# Computing closest stable non-negative matrices

Yu. Nesterov* and V.Yu. Protasov†

August 22, 2017

**Abstract**

Problem of finding the closest stable matrix for a dynamical system has many applications. It is well studied both for continuous and discrete-time systems, and the corresponding optimization problems are formulated for various matrix norms. As a rule, non-convexity of these formulations does not allow finding their global solutions. In this paper, we analyse positive discrete-time systems. They also suffer from non-convexity of the stability region, and the problem in the Frobenius norm or in Euclidean norm remains hard for them. However, it turns out that for certain polyhedral norms, the situation is much better. We show, that for the distances measured in the max-norm, we can find exact solution of the corresponding nonconvex projection problems in polynomial time. For the distance measured in the operator $\ell_\infty$-norm or $\ell_1$-norm, the exact solution is also efficiently found. To this end, we develop a modification of the recently introduced spectral simplex method. On the other hand, for all these three norms, we obtain exact descriptions of the region of stability around a given stable matrix. In the case of max-norm, this can be seen as an analogue of Kharitonov's theorem for non-negative matrices.

**Keywords:** non-negative matrix, spectral radius, Schur stability, iterative optimization method, polyhedral norm, non-symmetric eigenvalue problem

**AMS 2010** *subject classification: 15B48, 34K20, 90C26, 65F15*

## 1 Introduction

We address the problem of finding the closest stable or closest unstable non-negative matrix to a given matrix $A$. The stability is considered in the sense of Schur: a matrix is stable if all its eigenvalues are strictly less by modulo than one. If the matrix $A$ is stable, then an interesting problem is to find for it the closest unstable matrix, i.e., the

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL) and National Research University Higher School of Economics; e-mail: Yurii.Nesterov@uclouvain.be

†DISIM of University of L'Aquila, Department of Computer Science of Higher School of Economics, E-mail: v-protassov@yandex.ru. Scientific results of this paper were obtained with support of RSF Grant 17-11-01927.

closest to $A$ matrix $X$ such that $\rho(X) = 1$, where $\rho$ denotes the spectral radius. If $A$ is unstable, i.e., $\rho(A) > 1$, then the closest stable matrix does not exist, because the set of stable matrices is open. Hence, by the closest stable matrix we understand the closest matrix $X$ with $\rho(X) = 1$ (although $X$ is actually unstable). Sometimes a matrix with spectral radius one is referred to as *weakly stable*. So, if $A$ is unstable, then the problem is to find the closest weakly stable matrix.

In all the problems above, the choice of the matrix norm plays a crucial role. In this paper, we consider three polyhedral norms:

- max-norm $\|X\|_{\max} = \max\limits_{(i,j)} |X^{(i,j)}|$;
- $\ell_\infty$ operator norm: $\|X\|_\infty = \sup\limits_{u \neq 0} \frac{\|Xu\|_\infty}{\|u\|_\infty}$, where $\|u\|_\infty = \max\limits_{1 \leq i \leq n} |u^{(i)}|$;
- $\ell_1$ operator norm: $\|X\|_1 = \sup\limits_{u \neq 0} \frac{\|Xu\|_1}{\|u\|_1}$, where $\|u\|_1 = \sum\limits_{i=1}^{n} |u^{(i)}|$. Note that $\|X\|_\infty = \max\limits_{1 \leq i \leq n} \|X^T e_i\|_1$, where $e_i$ is the $i$th coordinate vector in $\mathbb{R}^n$.

Thus, we actually consider six problems of fining closest stable/unstable matrix in these three norms. For all problems, we characterize the optimal matrix and construct efficient algorithms for finding the solution. For the max norm, we explicitly find the closest unstable matrix and present an algorithm based on bisection for computing the closest stable matrix. For the $\ell_\infty-$ and $\ell_1$-norms, we also characterize the optimal matrices. For the closest unstable matrix, the solution is found explicitly, while for finding the closest stable matrix, we use the concept of product families and apply the spectral simplex method, which can optimize the spectral radius over such families [14, 16]. To this end, we develop a modification of this algorithm, the *greedy spectral simplex method*, which may be of some independent interest.

**Motivation.** The problem of finding the closest stable or unstable matrix plays an important role in the analysis of differential equations, linear dynamical systems, electrodynamics, etc. This problem is notoriously hard due to properties of the spectral radius as a function of matrix: it is neither convex nor concave, it may loose Lipschitz continuity at some points, etc. That is why the majority of methods for this problem find only local minima. Nevertheless, we are going to see that for some classes of matrices and matrix norms, this problem is efficiently solvable even for absolute minima. We analyze the case of non-negative matrices. They correspond to positive linear systems arising naturally in problems of combinatorics, mathematical economics, population dynamics, etc. We show that on the set of non-negative matrices equipped either with the max norm (entry-wise maximum), or with the $\ell_\infty$ or $\ell_1$ operator norms, the closest stable and unstable matrices admit explicit descriptions and can be found by efficient algorithms.

Finally, let us note that in the problem of finding the closest stable/unstable *non-negative* matrix to a matrix $A$, the matrix $A$ itself does not have to be non-negative. For any real-valued matrix $A$, this problem can be reduced to the case of non-negative $A$. Indeed, if we denote $A_+ = \max\{A, 0\}$ (the entrywise maximum), then we see that the closest stable matrices to the matrices $A$ and $A_+$ are the same. Similar reasoning works also for the closest unstable matrix. Therefore, in what follows we assume the initial matrix $A$ is non-negative.

**Contents.** We start with solving the problems of the closest stable/unstable non-negative matrix in the max-norm (Section 2). We show that global minima for both problems admit explicit description and can be found by polynomial algorithms. To make them more efficient, we take a close look at the problem of computing the largest eigenvalue of a non-negative matrix. In Section 3 we develop a new method with a local quadratic rate of convergence. In Section 4 we address the problems of the closest stable/unstable non-negative matrix in the $\ell_\infty$ and $\ell_1$ norms. We show that the closest unstable matrix admits an explicit description and can be computed within polynomial time, while for finding the closest stable matrix we develop a new *greedy spectral simplex method*. In both problems we apply the method of computing the Perron eigenvalue (see Section 3) and show that the corresponding algorithms have local quadratic convergence. Note that the greedy spectral simplex method has a much wider range of applications and, probably, is of independent interest.

**Notation.** In what follows, we denote by $\mathbb{R}^{n\times n}$ the set of real $n\times n$-matrices, and by $\mathbb{R}^{n\times n}_+$ the set of non-negative matrices. For $A\in\mathbb{R}^{n\times n}_+$ and $x\in\mathbb{R}^n_+$, denote

$$\operatorname{supp}(A) \;=\; \{(i,j)\mid\; A^{(i,j)}>0\},, \qquad \operatorname{supp}(x) \;=\; \{i\mid\; x^{(i)}>0\}.$$

For two vectors $x,y\in\mathbb{R}^n_+$, we denote $x\geq y$ if $x-y\;\in\;\mathbb{R}^n_+$. The *active set* of this equality is $\{i\mid x^{(i)}\;=\;y^{(i)}\}$.

We denote $\Omega=\{1,\dots,n\}$, and for any nonempty subset $\mathcal{I}\subset\Omega$, let $V_{\mathcal{I}}\;=\;\operatorname{span}\{e_i\mid i\in\mathcal{I}\}$. So, $V_{\mathcal{I}}$ is the coordinate subspace spanned by the basis vectors with indices from $\mathcal{I}$. Finally, we use notation $I_n$ for the unit $n\times n$-matrix, and $J_n\in\mathbb{R}^{n\times n}$ for the matrix of all ones.

Let $A\in\mathbb{R}^{n\times n}$ be a real square matrix with spectrum $\Lambda(A)\stackrel{\rm def}{=}\{\lambda_1,\dots,\lambda_n\}\subset\mathbb{C}$. Denote by $\rho(A)$ its spectral radius:

$$\rho(A) \;\;=\;\; \max_{\lambda\in\Lambda(A)}|\lambda|.$$

In this case, by Perron-Frobenius theorem, $\rho(A)\in\Lambda(A)$. So, there exists a positive eigenvalue equal to the spectral radius. This eigenvalue will be denoted by $\lambda_{\max}$ and referred to as *the leading eigenvalue*. An arbitrary non-negative eigenvector $v\neq 0$ with eigenvalue $\lambda_{\max}$ is called the *leading eigenvector*. By the same Perron-Frobenius theorem, every non-negative matrix has at least one leading eigenvector [7, chapter 8].

# 2 Problem with distances measured in max-norm

In this case, the problem is rather simple and admits very efficient solutions. For a non-negative $n\times n$ matrix $A$, we consider the following problems:

1) If $\rho(A)<1$, then we find the closest unstable matrix:

$$\|X-A\|_{\max}\;\to\;\min:\quad \rho(X)\;=\;1,\;X\;\geq\;0\,. \tag{1}$$

2) If $\rho(A)>1$, then we are interested in finding the closest stable matrix. The corresponding problem looks exactly as (1).

## 2.1 Closest unstable matrix

The spectral radius of a non-negative matrix $A$ can be represented in the following minimax form:

$$\rho(A) \quad = \quad \inf_{x>0} \max_{1 \le i \le n} \frac{1}{x^{(i)}} \langle e_i, Ax \rangle. \tag{2}$$

An important consequence of this representation is monotonicity of this function:

$$A \ge B \in \mathbb{R}^{n \times n}_+ \quad \Rightarrow \quad \rho(A) \ge \rho(B). \tag{3}$$

Sometimes we need conditions for strict monotonicity.

**Lemma 1** *Let $A, B \in \mathbb{R}^{n \times n}_+$. If for some $\gamma > 0$ we have*

$$A^{(i,j)} \quad \le \quad \gamma B^{(i,j)} \quad \forall (i,j) \in \sigma(A),$$

*then $\rho(A) \le \gamma \rho(B)$.*

**Proof:**
It follows immediately from the definition (2). □

**Remark 1** *Assumptions of Lemma 1 ensure strict monotonicity of spectral radius only if $A + B$ is an irreducible matrix. This condition cannot be dropped, and the corresponding examples are well known.* □

Consider the set of weakly stable non-negative matrices

$$\mathcal{S}_n \quad = \quad \{A \in \mathbb{R}^{n \times n}_+ : \ \rho(A) \le 1\},$$

and denote by $\mathcal{S}^0_n$ the set of stable matrices, for which inequality in the above definition is strict. In what follows, we often use a simple criterion for stable matrices.

**Lemma 2** *Non-negative matrix $A$ is stable if and only if the matrix $(I_n - A)^{-1}$ is well defined and non-negative.*

**Proof:**
Indeed, if $\rho(A) < 1$, then the matrix $(I_n - A)^{-1}$ can be represented by a convergent series $\sum\limits_{k=0}^{\infty} A^k$, which is a non-negative matrix.

Let $Y \stackrel{\text{def}}{=} (I_n - A)^{-1}$ be well defined and non-negative. Since it is non-degenerate, it has the same system of eigenvectors as matrix $A$. Hence, for the leading eigenvector $s \in \mathbb{R}^n_+$ of matrix $A$ we have $Ys = \frac{1}{1-\rho(A)} s$. Since $Y \ge 0$, we conclude that $\rho(A) < 1$. □

This simple fact helps us in computing the distance between a stable matrix and the boundary of the set of unstable matrices. Let us prove first an auxiliary statement.

**Lemma 3** *Let $A \in \mathcal{S}_n^0$ and $H \in \mathbb{R}_+^{n\times n}$, $H \neq 0$. Denote $\xi(A,H) = \rho((I_n - A)^{-1}H)$. Then*

$$A + \alpha H \quad \in \quad \mathcal{S}_n^0, \quad 0 \leq \alpha < \tfrac{1}{\xi(A,H)}, \tag{4}$$

*and $\rho\left(A + \frac{H}{\xi(A,H)}\right) = 1$.*

**Proof:**
Denote $B = (I_n - A)^{-1}H \geq 0$. Note that equality $H = (I_n - A)B$ implies that $I_n - A - \alpha H = (I_n - A) - \alpha\,(I_n - A)B$. Therefore,

$$W(\alpha) \stackrel{\text{def}}{=} I_n - (A + \alpha H) \quad = \quad (I_n - A)(I_n - \alpha B), \tag{5}$$

and $\rho(\alpha B) < 1$ for all $\alpha \in \left[0, \frac{1}{\rho(B)}\right)$. Consequently, all matrices $W(\alpha)$ are well defined and $W^{-1}(\alpha)$ are non-negative as a product of non-negative matrices. Hence, by Lemma 2, all matrices $W(\alpha)$ are stable.

On the other hand, if matrix $H$ is strictly positive, then its leading eigenvector $v$ is also strictly positive and $W\left(\frac{1}{\rho(B)}\right) v \stackrel{(5)}{=} 0$. Hence, by continuity of spectral radius, we conclude that $\rho\left(A + \frac{H}{\rho(B)}\right) = 1$. □

**Corollary 1** *Let $A \in \mathcal{S}_n^0$ and $H \in \mathbb{R}_+^{n\times n}$. Then all matrices from the set*

$$\left\{X \in \mathbb{R}^{n\times n} : \; 0 \leq X < A + \tfrac{H}{\rho((I_n - A)^{-1}H)}\right\}$$

*are stable.*

**Proof:**
This is a direct consequence of Lemma 3 and of monotonicity of spectral radius. □

We conclude this section by a variant of Corollary 1 for the special case $H = J_n$. It gives an explicit formula for the closest (in the max-norm) unstable matrix.

**Theorem 1** *Let $A \in \mathcal{S}_n^0$ and $e \in \mathbb{R}^n$ be the vector of all ones. Then all matrices from the set*

$$\left\{X \in \mathbb{R}^{n\times n} : \; 0 \leq X < A + \tfrac{J_n}{\langle (I_n - A)^{-1}e, e\rangle}\right\}, \tag{6}$$

*are stable. At the same time, $\rho\left(A + \frac{J_n}{\langle (I_n - A)^{-1}e, e\rangle}\right) = 1$.*

**Proof:**
It is enough to note that $J_n = ee^T$. Therefore

$$\rho((I_n - A)^{-1}J_n) \quad = \quad \rho((I_n - A)^{-1}ee^T) \; = \; \langle (I_n - A)^{-1}e, e\rangle.$$

□

The above statement can be seen as an analog for non-negative matrices of the well-known Kharitonov theorem, describing an $\ell_\infty$-neighborhood of a vector of coefficients, which belongs to the set of stable polynomials [8].

## 2.2 Closest stable matrix

In the previous subsection, we characterized the distance from a *stable* matrix to the boundary of stability. In this section, we consider another group of questions related to the distance from an *unstable* matrix to the nonconvex set of weakly stable matrices $\mathcal{S}_n$. As above, the distance is measured in the max-norm $\|X-A\|_{\max} = \max\limits_{1\le i,j\le n} |X^{(i,j)} - A^{(i,j)}|$.

Let $A \in \mathbb{R}^{n\times n}_+$. Consider the following parametric family of minimization problems:

$$\min_{X\in\mathbb{R}^{n\times n}_+} \{\rho(X) : \ \|X-A\|_{\max} \le \tau\}, \quad \tau \ge 0. \tag{7}$$

**Lemma 4** *The optimal solution of problem (7) is a matrix $A(\tau)$ with the following elements:*

$$A^{(i,j)}(\tau) \ = \ \max\{0, A^{(i,j)} - \tau\}, \quad i,j = 1,\dots,n.$$

**Proof:**
Indeed, matrix $A(\tau)$ is feasible for problem (7). On the other hand, for any other feasible solution $X$, we have $X \ge A(\tau)$. Thus, $A(\tau)$ is optimal for (7) in view of monotonicity of spectral radius (3). □

Consider now the following projection problem:

$$\tau_A \ = \ \min_{X\in\mathcal{S}_n} \|X-A\|_{\max}. \tag{8}$$

**Lemma 5** *Value $\tau_A$ is the unique root of equation $\rho(A(\tau)) = 1$.*

**Proof:**
Indeed, in view of Lemma 1, the function $\rho(A(\tau))$ is monotonically decreasing. □

**Example 1** *Consider the following Sudoku matrix:*

$$A \ = \ \left[\begin{array}{ccc|ccc|ccc}
5 & 3 & 4 & 6 & 7 & 8 & 9 & 1 & 2 \\
6 & 7 & 2 & 1 & 9 & 5 & 3 & 4 & 8 \\
1 & 9 & 8 & 3 & 4 & 2 & 5 & 6 & 7 \\ \hline
8 & 5 & 9 & 7 & 6 & 1 & 4 & 2 & 3 \\
4 & 2 & 6 & 8 & 5 & 3 & 7 & 9 & 1 \\
7 & 1 & 3 & 9 & 2 & 4 & 8 & 5 & 6 \\ \hline
9 & 6 & 1 & 5 & 3 & 7 & 2 & 8 & 4 \\
2 & 8 & 7 & 4 & 1 & 9 & 6 & 3 & 5 \\
3 & 4 & 5 & 2 & 8 & 6 & 1 & 7 & 9
\end{array}\right]$$

*In accordance to Lemma 5, we can easily find its $\ell_\infty$-projection onto the set of stable*

*matrices:*

$$X^* = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

*with* $\tau_A = 8$. □

Let $A \in \mathbb{R}^{n\times n}_+$. Assume that $\rho(A) > 1$. Our strategy of solving the problem (8) is as follows.

1. Sort all elements of matrix $A$ in an increasing order.
2. Using the criterion of Lemma 2, find by bisection in the element number the value $\tau_1$, which is the largest between all $A^{(i,j)}$ and zero, having $\rho(A(\cdot)) \geq 1$, and value $\tau_2$, which is the smallest element of $A$ with $\rho(A(A^{(i,j)})) < 1$.
3. Form the matrix $H$, with elements $H^{(i,j)} = \begin{cases} 1, & \text{if } (i,j) \in \text{supp}\,(A(\tau_1)), \\ 0, & \text{otherwise} \end{cases}$.
4. Compute the output as $\tau_A = \tau_2 - \frac{1}{\rho((I_n - A(\tau_2))^{-1}H)}$.

(9)

**Theorem 2** *Algorithm (9) computes an optimal solution of problem (8).*

**Proof:**
First of all, let us show that the Algorithm (9) is well-defined. Indeed,

$$A\left(\max_{1\leq i,j\leq n} A^{(i,j)}\right) \;=\; 0,$$

and $\rho(A(0)) = \rho(A) > 1$. Thus, we can find two values $\tau_1 < \tau_2$ from the set

$$\{0\} \bigcup \{A^{(i,j)}\}_{i,j=1}^n$$

such that $\rho(A(\tau_1)) \geq 1$, $\rho(A(\tau_2)) < 1$, and $A(\tau)$ is linear for $\tau \in [\tau_1, \tau_2]$. Hence, for $\tau$ from this interval we have

$$A(\tau) \;=\; A(\tau_2) + \tfrac{\tau_2 - \tau}{\tau_2 - \tau_1}(A(\tau_1) - A(\tau_2)) \;=\; A(\tau_2) + (\tau_2 - \tau)H.$$

It remains to apply Lemma 3. □

Let us discuss the computational complexity of Algorithm (9). Implementation of Step 1 needs $O(n^2 \log_2 n)$ operations. Step 2 requires $O(n^3 \log_2 n)$ operations. Step 3 needs $O(n^2)$ operations. And only Step 4, at which we have to compute the spectral radius of a non-negative matrix $(I_n - A(\tau_2))^{-1}H$ needs an iterative procedure, which rate of convergence may depend on the particular data. This is the reason why we analyze in Section 3 a computational method for approaching the spectral radius of a square matrix. If this method is used at Step 4 of Algorithm (9), then the whole procedure will have polynomial-time complexity.

# 3 Computing the largest eigenvalue

Let $A$ be a squared real matrix with spectrum $\Lambda(A)$. One of the most popular procedure for approaching its leading eigenvalue is the Power Method:

$$x_{k+1} \quad = \quad \frac{Ax_k}{\|Ax_k\|}, \quad k \geq 0, \tag{10}$$

where $\|\cdot\|$ is an arbitrary norm for $\mathbb{R}^n$. This method has two important advantages.

- Its iteration is very simple.
- Under some conditions, it has linear rate of convergence.

However, after a close look at this scheme, we can see that these advantages are not very convincing. Indeed, if matrix $A$ is dense, then each iteration of method (10) needs $O(n^2)$ operations. Moreover, the rate of convergence of this method depends on the gap between the magnitudes of the leading eigenvalue and of all others. The smaller is this gap, the slower is the rate of convergence. Hence, for method (10) it is impossible to derive worst-case polynomial-time complexity bounds.

It this section, we present a scheme which has much better theoretical guarantees. It is based on the interpretation of the leading eigenvalue of matrix $A$ as a root of the polynomial $p(\tau) = \det(\tau I_n - A)$.

We need to introduce the following notion.

**Definition 1** *A real polynomial $p$ has a semi-dominant real root $\tau_*$ if $p(\tau_*) = 0$ and*

$$\tau_* \quad \geq \quad \mathrm{Re}\,\lambda, \tag{11}$$

*where $\lambda \in \mathbb{C}$ is any other root of this polynomial.*

**Example 2** 1. Let $A$ be a real symmetric matrix. Then $\lambda_{\max}(A)$ is a a semi-dominant root of the polynomial $p(\tau) = \det(\tau I_n - A)$.
2. By the Perron-Frobenius theorem (see, for instance, [7, chapter 8]), for a non-negative matrix $A$, its spectral radius is a semi-dominant real root of the polynomial $p(\tau) = \det(\tau I_n - A)$.

Our interest to polynomials with semi-dominant real roots can be explained by the following property.

**Lemma 6** *Let monic polynomial $p$ have a semi-dominant real root $\tau_*$. Then the function $p(t)$ is a strictly increasing non-negative convex function on the set $[\tau_*, +\infty)$. Moreover, on this half-line all its derivatives are non-negative and for $\tau \geq \tau_*$ we have*

$$p(\tau) \quad \geq \quad (\tau - \tau_*)^n, \tag{12}$$

$$p(\tau) \quad \geq \quad \tfrac{1}{n} p'(\tau)(\tau - \tau_*). \tag{13}$$

**Proof:**
Denote by $\mathcal{R}(p)$ the set of all real roots of polynomial $p$, and by $\mathcal{C}(p)$ the set of all its complex roots. Further, for a real root $x \in \mathbb{R}$, define function $\xi_x(\tau) = \tau - x$, and for a complex root $\lambda \in \mathbb{C}$ define function $\psi_\lambda(\tau) = (\tau - \mathrm{Re}\ \lambda)^2 + (\mathrm{Im}\ \lambda)^2$. Then

$$p(\tau) \quad = \quad \left( \prod_{x \in \mathcal{R}(p)} \xi_x(\tau) \right) \cdot \left( \prod_{\lambda \in \mathcal{C}(p)} \psi_\lambda(\tau) \right). \tag{14}$$

In view of Definition 1, the polynomial $p$ is a product of functions, for which all derivatives are non-negative on the set $[\tau_*, +\infty)$ (we treat the value of function as derivative of degree zero). Hence, the same is true for the polynomial itself.

Further, for $\tau \geq \tau_*$ we have

$$\xi_x(\tau) \quad \stackrel{(11)}{\geq} \quad \tau - \tau_*, \ x \in \mathcal{R}(p), \quad \psi_\lambda(\tau) \quad \stackrel{(11)}{\geq} \quad (\tau - \tau_*)^2, \ \lambda \in \mathcal{C}(p).$$

Hence, (12) follows from representation (14).

In order to prove inequality (13), note that

$$\tfrac{p'(\tau)(\tau - \tau^*)}{p(\tau)} \quad = \quad \sum_{x \in \mathcal{R}(p)} \tfrac{\tau - \tau^*}{\tau - x} + 2 \sum_{\lambda \in \mathcal{C}(p)} \tfrac{(\tau - \mathrm{Re}\ \lambda)(\tau - \tau_*)}{(\tau - \mathrm{Re}\ \lambda)^2 + (\mathrm{Im}\ \lambda)^2}.$$

In view of condition (11), each term in the above sums is smaller than one. Hence, (13) follows. □

Let us show that the Newton Method is especially efficient in finding the maximal roots of increasing convex univariate functions.

Consider a convex univariate function $f$ such that

$$f(\tau_*) \quad = \quad 0, \quad f(\tau) \ > \ 0, \text{ for } \tau > \tau_*. \tag{15}$$

Let us choose $\tau_0 > \tau_*$. Consider the following Newton process:

$$\tau_{k+1} \quad = \quad \tau_k - \tfrac{f(\tau_k)}{g_k}, \tag{16}$$

where $g_k \in \partial f(\tau_k)$. Thus, we do not assume $f$ to be differentiable for $\tau \geq \tau_*$.

**Theorem 3** *Method (16) is well defined. For any $k \geq 0$ we have*

$$f(\tau_{k+1}) g_{k+1} \quad \leq \quad \tfrac{1}{4} f(\tau_k) g_k. \tag{17}$$

*Thus, $f(x_k) \leq \left(\tfrac{1}{2}\right)^k g_0(\tau_0 - \tau_*)$.*

**Proof:**
Denote $f_k = f(\tau_k)$. Let us assume that $f_k > 0$ for all $k \geq 0$. Since $f$ is convex, $0 = f(\tau_*) \geq f_k + g_k(\tau_* - \tau_k)$. Thus,

$$g_k(\tau_k - \tau_*) \;\geq\; f_k \;>\; 0. \tag{18}$$

This means that $g_k > 0$ and $\tau_{k+1} \in (\tau_*, \tau_k)$. In particular, we conclude that

$$\tau_k - \tau_* \;\leq\; \tau_0 - \tau_*. \tag{19}$$

Further, for any $k \geq 0$ we have:

$$f_k \;\geq\; f_{k+1} + g_{k+1}(\tau_k - \tau_{k+1}) \overset{(16)}{=} f_{k+1} + \tfrac{f_k g_{k+1}}{g_k}.$$

Thus, $1 \geq \frac{f_{k+1}}{f_k} + \frac{g_{k+1}}{g_k} \geq 2\sqrt{\frac{f_{k+1} g_{k+1}}{f_k g_k}}$, and this is (17). Finally, since $f$ is convex, we have

$$g_0 \overset{(18)}{\geq} \sqrt{\tfrac{f_0 g_0}{\tau_0 - \tau_*}} \overset{(17)}{\geq} 2^k \sqrt{\tfrac{f_k g_k}{\tau_0 - \tau_*}} \overset{(18)}{\geq} 2^k \sqrt{\tfrac{f_k^2}{(\tau_0 - \tau_*)(\tau_k - \tau_*)}} \overset{(19)}{\geq} 2^k \tfrac{f_k}{\tau_0 - \tau_*}.$$

□

For a polynomial with semi-dominant root, we can guarantee a linear rate of convergence in the argument.

**Theorem 4** *Let a polynomial $p$ has semi-dominant real root. Then for the sequence $\{\tau_k\}_{k \geq 0}$, generated by method (16) we have*

$$\tau_k - \tau_* \;\leq (1 - \tfrac{1}{n})^k (\tau_0 - \tau_*), \quad k \geq 0. \tag{20}$$

**Proof:**
Indeed, it is enough to combine inequality (13) with the step-size rule of method (16). □

In view of Theorem 4, method (16) can be equipped with a reliable stopping criterion. Indeed, if we need to achieve accuracy $\epsilon > 0$ in the argument, we can use the right-hand side of inequality

$$\tau_k - \tau^* \;\overset{(13)}{\leq}\; \tfrac{n f(\tau_k)}{f'(\tau_k)} \;\leq\; \epsilon \tag{21}$$

as a stopping rule. Since

$$\tfrac{n f(\tau_k)}{f'(\tau_k)} \;\overset{(16)}{=}\; n(\tau_k - \tau_{k+1}) \;\leq\; n(\tau_k - \tau^*) \;\overset{(20)}{\leq}\; n e^{-k/n} (\tau_0 - \tau_*),$$

this criterion will be satisfied after

$$n \left\lceil \ln \tfrac{n(\tau_0 - \tau^*)}{\epsilon} \right\rceil$$

iterations at most.

Thus, we have seen that method (16) has linear rate of convergence, which does not depend on the particular properties of function $f$. Let us show that in non-degenerate situation this method has local quadratic convergence (this never happens with the Power Method (10)).

**Theorem 5** *Let convex function $f$ be twice differentiable. Assume that it satisfies the conditions (15) and its second derivative increases for $\tau \geq \tau_*$. Then for any $k \geq 0$ we have*

$$f(\tau_{k+1}) \quad \leq \quad \frac{f''(\tau_k)}{2(f'(\tau_k))^2} \cdot f^2(\tau_k). \tag{22}$$

*If the root $\tau_*$ is non-degenerate:*

$$f'(\tau_*) \quad > \quad 0, \tag{23}$$

*then $f(\tau_{k+1}) \leq \frac{f''(\tau_0)}{2(f'(\tau_*))^2} \cdot f^2(\tau_k)$.*

**Proof:**
In view of conditions of the theorem, $f''(\tau) \leq f''(\tau_k)$ for all $\tau \in [\tau_{k+1}, \tau_k]$. Therefore,

$$f(\tau_{k+1}) \quad \leq \quad f(\tau_k) + f'(\tau_k)(\tau_{k+1} - \tau_k) + \tfrac{1}{2} f''(\tau_k)(\tau_{k+1} - \tau_k)^2 \quad \stackrel{(16)}{=} \quad \tfrac{1}{2} f''(\tau_k) \tfrac{f^2(\tau_k)}{(f'(\tau_k))^2}.$$

For proving the last statement, it remains to note that $f''(\tau_k) \leq f''(\tau_0)$ and $f'(\tau_k) \geq f'(\tau_*)$. □

**Corollary 2** *If $f$ is a monic polynomial of degree $n$ with real roots, then*

$$f(\tau_{k+1}) \quad \leq \quad \tfrac{n-1}{2n} f(\tau_k), \quad k \geq 0. \tag{24}$$

**Proof:**
Indeed, in this case $f(t) = \prod_{i=1}^{n} (t - x_i)$ with $x_i \in \mathbb{R}$, $i = 1, \ldots, n$. Therefore,

$$f'(t) \quad = \quad f(t) \sum_{i=1}^{n} \tfrac{1}{t - x_i},$$

$$f''(t) \quad = \quad f(t) \left[ \left( \sum_{i=1}^{n} \tfrac{1}{t - x_i} \right)^2 - \sum_{i=1}^{n} \tfrac{1}{(t - x_i)^2} \right] \quad \leq \quad \left(1 - \tfrac{1}{n}\right) f(t) \left( \sum_{i=1}^{n} \tfrac{1}{t - x_i} \right)^2.$$

It remains to use inequality (22). □

Note that both Theorems 3 and 5 are applicable to our main object of interest, the polynomial $p(\tau) = \det(\tau I_n - A)$, where $A$ is non-negative $n \times n$-matrix. However, the direct application of method (16) to this polynomial is expensive since at each iteration we need to compute a determinant of $n \times n$-matrix. This computation needs $O(n^3)$ operations. However, we can significantly reduce this cost by transforming matrix $A$ in a special *Hessenberg form*.

Recall that matrix $A$ has a lower Hessenberg form if

$$A^{(i,j)} \quad = \quad 0, \quad \forall j \geq i + 2, \quad i, j = 1, \ldots, n.$$

Thus, it has the following structure:

$$A_n(a, b, L) = \left( \begin{array}{c|c} a & \frac{L}{b^T} \end{array} \right),$$

where $a \in \mathbb{R}^n$, $b \in \mathbb{R}^{n-1}$, and $L$ is a lower-triangular $(n-1) \times (n-1)$-matrix. Any matrix can be represented in this form by transformation

$$A \quad \rightarrow \quad U^T A U,$$

where $U \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. This transformation is standard and it can be computed in $O(n^3)$ operations. At the same time, it does not change the polynomial $p(\tau) = \det(\tau I_n - A)$. Hence, let us assume that we already have matrix $A$ in the lower Hessenberg form.

In this case, all matrices $B(\tau) \stackrel{\text{def}}{=} \tau I_n - A$ have also the Hessenberg structure. Let us show that their determinants can be easily computed.

**Lemma 7** *Let matrix $B \in \mathbb{R}^{n \times n}$ have a lower Hessenberg form:*

$$B = \left( \begin{array}{c|c|c} \alpha & \beta & 0 \dots 0 \\ \hline a_1 & a_2 & \begin{array}{c} L \\ b^T \end{array} \end{array} \right),$$

*where $a_1, a_2 \in \mathbb{R}^{n-1}$, $L$ is a lower-triangular $(n-2) \times (n-2)$-matrix, and $b \in \mathbb{R}^{n-2}$ Then*

$$\det B \quad = \quad \det A_{n-1}(\alpha a_2 - \beta a_1, b, L). \tag{25}$$

**Proof:**
For $x \in \mathbb{R}^{n-1}$, consider the function $d(x) = \det A_{n-1}(x, b, L)$. Note that this function is linear in $x$. Therefore, by applying Laplace formula to the first row of matrix $B$, we get $\det B = \alpha d(a_2) - \beta d(a_1) = d(\alpha a_2 - \beta a_1)$. □

Thus, using the recursion (25), the value of polynomial $p(\tau) = \det B(\tau)$ can be computed in $\sum\limits_{k=1}^{n-1} 2(n-k) = n(n-1)$ multiplications. Clearly, its derivative can be also computed in $O(n^2)$ operations.

Note that the above procedure has a hidden drawback. Indeed, for a high dimension we can expect the value of polynomial $p(\tau) = \det(\tau I_n - A)$ to be very big. Therefore, the computation of its value and its derivative is computationally unstable. However, note that in the Newton method

$$\tau_{k+1} \quad = \quad \tau_k - \frac{p(\tau_k)}{p'(\tau_k)}, \quad k \geq 0, \tag{26}$$

the step size is given by a *ratio of polynomials*, which can be computed in a stable way.

Indeed, let us assume that our polynomial is represented in a multiplicative form: $p(\tau) = \prod\limits_{k=1}^{m} f_k(\tau)$, where $f_k$ are some functions defined in a neighborhood of $\tau \in \mathbb{R}$. Then

$$\frac{p'(\tau)}{p(\tau)} \quad = \quad \sum_{k=1}^{m} \frac{f_k'(\tau)}{f_k(\tau)}. \tag{27}$$

Thus, any multiplicative representation of polynomial $p$ allows a direct computation of the Newton step in an additive form, which is much more stable. Let us show how this representation can be computed for a Hessenberg matrix.

Our procedure is based on the recursion described in Lemma 7. However, we introduce in it some *scaling functions*, which prevent the growth of intermediate coefficients.

We generate a sequence of Hessenberg matrices $H_k$ of decreasing dimension. Let us choose $\tau_0 \in \mathbb{R}$ and define $H_0(\tau) = \tau I_n - A$. At iteration $k$, we assume that our matrix has the following structure:

$$H_k(\tau) = \left( \begin{array}{c|c|c} \alpha_k(\tau) & \beta_k(\tau) & 0 \ldots 0 \\ \hline a_k(\tau) & b_k(\tau) & \begin{array}{c} L_k(\tau) \\ c_k^T(\tau) \end{array} \end{array} \right) \in \mathbb{R}^{(n-k)\times(n-k)},$$

where $a_k(\tau), b_k(\tau) \in \mathbb{R}^{n-k-1}$, $L_k(\tau)$ is a lower-triangular $(n-k-2)\times(n-k-2)$-matrix, and $c_k(\tau) \in \mathbb{R}^{n-k-2}$. Let us define an arbitrary function of two variable $f_k(\alpha, \beta)$, which is analytic in the neighborhood of point $(\alpha_k(\tau_0), \beta_k(\tau_0))^T \in \mathbb{R}^2$. Then, for the next iteration we define

$$H_{k+1}(\tau) \;=\; A_{n-k-1}\left(\tfrac{\alpha_k(\tau)b_k(\tau)-\beta_k(\tau)a_k(\tau)}{f_k(\alpha_k(\tau),\beta_k(\tau))}, c_k(\tau), L_k(\tau)\right).$$

In this process, the last generated matrix will be $H_{n-2}(\tau) \in \mathbb{R}^{2\times 2}$. At this moment, we define

$$f_{n-2}(\tau) \;=\; \alpha_{n-2}(\tau)b_{n-2}(\tau) - \beta_{n-2}(\tau)a_{n-2}(\tau)$$

(in this case, $a_{n-2}(\tau)$ and $b_{n-2}(\tau)$ are real values). Under this convention, by Lemma 7 we have $p(\tau) = \prod_{k=0}^{n-2} f_k(\tau)$.

In the above process, it is reasonable to choose functions $f_k$, $0 \le k \le n-3$, in the simplest form. For example, they could be linear functions of two variables with coefficients $\pm 1$, ensuring the condition

$$f_k(\tau_0) \;=\; |\alpha_k(\tau_0)| + |\beta_k(\tau_0)|.$$

In this case, all matrices $H_k(\tau)$ will be some rational functions of $\tau$, well defined in a neighborhood of $\tau_0$ (since in the process (26) we cannot have $|\alpha_k(\tau_0)| + |\beta_k(\tau_0)| = 0$). Therefore, the derivatives of functions $f_k$ at $\tau_0$ can be easily computed by forward differentiation of the recursion formulas. The total complexity of this process will be of the order $O(n^2)$.

# 4 Problems with distances measured in $\ell_\infty$- and $\ell_1$-norms

As we know, the $\ell_\infty$-norm $\|X\|_\infty = \max\limits_{1\le i\le n} \sum_{j=1}^n |X^{(i,j)}|$ is dual to the $\ell_1$-norm $\|X\|_1 = \max\limits_{1\le j\le n} \sum_{i=1}^n |X^{(i,j)}|$, in particular, $\|X\|_\infty = \|X^T\|_1$. Therefore, if $X$ is a closest stable matrix for $A$ in the $\ell_\infty$-norm, then $X^T$ is a closest stable matrix for $A^T$ in the $\ell_1$-norm. Thus, the problem in the $\ell_1$-norm is equivalent to the same problem in $\ell_\infty$-norm with replacement of rows by columns. Therefore, we will deal with the $\ell_\infty$-norm only Thus, for a non-negative $n \times n$ matrix $A$, we consider the following two problems.

1) If $\rho(A) < 1$, then we find the closest unstable matrix:

$$\|X - A\|_\infty \;\to\; \min : \quad \rho(X) \;=\; 1,\; X \;\ge\; 0\,. \tag{28}$$

2) If $\rho(A) > 1$, then we find the closest stable matrix. Its mathematical formulation is the same as (28)

We will solve problems 1) and 2) by applying the technique of optimizing the spectral radius over product families of matrices with row uncertainties. This is possible since any ball in the space of matrices equipped with the $\ell_\infty$-norm forms a product family. For implementing this strategy, we will develop a greedy spectral simplex method, which minimizes the spectral radius over the matrix sets with polyhedral row uncertainties. All necessary definitions will be given later. Now we need to prove several auxiliary results on the spectral radius of non-negative matrices.

## 4.1 Some inequalities for spectral radius

**Lemma 8** *Let $A \in \mathbb{R}^{n\times n}_+$, $u \geq 0$, be a vector and $\lambda \geq 0$ be a real number. Then $Au \geq \lambda u$ implies that $\rho(A) \geq \lambda$. If for a strictly positive vector $v$ we have $Av \leq \lambda v$, then $\rho(A) \leq \lambda$.*

**Proof:**
If $Au \geq \lambda u$, then $A^k u \geq \lambda^k u$ for each $k$. Therefore, $\|A^k\| \geq \lambda^k$ for all $k$, and so $\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k} \geq \lambda$. The second statement is a simple consequence of the representation (2). □

**Corollary 3** *Let $A \in \mathbb{R}^{n\times n}_+$, $u \geq 0$ be a vector, and $\lambda \geq 0$ be a real number. If $Au > \lambda u$, then $\rho(A) > \lambda$. If $Au < \lambda u$, then $\rho(A) < \lambda$.*

For formulating the next auxiliary result, let us recall that the active set of a vector inequality $x \geq y$ is the set of indices for which this becomes an equality: $\mathcal{I} = \{i \mid x^{(i)} = y^{(i)}\}$.

**Lemma 9** *Let $A \in \mathbb{R}^{n\times n}_+$, $\rho(A) = 1$, and $u \in \mathbb{R}^n$ be a strictly positive vector such that $Au \leq u$ (or $Au \geq u$). Let $\mathcal{I}$ be the active set of this inequality. Then there is a nonempty subset $\mathcal{I}' \subset \mathcal{I}$ such that the subspace $V = V_{\mathcal{I}'}$ is invariant for $A^T$ and $\rho(A^T|_V) = 1$.*

**Proof:**
Consider the case $Au \leq u$ (the proof for $Au \geq u$ is literally the same). Denote $\mathcal{B} = \{b_i = A^{(i)} - e_i \mid i \in \mathcal{I}\}$, where $A^{(i)}$ is the $i$th row of $A$. If the system of inequalities $\langle b_i, h\rangle < 0$, $b_i \in \mathcal{B}$, has a solution $h \in \mathbb{R}^n$, then for all sufficiently small numbers $t > 0$, we have $A(u + th) < u + th$. Therefore, by Corollary 3, $\rho(A) < 1$, which contradicts to the assumption. Hence, this system does not have a solution, which by Farkas lemma [17] implies that $0 \in \mathrm{Conv}\,\{\mathcal{B}\}$. So, $\sum\limits_{i\in\mathcal{I}} \tau_i(A^{(i)} - e_i) = 0$ for some numbers $\tau_i \geq 0$, $\sum\limits_i \tau_i = 1$. If $v$ is the vector from $\mathbb{R}^n$ such that $v^{(i)} = \tau_i$ for $i \in \mathcal{I}$, and $v^{(i)} = 0$ otherwise, then $A^T v = v$ and $\mathrm{supp}\,(v) \subset \mathcal{I}$. Then the subspace $V_{\mathcal{I}'}$ with $\mathcal{I}' = \mathrm{supp}\,(v)$ is invariant for the matrix $A^T$ and $\rho(A^T|_{\mathcal{I}'}) \geq 1$ (Lemma 8). On the other hand, $\rho(A^T|_{\mathcal{I}'}) \leq \rho(A^T) = 1$, and therefore $\rho(A^T|_{\mathcal{I}'}) = 1$. □

## 4.2 Optimizing the spectral radius over product families

Consider one important class of matrices for which the problem of optimizing the spectral radius admits an efficient solution.

**Definition 2** *A family $\mathcal{F}$ of non-negative $n\times n$-matrices is called a product family if there exist compact sets $\mathcal{F}^{(i)} \subset \mathbb{R}^n_+$, $i = 1,\dots,n$, such that $\mathcal{F}$ consists of all possible matrices with $i$-th row from $\mathcal{F}^{(i)}$, for all $i = 1,\dots,n$.*

The sets $\mathcal{F}^{(i)}$ are called the *uncertainty sets*. Thus, product families are sets of matrices with independent row uncertainties: their rows are independently chosen from the sets $\mathcal{F}^{(i)}$. Topologically, they are indeed products of the uncertainty sets: $\mathcal{F} = \mathcal{F}^{(1)} \times \cdots \times \mathcal{F}^{(n)}$. Such families have been studied in the literature due to applications in spectral graph theory, asynchronous systems, mathematical economics, population dynamics, etc. (see [1, 3, 9, 13, 14, 19] and the references therein).

Product families have many remarkable properties. In particular, their joint and lower spectral radii are always attained at one matrix [1]. Moreover, the problems of minimizing and maximizing the spectral radius of a matrix over some compact set of matrices, being notoriously hard in general, becomes efficiently solvable over product sets. Recent paper [14] develops such methods in the case of *polyhedral* uncertainly sets, when each $\mathcal{F}^{(i)}$ is either a polytope given by vertices or a polyhedron given by a system of linear inequalities. Our crucial observation is

*For each $A \in \mathbb{R}^{n\times n}_+$ and for each $\tau > 0$, the set $\mathcal{B}_\tau(A) = \{X \in \mathbb{R}^{n\times n}_+ \mid \|X-A\|_\infty \le \tau\}$ is a product family with polyhedral row uncertainty sets.*

Thus, the positive part of any $\ell_\infty$-ball $\mathcal{B}_\tau(A)$ is a product family. Therefore, using methods of optimizing the spectral radius over product families, one can solve the problem $\rho(X) \to \min / \max$ over the set $X \in \mathcal{B}_\tau(A)$ and then try to adapt $\tau$ by a bisection procedure. The minimal $\tau$ such that $\min\limits_{X\in\mathcal{B}_\tau(A)} \rho(A) \le 1$, is the distance to the closest stable matrix, the minimal $\tau$ such that $\max\limits_{X\in\mathcal{B}_\tau(A)} \rho(A) \ge 1$, is the distance to the closest unstable matrix. For implementing this strategy, we modify some methods from [14] and [16] and apply them to the specific polyhedral uncertainly sets

$$\mathcal{F}^{(i)} = \mathcal{B}_\tau(A^{(i)}) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : \ \|x - A^{(i)}\|_1 \le \tau\}, \quad i = 1,\dots,n,$$

Our methods for optimizing the spectral radius over product families are based on the following simple fact. Let $A$ be a matrix from product family $\mathcal{F}$, and $v \in \mathbb{R}^n_+$ be its leading eigenvector. We say that $A$ is *minimal in each row* (with respect to $v$) if $\langle v, A^{(i)}\rangle = \min\limits_{x\in\mathcal{F}^{(i)}} \langle v, x\rangle$ for all $i = 1,\dots,n$. Similar definition is used for maximality in each row.

**Proposition 1** *Suppose $A$ belongs to a product family $\mathcal{F}$ and $v \in \mathbb{R}^n_+$ be its leading eigenvector. Then*

*1) if $A$ is minimal in each row with respect to $v$, then $\rho(A) = \min\limits_{X\in\mathcal{F}} \rho(X)$.*

*2) if $v > 0$ and $A$ is maximal in each row with respect to $v$, then $\rho(A) = \max\limits_{X\in\mathcal{F}} \rho(X)$.*

**Proof:**
The statement directly follows from Lemma 8. □

Thus, if a matrix from product family is optimal in each row, then it provides the global optimum for the spectral radius. For strictly positive matrices, the converse is also true. Indeed, applying Lemma 9, we obtain

**Corollary 4** *If matrix $A \in \mathcal{F}$ is strictly positive, then it has the minimal spectral radius in $\mathcal{F}$ precisely when $A$ is minimal in each row with respect to its (unique) leading eigenvector. The same is true for maximumization.*

However, if $A$ has some zero entries, then the converse to Proposition 1 may fail. Not every matrix from $\mathcal{F}$ with the minimal (maximal) spectral radius is minimal (respectively, maximal) in each row. Nevertheless, at least one matrix with this property always exists as the following proposition states.

**Proposition 2** *In every product family, there exists a matrix, which is minimal (maximal) in each row with respect to one of its leading eigenvectors.*

**Proof:**
For a given $\varepsilon > 0$ consider $\varepsilon$-shifted uncertainty sets $\mathcal{F}_\varepsilon^{(i)} = \mathcal{F}^{(i)} + \varepsilon e$ and the correspoding product family $\mathcal{F}_\varepsilon = \mathcal{F}_\varepsilon^{(1)} \times \cdots \times \mathcal{F}_\varepsilon^{(n)} = \mathcal{F} + \varepsilon J_n$. Let $A_\varepsilon \in \mathcal{F}_\varepsilon$ be the matrix with the minimal spectral radius. Since matrix $A_\varepsilon$ is strictly positive, Corollary 4 implies that $A_\varepsilon$ is minimal in each row. To any $\varepsilon$ we associate one of such matrices $A_\varepsilon$. By compactness, there is a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ such that $\varepsilon_k \to 0$ as $k \to \infty$, the matrices $A_{\varepsilon_k}$ converge to a matrix $A \in \mathcal{F}$ and their leading eigenvectors converge to a nonzero vector $v$. Then by continuity, $v$ is an eigenvector of $A$, and $A$ is minimal in each row with respect to $v$.
The proof for maximization is the same. □

Propositions 1 and 2 offer a method for optimizing the spectral radius over the product families by finding the optimal matrices in each row. This strategy was used in [14] for developing two optimization algorithms. One of them is the *spectral simplex method*. It consists in consecutive increasing of the spectral radius by one-row corrections of a matrix. The main idea is the following. We take a matrix $A$ from a product family $\mathcal{F}$ and compute its leading eigenvector $v$. Then, for each $i = 1, \ldots, n$, we try to maximize the scalar product of $v$ with rows from the uncertainty set $\mathcal{F}^{(i)}$. If for all $i$, the maximums are attained at the rows of $A$, then $A$ is maximal in each row and hence has the maximal spectral radius in $\mathcal{F}$. Otherwise, we replace one row of $A$, say, the $i$th one, with the row from $\mathcal{F}^{(i)}$ maximizing the scalar product. We obtain a new matrix. We compute its leading eigenvector, optimize the scalar products of rows with this eigenvector, etc.

The advantage of this method is that it is applicable for both maximizing and minimizing the spectral radius. However, its significant shortcoming is that it works efficiently only for strictly positive matrices. If some row from $\mathcal{F}^{(i)}$ has a zero entry, then the algorithm may cycle. Even if it does not cycle, the terminal matrix may not provide a solution. The idea of making all matrices positive by slight perturbations may cause instability, which is difficult to control. In high dimensions, even a very small perturbation

of coefficients may significantly change the spectral radius (see, for instance, [18] for the corresponding analysis). The modified spectral simplex method which avoids these troubles and is applicable for all non-negative matrices, was developed in [16]. The spectral simplex method demonstrates its exceptional efficiency even for matrices of relatively big size. In this paper, we present another modification, the *greedy spectral simplex method*, which speed up the convergence rate in the case of simply structured uncertainty sets, when the minimization problem $\langle v, x\rangle \to \min,\ x \in \mathcal{F}^{(i)}$, can be easily solved. We are going to show in Section 6 that the $\ell_\infty$-balls $B^{(i)}(\tau)$ possess this property.

## 4.3 Closest unstable matrix

For an arbitrary non-negative $n \times n$-matrix $A$ with $\rho(A) < 1$ we consider the problem of finding a closest unstable matrix to $A$ in the operator $\ell_\infty$-norm:

$$\|X - A\|_\infty \ \to\ \min : \quad \rho(X) = 1. \tag{29}$$

Denote the minimal norm in problem (29) by $\tau_*$. It is shown easily that the closest unstable matrix $X$ is non-negative and, moreover, it is elementwise bigger than or equal to the matrix $A$. So, the matrix $X - A$ is non-negative and has the sum of elements in each row at most $\tau_*$. Thus, for an arbitrary matrix $A \geq 0$ with $\rho(A) < 1$, the general problem (29) is equivalent to the following

$$\|X - A\|_\infty \ \to\ \min : \quad X \geq A\,,\ \rho(X) \ = \ 1. \tag{30}$$

For characterizing the optimal solution $X$, we introduce notation $E_k = e\, e_k^T$ for the matrix with $k$th column composed by ones and all other elements being zeros.

**Theorem 6** *Optimal value $\tau_*$ of problem (29) is reciprocal to the biggest component of the vector $(I - A)^{-1}e$. Let $k$ be the index of this component. Then the optimal solution of this problem is the matrix*

$$X_* \ = \ A \ + \ \tau_*\, E_k. \tag{31}$$

**Remark 2** *The main conclusion of Theorem 6 can be formulated as follows: if we want to increase the spectral radius of matrix $A$ as much as possible, having the sum of changes of entries in each row do not exceeding a fixed number $\tau > 0$, then we have to change all entries in one column (add $\tau$ to each entry). This "steepest growth" column of $A$ corresponds to the biggest component of the vector $(I - A)^{-1}e$.*

**Proof:**
The optimal matrix $X_*$ in problem (30) is also a solution to the maximization problem

$$\rho(X) \ \to\ \max : \quad \|X - A\|_\infty \ \leq\ \tau,\ X \ \geq\ A \tag{32}$$

for $\tau = \tau_*$. Let us characterize this matrix for arbitrary $\tau$. By Propositions 1 and 2, $X$ is maximal in each row for the product family with the uncertainty sets

$$\begin{aligned} \mathcal{B}_\tau^+(A) \ &= \ \mathcal{B}_\tau(A) \cap \{X \in \mathbb{R}^{n\times n} \mid X \geq A\} \\ &= \ \Big\{X \in \mathbb{R}^{n\times n} :\ X \geq A,\ \langle (X - A)e, e_i\rangle \leq \tau,\ i = 1, \ldots, n\Big\}. \end{aligned}$$

Conversely, every matrix $X$, which is maximal in each row with respect to a strictly positive leading eigenvector and such that $\rho(X) = 1$, is the closest unstable matrix for $A$.

Any matrix $X \in \mathcal{B}_\tau^+(A)$ with leading eigenvector $v$ is optimal in the $i$th row if and only if the scalar product $\langle X^{(i)} - A^{(i)}, v\rangle$ is maximal under the condition $\langle X^{(i)} - A^{(i)}, e\rangle = \tau$. This maximum is equal to $r\,\tau$, where $r$ is the maximal component of vector $v$. Denote the index of this component by $k$: $v^{(k)} = r$. Then

$$X^{(i)} \;-\; A^{(i)} \;=\; \tau\, e_k\,, \qquad i = 1, \ldots, n.$$

Hence, if $X$ is maximal in each row, then $X = A + \tau\, e\, e_k^T = A + \tau E_k$. Furthermore, since each set $\mathcal{B}_\tau^+(A^{(i)})$ contains a strictly positive point, it follows that

$$v^{(i)} \;=\; (Xv)^{(i)} = \max_{x \in \mathcal{B}_\tau(A^{(i)})} \langle x, v\rangle \;>\; 0.$$

Hence, vector $v$ is strictly positive, and by Proposition 1, the matrix $X$ has the biggest spectral radius on the set $\mathcal{B}_\tau^+(A^{(i)})$.

Thus, the optimal matrix has the form (31) for some $k$ and $\|X - A\|_\infty = \tau_*$. It remains to find $k \in \{1, \ldots, n\}$ for which the value of $\tau$ is minimal. Since $\rho(A + \tau_*\, E_k) = 1$, it follows that $\tau_*$ is the smallest positive root of the equation

$$\det\,(A \;-\; I \;+\; \tau\, E_k) \;=\; 0\,. \tag{33}$$

Since $\rho(A) < 1$, we have $(I - A)^{-1} \;=\; \sum\limits_{j=0}^{\infty} A^j \;\geq\; 0$. Multiplying equation (33) by $\det(-\,(I - A)^{-1})$, we obtain

$$\det\,(I \;-\; \tau\,(I - A)^{-1}\, E_k) \;=\; 0\,. \tag{34}$$

The matrix $\tau\,(I - A)^{-1}\, E_k$ has only one nonzero column. This is the $k$th column equal to $(I - A)^{-1} e$. Hence

$$\det\,(I \;-\; \tau\,(I - A)^{-1}\, E_k) \quad = \quad 1 \;-\; \tau\,[(I - A)^{-1} e]^{(k)}\,.$$

Thus, $\tau_*$ is the reciprocal to the $k$th component of the vector $(I - A)^{-1} e$. Hence the minimal $\tau$ corresponds to the biggest component of this vector. □

**Remark 3** *Vector $x = (I - A)^{-1} e$ needed in Theorem 2 can be found by solving the linear system $(I - A)x = e$. It suffices to find an approximate solution, because we actually need only the index of the biggest component of $x$. This can be done by the Power Method. Indeed, since $(I - A)^{-1} e \;=\; \sum\limits_{j=0}^{\infty} A^j e$, we see that the vector $(I - A)^{-1} e$ is the limit of the following recursive sequence: $x_0 = e$, $x_{j+1} = A\, x_j \;+\; e$, $j = 0, 1, \ldots$. This Power Method converges with the linear rate $O(\rho^k(A))$. Having an approximate value of the limiting vector, we can find $k$ as the index of its biggest component and $\tau_*$ as a reciprocal to this component. After that, the closest unstable matrix $X_*$ can be approximated by the formula (31).*

## 4.4 Closest stable matrix

For an arbitrary non-negative $n \times n$-matrix $A$ with $\rho(A) > 1$, consider the problem of finding a closest in the operator $\ell_\infty$-norm matrix to $A$, which has the spectral radius equal to one. Thus, we consider the same formulation (29), but for the case $\rho(A) > 1$. This problem can be written as follows

$$\|X - A\|_\infty \;\to\; \min : \quad X \geq 0\,,\; \rho(X) \;=\; 1. \tag{35}$$

This case is more difficult than finding the closest unstable matrix because now we have to respect the non-negativity conditions for the matrix $X$, which were actually redundant in the former case, but now becomes a serious restriction. That is why the optimal solution $X$ is usually found not by a formula but by an iterative procedure. The main idea is to solve the related problem

$$\left\{ \begin{array}{l} \rho(X) \;\to\; \min : \quad \|X - A\|_\infty \;\leq\; \tau \\ 0 \;\leq\; X \;\leq\; A, \end{array} \right. \tag{36}$$

and then apply a bisection in $\tau$ for finding the value of the parameter ensuring $\rho(X) = 1$. In fact, the algorithm works much faster by using a kind of mixed strategy. After several iterations of the bisection we can find $\tau$ by an explicit formula (see Section 6 for details).

Our main goal now is to solve (36) for a particular $\tau$. For this we develop a *greedy spectral simplex method*, which is a natural extension of the spectral simplex method presented and studied in [14, 16]. Let us start with some notation and auxiliary results.

Matrix $A \geq 0$ is called *irreducible* if it does not have a nontrivial invariant coordinate subspace, i.e., a subspace spanned by some elements $e_i$ of the canonical basis. A matrix is irreducible if and only if its graph $G$ is strongly connected, i.e., for every pair of vertices $i, j$, there is a path from $i$ to $j$.

Reducibility means that there is a proper nonempty subset $\Lambda \subset \Omega$ such that for each $i \in \Lambda$, the support of the $i$th column of $A$ is contained in $\Lambda$.

For every matrix $A \geq 0$, there exists a suitable permutation $P$ of the basis of $\mathbb{R}^n$, after which $A$ gets a block upper-triangular form with $r \geq 1$ diagonal blocks $A_j$ of sizes $d_j$, $j = 1, \ldots, r$, called the *Frobenius factorization*:

$$P^{-1}AP \quad = \quad \left( \begin{array}{cccc} A_1 & * & \ldots & * \\ 0 & A_2 & * & \vdots \\ \vdots & & \ddots & * \\ 0 & \ldots & 0 & A_r \end{array} \right). \tag{37}$$

For each $j = 1, \ldots, r$, the matrix $A_j$ in the $j$th diagonal block is irreducible. Any non-negative matrix possesses a unique Frobenius factorization up to a permutation of blocks.

The following fact of the Perron-Frobenius theory is well-known (e.g. [7, chapter 8]).

**Lemma 10** *An irreducible matrix has a simple leading eigenvalue.*

The converse is not true: a matrix with a simple leading eigenvalue can be reducible.

Let $A$ be $n \times n$ non-negative matrix. Its leading eigenvector $v$ is called *minimal* if there is no other leading eigenvector that possesses a strictly smaller (by inclusion) support. A minimal leading eigenvector can be found by Frobenius factorization (37).

The case of strictly positive leading eigenvector, when $v$ possesses a full support, is characterized by the following statement.

**Proposition 3** *If a non-negative $n \times n$ matrix $A$ has a strictly positive minimal leading eigenvector $v$, then the leading eigenvalue $\lambda_{\max}$ is simple and there exists a permutation $P$ of the basic vectors such that $A$ gets the block upper triangular form*

$$P^{-1}AP \quad = \quad \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}, \tag{38}$$

*where $B$ and $C$ are square matrices such that $C$ is irreducible with $\rho(C) = \lambda_{\max}$, and $\rho(B) < \lambda_{\max}$ (the block $B$ may be empty, in which case $P = I_n$ and $C = A$).*

**Proof:**
Without loss of generality it can be assumed that $\lambda_{\max} = 1$. Since $A^k v = v$ for all $k$, and $v$ is positive, it follows that the sequence $\|A^k\|$, $k \in \mathbb{N}$, is bounded and hence the eigenvalue 1 has only one-element Jordan blocks. If there are at least two of those blocks, then $A$ has at least two leading eigenvectors $v_1$ and $v_2$. Denoting $\alpha = \min \left\{ \frac{v_1^{(i)}}{v_2^{(i)}} \;\middle|\; v_2^{(i)} > 0 \right\}$ we see that $v_1 - \alpha v_2$ is a leading eigenvector, which has a zero component. This contradicts to the minimality of $v$. Therefore, the leading eigenvalue has a unique one-elements Jordan block, i.e., it is simple. Further, consider the Frobenius factorization of $A$ generated by a suitable permutation of the basis vectors $P$. In this factorization, matrix $P^{-1}AP$ has an upper-triangular block form with irreducible blocks. Since, $\lambda_{\max}$ is simple, there exists a unique block with this leading eigenvalue. Since the leading eigenvector of $A$ is strictly positive, it follows that this block takes the last position in the diagonal (i.e. in the lower right corner of the matrix). It remains to denote this block by $C$ and the union of all other blocks by $B$. □

The basis vectors corresponding to the block $C$ in factorization (38) span an invariant coordinate subspace of matrix $A^T$, on which this matrix is irreducible with spectral radius equal to $\lambda_{\max}$. Thus, we obtain the following consequence.

**Corollary 5** *If a non-negative $n \times n$-matrix $A$ has a minimal leading eigenvector $v > 0$, then there exists a unique nonempty subset $\mathcal{H} \subset \Omega$ such that $V_{\mathcal{H}}$ is an invariant subspace of $A^T$ on which this matrix is irreducible and has the spectral radius equal to $\rho(A)$.*

We call the subset $\mathcal{H} \subset \Omega$ from Corollary 5 the *basic set* of the matrix $A$ and $V_{\mathcal{H}}$ the basic subspace. Thus, matrix with a strictly positive minimal leading eigenvector possesses a unique basic set. By Proposition 3, the permutation $P$ maps the set $\{n - |\mathcal{H}| + 1, \ldots, n\}$ to the set $\mathcal{H}$.

# 5 Greedy spectral simplex method

For every $\tau > 0$, problem (36) can be soled by the greedy spectral simplex method presented in this section. We describe and analyse the algorithm for a more general

problem of minimizing the spectral radius over a product family $\mathcal{F} = \mathcal{F}^{(1)} \times \cdots \times \mathcal{F}^{(n)}$ with arbitrary polyhedral uncertainty sets $\mathcal{F}^{(i)} \subset \mathbb{R}^n_+$:

$$\rho(X) \ \to \ \min : \quad X \in \mathcal{F}. \tag{39}$$

For finding the closest stable matrix, we set $\mathcal{F}^{(i)} = \mathcal{B}^-_\tau(A^{(i)}) = \mathcal{B}_\tau(A^{(i)}) \cap \mathbb{R}^{n\times n}_+$ and obtain problem (36).

The idea of the greedy spectral simplex method naturally follows form Propositions 1 and 2. Let us take arbitrary matrix $X_0 \in \mathcal{F}$ and start the iterative scheme. In the beginning of $k$th iteration, $k \geq 0$, we have a matrix $X_k$. Let us find its leading eigenvector $v_k$ and for every $i = 1, \ldots, n$, solve the problem $\langle x, v_k\rangle \to \min\limits_{x\in\mathcal{F}^{(i)}}$. This can be done using the standard linear programming technique. In particular, the solution $x$ is always attained at a vertex of the polyherdon $\mathcal{F}^{(i)}$. Denote this solution (vertex of $\mathcal{F}^{(i)}$) by $X^{(i)}_{k+1}$ and compose the next matrix $A_{k+1}$ by the optimal rows $X^{(i)}_{k+1}$, $i = 1, \ldots, n$. Then compute the leading eigenvector of the new matrix, do the next iteration, etc. The algorithm terminates when the matrix $X_k$ is optimal in each row. In this case, we can set $X_{N+1} = X_N$. By Proposition 1, $X_N$ provides a global minimum to the problem (39).

Applying Corollary 4, we come to the following conclusion:

**Corollary 6** *If all the uncertainty sets $\mathcal{F}^{(j)}$ are strictly positive, then the spectral radius $\rho(X_k)$ of the sequence of matrices, arising in the greedy spectral simplex method, decreases in $k$. In particular, the algorithms never cycles.*

On the other hand, since each row $X^{(i)}_k$ is a vertex of the polyhedron $\mathcal{F}^{(j)}$, the total number of states is finite. Hence the algorithm finds the global minimum in a finite number of iterations. Thus, we have proved the following

**Theorem 7** *If all the uncertainty sets $\mathcal{F}^{(j)}$ are strictly positive, then the greedy spectral simplex method finds the optimal solution in finite number of iterations.*

However, if some vectors from $\mathcal{F}^{(j)}$ have zero entries, then the spectral radius $\rho(X_k)$ may not be strictly decreasing in $k$. In this case, $\rho(X_k)$ may stay unchanged for many iterations and the algorithm may cycle [16]. Moreover, without the positivity assumption, matrices $X_k$ may have multiple leading eigenvalues, which complicates their computation and causes an uncertainty in choosing the leading eigenvector $v_k$ from the corresponding root subspace. This is the reason why the greedy spectral simplex method needs to be modified for avoiding these drawbacks. We present below its modified version, which works efficiently for all non-negative polyhedral uncertainty sets including the case of sparse matrices.

**Algorithm 1**

.

`Initialization.` Let $\mathcal{F}^{(i)} \subset \mathbb{R}^n_+$, $i = 1, \ldots, n$, be the polyhedral uncertainty sets, and $\mathcal{F} = \mathcal{F}^{(1)} \times \cdots \times \mathcal{F}^{(n)}$ is the corresponding matrix family. Each $\mathcal{F}^{(i)}$ is given either by a finite set of vertices or by a system of linear inequalities. Choose arbitrary $X_1 \in \mathcal{F}$.

(*) $k$`th iteration.` For a non-negative $n \times n$-matrix $X_k$, compute its minimal leading eigenvector $v$ (take any of them, if there are several ones), set $\mathcal{S} = \mathcal{S}_k = \operatorname{supp}(v)$, $X = X_k|_{\mathcal{S}}$, and go to (**).

(**) `Main loop.` We have a set $\mathcal{S} \subset \Omega$, a square non-negative matrix $X$ on this set, and its minimal leading eigenvector $v > 0$. Denote by $\mathcal{H}$ the basic set of $X$. For each $i \in \mathcal{S}$ solve the problem

$$\langle x, v\rangle \ \to \ \min : \quad x \ \in \ \mathcal{F}^{(i)}. \tag{40}$$

Denote by $\mathcal{I}$ the set of indices $i$ such that the $i$th row of matrix $X$ provides the global minimum for this problem: $\mathcal{I} = \{i \in \mathcal{S} : \ \langle X^{(i)}, v\rangle = \min\limits_{x \in \mathcal{F}^{(i)}} \langle x, v\rangle\}$.

**If** $\mathcal{I} = \mathcal{S}$, then $\rho(X) = \min\limits_{Y \in \mathcal{F}} \rho(Y)$, and we **STOP** Algorithm 1. Otherwise, define the next matrix $X'$ as follows:

$$X^{(i)\prime} \quad = \quad \begin{cases} X^{(i)} & , \quad i \ \in \mathcal{I} \\ \arg\min\limits_{x \in \mathcal{F}^{(i)}} \langle x, v\rangle & , \quad i \notin \mathcal{I} \end{cases} \qquad i = 1, \ldots, n. \tag{41}$$

Thus, we leave all optimal rows of $X$ untouched and replace all other rows by solutions of problem (41).

**If** $\mathcal{H} \subset \mathcal{I}$, then $\rho(X') = \rho(X)$ and the leading eigenvalue of $X'$ is simple and is attained on $V_{\mathcal{H}}$. We compute the leading eigenvector $v'$ of $X'$.

If $v' > 0$, then we set $X = X', v = v'$. Otherwise, $v$ has zero entries and we set $\mathcal{S} = \operatorname{supp}(v')$, $X = X'|_{\mathcal{S}}$, and $v = v'|_{\mathcal{S}}$. In any case, we go to (**).

**If** $\mathcal{H} \not\subset \mathcal{I}$, then $\rho(X') < \rho(X)$. We define the next matrix $X_{k+1}$ as follows:

$$(X_{k+1})^{(i)} \quad = \quad \begin{cases} X^{(i)\prime} & , \quad i \ \in \mathcal{S} \\ X_k^{(i)} & , \quad i \notin \mathcal{S}\,. \end{cases} \qquad i = 1, \ldots, n. \tag{42}$$

and go to the next $(k+1)$st iteration (*).

This algorithmic procedure is justified by the following statement.

**Theorem 8** *Algorithm 1 is well-defined. It finds the global solution of problem (39) in finite number of steps.*

The well-definedness means that at each iteration matrix $X'$ has a leading eigenvector, which is unique up to a normalization . We are proving more: $X'$ has a simple leading eigenvalue. The finite-time termination means that the algorithm does not cycle. For proving both properties, we need one auxiliary result.

**Proposition 4** *Let a non-negative matrix $A$ have a minimal leading eigenvector $v > 0$ and let $\mathcal{H}$ be the corresponding basic set. Let a non-negative matrix $A'$ and a set $\mathcal{I} \subset \Omega$ be such that*

$$\begin{cases} A^{(i)\prime} = A^{(i)} & , \quad i \in \mathcal{I} \\ \langle A^{(i)\prime}, v\rangle < \langle A^{(i)}, v\rangle & , \quad i \notin \mathcal{I}. \end{cases}$$

*Then $\rho(A') \le \rho(A)$, and the equality $\rho(A') = \rho(A)$ holds if and only if $\mathcal{H} \subset \mathcal{I}$. In this case, matrix $A'$ has a block upper triangular form (38) in the same basis with $C' = C$ and $\rho(B') < \rho(A)$.*

**Proof:**
Since $A'v \le Av$ with $v > 0$, it follows that $\rho(A') \le \rho(A)$. The set of active inequalities in the system $A'v \le Av$ coincides with $\mathcal{I}$. Hence, by Lemma 9, if $\rho(A') = \rho(A)$, then there is a subset $\mathcal{I}' \subset \mathcal{I}$ such that the subspace $V = V_{\mathcal{I}'}$ is invariant for $A^T$ and $\rho(A^T|_V) = \rho(A)$.

By Corollary 5, we see that $\mathcal{I}'$ contains $\mathcal{H}$ and hence $\mathcal{H} \subset \mathcal{I}$. Therefore, $A^{(i)\prime} = A^{(i)}$ for all $i \in \mathcal{I}$. So, matrix $A'$ has the block upper triangular form (38) in the same basis with $C' = C$.

It remains to prove that $\rho(B') < \rho(A)$. Without loss of generality we assume $\rho(A) = 1$. Denote by $u$ the part of the vector $v$ supported on the set $\Omega \setminus \mathcal{H}$. Since $\langle B^{(i)\prime}, u\rangle \le \langle A^{(i)\prime}, v\rangle$ for all $i$, we see that the set of active inequalities for $Bu \le u$ is a subset of $\mathcal{I}$, which does not intersect $H$. Applying Lemma 9 again, we see that this subset must contain $\mathcal{H}$. This contradiction completes the proof. □

**Proof of Theorem 8:**

We need to establish two properties.

1) (well-definiteness) Every matrix $X$ has a unique simple leading eigenvalue.

2) (finite termination) The algorithm does not cycle.

The first statement follows directly from Proposition 3. For proving non-cyclicity, we note that the spectral radii $\rho(X_k)$ strictly decrease in $k$. Hence, it suffices to show that the algorithm cannot cycle within one iteration. Furthermore, the sets $\mathcal{S}$ form a non-increasing embedded sequence. Therefore, cycling may happen only within one set $\mathcal{S} = \mathcal{S}_k$ on $k$th iteration. In this case, the greedy spectral simplex method generates a sequence of matrices $X$ on the set $\mathcal{S}$. Denote these matrices by $X_{k,1}, , X_{k,2}, \ldots$. Each of these matrices $X_{k,j}$ has a simple leading eigenvalue $\lambda_{\max}$, same for all $j$.

If this sequence is cycling, then the algorithm for a perturbed family $\mathcal{F}_\varepsilon = \{Y + \varepsilon J_n \mid Y \in \mathcal{F}\}$ is also cycling, whenever $\varepsilon > 0$ is small enough. Indeed, all the rows $X^{(i)}_{k,j}$, $j \in \mathbb{N}$, run over the finite set of vertices of the polytope $\mathcal{F}^{(i)}$. Hence, all $X_{k,j}$, $j \in \mathbb{N}$ run over a finite set of matrices $\operatorname{extr}(\mathcal{F})$. Same is true for the perturbed family $\mathcal{F}_\varepsilon$: the matrices run over the finite set of vertices $\operatorname{extr}\mathcal{F}_\varepsilon$.

Furthermore, the leading eigenvector of $X_{k,j}$ corresponds to the simple eigenvalue $\lambda_{\max}$ and hence it depends continuously on the coefficients of $X_{k,j}$. Since the total set of matrices $X_{k,j}$ is finite, all their leading eigenvectors $v_{k,j}$ are uniformly close to the leading eigenvectors of the perturbed matrices $X_{k,j,\varepsilon}$, whenever $\varepsilon$ is small enough. Therefore, all strict inequalities $\langle X'^{(i)}, v\rangle < \langle X^{(i)}, v\rangle$, for $X = X_{k,j}$, $X' = X_{k,j+1}$, $v = v_{k,j}$, involved in the construction of matrix $X' = X_{k,j+1}$ by formula (42), remain strict after the $\varepsilon$-perturbation. Hence, the perturbed algorithm runs over the same sequence of perturbed matrices $X_{j,\varepsilon}$. If the algorithm cycles, it follows that $X_j = X_{j+m}$ for some $j$ and $m$, and hence $X_{j,\varepsilon} = X_j + \varepsilon J_n = X_{j+m} + \varepsilon J_m = X_{j,\varepsilon}$. However, the algorithm, as applied to strictly positive matrices, does not cycle (Theorem 7). Hence, the equality $X_{j+m,\varepsilon} = X_{j,\varepsilon}$ is impossible. □

# 6 Implementation details of Algorithm 1

Each step of Algorithm 1 involves one computation of the minimal leading eigenvector of a square matrix $X$ and the solution of minimization problem (40) for each row of $X$. The size $m$ of this matrix is equal to $|\mathcal{S}|$, where $\mathcal{S}$ is the support of the leading eigenvector of the matrix obtained at the previous step. Let us look at these operations.

`Computing the leading eigenvector of` $X$ is the most expensive operation. It can be done in two steps: Frobenius factorization of $X$ ($O(m^2)$ operations) and computing the leading eigenvalues of the blocks. Note that by the construction of the algorithm, the leading eigenvalue $\lambda_{\max}$ of $X$ is simple and hence $\lambda_{\max}$ is Lipschitz continuous in matrix coefficients. The computation of $\lambda_{\max}$ can be done as suggested in Section 3.

`Solving the problem (40) in each row of` $X$ can be implemented for every polyhedral set $\mathcal{F}^{(i)}$ as a usual linear programming problem, or just by inspection of the finite number of vertices. If $\mathcal{F}^{(i)}$ is $\ell_\infty$-ball, then it can be done much simpler. We show this in the next subsection.

Let us look now at the implementation details for the problem of finding the closest non-negative stable matrix (35). Assume that $A \geq 0$ and $\rho(A) > 1$. We set $\tau_0 = \frac{1}{2}\|A\|$ and start the bisection method in $\tau$. For each $\tau$, we solve problem (39) for the uncertainty sets being positive parts of $\ell_\infty$-balls of radius $\tau$ centered at the rows of matrix $A$. Thus, $\mathcal{F}^{(i)} = \mathcal{B}_\tau^-(A^{(i)}) = \mathcal{B}_\tau(A^{(i)}) \cap \mathbb{R}_+^{n\times n}$.

We apply Algorithm 1 for this problem. Its implementation is basically the same as for the usual polyhedral sets. However, there are some simplifications.

1. `Realization of Algorithm 1 for` $\ell_\infty$`-balls` $\mathcal{F}^{(i)} = \mathcal{B}_\tau^-(A^{(i)})$. Solving minimization problem (40) at each iteration can be done explicitly. Firstly, we order the entries of the leading eigenvector $v$ with indices from $\mathcal{S}$: $v^{(j_1)} \geq \ldots \geq v^{(j_m)}$, where $\{j_1, \ldots, j_m\} = \mathcal{S}$. Then the problem (40) becomes as follows.

$$\sum_{k=1}^m v^{(j_k)} x^{(j_k)} \;\to\; \min : \quad \sum_{k=1}^m x^{(j_k)} \;\geq\; -\tau + \sum_{k=1}^m A^{(i,j_k)}. \tag{43}$$

Define by $\ell = \ell(\tau)$ the minimal index such that $\sum_{s=1}^{\ell} A^{(i,j_s)} > \tau$. If $\sum_{s=1}^m A^{(i,j_s)} \leq \tau$, then we set $\ell = m+1$. The solution to the problem (43) is then

$$x^{(j_k)} \;=\; \begin{cases} 0 & k < \ell \\ -\tau + \sum_{s=1}^{\ell} A^{(i,j_s)} & k = \ell \\ A^{(i,j_k)} & k > \ell \end{cases} \tag{44}$$

If $\ell = m+1$, then $x = 0$.

Applying bisection, we produce a sequence $\{\tau_i\}_{i\geq 0}$ converging to the optimal point. For each $i$, we minimize the spectral radius $\rho(X)$ on the $\ell_\infty$-ball $\mathcal{B}_{\tau_i}^-(A)$ by applying Algorithm 1. When the step length of the bisection $|\tau_{i+1} - \tau_i|$ becomes small enough, we can stop it and find the exact solution in one step. The following method can be applied when either the step length of the bisection becomes small or when the ordering of entries of $v$ stays unchanged for several $\tau_i$.

2. `Finding` $\min \tau$ `for which` $\rho(X) = 1$, $\|X - A\|_\infty = \tau$. We assume that the ordering of entries of the current leading eigenvector $v$ coincides with the ordering for the final $v$ (of the optimal matrix $X$). Consequently, we try to obtain the exact value

of $\tau_*$ within one iteration by assuming $\rho(X) = 1$ for the matrix $X$ constructed by the formula (44). We have $X = C - \tau R$, where

$$C^{(i,j_k)} \;=\; \begin{cases} 0 & k < \ell_i \\ \sum_{s=1}^{\ell_i} A^{(i,j_s)} & k = \ell_i \\ A^{(i,j_k)} & k > \ell_i \end{cases} \quad ; \qquad i,\, j_k \in \mathcal{S},\; k = 1, \ldots, m\,. \tag{45}$$

Here $\ell_i$ is the smallest index such that $\sum_{s=1}^{\ell_i} A^{(i,j_s)} > \tau$ (if $\sum_{s=1}^{m} A^{(i,j_s)} \le \tau$, then we set $\ell_i = m+1$, and $C^{(i,j_k)} = 0$ for all $k = 1, \ldots, m$), $R$ is a Boolean matrix, which has in $i$th row all zeros except a single 1 at position $\ell$ (provided that $\ell \le m$), and all zeros otherwise.

Denote $\tau_1 = \min\limits_{i \in \mathcal{S}} \sum_{s=1}^{\ell_i} A^{(i,j_s)}$. By construction, we have $C - \tau_1 R \ge 0$ and $\tau_1 > \tau$. Hence $\rho(C - \tau_1 R) < 1$. Since $1 = \rho(C - \tau R) = \rho\,(C - \tau_1 R + (\tau_1 - \tau)R)$, it follows that $\det\,(I - (C - \tau_1 R) \,-\, (\tau_1 - \tau)R) = 0$, and consequently

$$\det\,\Big(\, \tfrac{1}{\tau_1 - \tau}\, I \,-\, [\, I - (C - \tau_1 R)\,]^{-1} R \,\Big) \;\;=\;\; 0. \tag{46}$$

Note that $[\, I - (C - \tau_1 R)\,]^{-1} \;=\; \sum_{k=0}^{\infty} (C - \tau_1 R)^k \;\ge\; 0$. Hence, equation (46) means that the number $\frac{\lambda}{\tau_1 - \tau}$ is the leading eigenvalue of the non-negative matrix $[\, I - (C - \tau_1 R)\,]^{-1} R$. Hence, $\lambda$ can be found numerically be the method described in Section 3.

# 7 Conclusion

We present numerical methods for finding the closest stable or closest unstable non-negative matrix to a given matrix $A$. Three possible cases of measuring distances are considered: matrix max-norm (the maximal absolute value of entries), $\ell_\infty$ operator norm (the maximal sum of elements of rows), and $\ell_1$ operator norm (the maximal sum of elements of columns). We show that in all these cases, the absolute minimum can be found efficiently. The closest unstable matrix is computed by explicit formulas; the closest stable matrix can be found by an iterative relaxation scheme that makes use of recent "spectral simplex method".

From the practical point of view, we arrived to an interesting conclusion: for increasing the spectral radius so that the sum of entries in each row of the matrix increases by at most $a$, one needs to change by $a$ all elements of one column . That "most sensitive" column corresponds to the maximal component of the vector $(I - A)^{-1} e$. In the Leontief input-output model [11], this principle means that the economy suffers in a worst way when only one sector is perturbed. Moreover, this sector can be easily identified. In the matrix models of population dynamics (see, for instance, [13]), the same principle means that if an ecological system (say, a forest) is going to die, then for improving the situation, it is enough to support only one type of plants without touching the others.

# References


[1] V.D.Blondel and Y.Nesterov, *Polynomial-time computation of the joint spectral radius for some sets of nonnegative matrices*, SIAM J. Matrix Anal. Appl. 31 (2009), no 3, 865–876.

[2] A.Berman and R.J.Plemmons, *Nonnegative matrices in the mathematical sciences*, Acedemic Press, Now York, 1979.

[3] D.G.Cantor and S.A.Lippman, *Optimal investment selection with a multiple of projects*, Econometrica, 63 (1995), 1231–1240.

[4] L.Fainshil and M.Margaliot, *A maximum principle for the stability analysis of positive bilinear control systems with applications to positive linear switched systems*, SIAM J. Control Optim. 50 (2012), no. 4, 2193–2215.

[5] S.Friedland, *The maximal eigenvalue of* 0-1 *matrices with prescribed number of ones*, Linear Alg. Appl., 69 (1985), 33–69.

[6] R.Jungers, V.Yu.Protasov, and V.Blondel, *Efficient algorithms for deciding the type of growth of products of integer matrices*, Linear Alg. Appl., 428 (2008), no 10, 2296–2312.

[7] R.A.Horn and C.R.Johnson, *Matrix analysis*, Cambridge University Press, 1990.

[8] V. L. Kharitonov, *Asymptotic stability of an equilibrium position of a family of systems of differential equations*, Differentsialnye uravneniya, 14 (1978), 2086-2088 (in Russian).

[9] V.S.Kozyakin, *A short introduction to asynchronous systems*, in Proceedings of the Sixth International Conference on Difference Equations (Augsburg, Germany 2001): New Progress in Difference Equations, B. Aulbach, S. Elaydi, and G. Ladas, eds., CRC Press, Boca Raton, FL (2004), 153–166.

[10] W.Leontief, *Input-output economics*, 2nd ed., Oxford Uni. Press, NY, 1986.

[11] D.Liberzon, *Switching in systems and control*, Birkhauser, Boston, MA, 2003.

[12] H. Lin and P.J. Antsaklis, *Stability and stabilizability of switched linear systems: a survey of recent results*, IEEE Trans. Autom. Contr., 54 (2009), no 2, 308–322.

[13] D.O. Logofet, *Matrices and Graphs: Stability Problems in Mathematical Ecology*, CRC Press, Boca Raton, 1993.

[14] Y.Nesterov, V.Yu.Protasov, *Polynomial-time computation of the joint spectral radius for some sets of nonnegative matrices*, SIAM J. Matrix Anal. Appl. 34 (2013), no 3, 999–1013.

[15] F.X.Orbandexivry, Y.Nesterov, and P.Van Dooren, *Nearest stable system using successive convex approximations*, Automatica, 49 (2013), pp. 1195 1203.

[16] V.Yu.Protasov, *The spectral simplex method*, Math. Prog., 156 (2016), 485-511

[17] V.Schvatal *Linear programming*, W. H. Freeman, 1983.

[18] G.W.Stewart and J.G.Sun, *Matrix perturbation theory*, Acad. Press, NY, 1990.

[19] A. G. Vladimirov, N. Grechishkina, V. Kozyakin, N. Kuznetsov, A. Pokrovskii, and D. Rachinskii, *Asynchronous systems: Theory and practice*, Inform. Processes, 11 (2011), 1–45.