Vladimir Protasov

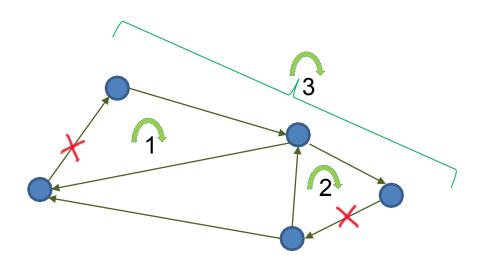
University of L'Aquila (Italy), MSU, HSE (Russia)

The Maximal Acyclic Subgraph and Stability of Linear Systems

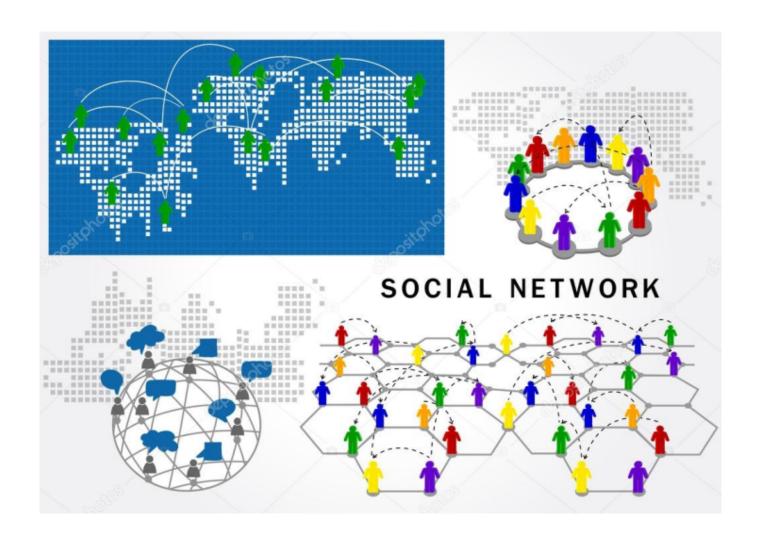
The Maximal Acyclic Subgraph

problem, Mas ven directed graph.

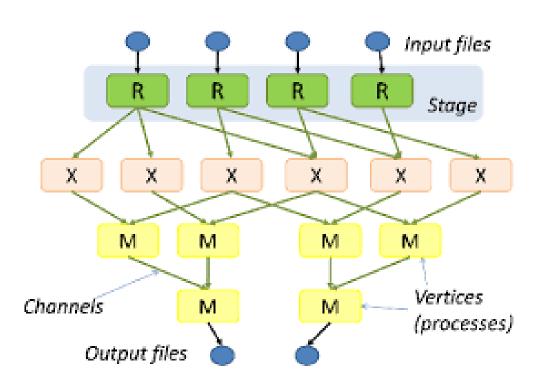
Find its acyclic subgraph G' = (V, E') for which $|E'| \rightarrow \max$



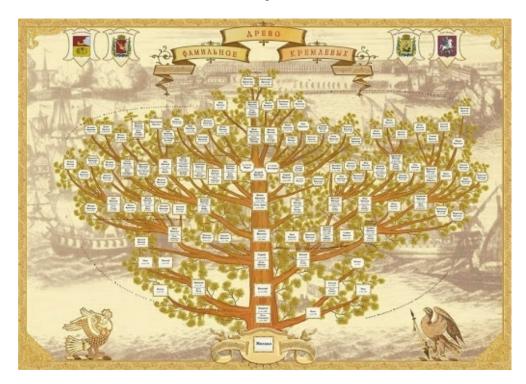
Answer: $\max |E'| = 5$



Computer architecture, parallel computing



A family tree



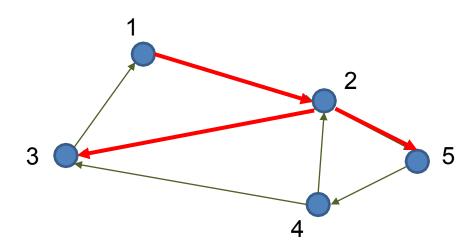
A family tree is not a tree!

One man has 40 generations of ancestors for 1000 years. $2^{40} > 10^{12}$ ancestors. More than the total population of the Earth!

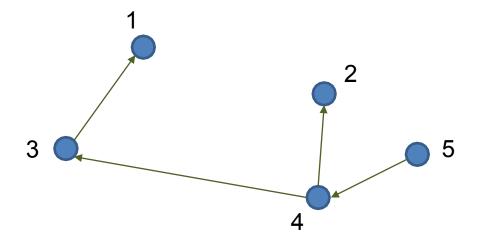
The Pedigree Phenomenon

A family tree is an acyclic graph

The simplest method: make some ordering of vertices, then take those edges E' directed in the increasing order (or decreasing). Then G' = (V, E') is acyclic.



We have |E'| = 3



For the decreasing edges |E'| = 4

At least one of these two sets of edges contains $\geq \frac{1}{2} |E|$ edges.

Therefore, this simple method gives an approximate answer with factor $\geq \frac{1}{2}$

This is still the best approximation obtained by a polynomial algorithm.

No polynomial algorithm is known with approximation factor $\frac{1}{2} + \varepsilon$

Finding max |E'| is NP complete

The MAS problem is in the list of 21 NP-complete problems by R.Karp (1973)

Finding an acycling subgraph with
$$\geq \left(\frac{65}{66} + \varepsilon\right) \max |E'|$$
 is NP hard

There are algorithms that give with approximation factor $\frac{1}{2} + \varepsilon$ for the vast majority of graphs.

We trace the connection between the MAS problem and the problem of the closest positive stable dynamical system.

This will give:

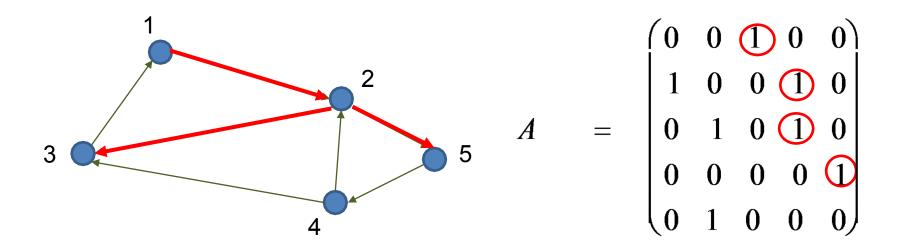
- 1. A fast method providing estimates not worse than existing methods
- 2. Exact solutions for some reformulations of MAS problem.

The main approach is to work with spectral properties of the adjacency matrix A.

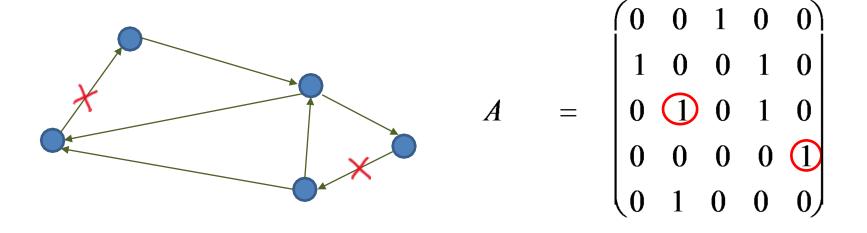
 $(A)_{ij} = 1$ if and only if there is an edge from j to i.

Observation:

A graph is acyclic $\Leftrightarrow \rho(A) = 0$



Observation: A graph is acyclic $\Leftrightarrow \rho(A) = 0$



Observation:

A graph is acyclic
$$\Leftrightarrow \rho(A) = 0$$

Proof. If G is acyclic, then all paths allong G have lengths

at most
$$n-1 \implies A^n = 0 \implies \rho(A) = 0$$
.

Conversely. If there exists a cycle, then $||A^k|| \ge 1$ for each $k \implies$

(Gelfand's formula)
$$\rho(A) = \lim_{k \to \infty} \left| \left| A^k \right| \right|^{1/k} \ge 1.$$

Observation:

If X is a non-negative $d \times d$ matrix with $\rho(X) = 0$, then the corresponding Boolean matrix \tilde{X} :

$$(\tilde{X})_{ij} = 1 \text{ if } (X)_{ij} > 0$$

$$(\tilde{X})_{ij} = 0$$
 if $(X)_{ij} = 0$

possesses the same property $\rho(\tilde{X}) = 0$

If we replace X by \tilde{X} , the distance $||A-X||_2$ decreases.

$$||Y||_2 = \left(\sum_{i,j} |Y_{ij}|^2\right)^{1/2}$$
 is the Frobenius norm.

X is the MAS, if and only if X is the sulution of the problem

$$\left\|A - X\right\|_2 \to \min$$

$$\rho(X) = 0$$

This problem is closely related to the stabilising problem for positive dynamical systems:

$$||A - X||_2 \to \min$$

$$\rho(X) = 1$$

A linear dynamical system with discrete time:

$$x_{k+1} = A x_k, \quad k \in \mathbb{Z}_+ \quad x_0 \text{ is given}$$

A system is stable if all trajectories tend to zero (Schur stability)

$$|\lambda_k| < 1$$
, $\lambda_k \in sp(A)$

Positive linear systems.

How to find the closest stable/unstable system?

Applications

Mathematical economics (Leontief input-output model)

Population dynamics in mathematical biology

Epidemiological dynamics

Fluid network control

Blanchini, Colaneri, Valker "Switched positive linear systems", 2015

Krause, "Swarm dynamics and positive dynamical systems", 2013

Let A be a $d \times d$ matrix, $\rho(A)$ be its spectral radius.

If $\lambda_1, ..., \lambda_d$ are eigenvalues of A and $|\lambda_1| \ge ... \ge |\lambda_d|$, then $\rho(A) = |\lambda_1|$.

Theorem 1. (Perron (1906), Frobenius (1913)). If $A \ge 0$, then the spectral radius is attained at a real positive eigenvalue $\rho(A) = \lambda_1 \ge 0$.

There is an eigenvector $v \ge 0$ such that $Av = \lambda_1 v$.

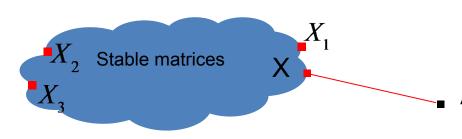
If A > 0, then the largest by modulo eigenvalue is unique and simple, and so is the corresponding eigenvector.

$$\rho(A) = \lambda_1$$

We call $\lambda_{\text{max}} = \lambda_1 = \rho(A)$ the leading eigenvalue and v the leading eigenvector.

$$||X - A|| \to \min$$

$$\rho(X) = 1$$



 $X \in abs \min$ $X, X_1, X_2, X_3 \in loc \min$

A lot depend on the norm in \mathbb{R}^d

We consider the Frobenius norm

$$(A, B) = tr A^T B;$$
 $||A||^2 = (A, A) = \sum_{i,j=1}^d a_{ij}^2$

R. Byers (1988)

F.X. Orbandexivry, Y. Nesterov, and P. Van Dooren, (2013)

C. Mehl, V. Mehrmann, and P. Sharma (2016)

N. Gillis and P. Sharma (2017)

J. Anderson (2017)

N.Guglielmi, C.Lubic (2019)

The problem is nasty. No efficient algorithms are available.

There are examples of d x d matrices that have at least 2^d locally closest stable matrices (Guglielmi, V.P., 2018).

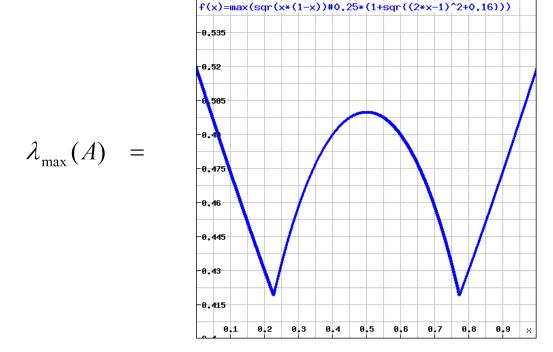
Reasons:

- The spectral radius is neither convex nor concave in matrices
- The spectral radius is non-Lipschitz, if the leading eigenvalue is multiple.

Example 1. For the set $M = [A_1, A_2] = co\{A_1, A_2\},\$

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.1 \\ 0 & 0 & 0.1 & 0 \end{pmatrix} \quad ; \qquad A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0.1 & 0.5 \end{pmatrix}$$

We have $A = (1-x)A_1 + x A_2$, $x \in [0,1]$. "Matrix segment"



The algorithm of alternating relaxations

(N.Guglielmi, V.P., 2018)

Let A be a non-stable matrix, $\rho(A) > 1$.

1) Take a matrix X_0 , $\rho(X_0) = 1$, v_0 is the right leading eigenvector: $X_0 v_0 = v_0$

 X_1 is the solution of the problem $||X - A|| \rightarrow \min$ $Xv_0 = v_0$

2) Take u_1 be the left leading eigenvector: $u_1X_1 = u_1$

 X_2 is the solution of the problem $||X - A|| \rightarrow \min$ $u_1 X = u_1$

$$X \in abs \min$$
Stable matrices

Then we alternate left and right leading eigenvectors. The distance to A decreases each step.

Theorem. If the algorithm converges to a positive matrix X, then X is a global minimum. In this case the convergence is linear and the rate can be estimated from above.

If the limit matrix X have some zero components, then

1) If X is irreducible, then X is a local minimum;

- $X = \left(\begin{array}{cc} X_1 & * \\ 0 & X_2 \end{array}\right)$
- 2) If X is reducible, then the we obtain several problems of smaller dimensions. We solve them separately

Theorem. If the algorithm always converges to a local minimum with a linear rate.

Example.

$$A = \begin{pmatrix} 0.7 & 0.2 & 0.1 & 0.5 & 1.0 \\ 0.3 & 0.6 & 0.2 & 0.8 & 0.3 \\ 0.5 & 0.7 & 0.9 & 1.0 & 0.5 \\ 0.1 & 0.1 & 0.3 & 0.8 & 0.3 \\ 0.8 & 0.2 & 0.9 & 0.3 & 0.2 \end{pmatrix}, \quad \text{with} \quad \rho(A) = 2.4031.$$

$$X_2 = \begin{pmatrix} 0 & 0.4204 & 0.1759 & 0.6770 & 0.3 \\ 0.7343 & 0.3796 & 0.1797 & 0 & 0.5 \\ 0.0274 & 0 & 0.5791 & 0.0069 & 0.8 \\ 0.1334 & 0.0580 & 0.6719 & 0.6403 & 1 \\ 0 & 0 & 0 & 0 & 0.8 \end{pmatrix}$$

The same algorithm can be applied to the MAS problem:

$$\left\|A - X\right\|_2 \to \min$$

$$\rho(X) = 0$$

However, it may terminate at a local minimum

What to do?

Change the norm!

Y.Nesterov, V.P (2017) The nonnegative stability problem

We consider two polyhedral matrix norms in \mathbb{R}^d :

$$L_{\infty}$$
 norm: $||A||_{\infty} = \max_{i=1,\dots,d} \sum_{j=1}^{d} |a_{ij}|$

$$L_1 \text{ norm: } ||A||_1 = \max_{i=1,...,d} \sum_{i=1}^d |a_{ij}|$$

In those norms the absolute minimum can be efficiently found

The problem becomes:

Find the minimal $\tau > 0$ with the following property:

The matrix can be made stable/unstable by chanding elements of each row by at most τ in the sum

Note that rows of the matrix can be changed independently of each other. Every row runs over an L_1 — ball independently of others.

Optimizing the spectral radius over product families

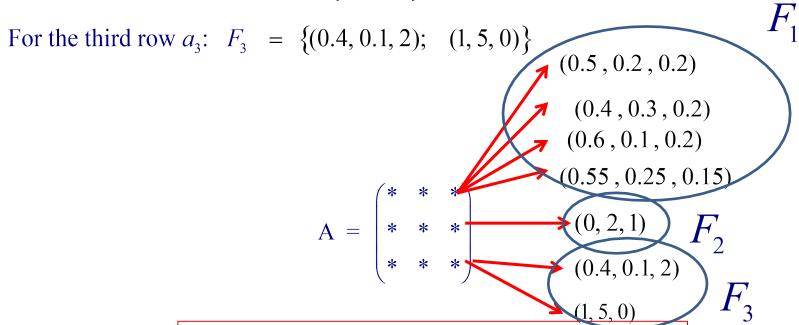
Definition 1. A family of matrices is called a *product family*, if the rows of matrices are chosen independently from given sets (*uncertainty sets*) F_i , i = 1, ..., d.

Example 2. A family of 3x3-matrices. The uncertainty sets are

For the first row a_1 :

$$F_1 = \{(0.5, 0.2, 0.2); (0.4, 0.3, 0.2); (0.6, 0.1, 0.2); (0.55, 0.25, 0.15)\}$$

For the second row a_2 : $F_2 = \{(0, 2, 1)\};$



We obtain the family M of 4x1x2 = 8 matrices

We have minimized the spectral radius over the set of eight matrices

Suppose we have four rows for each line 1, 2, and 3.

In this case we have 4x4x4 = 64 matrices. We choose one with the smallest leading eigenvalue.

Curse of dimensionality

If we have d = 50 and just TWO lines in each uncertainty set, then the total number of matrices is $2^{50} > 10^{15}$.

Moreover, the set of rows may be polyhedral (a subset of R^d defined by a system of linear inequalities).

Product families with row uncertainties

V.Kozyakin (2004) V.Blondel, Y.Nesterov (2009) Y.Nesterov, V.P. (2013)

Applications:

- Leontief model (mathematical economics)
- Structured population dynamics, mathematical ecology
- Spectral graph theory
- Asyncronouos systems



Wassily Wassilievich Leontief (1906 - 1999) Василий Васильевич Леонтьев

- 1906 born in the family of W.Leontief (from an old-believer Russian orthodox family) and Genya Leontief (Becker) from a rich Jewish merchant family from Odessa.
- 1924 Masters degree in Economics, University of Leningrad (St. Petersburg). Was persecuted and detained several times by Soviet authorities.

 1925 was allowed to leave Soviet Union
- 1932 1975 affiliated with Harward, from 1975 is with the New York University.
- 1973 the Nobel Prize in Economics.

The Leontief input-output model (1966, Nobel Prize 1973)

expresses inter-industry relationships in linear algebraic terms.

Suppose the economy has *d* sectors.

The *i*th sector produces x_i units of some single homogeneous good.

To produce one unit it consumes a_{ij} units from sector j.

The final demand:

In addition, each sector must leave b_i units of its output to consumers

For every i = 1, ..., d, we have the equality

$$x_i = b_i + \sum_{j=1}^d a_{ij} x_j$$

Around 1949, Leontief used the primitive computer systems at Harvard to model data provided by the U.S. Bureau of Labor Statistics.

He divided the U.S. economy into d = 500 sectors.

$$x_i = b_i + \sum_{j=1}^d a_{ij} x_j, \quad i = 1, ..., d.$$

In the matrix form: $A = (a_{ij}), i, j = 1, ..., d$:

$$(1) x = Ax + b$$

Definition 1. The economy is productive if it is able to provide any final demand.



Equation (1) has a nonnegative solution x for every nonnegative b.

Theorem 6. (W.Leontief). The economy is productive if and only if $\rho(A) < 1$.

Closed stable matrix: how to change economy to become productive? The uncertainty sets (rows of the Leontief matrix) are the technologies.

Positive systems in Population Dynamics

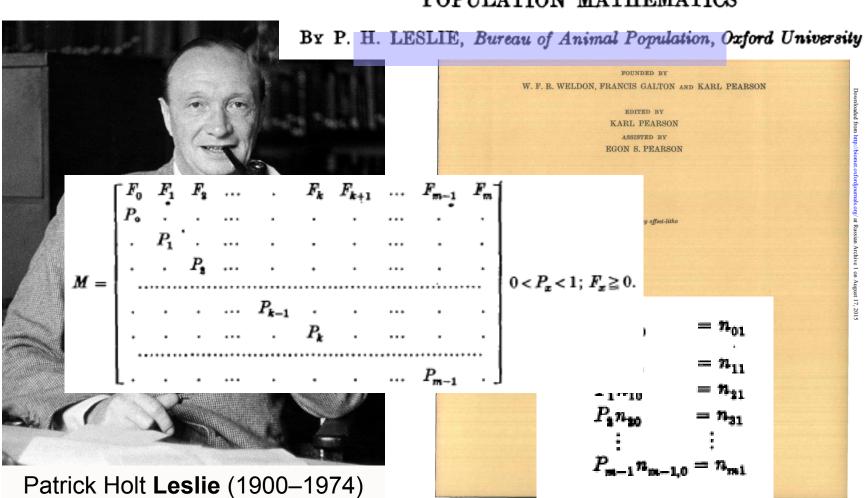
Matrix Population Models

November 1945

Vol. XXXIII. Part III

November 194

ON THE USE OF MATRICES IN CERTAIN POPULATION MATHEMATICS



Matrix Population Models



Hans J. Math. Biol. 44, 450–462 (2002)

Applications of Perron–Frobenius theory to population dynamics



Optimizing the spectral radius for product families

Studied in: Y.Nesterov, V.P. (2013), V.P. (2015)

The spectral simplex method

Definition 2. A *one-line correction*, of a matrix is a replacement of one of its lines.

Example 3. A correction of the first line. We replace the row a_1 by some row a'_1 .

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow A' = \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$a'_1 = (a'_{11}, a'_{12}, a'_{13})$$

Theorem 2. Let M be a product family of strictly positive matrices, $F_1, ..., F_d$ be uncertainty sets. For every $A \in M$ with the leading eigenvalue λ and eigenvector v, we have

a) If there is $a'_i \in F_i$ such that $(v, a'_i) > (v, a_i)$, then after the one-line correction we have

$$\lambda_{\max}(A') > \lambda_{\max}(A)$$

b) If the matrix A is maximal in each row with respect to v, i.e.,

$$(v, a_i) = \max_{a'_i \in F_i} (v, a'_i), \qquad i = 1,..., d, \text{ then}$$

$$\lambda_{\max}(A) = \max_{A' \in M} \lambda_{\max}(A')$$

The spectral simplex method

Initialization. Take an arbitrary matrix $A_1 \in M$.

Main loop. We have a matrix A_k and its leading eigenvector $v_k > 0$.

For every i = 1,...,d do:

Step *i*. Find
$$a'_i = \underset{b_i \in F_i}{\operatorname{arg\,max}} (v, b_i)$$
.

If $a'_i = a_i$, then set $A_{k+1} = A_k$ and go to the step $i+1$.

Otherwise, we have $(v, a'_i) \ge (v, a_i)$.

Make the one-line correction in the *i*th line.

Theorem 3 implies that $\rho(A'_k) > \rho(A_k)$.

Put $A_{k+1} = A'_{k}$. We have $\rho(A_{k+1}) > \rho(A_{k})$.

Compute the leading eigenvector $v_{k+1} > 0$ of A_{k+1} .

Go to step i = 1.

If the dth step is over, then END.

Theorem 3. For strictly positive matrices, the spectral simplex method is well-defined, does not cycle, and finds the solution within finite time.

Theorem 3. For strictly positive matrices, the spectral simplex method is well-defined, does not cycle, and finds the solution within finite time.

In many problems, the matrices are sparse. In this case we are in trouble.

- \bullet The leading eigenvector v of a matrix A may not be unique.
- The spectral radius is not strictly increasing with iteration, but just non-decreasing

The algorithm may cycle

Example 4.
$$F_1 = \{(0.5, 1, 0.5), (0.5, 0.5, 1)\}, F_2 = \{(0, 1, 0)\}, F_2 = \{(0, 0, 1)\}$$

$$\begin{pmatrix} 0.5 & 1 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \longleftarrow \qquad \begin{pmatrix} 0.5 & 0.5 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For sparse matrices, the algorithm cycles very often

Theorem 4. Assume a nonnegative matrix A has a simple leading eigenvector $v \ge 0$, ||v|| = 1. Then after an arbitrary one-line correction such that $(v, a'_i) > (v, a_i)$, the matrix A' possesses the same property.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow A' = \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$a'_{1} = (a'_{11}, a'_{12}, a'_{13}) \text{ such that } (v, a'_{1}) > (v, a_{1}).$$

Theorem 5. If the initial matrix A_1 of the spectral simplex method has a simple leading eigenvector, then all matrices in all iterations possess the same property, and the algorithm does not cycle.

How to choose A_1 to possess a unique leading eigenvector?

For instance to take the kth row of A_1 to be the arithmetic mean of all rows from the uncertainty set F_k , for each k = 1, ..., d.

The numerical efficiency of the spectral simplex method

Table 1 The number of iterations for maximizing the spectral radius of positive $d \times d$ matrices

d/n	2	5	10	50	100
5	3	6	8	10	13
10	7	12	14	18	23
50	29	48	58	92	109
100	56	99	131	197	213
500	274	542	701	884	1034

The sets F_i are finite, each has n elements

For d = 100, n = 2, we have the 100-dimensional Boolean cube.

The number of vertices is 2^{100} . However, the algorithm performs only 56 one-line corrections.

$$t = 12 s.$$

Theorem 6. For a product family M of strictly positive matrices, there are constants C > 0, $q \in (0,1)$, such that

$$\left|\rho(A_N) - \rho(A_*)\right| \leq C q^N,$$

where A_* is the optimal matrix, A_N is the matrix obtained in the Nth iteration of the spectral simpex method.

What happens if we optimize not one row but all rows simultaneously?

The greedy method

One iteration: find the leading eigenvector v_k of A_k and optimize ALL ROWS (not one) with respect to v_k $(A_{k+1})_j = \min_{x \in F_i} (v_k, x), \quad j = 1, ..., d$

M.Akian, S.Gaubert, J.Grand-Clément, and J.Guillaud, The operator approach to entropy games (2017)

Y.Nesterov, V.P., Computing closest stable non-negative matrix (2017), to appear in SIMAX

For small dimensions (d=2,3) we got worse results (3-4 iterations). This is natural. But the complexity almost does not grow with the dimension

In high dimensions:

The greedy algorithm converges with a fantastic rate!

$d \setminus N$	50	100	250
25	5.5	6.2	6.3
100	4.2	4.5	4.6
500	4.1	4.3	4.3
2000	4.1	4.3	4.1

Bless of dimensionality?

Theorem (A.Cvetkovic, V.P., 2018) The greedy algorithm has a quadratic rate of convergence, provided all the uncertainty sets are strictly positive.

$$||X_{k+1} - X|| \leq B||X_k - X||^2$$

 $B \leq \frac{CR}{2r\rho(X)}$, where R and r are maximal and minimal curvature radii of two-dimensional cross-sections of ∂F_i

Finding the closest stable matrix by the greedy method

Solution. Take arbitrary R > 0 and solve the probem

$$\rho(X) \to \min$$

$$\left| \left| X - A \right| \right|_{\infty} \le R$$

Then by bisection find minimal R such that $\rho(X) = 1$, $|X - A|_{\infty} = R$

Replace 1 by 0. We obtain:

Solution. Take arbitrary R > 0 and solve the probem

$$\rho(X) \to \min$$

$$\left| \left| X - A \right| \right|_{\infty} \le R$$

Then by bisection find minimal R such that $\rho(X) = 0$, $|X - A|_{\infty} = R$

This is a solution to the following problem:

Given a graph G= (V,E). Find the minimal R such that one can make G acyclic by removing at most R outgoing edges from each vertex.

Another reformulation of the MAS problem:

Problem. Given a graph G with n vertices and numbers $r_1,...,r_n$.

Make G acyclic by removing at most r_k edges going from the kth vertex, k = 1,...,n.

The problem admits a fast solution provided by the greedy method:

$$\rho(X) \to \min$$

$$\sum_{i=1}^{n} |x_{ki} - a_{ki}| \le r_k, \quad k = 1, ..., n$$

If the answer is $\rho(X) = 0$, we are done.

If the answer is $\rho(X) > 0$, the required is impossible.

Solving the MAS problem by the greedy method

Solution. Take arbitrary R > 0 and solve the probem

$$\rho(X) \to \min$$
 $||X - A||_{\infty} \le R$

Then by bisection find minimal R such that $\rho(X) = 0$, $||X - A||_{\infty} = R$

We have omitted the maximal possible number of outgoing edges from each vertex. Thus, we omitted much more edges than needed.

But we have obtained the desired ordering (in which the matrix X has a diagonal form). Then we restore all ones over the main diagonal.

The approximation is more or less the same as for the existing algorithms. Works extremely fast.

Approximate solution of MAS by the greedy method

				1000	
time	0.36s	8.1s	66.42s	622.43s	2860.79s
$\#\operatorname{steps}$	17	34.6	38.5	$44.7 \\ 0.616$	50.3
γ	0.644	0.621	0.615	0.616	0.616

Table 1. Solving the max-MAS and approximating MAS for graphs with sparsity 9-51%

				1000	
time	0.35s	6.56s	61.06s	605.73s	2614.02s
$\#\operatorname{steps}$	18.9	32.1	41.8	43.1	2614.02s 43.7 0.592
γ	0.6	0.592	0.592	0.593	0.592

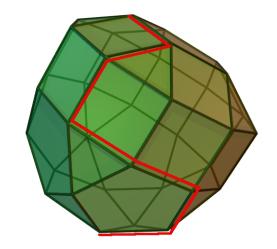
Table 2. Solving the max-MAS and approximating MAS for graphs with sparsity 26-95%

The classical simplex method (for linear programming, G.Dantzig, 1947).

LP problem:
$$\begin{cases} (c,x) \rightarrow \max \\ (a_i,x) \leq b_i, & i=1,...,N \end{cases}$$

Step-by-step increasing of the objective function (c, x) going along the edges of the polyhedron

$$G = \{ (a_i, x) \leq b_i, i = 1, ..., N \}.$$



In practice, converges extremely fast.

G.Dantzig believed that the number of steps is linear in N and d.

1972. V.Klee and G.Minty constructed an example with 2^N iterations.

In average, the number of iteration is indeed linear in N and d (S.Smale, 1983).

What is the theoretical complexity of the **Spectral simplex method?**

Thank you!