Applications of numerical linear algebra for the study of complex networks and systems

A necessary and sufficient condition for consensus. Rate of convergence. Introduction to time varying averaging systems.

enrico.bozzo@uniud.it

Roma Tor Vergata, November 22-23-24 2022

Summary

- Some results on substochastic matrices are a good starting point for proving a necessary and sufficient condition for consensus.
- If A is row stochastic, the rate of convergence of $x(k+1) = Ax(k)$ is related to the essential spectral radius of A.
- If the matrices $A(k)$ are row stochastic then $x(k+1) = A(k)x(k)$ is defined time varying averaging system.
- Convergence and rate of convergence of time varying averaging systems are simpler to study when the graphs $D(A(k))$ have good connection properties.
- The Massey method has been proposed for ranking college football teams. A time aware variant of the method leads to a time varying averaging system.

Substochastic matrices

Definition

A $n \times n$ matrix $A \geq 0$ is row substochastic if $Ae \leq e$ and there exists an index $i \in \{1, \ldots, n\}$ such that $e_i^T Ae < 1$; is column substochastic if A^T is row substochastic.

For definiteness, we will treat the case of row substochastic matrices. Notice that $e_i^T Ae$ is the weighted out degree of the node i in the graph $G = D(A)$. Of course, if A is row substochastic and irreducible then

- if $\max Ae < 1$, since $\rho(A) \leq \max Ae$ the matrix is convergent;
- if $\max Ae = 1$, since $\min Ae < \max Ae$ we have $\rho(A) < \max Ae$ and the matrix is convergent as well.

A stronger result about convergent substochastic matrices

It follows that irreducibility is a sufficient condition for convergence for a substochastic matrix. The following result is stronger since it individuates a necessary and sufficient condition.

Theorem

If A is row substochastic then it is convergent if and only if there is a path from each other node of $G = D(A)$ to some of the nodes having out-degree less than one.

Preliminaries

Let A be row substochastic.

- If $Ae \leq e$ then $A^2e = AAe \leq Ae \leq e$, so that $A^ke \leq e$.
- If $e_i^T Ae < 1$ then $e_i^T A^2 e = e_i^T AAe \leq e_i^T Ae < 1$ so that $e_i^T A^k e < 1$.
- It follows that A^k is row substochastic for $k \geq 1$.
- Let us assume that $a_{i,j} > 0$ and $e_j^T Ae < 1$. Then

$$e_i^T A^2 e = e_i^T AAe = e_i^T A \sum_k e_k^T Ae\, e_k = \sum_k e_k^T Ae\, a_{i,k}.$$

But

$$\sum_k e_k^T Ae\, a_{i,k} \leq \sum_{k \neq j} a_{i,k} + e_j^T Ae\, a_{i,j} < e_i^T Ae.$$

It follows that $e_i^T A^2 e < 1$.

Preliminaries

What we observed leads us to the following statement.

Lemma

Let A be row substochastic and $G = D(A)$. If in G there is a simple path of length p between a node i and a node j such that $e_j^T Ae < 1$ then $e_i^T A^p e < 1$.

Proof The proof, that can be obtained by induction, is left as an exercise.

Convergent Substochastic matrices

Proof of the theorem If there is a path from each other node to a node having out-degree less than one, by using the lemma we obtain that there exists an integer m such that $\max A^m e < 1$. Hence $\rho(A^m) = \rho(A)^m < 1$ so that $\rho(A) < 1$. On the other hand, let A be convergent and let us assume by contradiction that there exists a set $S \neq \emptyset$ of nodes of G not connected to nodes having out-degree less than one. Hence we can determine a permutation matrix P such that separate the nodes of S from the remaining ones $PAP^T = \begin{bmatrix} A_{1,1} & O \\ A_{2,1} & A_{2,2} \end{bmatrix}$ where $A_{1,1}$ and $A_{2,2}$ are square and $A_{1,1}e = e$. This implies $1 = \rho(A_{1,1}) \leq \rho(A) \leq 1$ so that A cannot be convergent.

Example: Leslie population model

If $\sum_{k=1}^{n} \alpha_k < 1$ and $\beta_i \leq 1$ for $i = 1, \ldots, n-1$ then A_L is convergent even if it fails to be irreducible.

A necessary and sufficient condition to achieve consensus

Remember that if A is row stochastic and primitive then the linear system $x(k) = Ax(k-1)$ achieves consensus. However primitivity is just a sufficient condition while the following theorem states a necessary and sufficient condition.

Theorem

Let A be a row stochastic matrix and let $G = D(A)$. The following statements are equivalent:

(1) $\rho(A) = 1$ is a simple eigenvalue of A and if $\mu \in \sigma(A)$ and $\mu \neq 1$ then $|\mu| < 1$.

(2) $\lim_{k \to +\infty} A^k = ew^T$ where $w \geq 0$ and $w^T e = 1$.

(3) A has a power with a positive column.

(4) G contains a globally reachable node ad the subgraph of globally reachable nodes is aperiodic.

Comments

- It is not difficult to verify that the four statements are always satisfied for a primitive matrix A.
- For stochastic matrices, the theorem can be seen as an extension of Perron theorem to non primitive and even reducible matrices such as, for example,
$$A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$
- A digraph has a globally reachable node if and only if there is a unique sink in its condensation digraph $C(G)$. Hence consensus is reached if and only if in $C(G)$ there is a unique sink that corresponds to an aperiodic subgraph of G.
- The theorem can be extended to the case where $C(G)$ has $M \geq 2$ sinks corresponding to aperiodic subgraphs of G. This can be shown to be equivalent to the case where 1 is a semi simple eigenvalue of A of multiplicity M and the remaining eigenvalues have modulus < 1.

(1)⇔(2)

First of all let us observe that if $\lim_{k\to+\infty} A^k = ab^T$, with $a \neq 0 \neq b$ then $Aa = a$, $b^T A = b^T$ and $b^T a = 1$.

As we know, A is semi-convergent if and only if 1 is a semi simple eigenvalue and all other eigenvalues have modulus less than 1. However, $\lim_{k\to+\infty} A^k$ has rank one if and only if 1 is simple and the other eigenvalues have modulus < 1.

(2)$\Rightarrow$(3)

Since $w^T e = 1$ the vector w has some positive entry. Let j be such that $w_j > 0$, then $\lim_{k\to+\infty} A^k e_j = ew^T e_j = w_j e > 0$. This means that there exists an integer K such that $A^K e_j > 0$.

(3)⇒(4)

Let K be such that $A^K e_j > 0$. Then the j-th node of $G = D(A)$ is globally reachable. Moreover, since A is row stochastic $A^k e_j > 0$ for $k \geq K$. Let us consider the subgraph of G made of the globally reachable nodes and let us assume, by contradiction, that it is periodic of period $p > 1$. Since $A^K_{j,j} > 0$ and $A^{K+1}_{j,j} > 0$ there are two paths of $K + 1$ and $K + 2$ ordered nodes starting and ending with node j, that, by virtue of our assumption, cannot have a loop. These nodes are the endpoints of K and $K + 1$ edges respectively. We can trasform these two paths in simple paths by eliminating from them all the cycles connecting intermediate repeated nodes thus obtaining two cycles of $K - \alpha p$ and $K + 1 - \beta p$ edges, and distinct nodes, where α and β are suitable integers. Clearly $p > 1$ cannot be a common divisor of these two numbers, since their difference is $1 + (\alpha - \beta)p$. It follows that necessarily $p = 1$.

(4)⇒(2)

Let us assume that $G = D(A)$ contains a globally reachable node and the subgraph of the globally reachable nodes is aperiodic. Of course if all the nodes of G would be globally reachable then G would be strongly connected, and being aperiodic, it follows that A would be primitive and (2) would be certainly satisfied. Hence let us assume that there are nodes of G not globally reachable, so that A is reducible and there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{1,1} & O \\ A_{2,1} & A_{2,2} \end{bmatrix},$$

where $A_{1,1}$ is the adjacency matrix of the subgraph of the globally reachable nodes. Since $A_{1,1}$ is irreducible and aperiodic it is primitive, and being row stochastic $\lim_{k\to+\infty} A_{1,1}^k = ew_1^T$ where $w_1 > 0$ and $w_1^T e = 1$.

(4)⇒(2)

Clearly $A_{2,1}$ has some positive entry so that $A_{2,2}$ is row substochastic. In addition, let us consider the subgraph of G whose adjacency matrix is $A_{2,2}$. Every node of this subgraph has a path towards a node having out-degree less that one because these are the nodes that give access to the globally reachable nodes. It follows that $A_{2,2}$ is convergent so that PAP^T is semi convergent and

$$\lim_{k\to+\infty}(PAP^T)^k = \lim_{k\to+\infty}(PA^kP^T) = \lim_{k\to+\infty}\begin{bmatrix} A_{1,1}^k & O \\ X(k) & A_{2,2}^k \end{bmatrix} = \begin{bmatrix} ew_1^T & O \\ X & O \end{bmatrix}.$$

Being $A_{2,2}$ convergent and $A_{1,1}$ primitive the limit has to be a row stochastic matrix of rank one and the only possibility is

$$\lim_{k\to+\infty}(PAP^T)^k = ew^T$$

where $w = \begin{bmatrix} w_1 \\ O \end{bmatrix}$.

Nodes not globally reachable

It is interesting to observe that the entries of w corresponding to the nodes globally reachable are positive while the remaining are equal to zero. If $x(k) = Ax(k-1)$ then

$$\lim_{k \to +\infty} x(k) = (w^T x(0))e$$

so that the initial values of $x_i(0)$ of all nodes i not globally reachable have no effect on the final convergence value.

Example: Leslie population model

Let us assume $\beta_i = 1$ for $i = 1, \ldots, n-1$, $\sum_{k=1}^{n} \alpha_k = 1$, $\alpha_q \neq 0 \neq \alpha_p$, with $q < p$ coprimes, $\alpha_{p+1} = \ldots = \alpha_n = 0$. Hence the nodes from 1 to p are globally reachable and form an aperiodic subgraph of $G(A_L)$. The first p entries of w are different from zero while the remaining are equal to zero. This means that only the first p entries of $x(0)$ have influence on the consensus reached by the dynamical system $x(k) = A_L x(k-1)$.

Rate of convergence

Of course, besides convergence, the rate of convergence is fundamental in applications. For simplicity let us consider the case where 1 is a simple eigenvalue of A, the other eigenvalues have modulus less than 1 and A is diagonalizable. Then

$$A = V \begin{bmatrix} 1 & O \\ O & \Lambda \end{bmatrix} V^{-1},$$

where Λ is a diagonal matrix of order $n-1$ containing the eigenvalues of A different from 1. Clearly

$$A^k = V \begin{bmatrix} 1 & O \\ O & \Lambda^k \end{bmatrix} V^{-1}.$$

Rate of convergence

Let $V = [e, *]$ and $(V^{-1})^T = [w, *]$. Then

$$A^k = V \begin{bmatrix} 1 & O \\ O & O \end{bmatrix} V^{-1} + V \begin{bmatrix} O & O \\ O & \Lambda^k \end{bmatrix} V^{-1} = ew^T + V \begin{bmatrix} O & O \\ O & \Lambda^k \end{bmatrix} V^{-1},$$

so that

$$\|A^k - ew^T\|_2 = \|V \begin{bmatrix} O & O \\ O & \Lambda^k \end{bmatrix} V^{-1}\|_2 \le \|V\|_2 \|V^{-1}\|_2 \|\Lambda\|_2^k.$$

We define the essential spectral radius of A as $\rho_{\mathrm{ess}}(A) = \rho(\Lambda)$ hence

$$\|A^k - ew^T\|_2 \le \|V\|_2 \|V^{-1}\|_2 \rho_{\mathrm{ess}}(A)^k.$$

Time-varying averaging systems

The dynamical system

$$x(k+1) = A(k)x(k), \qquad k \geq 0$$

where the matrices $A(k)$ are row stochastic and $x(0) \geq 0$ is known as time-varying averaging system. Notice that, $A(k)e = e$ so that $x(0) = e$ implies $x(k) = e$ for every $k \geq 0$. For these systems it is important to obtain:

- conditions for convergence to consensus, i.e., convergence of $x(k)$ to a suitable multiple of e;
- some form of estimate of the asymptotic rate of convergence.

The simplest theorems involve properties of connection of the graphs of each of the $A(k)$. However, if these properties are not satisfied, it is necessary to examine also the products of the $A(k)$.

The case where $A(k)$ are symmetric

Theorem

For $k \geq 0$ let $A(k)$ be a sequence of symmetric row stochastic matrices of order n such that:

(1) there exists $\epsilon > 0$ such that, for every k, all the nonzero entries of $A(k)$ belong to the interval $[\epsilon, 1]$;

(2) for $k \geq 0$ the digraph $D(A(k))$ is strongly connected and aperiodic (namely $A(k)$ is primitive).

Then the time-varying averaging system $x(k+1) = A(k)x(k)$ converges to average consensus $\frac{1}{n}e^T x(0)e$ and the rate of convergence is the supremum of the set of the essential spectral radii of all the possible $A(k)$.

The case where $A(k)$ are symmetric

Proof Let us start with the algebraic part of the theorem. We perform the change of variable $\delta(k) = x(k) - \frac{e^T x(0)}{n} e$. Then

$$\begin{aligned} & x(k+1) = A(k)x(k) \\ \Leftrightarrow \quad & x(k+1) - \frac{e^T x(0)}{n} e = A(k)(x(k) - \frac{e^T x(0)}{n} e) \\ \Leftrightarrow \quad & \delta(k+1) = A(k)\delta(k). \end{aligned}$$

Since $A(k)$ is symmetric $e^T A(k) = e^T$ so that

$$e^T x(k) = e^T A(k-1)x(k-1) = e^T x(k-1) = \ldots = e^T x(0),$$

so that

$$e^T \delta(k) = e^T x(k) - e^T x(0) = 0,$$

hence

$$\delta(k+1) = (A(k) - \frac{1}{n} e e^T)\delta(k).$$

The case where $A(k)$ are symmetric

Since $A(k)$ is symmetric and primitive it turns out that there exists an orthogonal matrix $V(k)$ such that $A(k) = V(k)\Lambda(k)V^T(k)$ where $V(k)e_1 = \frac{e}{\sqrt{n}}$ and

$$\Lambda(k) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \lambda_2(k) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n(k) \end{pmatrix},$$

where $\rho_{ess}(A(k)) = \max\{\lambda_2(k), \ldots, \lambda_n(k)\} < 1$.

The case where $A(k)$ are symmetric

Hence

$$\|A(k) - \frac{1}{n} ee^T\|_2 = \|V(k)\Lambda(k)V(k)^T - V(k)e_1 e_1^T V(k)^T\|_2$$

$$= \quad \|\Lambda - e_1 e_1^T\|_2 = \| \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2(k) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n(k) \end{pmatrix} \|_2 = \rho_{ess}(A(k)).$$

Hence

$$\|\delta(k+1)\|_2 = \|(A(k) - \frac{1}{n} ee^T)\delta(k)\|_2 \leq \rho_{ess}(A(k))\|\delta(k)\|_2.$$

Now we need to show that there is a constant $c < 1$ such that $\sup_k \rho_{ess}(A(k)) < c$.

The case where $A(k)$ are symmetric

Because each $D(A(k))$ is strongly connected and aperiodic, $A(k)$ is primitive and has essential spectral radius strictly less than 1.
For fixed n, there exists only a finite number of possible unweighted strongly connected aperiodic graphs with n nodes, and for each given graph the set of the symmetric matrices with weights in the interval $[\epsilon, 1]$ is closed and limited, and the finite union closed and limited sets is again closed an limited. Each of these matrices essential has spectral radius strictly less than 1.
It is known that the function from the entries of a matrix and its eigenvalues and the function from $n-1$ numbers and their maximum are continuous. Hence ρ_{ess} attains on the matrices of the set a maximum value and this maximum of has to be less than 1.
It follows that there exists $c < 1$ such that $\sup_k \rho_{ess}(A(k)) \leq c$.

Union of digraphs

Given two digraphs $G = (V, E)$ and $G' = (V', E')$ their union is defined as $G \cup G' = (V \cup V', E \cup E')$. Clearly if $V = V'$ then the union is essentially defined by the union of the edge sets. If A and A' are two non-negative matrices of the same order with positive diagonal entries and $G = D(A)$, $G' = D(A')$ then $E \cup E'$ is contained in the edge set of $D(AA')$.
The following theorem has mainly theoretical interest: it shows that, if the graphs $D(A(k))$ are not connected, the union of their digraphs can nevertheless have useful connection properties.

Connection over times

In the following theorem a suitable hypothesis of connection is made on the union of the digraphs of the matrices.

Theorem

Let $A(k)$ for $k \geq 0$ be a sequence of row stochastic matrices of order n with positive diagonal entries and let $G(k) = D(A(k))$. Let us assume that:

(1) there exists $\epsilon > 0$ such that for $k \geq 0$ each non-zero entry of $A(k)$ belongs to the interval $[\epsilon, 1]$;

(2) there exists an integer $\delta > 0$ such that for each $k \geq 0$ the digraph $G(k) \cup \ldots \cup G(k + \delta - 1)$ contains a globally reachable node.

Then the time-varying averaging system $x(k+1) = A(k)x(k)$ converges to consensus, and to average consensus if the $A(k)$ are doubly stochastic.

The Massey method

In 1997, Kenneth Massey proposed a method for ranking college football teams. Massey's method rates a given team according to the point spread of the team (the difference between points for and against the team) and the ratings of the other teams matched so far in a tournament.

The Massey method

Let d_i be the number of the matches played by team i, let $a_{i,j}$ the number of the matches played by team i against team j and let p_i be the difference between the points made and suffered by the team i. Then the rating r_i of team i according to Massey's method is given by

$$r_i = \frac{1}{d_i} \sum_j a_{i,j} r_j + \frac{p_i}{d_i}.$$

Notice that r_i is the sum of two components

- the mean rating of teams that i has matched:

$$r_i^{(1)} = \frac{1}{d_i} \sum_j a_{i,j} r_j;$$

- the mean point spread of team i in the played matches:

$$r_i^{(2)} = \frac{p_i}{d_i}.$$

The time-aware Massey method

Recently, a time-aware version of the Massey method has been proposed. The idea of for incorporating time in the method is simple: to take into account the strength of the teams at the time when the match is played. In order to simplify the presentation let us consider the case where the tournament is arranged as in the italian serie A competition: if there are n teams, n assumed to be even for simplicity, at each day k of the sport season, for $k = 1, \ldots, n-1$, each team competes against a team not matched before.

The time-aware Massey method

The rating of team i at season day k is:

$$r_i(k) = \frac{1}{k} \sum_{j=1}^{k} r_{\pi_i(j)}(j-1) + \frac{p_i(k)}{k},$$

where $1 \leq k \leq n-1$, $\pi_i(j)$ is the team that competes against i at day j, $p_i(k)$ is the point spread of team i at day k, and $r_i(0) = 0$ for all teams. This means that the rating $r_i(k)$ of team i at day k is the sum $r_i^{(1)}(k) + r_i^{(2)}(k)$ of two components:

- the mean historical rating of teams that i has matched:

$$r_i^{(1)}(k) = \frac{1}{k} \sum_{j=1}^{k} r_{\pi_i(j)}(j-1);$$

- the mean point spread of team i at day k:

$$r_i^{(2)}(k) = \frac{p_i(k)}{k}.$$

The time-aware Massey method in matrix form

It is convenient to write the previous equation in an alternative compact matrix form. First of all we need a family of symmetric permutation matrices $P(k)$, $k = 1, \ldots, n-1$ that embody the tournament calendar. Since in a round-robin tournament each team must compete against every other team the permutation matrices $P(k)$ satisfy the constraint

$$\sum_{k=1}^{n-1} P(k) = ee^T - I = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}.$$

The time-aware Massey method in matrix form

Now, let $s(t)$ be a vector such that $s_i(t)$ is equal to the points realized by the team i in the match played at time t. We can describe the temporal Massey method as

$$r(k) = \frac{1}{k}\sum_{t=1}^{k} P(t)r(t-1) + \frac{1}{k}\sum_{t=1}^{k}(I - P(t))s(t),$$

where $k = 1, \ldots, n-1$. Here we assume $r(0) = [0, \ldots, 0]^T$, i.e., zero initial rating for each team, but other choices could be considered.

The time-aware Massey method in matrix form

It is possible rewrite the last equation in order to express $r(k)$ as a function of $r(k-1)$. Observe that, for $k = 1, \ldots, n-1$:

$$
\begin{aligned}
r(k) &= \frac{1}{k}\sum_{t=1}^{k}\big(P(t)r(t-1) + (I - P(t))s(t)\big) \\
&= \frac{k-1}{k}\frac{1}{k-1}\sum_{t=1}^{k-1}\big(P(t)r(t-1) + (I - P(t))s(t)\big) \\
&\qquad + \frac{1}{k}\big(P(k)r(k-1) + (I - P(k))s(k)\big) \\
&= \frac{k-1}{k}r(k-1) + \frac{1}{k}\big(P(k)r(k-1) + (I - P(k))s(k)\big) \\
&= \frac{1}{k}\big((k-1)I + P(k)\big)r(k-1) + \frac{1}{k}(I - P(k))s(k).
\end{aligned}
$$

The time-aware Massey method in matrix form

Let us set

$$C(k) = \frac{1}{k}P(k) + \frac{k-1}{k}I.$$

Notice that $C(k) \geq 0$ and

$$C(k)e = \frac{1}{k}P(k)e + \frac{k-1}{k}e = \frac{1}{k}e + \frac{k-1}{k}e = e$$

hence these matrices are row stochastic, besides being symmetric due to the symmetry of the permutations $P(k)$.

The time-aware Massey method in matrix form

By using the $C(k)$ we obtain

$$\begin{aligned} r(k) &= \frac{1}{k}\big((k-1)I + P(k)\big)r(k-1) + \frac{1}{k}(I - P(k))s(k) \\ &= C(k)r(k-1) - C(k)s(k) + s(k). \end{aligned}$$

This suggests the change of variable $q(k) = r(k) - s(k)$ that yields

$$q(k) = C(k)q(k-1) + C(k)(s(k-1) - s(k)).$$

The perfect season

The dynamical system

$$q(k) = C(k)q(k-1) + C(k)(s(k-1) - s(k))$$

falls outside the framework discussed so far. We discuss an important particular case, theoretically very interesting, known as perfect season. In the perfect season, if i competes with j then the point difference is $j - i$ (hence i defeats j in the case where $i < j$). In order to model these results it is sufficient to set $s_i(t) = n - i$. With this assumption $s(t) = s = [n-1, n-2, \ldots, 0]^T$ becomes independent from t and $s(k-1) - s(k) = 0$. Hence in the perfect season we have to study the time-varying averaging system

$$q(k) = C(k)q(k-1) \quad \text{where} \quad q(0) = s.$$

Why the perfect season is important?

The perfect season has been introduced in the literature as a first step for the study of the sensitivity (variation of the output as a consequence of a small variation of the input) of a ranking method. For example, it can be shown that the original (not time aware) Massey method at the end of the perfect season produces the expected ranking, where team i precedes team $i+1$ for $i = 1, \ldots, n-1$, and in addition the rating of the teams are evenly spaced. Actually, a more detailed analysis shows that the original Massey's method has excellent sensitivity properties. Can we obtain the same results for time-aware Massey method? Let us start from an asymptotic result that we can deduce without too many effort from the theory developed so far on time-varying averaging systems.

Asymptotic convergence

For the matrices $C(k)$ unfortunately $G(k) = D(C(k))$ is not connected, so that we are forced to consider their products. Notice that

(1) The smallest nonzero entry of the $C(k)$ is $\frac{1}{n-1}$.

(2) Each of the $G(k)$ has a self loop at each node.

(3) The digraph $G(1) \cup \ldots \cup G(n-1)$ not only contains a globally reachable node but actually it is complete.

We conclude that if the tournament were repeated many times then $q(k)$ would converge to average consensus $\frac{1}{n}e^T q(0)e = -\frac{1}{n}e^T se = \frac{1-n}{2}e$. Since $r(k) = q(k) + s$ we obtain that $r(k)$ converges to $s + \frac{1-n}{2}e$ a vector of evenly spaced ratings.

Rating and ranking at the end of one tournament

It is clear that here an asymptotic result is not satisfying and that we need to obtain some result that can be compared with what known for the classical Massey method, that holds at the end of just one tournament. A first simple observation is that the ratings depend from the sequence of the matches. In addition, if for $n = 4$ and $n = 6$ the time aware method always produces the expected ranking, for $n = 8$ there are tournaments that not always end with this result.

Example

For $k = 1, \ldots, 8$, let the row k of the following table represent the match calendar of the team k

$$\begin{matrix} 4 & 8 & 5 & 7 & 6 & 2 & 3 \\ 5 & 6 & 3 & 8 & 7 & 1 & 4 \\ 8 & 7 & 2 & 4 & 5 & 6 & 1 \\ 1 & 5 & 6 & 3 & 8 & 7 & 2 \\ 2 & 4 & 1 & 6 & 3 & 8 & 7 \\ 7 & 2 & 4 & 5 & 1 & 3 & 8 \\ 6 & 3 & 8 & 1 & 2 & 4 & 5 \\ 3 & 1 & 7 & 2 & 4 & 5 & 6 \end{matrix},$$

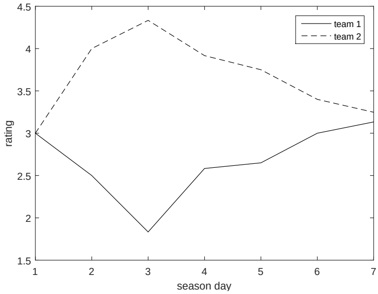
then the obtained ratings approximated to the second decimal digit are respectively

$$3.13 \quad 3.24 \quad 1.80 \quad 0.486 \quad -1.19 \quad -1.86 \quad -2.14 \quad -3.47 \; ,$$

so that team 2 wins the perfect season despite the fact that is defeated by team 1. Why?

Example

The reason is that the match between team 1 and team 2 is located towards the end of the tournament (more precisely in the sixth day of the perfect season). Team 2 arrives to the match with a higher rating since it defeats in the first days of the season teams that are stronger of the teams defeated by 1 in the same days. When team 2 is defeated by team 1, the difference between the two ratings decreases, but the rating of 2 remains higher.

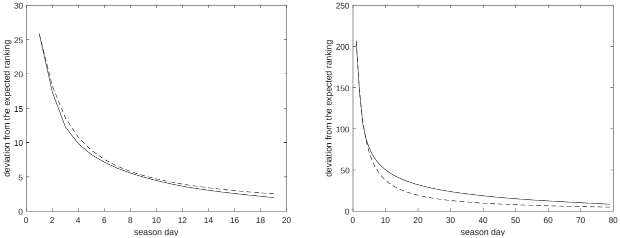


Rate of convergence

What obtained so far suggests that the dynamical system $q(k) = C(k)q(k-1)$ converges but the convergence speed does not guarantee that the expected ranking is reached at the end of one tournament. In order to characterize the convergence speed it is useful to introduce the concept of paracontraction. A matrix A is a paracontraction with respect to a given norm $\| \circ \|$ if $Ax \neq x$ implies $\|Ax\| < \|x\|$. This means that there are only two possibilities for a vector x: or $Ax = x$ or the norm of x is reduced by the application of A. It is not difficult to show that the matrices $C(k)$ are paracontractions with respect to the Euclidean norm $\| \circ \|_2$.

Rate of convergence

By exploiting this property it is possible to obtain a good description of the observed deviation from the expected ranking.
In the figure the solid line is a measure of the deviation from the expected ranking obtained experimentally by a mean over 1000 perfect seasons of 20 teams (left) and 80 teams (right), while the dashed line is obtained by exploiting paracontractivity.



Conclusions

- A substochastic matrix A is convergent if and only if in $D(A)$ there is a path from every node to a node having out degree less than 1 (condition weaker than irreducibility).
- A dynamical system $x(k+1) = Ax(k)$ with A row stochastic, reach consensus if and only if $D(A)$ contains a globally reachable nodes and the subset of globally reachable nodes is aperiodic (condition weaker than primitivity).
- Rate of convergence depends on essential spectral radius of A.

Conclusions

- For time varying averaging systems $x(k) = A(k)x(k-1)$, obtain conditions for convergence and estimate for the rate of convergence can be difficult.
- The simplest theorems involve properties of connection of the graphs of each of the $A(k)$. However, if these properties are not satisfied, it is necessary to examine also the products of the $A(k)$.
- Time aware Massey method leads, in the simple setting of perfect season, to a time varying averaging system.
- It is simple to prove the asymptotic convergence of the method. Only preliminary results on the rate of convergence have been obtained.

Suggested readings

F. Bullo
Lectures on network systems
Create space, 2018.

T. Chartier, E. Kreutzer, A. Langville, and K. Pendings
Sensitivity and stability of ranking vectors
SIAM Journal on Scientific and Statistical Computing, 33, pp. 1077–1102, 2011.

M. Franceschet, E. B., P. Vidoni
The temporalized Massey's method
Journal of quantitative analysis of sports, 13(2), pp. 37–48, 2017.

E. B., P. Vidoni, M. Franceschet,
A parametric family of Massey-type methods: inference, prediction and sensitivity
Journal of Quantitative Analysis of Sports, 16(3), 2020, pp. 255–269.

Exercises

1. Prove the lemma on substochastic matrices.
2. Let $A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix} \geq 0$. Prove that $\rho(A) < 1$ if and only if A is row substochastic.
3. Prove that if $\lim_{k\to+\infty} A^k = ab^T$, with $a \neq 0 \neq b$ then $Aa = a$, $b^T A = b^T$ and $b^T a = 1$.
4. Notice that, in order to obtain the ratings in the original Massey method it is necessary to solve a linear system. (a) Write explicitly the system $Lr = p$ that relates ratings and points. (b) Prove that L is singular. (c) Write explicitly the system $Lr = p$ in the perfect season with n teams. (d) Assuming that the rank of L is $n-1$ show that the system is solvable and propose a strategy to solve it.
5. Prove that the matrices $C(k)$ are paracontractions in the Euclidean norm.