$ are each $1 - n\epsilon$
- Any non-negative matrix with identical row sums s has coefficient of ergodicity at most s.

Now, having paused to make these two claims, we present our main argument. Suppose we have chosen $(A_i)_{i \geq 0}$ as described at the beginning of this section. Define $\mathcal{B}_i = A_{in-1}A_{in-2}\cdots A_{(i-1)n}$ for $i \geq 1$. Then by the previous two claims, we have $\tau_1(\mathcal{B}_i) \leq 1 - n\delta^{n-1}$ for each i. Recall that τ_1 is submultiplicative and, for stochastic matrices, bounded above by 1. So if we choose i and let $m = \max\{j : jn - 1 \leq i\}$, then $\tau_1(\mathcal{A}_i) = \tau_1(A_iA_{i-1}\cdots A_{mn}\mathcal{B}_m\mathcal{B}_{m-1}\cdots\mathcal{B}_1) \leq \tau_1(\mathcal{B}_m\mathcal{B}_{m-1}\cdots\mathcal{B}_1) \leq \tau_1(\mathcal{B}_m)\tau_1(\mathcal{B}_{m-1})\cdots\tau_1(\mathcal{B}_1) \leq (1 - n\delta^{n-1})^m$. Of course, we then have $\lim_{i\to\infty}\tau_1(\mathcal{A}_i) \leq \lim_{i\to\infty}(1 - n\delta^{n-1})^m = 0$, and actually equality holds, since τ_1 can take only non-negative values.

However, a priori this is not enough to tell us that $\lim_{i\to\infty} \mathcal{A}_i$ even exists, much less that it is rank 1. To show that this is in fact the case, we use a Cauchy sequence argument, for which we require two more ingredients. Roughly speaking, these are (1) that for any i, j and sufficiently large m, the distance between $(\mathcal{A}_m)_{ij}$ and $(\mathcal{A}_m)_{1j}$ is small; and (2) that for all $k \geq 0$, the distance between $(\mathcal{A}_{m+k})_{ij}$ and $(\mathcal{A}_m)_{1j}$ is small. With some tinkering, these facts both follow quickly.

Using the triangle inequality to combine the two, we find that

$$|(\mathcal{A}_{m+k})_{ij} - (\mathcal{A}_m)_{i'j}|$$

can be made arbitrarily small for sufficiently large m.

This establishes that if we write $x^{(j)}_{kn+l} = (\mathcal{A}_{k+1})_{lj}$ for any j, then the sequence $(x^{(j)}_i)_{i\geq 0}$ is Cauchy, and hence converges to some real number p_j. We know that each subsequence must also converge to p_j. So, each of the n subsequences created by choosing the first term to be $x^{(j)}_i$ for some $i \in \{0, 1, \ldots, n-1\}$ and choosing after that only every n-th term from the original sequence must also converge to p_j. That is, $\lim_{k\to\infty}(\mathcal{A}_k)_{ij} = p_j$ for each i. Hence $\mathcal{A}_k$ converges to the rank 1 matrix

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_{n-1} \\ p_0 & p_1 & \cdots & p_{n-1} \\ \vdots & & \vdots & \\ p_0 & p_1 & \cdots & p_{n-1} \end{bmatrix}.$$

Thus, we arrive at the second main result of this presentation, which is stated as follows.

Theorem

If $\mathcal{A}_i$ is as described at the beginning of this section, then

$$\lim_{i\to\infty} \mathcal{A}_i = L,$$

where L is a rank 1 stochastic matrix. Hence if $\Pi^{(i)}$ is the corresponding sequence of polygons, we have

$$\lim_{i\to\infty} \Pi^{(i)} = L\Pi^{(0)},$$

and thus $(\Pi^{(i)})_{i\geq 0}$ converges to a point.

Unfortunately, we see no clear way to describe L explicitly, since Perron's theorem – the tool that allowed us to make an analogous description in the first generalization – is no longer applicable.

The Third Generalization

Up to this point, we have been concerned with sequences of polygons in which the i-th vertex of the $(k+1)$-th polygon lies along the open line segment between the i-th and $(i+1)$-th vertices of the k-th polygon. But what if we loosen this requirement? Here, we consider polygons derived from sequences of stochastic matrices of a more general class: those that we call "circulant-patterned," where we say a matrix A is circulant-patterned if, for some circulant matrix B, we have $(A)_{ij} = 0 \iff (B)_{ij} = 0$. By using this type of matrix, we may allow the vertices of a new polygon to be chosen from anywhere within the convex hull of the old polygon.

Since circulant-patterned matrices are in some sense similar to circulant matrices, it seems reasonable that knowledge about when products of circulant matrices converge could be useful in determining when circulant-patterned matrices converge. Indeed this turns out to be the case, and luckily Tollisen and Lengyel have already proven such a theorem. We present the relevant portion of the theorem on the next slide.

Theorem (Tollisen and Lengyel)

For any stochastic circulant matrix A, we have

$$(A^k)_{ij} \approx \begin{cases} \frac{gcd(n,g)}{n}, & \text{if } j - i \equiv ku \mod \gcd(n,g) \\ 0, & \text{otherwise} \end{cases}$$

as $k \to \infty$, *where* $u = \min\{i : (A)_{1i} > 0\}$ *and* $g = \gcd\{i - u : (A)_{1i} > 0\}$.

In addition to this useful theorem, we present two quick observations that together will allow us to derive a nice result.

But first, a quick note on terminology: we say two matrices A and B *share a zero pattern* if $(A)_{ij} = 0 \iff (B)_{ij} = 0$.

Now, the first observation is: if $(A_i)_{i\geq 0}$ and $(B_j)_{j\geq 0}$ are two sequences of nonnegative $n \times n$ matrices such that A_i and B_i have the same zero pattern for each i, then $A_k A_{k-1} \cdots A_0$ and $B_k B_{k-1} \cdots B_0$ share a zero pattern for any k. The proof is by induction.

The second observation is: choose any $k \in \mathbb{N}$. For each $i \in \{0, 1, \ldots, k-1\}$, let A_i be a nonnegative $n \times n$ matrix, and let $\mathcal{A}_{i+1} = A_i A_{i-1} \cdots A_0$. Suppose that $\epsilon > 0$ is such that there do not exist l, m, i for which $0 < (A_i)_{lm} < \epsilon$. Then there do not exist i, j for which $0 < (\mathcal{A}_k)_{ij} < \epsilon^k$. Though this proof is slightly more complicated it is still a fairly simple induction argument and we omit it for time.

Now, we have all the tools needed to prove our penultimate proposition, which we present on the following slide.

Let A be a stochastic circulant matrix such that $\lim_{k\to\infty} A^k = L$ for some rank 1 matrix L. Let $(A_i)_{i\geq 0}$ be a sequence of stochastic matrices, each with the same zero pattern as A. Suppose further that for some $\epsilon > 0$, there do not exist i, j, k for which $0 < (A_k)_{ij} < \epsilon$. Then $\lim_{k\to\infty} A_k A_{k-1} \cdots A_1 A_0 = L'$ for some rank 1 matrix L'.

We now present an outline of the proof. We know that the product of stochastic matrices is stochastic, and it can be shown that the product of circulant matrices is circulant. Hence A^k a circulant stochastic matrix for all k, and since we have assumed that L exists, it follows that it must be circulant stochastic as well.

From Tollisen and Lengyel's theorem, we know that each entry of L must be either 0 or $\frac{\gcd(n,g)}{n}$, where g is a constant. We can show that if $(L)_{1j} = 0$ for any j, then $\operatorname{rank}(A) \geq 2$, a contradiction. So, L must be positive, with all entries equal to $\frac{1}{n}$.

Now, since we have established that $L > 0$, there must be some k such that $A^k > 0$. Define $\mathcal{B}_i = A_{ik-1}A_{ik-2}\cdots A_{(i-1)k}$ for $i \geq 1$. Then by our first observation, each$\mathcal{B}_i$ has the same zero pattern as A^k. That is, $\mathcal{B}_i > 0$ for all i. What is more, by our second observation we know that $\mathcal{B}_i > \epsilon^k$.

From here, we can use an argument that is identical to the one used during our second generalization to show that $\lim_{i\to\infty} \tau_1(A_i A_{i-1} \cdots A_0) = 0$, and then the result follows from the same Cauchy sequence argument as before. Hence, our proof is complete.

The theorem we just proved is particularly useful to us because Tollisen and Lengyel's theorem gives us a simple way to determine whether the powers of a particular stochastic circulant matrix converge to a rank 1 matrix.

Indeed, we can quickly show from their theorem that if A is a stochastic circulant matrix, then A^k converges to a rank one matrix as $k \to \infty$ if and only if $\gcd(n, g) = 1$, where g is defined as before.

So, we can restate our theorem in the following, more convenient, form.

Theorem

Let $(A_i)_{i\geq 0}$ *be a sequence of stochastic, circulant-patterned,* $n \times n$ *matrices that all have a common zero pattern. Suppose that* $\gcd(n, g) = 1$*, where* $(a_0, a_1, \ldots, a_{n-1})$ *is the first row of* A_0*,* $u = \min\{i | A_i > 0\}$*, and* $g = \gcd\{i - u | a_i > 0\}$*. Suppose also that for some* $\epsilon > 0$ *there do not exist* i, j, k *for which* $0 < (A_k)_{ij} < \epsilon$*. Then*

$$\lim_{k\to\infty} A_k A_{k-1} \cdots A_0 = L$$

for some rank 1 *matrix* L*. Hence, if* $\Pi^{(k)} = A_k A_{k-1} \cdots A_1 \Pi^{(0)}$*, then*

$$\lim_{k\to\infty} \Pi^{(k)} = L\Pi^{(0)},$$

and thus $(\Pi^{(k)})_{k\geq 0}$ *converges to a point.*

Divergent Matrices

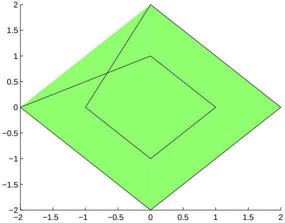
Let

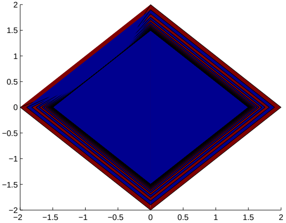
$$A = \begin{bmatrix} a_0 & 0 & 0 & \cdots & 0 & (1-a_0) & 0 & \cdots & 0 \\ 0 & a_1 & 0 & \cdots & 0 & 0 & (1-a_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & (1-a_{n-1}) & 0 & 0 & \cdots & a_{n-1} \end{bmatrix}$$

with $0 < a_i < 1$ for all $i = 0, 1, \ldots, n-1$. Note that $(A)_{ii} = a_i$ and $(A)_{ij} = 1 - a_i$ when $j = (i + g \bmod n)$ for some fixed $g \in \mathbb{Z}$ such that $1 \leq g < n$ for each $i \in \{0, 1, \ldots, n-1\}$.

When $\gcd(n, g) = 1$ we have convergence to a rank-1 matrix. When $\gcd(n, g) = m$ for some $m \in \mathbb{Z}$ such that $1 < m \leq g$, we have convergence to a rank-m matrix.

$$A = \begin{bmatrix} .1 & 0 & 0 & 0 & .9 & 0 & 0 & 0 \\ 0 & .1 & 0 & 0 & 0 & .9 & 0 & 0 \\ 0 & 0 & .1 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & .1 & 0 & 0 & 0 & .9 \\ .9 & 0 & 0 & 0 & .1 & 0 & 0 & 0 \\ 0 & .9 & 0 & 0 & 0 & .1 & 0 & 0 \\ 0 & 0 & .9 & 0 & 0 & 0 & .1 & 0 \\ 0 & 0 & 0 & .9 & 0 & 0 & 0 & .1 \end{bmatrix} \quad \text{with } g = 4, n = 8, \gcd(n, g) = 4$$





Theorem (7, Tollisen & Lengyel)

Let $A = circ[a_0, a_1, \ldots, a_{n-1}]$ *be a circulant stochastic matrix with* $L = \{i | a_i > 0\}, u = \min L = \min\{i | a_i > 0\}, L' = \{i - u | a_i > 0\}$, *and* $g = \gcd(L')$. *Partition the* n *positions around the circle into* $\gcd(n, g)$ *subsets:* $S_j = \{s : s \equiv j \bmod \gcd(n, g), 0 \leq s < n\}$, $j = 0, 1, \ldots, \gcd(n, g) - 1$, *and define the range of each* $\underline{x}_k = A^k \underline{x}_0$ *when restricted to the subset* S_j *to be*

$$R_j^{(k)} = \max\{b_i^{(k)} : i \in S_j\} - \min\{b_i^{(k)} : i \in S_j\}.$$

Then, for each j, $R_j^{(k)} \to 0$ *as* $k \to \infty$.

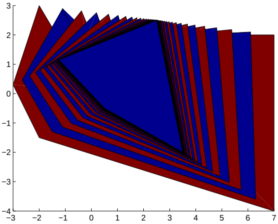
Examples

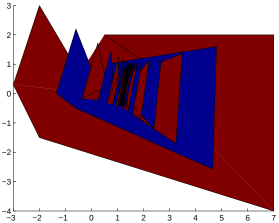
$A = circ[a_0, 0, 0, 0, a_4, 0, 0, 0]$

- $L = \{0, 4\}$, $u = 0$, $L' = \{0, 4\}$
- $g = \gcd(0, 4) = 4$
- $S_0 = \{0, 4\}$, $S_1 = \{1, 5\}$, $S_2 = \{2, 6\}$, $S_3 = \{3, 7\}$

$B = circ[0, a_1, 0, a_3, 0, a_5, 0, a_7]$

- $L = \{1, 3, 5, 7\}$, $u = 1$, $L' = \{0, 2, 4, 6\}$
- $g = \gcd(0, 2, 4, 6) = 2$
- $S_0 = \{0, 2, 4, 6\}$, $S_1 = \{1, 3, 5, 7\}$





Theorem (8, Tollisen & Lengyel)

For any circulant stochastic matrix A and any initial configuration $\underline{x}_0$, let u and g be defined as above. Then, the Markov chain with transition matrix A consists of $\gcd(n, g, u)$ recurrent classes, each with period $p = \frac{\gcd(n,g)}{\gcd(n,g,u)}$. In other words, the n positions around the circle can be partitioned into $\gcd(n, g, u)$ rotationally symmetric subsets where, on each subset either the coordinates of $\underline{x}_k$ converge (if $p = 1$) or asymptotically cycle through the values with (possibly non-fundamental) period p.

$A_1 = circ[a_0, 0, 0, 0, a_4, 0, 0, 0]$

- $L = \{0, 4\},\ u = 0,\ L' = \{0, 4\}$
- $g = \gcd(0, 4) = 4$
- $p = \gcd(n, g)/\gcd(n, g, u) = 4/4 = 1$ (static divergence)

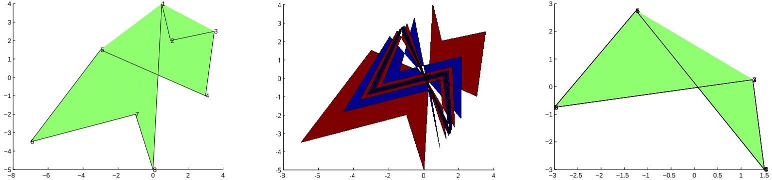
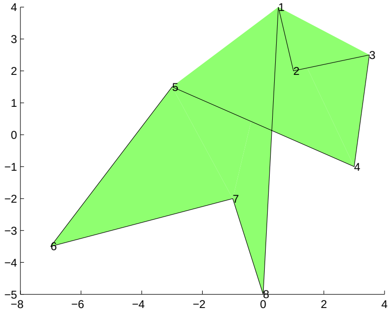


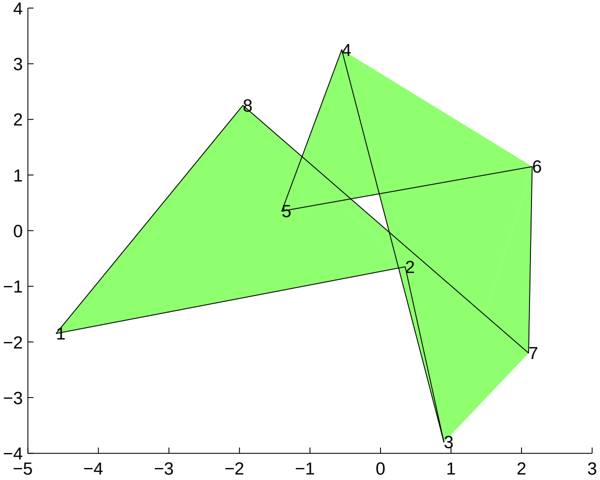
Figure : An 8-gon, its first 100 iterations, and its 100th iteration alone

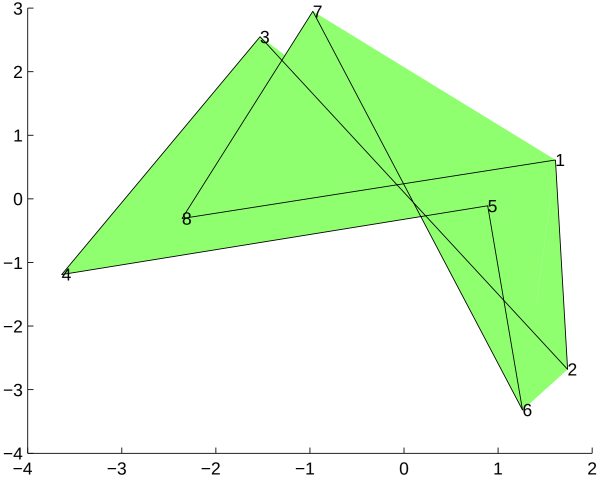
$A_2 = circ[0, a_1, 0, 0, 0, a_5, 0, 0]$

- $L = \{1, 5\}$, $u = 1$, $L' = \{0, 4\}$
- $g = \gcd(0, 4) = 4$
- $p = 4/1 = 4$ (rotating divergence)

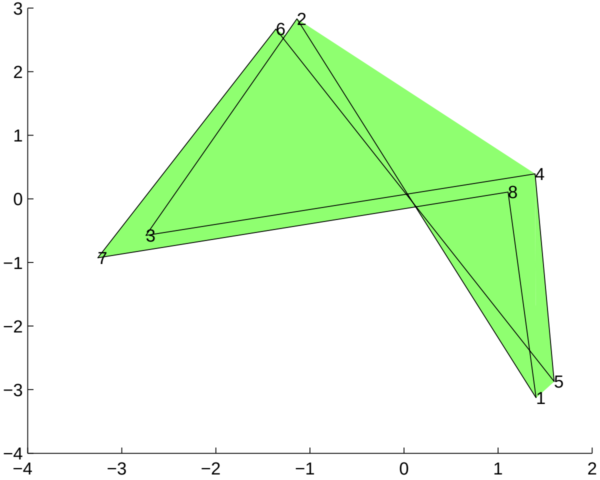


Divergent Matrices

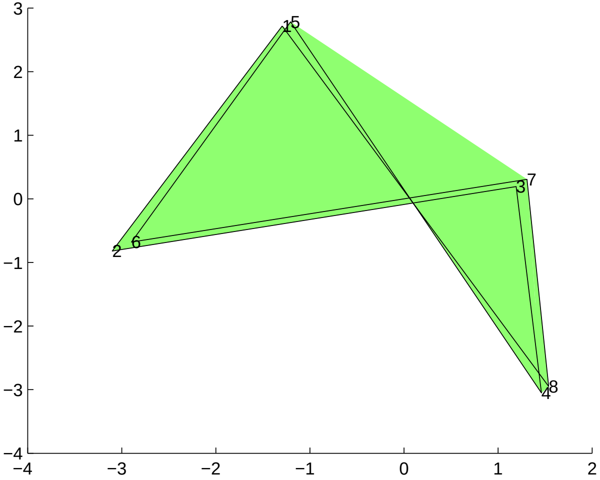



1
2
3
4
5
6
7
8
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2

Divergent Matrices

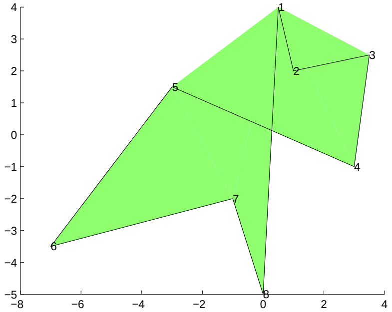


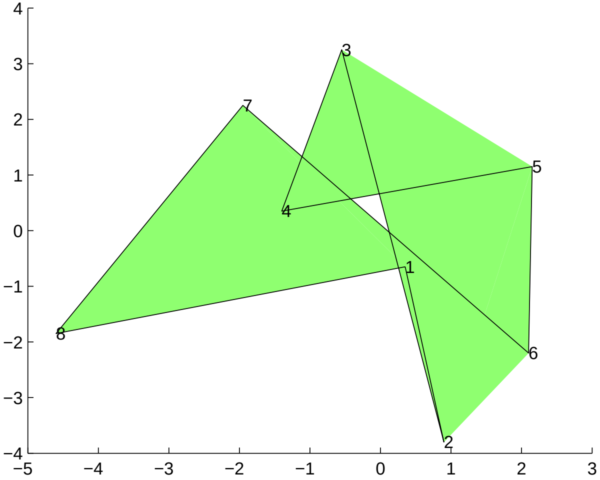
Divergent Matrices



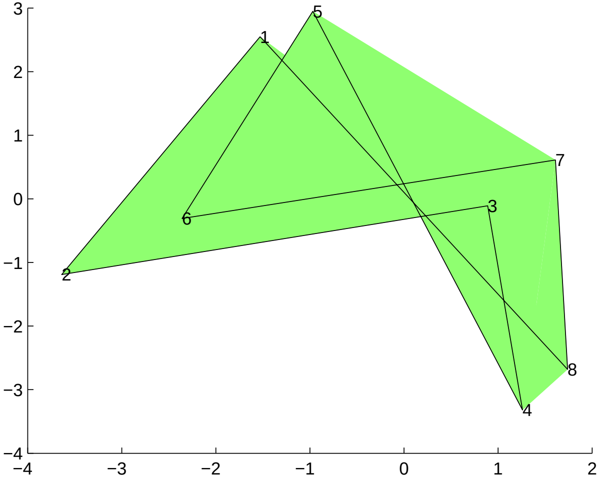
$A_3 = circ[0, 0, a_2, 0, 0, 0, a_6, 0]$

- $L = \{2, 6\}$, $u = 2$, $L' = \{0, 4\}$
- $g = \gcd(0, 4) = 4$
- $p = 4/2 = 2$ (switching divergence)

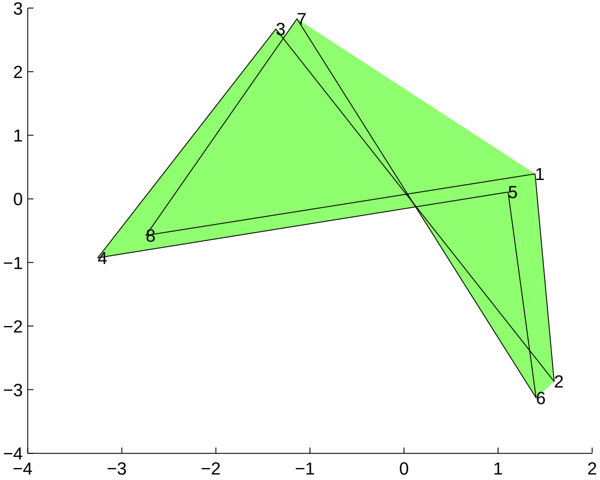



4
3
2
1
0
−1
−2
−3
−4
−5
−4
−3
−2
−1
0
1
2
3
1
2
3
4
5
6
7
8

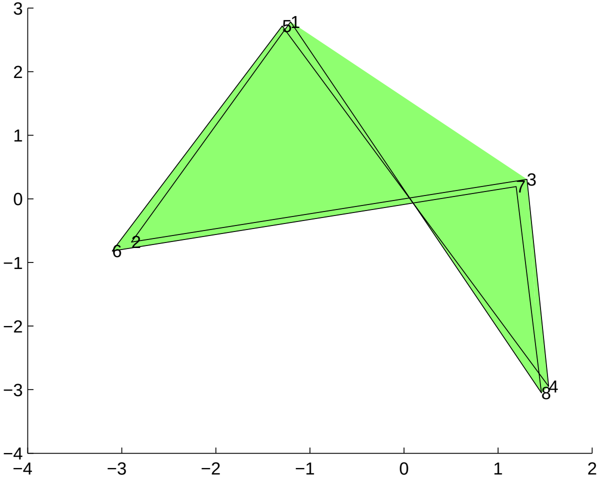
Divergent Matrices



Divergent Matrices



Divergent Matrices



Representative Polynomial Revisited

If for a matrix A defined as before with $g = 4$ and $n = 8$, we have:

Static Divergence
$(a_0, 0, 0, 0, a_4, 0, 0, 0)$

ν	$f(\omega_\nu)$
0	1
1	$a_0 - a_4$
2	1
3	$a_0 - a_4$
4	1
5	$a_0 - a_4$
6	1
7	$a_0 - a_4$

Rotating Divergence
$(0, a_1, 0, 0, 0, a_5, 0, 0)$

ν	$f(\omega_\nu)$
0	1
1	$a_1 e^{2\pi i \frac{1}{8}} + a_5 e^{2\pi i \frac{5}{8}}$
2	i
3	$a_1 e^{2\pi i \frac{3}{8}} + a_5 e^{2\pi i \frac{15}{8}}$
4	-1
5	$a_1 e^{2\pi i \frac{5}{8}} + a_5 e^{2\pi i \frac{25}{8}}$
6	-i
7	$a_1 e^{2\pi i \frac{7}{8}} + a_5 e^{2\pi i \frac{35}{8}}$

Switching Divergence
$(0, 0, a_2, 0, 0, 0, a_6, 0)$

ν	$f(\omega_\nu)$
0	1
1	$i(a_2 - a_6)$
2	-1
3	$i(a_6 - a_2)$
4	1
5	$i(a_2 - a_6)$
6	-1
7	$i(a_6 - a_2)$

Considering eigenvalues again, this is a good indicator of what our polygons are doing as we take more and more iterations.

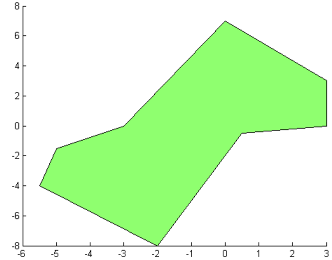
Further Investigation I: Block Matrices

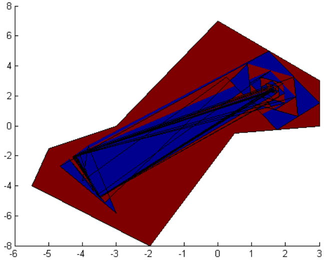
Suppose we have an action matrix A with r blocks, for example

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

This creates r independent systems that we can look at individually.

Using the matrix on the previous slide and the 8-gon below, we get 2 independent systems converging to 3 points.





Question: Can we make these systems interact?

Possible Solution: We introduce a circulant stochastic matrix B so we have $\Pi^{(k)} = A^{k-r-1} B A^r \Pi$ for some $r \in \mathbb{N}$ with $0 \leq r < k-1$. For example,

$$B = \begin{bmatrix} .4 & 0 & 0 & 0 & .6 & 0 & 0 & 0 \\ 0 & .4 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & .4 & 0 & 0 & 0 & .6 & 0 \\ 0 & 0 & 0 & .4 & 0 & 0 & 0 & .6 \\ .6 & 0 & 0 & 0 & .4 & 0 & 0 & 0 \\ 0 & .6 & 0 & 0 & 0 & .4 & 0 & 0 \\ 0 & 0 & .6 & 0 & 0 & 0 & .4 & 0 \\ 0 & 0 & 0 & .6 & 0 & 0 & 0 & .4 \end{bmatrix}.$$

Hopefully, this might break vertices out of their independent systems.

What is the goal in introducing B?

- *Change of shape of limiting polygon?*
  Unless B is the identity matrix or a matrix containing the same blocks, the application of B changes the limiting polygon in some way or another – typically, it shrinks.
- *Vertex rotation?*
  It is difficult to see the effect of B on intermediate polygons in the sequence, since the polygons following the application of B tend to be drastically different with each different B. However, if we want to rotate the vertices of the limiting polygon, we can indeed change B accordingly. And the more we apply B, the more rotations we get.

Goal?, cont'd

- *Re-partition of the vertices?*
  We have not yet found a matrix B that will do this. In practice, the subsets of vertices that converge together S_j stay together, even with more applications of B.

- *Convergence?*
  If we want the polygon to converge to a single point (i.e. $\text{rank}(\Pi^{(k)}) = 1$) after only one application of B, then B must be rank-1, since $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$ and $\lim_{k\to\infty} A^k \geq r$ and $\lim_{j\to\infty} B^j = \gcd(n, g)$. If we want convergence after several iterations, then B must be ergodic.

  Question: Exactly how many applications of B would we need for this kind of convergence?

Further Investigation II: Mixing Time

We know that a Markov chain with transition matrix P will have a unique stationary distribution π and that after a time $t_{\text{mix}}(\varepsilon)$ it will be "close enough" to π. We know, given a circulant, stochastic action matrix A, that $\lim_{k\to\infty} A^k = Q$ where Q is a matrix of rank $\gcd(n, g)$ (with period $p = \gcd(n, g, u)$). If A is ergodic then Q is a rank-1 matrix with all rows equal to $[\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}]$.

Question: At which iteration k will A^k be "close enough" to its stationary distribution Q?

Define

$$d(t) := \max_{x \in \Omega} \|P(x, \cdot)^t - \pi\|_{TV}$$

where Ω is our state space, our stationary distribution is π, and for two probability distributions μ and ν and an event A in Ω, *total variation distance* is defined as

$$\|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

We now define *mixing time*, denoted by $t_{\text{mix}}(\varepsilon)$, as

$$t_{\text{mix}}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}.$$

When we set our set of vertices Π as our state space Ω, our action matrix A as our probability matrix $P(x, \cdot)$, and our limiting matrix Q of rank $\gcd(n, g)$ as our stationary distribution π, what is our mixing time?

Some Approaches

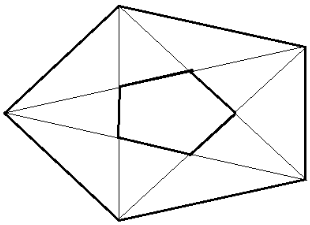
- begin with a basic example, the midpoint problem matrix $A = circ[\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0]$ with stationary distribution

$$Q = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \\ & \vdots & & \ddots & \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{bmatrix}$$

  and then branch out to other matrices
- explore a related topic: The parameter $S = \sum_{\nu=0}^{n-1} |z_\nu^{(k)} - c|^2$ measures collective distance from centroid $c = \frac{1}{n}\sum_{\nu=0}^{n-1} z_\nu^{(k)}$. If A is ergodic and $\lim_{k\to\infty} A^k = Q$ then $\Pi^{(k)} = (c, c, \ldots, c)^T$. Can we use this (or another parameter) to establish a rate of convergence for our polygon transformation sequence?

Further Investigation III: Schoenberg's Conjecture

Begin with a convex 5-gon with vertices $z_0, z_1, \ldots, z_4$. Connect z_0 to z_2, z_1 to $z_3, \ldots, z_4$ to z_1. Our new 5-gon is given by the area of these new edges. Show that the sequence of 5-gons converges to a point.



To our knowledge, this problem has not been solved... yet.

References

- Horn, Roger A. and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press (1999).
- Ipsen, Ilse C. F. and Teresa M. Selee, *Ergodicity Coefficients Defined by Vector Norms*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 153-200.
- Kra, Irwin and Santiago R. Simanca, *On Circulant Matrices*, Notices Amer. Math. Soc., 59 (2012), pp. 368-377.
- Ouyang, Charles, *A Problem Concerning the Dynamic Geometry of Polygons*, (2013).
- Schoenberg, I. J., *The Finite Fourier Series and Elementary Geometry*, Amer. Math. Monthly, 57 (1950), pp. 390-404.
- Tollisen, Gregory P. and Tamás Lengyel, *Intermediate and Limiting Behavior of Powers of Some Circulant Matrices*, Ars Combin., 88 (2008), pp. 229-255.

The End