FACULTEIT ECONOMISCHE EN SOCIALE
WETENSCHAPPEN & SOLVAY BUSINESS SCHOOL

Vrije Universiteit Brussel

ES-Working Paper no. 1

The Evolution of Eigenvalues under a Perturbation

Pauwelyn P. And Guerry M.

February 25th, 2016

Vrije Universiteit Brussel - Pleinlaan 2, 1050 Brussel - www.vub.ac.be - info@vub.ac.be

Printed in Belgium

Vrije Universiteit Brussel

Faculty of Economics, Social Sciences and Solvay Business School

B-1050 Brussel

Belgium

www.vub.ac.be

The Evolution of Eigenvalues under a perturbation A study on stochastic 3×3 matrices

Pauwelyn PJ., Guerry M.A.

1 Abstract

Abstract

In applied sciences, many transition processes are described by Markov models. Markov chains use mainly stochastic matrices. In this paper, the interest lies with the non-diagonalizable stochastic matrices. We will explicitly show that it is possible for every non-diagonalizable stochastic 3×3 matrix to be perturbed into a diagonalizable stochastic matrix with eigenvalues arbitrarily close to the original eigenvalues. We also show that the resulting differences on the eigenvalues are continuously dependent on the executed perturbations.

Keywords: Markov models, Stochastic matrices, Non-diagonalizable matrices, Perturbation theory, Manpower planning

JEL Classification: C02

2 Introduction

In applied sciences, Markov chains are often used to model transition processes, in many disciplines, for example in manpower planning [8] and the social sciences [9]. These Markov models often use stochastic matrices that are based upon estimates for the transition probabilities distilled from data. This means there is a difference between these estimated transition probabilities and the theoretical transition probabilities, that results in errors in further calculations. Unfortunately, this is a consequence of working with data. But we might be able to use this problem to our advantage. The estimated transition probability lies in a high percentage confidence interval. If we slightly increase or decrease this estimated transition probability, it will still lie in a high percentage confidence interval. This allows us to slightly alter the estimated transition probabilities. These alterations will be made by the use of perturbations [6]. In matrix theory, diagonalizable matrices are easier to use in calculations, contrary to the non-diagonalizable matrices. For example, calculating matrix powers is harder with non-diagonalizable matrices and this affects the difficulty of finding solutions to homogeneous systems of linear differential/difference equations with constant coefficients, as shown in the sections 3.3.1 and 3.3.2 in [5].
It might a great relief for many applied scientists, if we could prove that the usage of non-diagonalizable matrices is redundant. The goal of the paper is to find a way to perturb non-diagonalizable stochastic 3×3 matrices into diagonalizable stochastic 3×3 matrices with an arbitrarily small difference on the eigenvalues.

3 Literature

Hartfiel[3] already proved some interesting theorems to back up the idea of this paper. First of all, Hartfiel proved that the diagonalizable stochastic matrices are dense in the set of the stochastic matrices. This means that every stochastic matrix can be written as the limit of a sequence of diagonalizable matrices.

Theorem 1. *Let K be a convex set of matrices and K' those matrices in K which have distinct eigenvalues. If $K' \neq \emptyset$, then K' is dense in K.*

It is easy to see that the set of the stochastic matrices is a convex set. Suppose $A = (a_{ij})$ and $B = (b_{ij})$ are stochastic matrices. This means that the rowsums of A and B are all 1:

$$\sum_{j=1}^{n} a_{ij} = 1, \qquad \forall i \in \{1, 2, 3, \dots, n\}$$
$$\sum_{j=1}^{n} b_{ij} = 1, \qquad \forall i \in \{1, 2, 3, \dots, n\}$$

For the set to be convex, the matrix $C = sA + (1-s)B$ has to be stochastic for all $s \in]0,1[$.

Therefore, we look at the rowsums of the matrix C

$$\begin{aligned}\sum_{j=1}^{n} c_{ij} &= \sum_{j=1}^{n} s a_{ij} + (1-s) b_{ij} \\ &= \sum_{j=1}^{n} s a_{ij} + \sum_{j=1}^{n} (1-s) b_{ij} \\ &= s \sum_{j=1}^{n} a_{ij} + (1-s) \sum_{j=1}^{n} b_{ij} \\ &= s \cdot 1 + (1-s) \cdot 1 \\ &= 1\end{aligned}$$

This means that the set of the $n \times n$ stochastic matrices is convex. By the previous theorem, it follows that the set, consisting of the stochastic matrices with distinct eigenvalues, is dense in the set of the stochastic matrices. Since every matrix with distinct eigenvalues is also diagonalizable, see section 4.3 in [6], it follows that the set of the diagonalizable stochastic matrices contains the set of the stochastic matrices with distinct eigenvalues. Combining this fact with the theorem above, we obtain that set of the diagonalizable stochastic matrices is dense in the set of the stochastic matrices.

Definition 1. *An eigenvalue λ is stable if and only if $|\lambda| < 1$.*

Theorem 2 (Hartfiel,1)**.** *Let ϵ be a positive number and let P be a $n \times n$ matrix with non-stable eigenvalues and having linearly elementary divisors. Then there exists a neighborhood N of P such that if $P' \in N$ and P' has non-stable eigenvalues and corresponding eigenspaces exactly those of P, then:*

$$\left|\left|P^k - (P')^k\right|\right|_2 < \epsilon, \forall k \geq 0.$$

with $||.||_2$ the euclidean matrixnorm.

Definition 2. *An eigenvalue λ of a matrix P such that Re $\lambda \geq$ Re β for all eigenvalues β of P is called an r-maximal eigenvalue of P.*

Theorem 3 (Hartfiel,2)**.** *Let ϵ be a positive number and let P be a $n \times n$ matrix, whose r-maximal eigenvalues correspond to linear elementary divisors. Then there exists a neighborhood N of P such that if $P' \in N$ and P' has r-maximal eigenvalues, and corresponding eigenspaces, exactly the same as those of P, then:*

$$\frac{||P^k - (P')^k||_2}{||P^k||_2} < \epsilon, \forall k \geq 0.$$

Theorem 4 (Hartfiel,3)**.** *Let P be a $n \times n$ matrix such that all r-maximal eigenvalues of A have linear elementary divisors. Then, given any positive number ϵ, there is a neighborhood N of P such that if $P' \in N$ and P' has r-maximal eigenvalues and corresponding eigenspaces, exactly those of A, then:*

$$\left|\left|\frac{P^k}{||P^k||_1} - \frac{(P')^k}{||(P')^k||_1}\right|\right| < \epsilon, \forall k \geq 0.$$

with $||.||_1$ the Manhattan norm.

These last 3 theorems suggest that the spectrum (eigenvalues, eigenvectors and eigenspaces) of the perturbed matrices are behaving well after a perturbation, given the right properties of both matrices. These four theorems combined with the Perron-Frobenius-theorem might be sufficient to claim that for some set of non-diagonalizable matrices, there exists a perturbation on these

matrices, such that the properties[1] of the perturbed matrix are arbitrarily close to the properties of the original matrix.
These theorems don't necessarily say that the non-diagonalizable matrices are redundant, but they definitely hint this way.
The challenge arising from this idea is to determine an algorithm to find the perturbed diagonalizable stochastic matrix P' corresponding to a non-diagonalizable stochastic matrix P with analogue properties with arbitrarely small differences on the eigenvalues.

4 Perturbing non-diagonalizable matrices

The main interest in this paper is the non-diagonalizable stochastic matrices. In this paper, we only look at the 3-dimensional case. The 2-dimensional case is trivial, because there are no non-diagonalizable 2×2 stochastic matrices. A stochastic matrix has always an eigenvalue 1. So, the only possibility for a 2×2 matrix to be non-diagonalizable is for the matrix to have the eigenvalue 1 with multiplicity two. But Minc[7] showed that for a stochastic matrix the eigenvalue 1 can only have Jordanblocks of order 1×1. Thus a 2×2 stochastic matrix can not be non-diagonalizable.
Now consider a stochastic 3×3 matrix:

$$P = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 - a_{12} - a_{13} & a_{12} & a_{13} \\ 1 - a_{22} - a_{23} & a_{22} & a_{23} \\ 1 - a_{32} - a_{33} & a_{32} & a_{33} \end{pmatrix}$$

with nonnegative elements and rowsums equal to 1:

$$\begin{aligned} a_{11} + a_{12} + a_{13} &= 1 \\ a_{21} + a_{22} + a_{23} &= 1 \\ a_{31} + a_{32} + a_{33} &= 1 \end{aligned}$$

A matrix P is diagonalizable when there exists a diagonal matrix J and an invertible matrix Q such that $P = QJQ^{-1}$, or in other words, P is similar to a diagonal matrix. If a matrix is non-diagonalizable, it is also possible to rewrite the matrix P as $P = QJQ^{-1}$, but in this case J is a Jordanmatrix, in other words, P is not similar to a diagonal matrix. Since a stochastic matrix has always the eigenvalue 1, for a stochastic 3×3 matrix to be a non-diagonalizable matrix, the remaining 2 eigenvalues must be the same. Thus the non-diagonalizable matrix P must be similar to the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & b \end{pmatrix}$$

with $b \in \mathbb{R}$. A non-diagonalizable stochastic 3×3 matrix can not be similar to one of the following Jordan matrices:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

because the eigenvalue 1 is always an semi-simple eigenvalue[7]. Thus the only possibility for a non-diagonalizable stochastic 3×3 matrix is that it is similar to

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & b \end{pmatrix}$$

[1]Properties are eigenvalues, eigenvectors, eigenspaces, solutions to systems of equations corresponding to this matrix, and so on.

Starting from the matrix P, we determine it's characteristic equation and demand that P is non-diagonalizable. Let P be the stochastic, non-diagonalizable matrix, like above. The characteristic equation of P can be written as[4]:

$$-\lambda^3 + tr(P)\lambda^2 + [1 - tr(P) - det(P)]\,\lambda + det(P) = 0$$

This can be factorized to

$$(\lambda - 1)\left[-\lambda^2 + (tr(P) - 1)\lambda - det(P)\right] = 0$$

Since this second factor has only one zero, it's discriminant should be zero:

$$\begin{aligned}
0 = \Delta &= [tr(P) - 1]^2 - 4det(P)\\
&= [a_{11} + a_{22} + a_{33} - 1]^2 - 4\left(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}\right)\\
&\quad + 4\left(a_{11}a_{23}a_{32} + a_{22}a_{13}a_{31} + a_{33}a_{12}a_{21}\right)
\end{aligned}$$

P is a stochastic matrix, thus it's rowsums equal 1. Using this condition, we can eliminate three parameters by substituting one equation out of each of the following rows into the discriminant.

$$\begin{array}{llll}
1) & a_{11} = 1 - a_{12} - a_{13}, & a_{12} = 1 - a_{11} - a_{13}, & a_{13} = 1 - a_{11} - a_{12}\\
2) & a_{21} = 1 - a_{22} - a_{23}, & a_{22} = 1 - a_{21} - a_{23}, & a_{23} = 1 - a_{21} - a_{22}\\
3) & a_{31} = 1 - a_{32} - a_{33}, & a_{32} = 1 - a_{31} - a_{33}, & a_{33} = 1 - a_{31} - a_{32}
\end{array}$$

This gives 27 possibilities for substitutions in the discriminant. We are not going to check all of them, we limit ourselves to three cases, namely, substituting the first column, the second column and eventually the third column. Substituting the first column[2] gives:

$$\begin{aligned}
\Delta &= [a_{11} + a_{22} + a_{33} - 1]^2 - 4\left(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}\right)\\
&\quad + 4\left(a_{11}a_{23}a_{32} + a_{22}a_{13}a_{31} + a_{33}a_{12}a_{21}\right)\\
&= \left(a_{12} - a_{13} - a_{22} + a_{33}\right)^2 + 4\left(a_{13} - a_{23}\right)\left(a_{12} - a_{32}\right)
\end{aligned}$$

By substituting the second respectively the third column, we obtain:

$$\begin{aligned}
\Delta &= [a_{11} + a_{22} + a_{33} - 1]^2 - 4\left(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}\right)\\
&\quad + 4\left(a_{11}a_{23}a_{32} + a_{22}a_{13}a_{31} + a_{33}a_{12}a_{21}\right)\\
&= \left(a_{11} - a_{21} - a_{33} + a_{23}\right)^2 + 4\left(a_{23} - a_{13}\right)\left(a_{21} - a_{31}\right)
\end{aligned}$$

and

$$\begin{aligned}
\Delta &= [a_{11} + a_{22} + a_{33} - 1]^2 - 4\left(a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}\right)\\
&\quad + 4\left(a_{11}a_{23}a_{32} + a_{22}a_{13}a_{31} + a_{33}a_{12}a_{21}\right)\\
&= \left(a_{22} - a_{32} - a_{1} + a_{31}\right)^2 + 4\left(a_{31} - a_{21}\right)\left(a_{32} - a_{12}\right)
\end{aligned}$$

These three equations can be written as one equation:

$$\begin{aligned}
\Delta = &\left(a_{i,i+1} - a_{i,i+2} - a_{i+1,i+1} + a_{i+2,i+2}\right)^2\\
&+ 4\left(a_{i,i+2} - a_{i+1,i+2}\right)\left(a_{i,i+1} - a_{i+2,i+1}\right)
\end{aligned}$$

with $i \in \{1, 2, 3\}$ and we look at $i, i+1, i+2$ as elements[3] of $\mathbb{Z}_3 = \{1, 2, 3\}$, where $4 := 1$ and $5 := 2$. Since P is non-diagonalizable, Δ is 0. For P' to have distinct eigenvalues, Δ must be

[2]These are the equations: $a_{11} = 1 - a_{12} - a_{13}$, $a_{21} = 1 - a_{22} - a_{23}$ and $a_{31} = 1 - a_{32} - a_{33}$

[3]If $i = 2$, then $i + 2 = 4$, but we look at it if it was 1. Because $1 = 4 \bmod 3$.

nonzero. We demand that Δ is strictly positive, because we want the resulting eigenvalues also to be real, since the original eigenvalue was real. Therefore Δ must be positive. For negative Δ, we would introduce complex eigenvalues. This changes some of the properties of the matrix, while we are trying to conserve the properties of the matrix.
Now, Δ depends on 6 different parameters, namely $a_{i,i+1}; a_{i,i+2}; a_{i+1,i+1}; a_{i+1,i+2}; a_{i+2,i+1}$ and $a_{i+2,i+2}$. We want to perturb one or more of these 6 elements. Before we can do this, we need to know how the discriminant reacts on the perturbations. We will change the stochastic matrix P into the stochastic matrix P' by perturbation:

$$\begin{aligned}
P' &= \begin{pmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{pmatrix} = \begin{pmatrix} 1-a'_{12}-a'_{13} & a'_{12} & a'_{13} \\ 1-a'_{22}-a'_{23} & a'_{22} & a'_{23} \\ 1-a'_{32}-a'_{33} & a'_{32} & a'_{33} \end{pmatrix} \\
&= \begin{pmatrix} 1-(a_{12}+\epsilon_{12})-(a_{13}+\epsilon_{13}) & a_{12}+\epsilon_{12} & a_{13}+\epsilon_{13} \\ 1-(a_{22}+\epsilon_{22})-(a_{23}+\epsilon_{23}) & a_{22}+\epsilon_{22} & a_{23}+\epsilon_{23} \\ 1-(a_{32}+\epsilon_{32})-(a_{33}+\epsilon_{33}) & a_{32}+\epsilon_{32} & a_{33}+\epsilon_{33} \end{pmatrix} \\
&= P + \begin{pmatrix} -\epsilon_{12}-\epsilon_{13} & \epsilon_{12} & \epsilon_{13} \\ -\epsilon_{22}-\epsilon_{23} & \epsilon_{22} & \epsilon_{23} \\ -\epsilon_{32}-\epsilon_{33} & \epsilon_{32} & \epsilon_{33} \end{pmatrix} = P + E
\end{aligned}$$

Before we start calculating with the perturbations ϵ_{ij}, we have to apply some restrictions on them. First, the perturbations have to be small in comparison to their corresponding element. Second, the perturbations have to be small in comparison to the complement of their corresponding element.

$$|\epsilon_{ij}| \lll a_{ij} \tag{1}$$
$$|\epsilon_{ij}| \lll 1 - a_{ij} \tag{2}$$

These two inequalities are the mathematical translation of the small errors on the transition probabilities. As long as these perturbations are kept small in relation to the corresponding transition probability, the new transition probabilities $a'_{ij} = a_{ij} + \epsilon_{ij}$ are still within a (high-percentage) confidence interval for the original transition probability. These two rules imply three important things about the certainties[4]

1. Every 0 stays 0. (Since (1))
2. Every 1 stays 1. (Since (2))
3. For every $a_{ij} \in\ ]0,1[$, $a_{ij} + \epsilon_{ij}$ stays in this interval and does not become 0 or 1. Thus, non-certainties do not become certainties.

Under these perturbations, the matrix P changes and becomes a matrix P'. If P changes, Δ will change accordingly. Therefore, we want to know the influence on Δ of a perturbation on an element

[4]Certainties are events with probability 0 or 1.

of P. Therefore we will look at the partial derivatives of Δ:

$$\begin{aligned}
\frac{\partial \Delta}{\partial a_{i+1,i+1}} &= -2(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}) \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i+1,i+1}} > 0 \Leftrightarrow a_{i+2,i+2} + a_{i,i+1} < a_{i+1,i+1} + a_{i,i+2} \\
\frac{\partial \Delta}{\partial a_{i+2,i+2}} &= 2(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}) \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i+2,i+2}} > 0 \Leftrightarrow a_{i+2,i+2} + a_{i,i+1} > a_{i+1,i+1} + a_{i,i+2} \\
\frac{\partial \Delta}{\partial a_{i,i+1}} &= 2(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}) + 4(a_{i,i+2} - a_{i+1,i+2}) \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i,i+1}} > 0 \Leftrightarrow a_{i+2,i+2} + a_{i+1,i} > a_{i+1,i+2} + a_{i,i} \\
\frac{\partial \Delta}{\partial a_{i,i+2}} &= -2(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}) + 4(a_{i,i+1} - a_{i+2,i+1}) \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i,i+2}} > 0 \Leftrightarrow a_{i+1,i+1} + a_{i+2,i} > a_{i+2,i+1} + a_{i,i} \\
\frac{\partial \Delta}{\partial a_{i+2,i+1}} &= -4(a_{i,i+2} - a_{i+1,i+2}) \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i,i+2}} > 0 \Leftrightarrow a_{i,i+2} < a_{i+1,i+2} \\
\frac{\partial \Delta}{\partial a_{i+1,i+2}} &= -4(a_{i,i+1} - a_{i+2,i+1}) > 0 \\
&\hookrightarrow \quad \frac{\partial \Delta}{\partial a_{i+1,i+2}} > 0 \Leftrightarrow a_{i,i+1} < a_{i+2,i+1}
\end{aligned} \tag{3}$$

If these conditions are true, then the discriminant increases when we perturb the corresponding element with a positive quantity. This means we can use these conditions to determine which elements we should perturb on.

5 Perturbing stochastic matrices

To obtain a stochastic matrix P' from the original stochastic matrix P after perturbation, there are two choices for every row:

1. leave the row unchanged, so the rowsum stays 1.
2. or perturb at least two elements in a row, with their sum equal to zero. If we would perturb only one element in a row, the rowsum changes and P' is not stochastic.

This is necessary to keep the rowsums of P and P' one. The necessary conditions (3) to allow positive perturbing from the previous paragraph can be used to determine the sign of the perturbation ϵ_{ij}. The conditions were given such that we could determine which element we can perturb on. We can alter these conditions to determine the sign of the perturbation ϵ_{ij}. For example, the first one, the perturbation corresponding to $a_{i+1,i+1}$:

$$\frac{\partial \Delta}{\partial a_{i+1,i+1}} > 0 \Leftrightarrow a_{i+2,i+2} + a_{i,i+1} < a_{i+1,i+1} + a_{i,i+2}$$

This is rewritten as follows:

$$\frac{\partial \Delta}{\partial a_{i+1,i+1}} > 0 \Leftrightarrow a_{i+1,i+1} + a_{i,i+2} - a_{i+2,i+2} - a_{i,i+1} > 0$$

This equivalence shows that the partial derivative and the condition have the same sign, thus we can define the sign of the perturbation in the following sign-equation:

$$\mathrm{sgn}(\epsilon_{i+1,i+1}) = \mathrm{sgn}(a_{i+1,i+1} + a_{i,i+2} - a_{i+2,i+2} - a_{i,i+1})$$

We demand that the perturbation on a_{ij} has the same sign as the partial derivative of Δ with respect to that element a_{ij}. The reason hereof goes as follows: If the partial derivative of Δ with respect to a_{ij} is positive, then Δ increases if a_{ij} increases. If the partial derivative of Δ with respect to a_{ij} is negative, then Δ decreases if a_{ij} increases and Δ increases if a_{ij} decreases. We want that Δ becomes positive instead of staying zero. So we allow only the perturbations which make Δ increase. Therefore ϵ_{ij} must satisfy these sign-equations. Analogue sign-equations can be formulated for the other perturbations:

$$\begin{aligned}
\mathrm{sgn}(\epsilon_{i,i+1}) &= \mathrm{sgn}(a_{i+2,i+2} + a_{i+1,i} - a_{i+1,i+2} - a_{i,i}) \\
\mathrm{sgn}(\epsilon_{i,i+2}) &= \mathrm{sgn}(a_{i+1,i+1} + a_{i+2,i} - a_{i+2,i+1} - a_{i,i}) \\
\mathrm{sgn}(\epsilon_{i+1,i+2}) &= \mathrm{sgn}(a_{i+2,i+1} - a_{i,i+1}) \\
\mathrm{sgn}(\epsilon_{i+2,i+1}) &= \mathrm{sgn}(a_{i+1,i+2} - a_{i,i+2}) \\
\mathrm{sgn}(\epsilon_{i+2,i+2}) &= \mathrm{sgn}(a_{i+2,i+2} + a_{i,i+1} - a_{i+1,i+1} - a_{i,i+2})
\end{aligned}$$

We can perturb almost any matrix on the previous manner, we need only one of the six conditions (3) to be true to be able to find a perturbed matrix P'. The only case where we can not find a perturbation is when in every condition there is an equality. The question here is, which matrices satisfy all the conditions with an equality? Therefore we solve the system wherein all conditions are represented:

$$\begin{cases}
a_{i+1,i+1} + a_{i,i+2} - a_{i+2,i+2} - a_{i,i+1} = 0 \\
a_{i+2,i+2} + a_{i+1,i} - a_{i+1,i+2} - a_{i,i} = 0 \\
a_{i+1,i+1} + a_{i+2,i} - a_{i+2,i+1} - a_{i,i} = 0 \\
a_{i+2,i+1} - a_{i,i+1} = 0 \\
a_{i+1,i+2} - a_{i,i+2} = 0 \\
a_{i+2,i+2} + a_{i,i+1} - a_{i+1,i+1} - a_{i,i+2} = 0
\end{cases}$$

Solving this system gives the following solution, by using the stochastic property of P:

$$\begin{cases}
a_{i+1,i+1} = a_{i+2,i+1} + a_{i+2,i+2} - a_{i+1,i+2} \\
a_{i,i+1} = a_{i+2,i+1} \\
a_{i,i+2} = a_{i+1,i+2}
\end{cases}$$

If we put this solution back into matrix-form (for $i = 1$), we find all the matrices which can not be perturbed in a direct manner. For $i = 2$ and $i = 3$, we find the same set of matrices.

$$NPS := \left\{ \begin{pmatrix} 1-k-l & l & k \\ 1-l-m & l+m-k & k \\ 1-l-m & l & m \end{pmatrix} \middle| \, k, l, m, k+l, l+m, l+m-k \in [0,1] \right\}$$

This set NPS of the Non-Perturbable Stochastic matrices contains all the matrices which have an eigenvalue with algebraic multiplicity 2 and which can not be perturbed by the method above to obtain a diagonalizable matrix. It is interesting to see if these matrices are diagonalizable or not:

When we determine the Jordanform of all these matrices, then we find that the Jordanmatrix is the following:

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & m-k & 0 \\ 0 & 0 & m-k \end{pmatrix}$$

All matrices in NPS are diagonalizable, this fact leads to the next theorem:

Theorem 5. *Every stochastic 3×3 matrix either is diagonalizable or can be perturbed to a diagonalizable matrix.*

This theorem is a consequence of the fact that the diagonalizable stochastic matrices are dense in the set of the stochastic matrices. The interesting part is, with the explicit use of the perturbations, we can find a method to explicitly determine the perturbed matrix.

6 Eigenvalues of the perturbed matrix

In this section, we create a method to find the perturbed matrix, such that this perturbed matrix is diagonalizable. First, we determine the influence of the perturbations on the eigenvalues. The eigenvalues of the non-diagonalizable matrix P are:

$$\begin{aligned} 0 = det(P - \lambda I) &= -\lambda^3 + tr(P)\lambda^2 + [1 - tr(P) - det(P)]\,\lambda + det(P) \\ &= (\lambda - 1)\left[-\lambda^2 + (tr(P) - 1)\lambda - det(P)\right] \end{aligned}$$

This characteristic equation has two solution, 1 and λ_1, with:

$$\lambda_1 = \frac{1 - tr(P) + \sqrt{(tr(P) - 1)^2 - 4det(P)}}{-2} = \frac{tr(P) - 1}{2}$$

because the discriminant $\Delta_P = (tr(P) - 1)^2 - 4det(P) = 0$. For the eigenvalues of P', we can use the same formula:

$$\begin{aligned} 0 = det(P' - \lambda I) &= -\lambda^3 + tr(P')\lambda^2 + \left[1 - tr(P') - det(P')\right]\lambda + det(P') \\ &= (\lambda - 1)\left[-\lambda^2 + (tr(P') - 1)\lambda - det(P')\right] \end{aligned}$$

As we found before, the discriminant is:

$$\Delta_{P'} = \left(a'_{i+1,i+2} - a'_{i+1,i+1} + a'_{i,i+1} - a'_{i,i+2}\right)^2 + 4(a'_{i,i+1} - a'_{i+2,i+1})(a'_{i,i+2} - a'_{i+1,i+2})$$

Writing $a'_{ij} = a_{ij} + \epsilon_{ij}$ gives

$$\begin{aligned} \Delta_{P'} = &\left(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}\right)^2 \\ &+ 4(a_{i,i+1} - a_{i+2,i+1})(a_{i,i+2} - a_{i+1,i+2}) + \left(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2}\right)^2 \\ &+ 2\left(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}\right)\left(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2}\right) \\ &+ 4\left(a_{i,i+1} - a_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right) + 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(a_{i,i+2} - a_{i+1,i+2}\right) \\ &+ 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right) \\ = &\;\Delta_P + \epsilon_\Delta \\ = &\;\epsilon_\Delta \end{aligned}$$

with

$$\Delta_P = \left(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}\right)^2 + 4(a_{i,i+1} - a_{i+2,i+1})(a_{i,i+2} - a_{i+1,i+2})$$

and

$$\begin{aligned}\epsilon_\Delta = &(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2})^2 \\ &+ 2\left(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}\right)\left(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2}\right) \\ &+ 4\left(a_{i,i+1} - a_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right) + 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(a_{i,i+2} - a_{i+1,i+2}\right) \\ &+ 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right)\end{aligned}$$

For a non-diagonalizable matrix P, Δ_P is always zero. Now, we obtain a formula for the eigenvalues of the perturbed matrix P':

$$\lambda'_{1,2} = \frac{tr(P') - 1 \mp \sqrt{\Delta_{P'}}}{2} = \frac{tr(P') - 1 \mp \sqrt{\epsilon_\Delta}}{2}$$

We know that $\Delta_{P'}$ is positive, because we made it positive by performing a perturbation on P to obtain P' with a positive discriminant. Thus we know that both λ_1 and λ_2 are real numbers. The last thing to do is to check whether the difference in the corresponding eigenvalues is continuously dependent of the perturbations ϵ_{ij}. Therefore we consider the quantities $|\lambda'_1 - \lambda_1|$ and $|\lambda'_2 - \lambda_2|$:

$$\begin{aligned}\left|\lambda'_1 - \lambda_1\right| &= \left|\frac{1 - tr(P') + \sqrt{\Delta_P + \epsilon_\Delta}}{-2} - \frac{1 - tr(P)}{-2}\right| \\ &= \left|\frac{1 - [tr(P) - \epsilon_{i,i+1} - \epsilon_{i,i+2} + \epsilon_{i+1,i+1} + \epsilon_{i+2,i+2}] + \sqrt{\Delta_P + \epsilon_\Delta}}{-2} - \frac{1 - tr(P)}{-2}\right| \\ &= \left|\frac{\epsilon_{i,i+1} + \epsilon_{i,i+2} - \epsilon_{i+1,i+1} - \epsilon_{i+2,i+2} + \sqrt{\epsilon_\Delta}}{2}\right|\end{aligned}$$

and

$$\begin{aligned}\left|\lambda'_2 - \lambda_2\right| &= \left|\frac{1 - tr(P') - \sqrt{\Delta_P + \epsilon_\Delta}}{-2} - \frac{1 - tr(P)}{-2}\right| \\ &= \left|\frac{1 - [tr(P) - \epsilon_{i,i+1} - \epsilon_{i,i+2} + \epsilon_{i+1,i+1} + \epsilon_{i+2,i+2}] - \sqrt{\Delta_P + \epsilon_\Delta}}{-2} - \frac{1 - tr(P)}{-2}\right| \\ &= \left|\frac{\epsilon_{i,i+1} + \epsilon_{i,i+2} - \epsilon_{i+1,i+1} - \epsilon_{i+2,i+2} - \sqrt{\epsilon_\Delta}}{2}\right|\end{aligned}$$

Now, consider the function:

$$F_h : D \to \mathbb{R} : (\epsilon_{i,i+1}; \epsilon_{i,i+2}; \epsilon_{i+1,i+1}; \epsilon_{i+1,i+2}; \epsilon_{i+2,i+1}; \epsilon_{i+2,i+2}) \mapsto \left|\lambda'_h - \lambda_h\right|$$

where

$$\begin{aligned}D = &\operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i,i+1}) \cdot c\} \times \operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i,i+2}) \cdot c\} \times \operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i+1,i+1}) \cdot c\} \\ &\times \operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i+1,i+2}) \cdot c\} \times \operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i+2,i+1}) \cdot c\} \times \operatorname{conv}\{0, \operatorname{sgn}(\epsilon_{i+2,i+2}) \cdot c\}\end{aligned}$$

where conv stands for the convex hull and $c <<< 1$ is a cut-off value for the perturbations. Our goal is to show that F_h is continuous. Therefore, we write F_h as a combination of four other functions:

$$F_h = f_1 \circ (f_2 \pm f_3 \circ f_4)$$

with

$$
\begin{aligned}
&f_1 : \mathbb{R} \to \mathbb{R} : x \mapsto \left|\frac{x}{2}\right| \\
&f_2 : \mathbb{R}^6 \to \mathbb{R} : (x_1, x_2, x_3, x_4, x_5, x_6) \mapsto x_1 + x_2 - x_3 - x_6 \\
&f_3 : \mathbb{R}^+ \to \mathbb{R} : x \mapsto \sqrt{x} \\
&f_4 : \mathbb{R}^6 \mapsto \mathbb{R} : (\epsilon_{i,i+1}; \epsilon_{i,i+2}; \epsilon_{i+1,i+1}; \epsilon_{i+1,i+2}; \epsilon_{i+2,i+1}; \epsilon_{i+2,i+2}) \\
&\quad \mapsto \Delta_P + \left(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2}\right)^2 + 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right) \\
&\qquad + 2\left(a_{i+2,i+2} - a_{i+1,i+1} + a_{i,i+1} - a_{i,i+2}\right)\left(\epsilon_{i+2,i+2} - \epsilon_{i+1,i+1} + \epsilon_{i,i+1} - \epsilon_{i,i+2}\right) \\
&\qquad + 4\left(a_{i,i+1} - a_{i+2,i+1}\right)\left(\epsilon_{i,i+2} - \epsilon_{i+1,i+2}\right) + 4\left(\epsilon_{i,i+1} - \epsilon_{i+2,i+1}\right)\left(a_{i,i+2} - a_{i+1,i+2}\right)
\end{aligned}
$$

f_1 is clearly continuous everywhere. The same goes for f_2. f_3 is only continuous for all $x > 0$ and right continu in $x = 0$. f_4 is a polynomial in it's variables, because Δ_P is a constant[5] in this case. For the continuity of F_h, there is only one problem, we need f_4 to be a positive function, otherwise there would arise problem in the domain of $f_3 \circ f_4$. But we applied our perturbations such that $\Delta_P + \epsilon_\Delta$ is positive. Then F_h is a continuous function. Because F_h is continu, it means by the definition of continuity that the difference of the corresponding eigenvalues λ_h and λ'_h can be arbitrarily small, when the ϵ_{ij} are chosen small enough. The formulas for the eigenvalues of the perturbed matrices are:

$$
\begin{aligned}
&\lambda_0 = 1 \to \lambda'_0 = 1 \\
&\lambda_1 = \frac{1 - tr(P)}{-2} = \frac{a_{22} + a_{33} - a_{12} - a_{13}}{2} \\
&\to \lambda'_1 = \frac{1 - tr(P') + \sqrt{\epsilon_\Delta}}{-2} = \frac{1 - tr(P) + \epsilon_{i,i+1} + \epsilon_{i,i+2} - \epsilon_{i+1,i+1} - \epsilon_{i+2,i+2} + \sqrt{\epsilon_\Delta}}{-2} \\
&\lambda_2 = \frac{1 - tr(P)}{-2} = \frac{a_{22} + a_{33} - a_{12} - a_{13}}{2} \\
&\to \lambda'_2 = \frac{1 - tr(P') - \sqrt{\epsilon_\Delta}}{-2} = \frac{1 - tr(P) + \epsilon_{i,i+1} + \epsilon_{i,i+2} - \epsilon_{i+1,i+1} - \epsilon_{i+2,i+2} - \sqrt{\epsilon_\Delta}}{-2}
\end{aligned}
$$

This result implies that we can alter theorem 5 to:

Theorem 6. *Every stochastic 3×3 matrix either is diagonalizable or can be perturbed to a diagonalizable matrix, with the difference between the corresponding eigenvalues arbitrarily small.*

7 Example

An example of a non-diagonalizable stochastic matrix is:

$$
P = (a_{ij}) = \begin{pmatrix} \frac{7}{12} & \frac{1}{6} & \frac{1}{4} \\ \frac{1}{3} & \frac{5}{12} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}
$$

The corresponding Jordanform is:

$$
J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 1 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}
$$

[5] Δ is actually 0.

Thus P is clearly a non-diagonalizable stochastic matrix with eigenvalues 1, multiplicity 1 and $\frac{1}{4}$ with multiplicity 2. The left eigenvector corresponding to the eigenvalue 1 is the vector $\left(\frac{11}{9}, \frac{7}{9}, 1\right)$. The generalized eigenvectorspace corresponding to the eigenvalue $\frac{1}{4}$ is $\{(-z-y, y, z)|y, z \in \mathbb{R}\}$ and is spanned by the vectors $(-1, 1, 0)$ and $(-1, 0, 1)$.
In this case, P' is a perturbation of the matrix P,where we only perturb the element p_{13}. If we choose i to be 1, then we have check the inequality for the (i,i+2)th element:

$$\begin{aligned}\operatorname{sgn}(\epsilon_{i,i+2}) &= \operatorname{sgn}\left(a_{i+1,i+1} + a_{i+2,i} - a_{i+2,i+1} - a_{i,i}\right) = \operatorname{sgn}(a_{22} + a_{31} - a_{32} - a_{11}) \\ &= \operatorname{sgn}\left(\frac{5}{12} + \frac{1}{4} - \frac{1}{4} - \frac{7}{12}\right) = \operatorname{sgn}\left(-\frac{1}{6}\right)\end{aligned}$$

And $-1/6 < 0$, thus we should perturb the element p_{13} with a small negative constant. We choose for the perturbation $\epsilon_{13} = \frac{-1}{1000}$. If we perturb p_{13}, we also have to perturb p_{11} with the constant $\epsilon_{11} = \frac{1}{1000}$. This is necessary to conserve the rowsum. Let P' be

$$P' = P + \begin{pmatrix} \frac{1}{1000} & 0 & \frac{-1}{1000} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Then the eigenvalues of P' are 1; $\dfrac{1503+\sqrt{3009}}{6000}$ and $\dfrac{1503-\sqrt{3009}}{6000}$. The corresponding left eigenvectors are:

$$\begin{aligned} v_1 &= \left(\frac{2750}{2243}, \frac{1747}{2243}, 1\right) \\ v_{\frac{1503+\sqrt{3009}}{6000}} &= \left(-\frac{\sqrt{3009}+3}{6}, \frac{\sqrt{3009}-3}{6}, 1\right) \\ v_{\frac{1503-\sqrt{3009}}{6000}} &= \left(-\frac{\sqrt{3009}+3}{6}, \frac{\sqrt{3009}-3}{6}, 1\right) \end{aligned}$$

To express the difference is between the eigenspaces, we consider the angle between the spaces. For the eigenspaces corresponding to the eigenvalue 1, we compute the angle θ between both directions:

$$\begin{aligned} \cos(\theta) &= \frac{\vec{u} \bullet \vec{v}}{|\vec{u}| \cdot |\vec{v}|} = 0{,}9952024 \\ &\Rightarrow \theta = 5{,}615° \end{aligned}$$

For the other eigenspace, corresponding to the eigenvalue $\frac{1}{4}$ and the eigenvalues $\frac{1503+\sqrt{3009}}{6000}$ and $\frac{1503-\sqrt{3009}}{6000}$, we have to determine the function for both eigenspaces, then determine their normal and finally compute the angle between both normals. For the original matrix, the eigenspace is the following set of eigenvectors:

$$\{(-a-b, a, b)\,|a, b \in \mathbb{R}\} = \operatorname{Span}\{(-1, 1, 0)\,; (-1, 0, 1)\}$$

The normal of this eigenspace is the vector $\vec{N_1} = (1, 1, 1)$. The eigenspace corresponding to the eigenvalues $\frac{1503+\sqrt{3009}}{6000}$ and $\frac{1503-\sqrt{3009}}{6000}$ is spanned by:

$$\operatorname{Span}\left\{\left(\frac{-\sqrt{3009}-3}{6}, \frac{\sqrt{3009}-3}{6}, 1\right); \left(\frac{\sqrt{3009}-3}{6}, \frac{-\sqrt{3009}-3}{6}, 1\right)\right\}$$

The normal on this eigenspace is the vectorproduct of these two vectors, namely:

$$\vec{N_2} = \left(\frac{-\sqrt{3009}-3}{6}, \frac{\sqrt{3009}-3}{6}, 1\right) \times \left(\frac{\sqrt{3009}-3}{6}, \frac{-\sqrt{3009}-3}{6}, 1\right) = \left(\frac{\sqrt{3009}}{3}, \frac{\sqrt{3009}}{3}, \frac{\sqrt{3009}}{3}\right)$$

The angle between these two normals is:

$$\cos\left(\theta\left(\vec{N_1}, \vec{N_2}\right)\right) = \frac{\vec{N_1} \bullet \vec{N_2}}{|\vec{N_1}| \cdot |\vec{N_2}|} = \frac{\sqrt{3009}}{\sqrt{3}\sqrt{1003}} = 1$$
$$\Rightarrow \theta\left(\vec{N_1}, \vec{N_2}\right) = 0^\circ$$

This eigenspace is actually the same, despite the perturbation.
In this example, it becomes clear that is possible to find a perturbation such that the difference between the corresponding eigenvalues is small. In this example the difference is:

$$\left|\lambda_1' - \lambda_1\right| = \left|\frac{1503+\sqrt{3009}}{6000} - \frac{1}{4}\right| = \left|\frac{3+\sqrt{3009}}{6000}\right| \approx 0.009642$$
$$\left|\lambda_2' - \lambda_2\right| = \left|\frac{1503-\sqrt{3009}}{6000} - \frac{1}{4}\right| = \left|\frac{3-\sqrt{3009}}{6000}\right| \approx 0.008642$$

The deviation on the eigenvalues is less than $\frac{1}{100}$. And this deviation becomes even smaller when we had chosen a smaller ϵ_{13}.
In this example we have only used one perturbation namely ϵ_{13} and we found a good result for the eigenvalues (and even a decent result for the eigenspaces). Six perturbations seems to be too many perturbations, but if we want to involve the behaviour of the eigenvectors and eigenspaces, they all might become quite useful.

8 Conclusion and further research

We have calculated the influence of a stochastic perturbation on the eigenvalues of a stochastic matrix. By theorem 6, we showed that every non-diagonalizable stochastic matrix becomes diagonalizable after the right perturbation and the difference between the corresponding eigenvalues is arbitrarily small.
In the example, we saw that the eigenspaces were close to each other, and the 2 dimensional eigenspaces were even exactly the same space. It might be interesting for further research to look into the eigenvectors and the eigenspaces of the perturbed matrix in comparison to those of the original matrix.
Another idea for further research is the measurement of differences in eigenspaces by the use of angles. This method of measuring the difference between two eigenspaces is obvious in a 3-dimensional space, but for a higher dimensional space is less easy to interpret the value of the angle(s).

9 Bibliography

[1] Seneta E., Non-negative matrices, An introduction to theory and applications, 1973

[2] Hartfiel D.J., Dense sets of diagonalizable matrices, 1995

[3] Hartfiel D.J., Tracking in Matrix Systems, 1992

[**4**] Guerry M.A., The Embedding problem for monotone Markov chains, 2016 (submitted for publication)

[**5**] Saber Elaydi, An Introduction to Difference Equations, 2005

[**6**] Hogben L., Handbook of Linear Algebra, 2007

[**7**] Minc H., Nonnegative Matrices, 1988

[**8**] Bartholomew D.J., Forbes A.F., McClean S.I., Statistical Techniques for Manpower Planning, 1991.

[**9**] Bartholomew D.J., Stochastic Models for Social Processes, 1973