HAL
open science

# Direct Solution of the Inverse Stochastic Problem through Elementary Markov State Disaggregation

Lorenzo Ciampolini, Sylvain Meignen, Olivier Menut, Turgis David

▶ **To cite this version:**

Lorenzo Ciampolini, Sylvain Meignen, Olivier Menut, Turgis David. Direct Solution of the Inverse Stochastic Problem through Elementary Markov State Disaggregation. 2014. hal-01016804

**HAL Id: hal-01016804**

**https://hal.science/hal-01016804**

Preprint submitted on 1 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# DIRECT SOLUTION OF THE INVERSE STOCHASTIC PROBLEM THROUGH ELEMENTARY MARKOV STATE DISAGGREGATION

LORENZO CIAMPOLINI*, SYLVAIN MEIGNEN†, OLIVIER MENUT‡, AND DAVID TURGIS§

**Abstract.** Existing methods to find the eigenvalue spectrum (or a reasonable approximation to it) of square matrices can be extended to Stochastic Matrices (SM). The matter is more delicate for the Inverse Eigenvalue Problem (IEP), which consists in the reconstruction of a matrix from a given eigenvalue spectrum. In this work, we present a simple method to solve a real-valued IEP for SM by constructing step-by-step the solution matrix through an elementary Markov state disaggregation method named state splitting, and based on a matrix operator. After showing some results on how the splitting operator influences the steady-state distribution of the Markov chain associated with the SM, we demonstrate that the state splitting operator has a fundamental property: when applied to a SM A of size n-by-n, it yields a SM of size (n+1)-by-(n+1), whose eigenvalue spectrum is equal to that of A, plus an additional eigenvalue belonging to a bounded interval. We use a constructive method to prove that for any spectrum made of real and positive eigenvalues, one can build up an infinite number of SM sharing this spectrum. Finally, we present a new sufficient condition to test if a given set of real values can be the spectrum of a SM constructed by the proposed method.

**Key words.** Stochastic matrix, inverse eigenvalue problem, state disaggregation.

**AMS subject classifications.** 65F18, 15B51, 15A18, 15A29

**1. Introduction.** This work deals with the eigenvalue spectrum of a particular class of square matrices: stochastic matrices (SMs), for which all elements are non-negative and such that their sums along rows[1] are all equal to one [1]. SMs are important because they can describe the dynamics of Markov chains (MCs). In detail, the $i$-th row of a $n \times n$ SM describes the probability distribution (which sums up to unity) of the transition of the system towards $n$ possible states, assuming that the system is in the $i$-th state. If one writes a row state vector $\underline{u}_{k-1}$ holding a repartition distribution in $n$ different possible states, the dynamics of the system across consecutive time steps is described by the matrix equation

$$\underline{u}_k = \underline{u}_{k-1}\mathbf{P}. \tag{1.1}$$

where the transition matrix $\mathbf{P}$ is a SM. If one sets $\underline{u}_{k-1} = \underline{u}_k$, (1.1) becomes the unit eigenvalue equation, whose normalized solution $\underline{u}_k/|\underline{u}_k| = \underline{\rho}$ is the steady-state distribution of the MC associated with $\mathbf{P}$[2] .

Given an arbitrary square matrix, the direct eigenvalue problem (DEP) consists in finding its eigenvalue spectrum. Many methods exist to solve the DEP, numerically

*STMicroelectronics, TRD/CCDS/eSRAM, +33(0)4 7692 5032 Fax : +33(0)4 7692 5331; 850, Rue Jean Monnet, Crolles-38926 (lorenzo.ciampolini@st.com).

†MGMI team, LJK Laboratory LJK, University of Grenoble

‡STMicroelectronics, Crolles

§STMicroelectronics, Crolles

[1]This is the row-stochastic convention [13]. Some authors [12, 9, 10] use instead the column-stochastic convention and thus a transposed transition matrix $\mathbf{Q} = \mathbf{P}^T$, where the sum of the elements along a column is equal to one. It is an equivalent notation; If $\lambda_i\underline{u}_i = \mathbf{Q}\underline{u}_i$, then $\lambda_i\underline{v}_i = \lambda_i\underline{u}_i^T = \underline{u}_i^T\mathbf{Q}^T = \underline{u}_i^T\mathbf{P}$; $\mathbf{Q}$ and $\mathbf{P}$ share the same eigenvalues, but the (right) column eigenvectors $\underline{u}_i$ of $\mathbf{Q}$ are (left) row eigenvectors $\underline{u}_i^T$ of $\mathbf{P}$, and vice versa. Left and right eigenvectors coincide in general only for symmetrical or self-adjoint matrices.

[2]Instead of indicating $\underline{\rho}$ as the normalized left eigenvector relative to the principal eigenvalue of $\mathbf{P}$, this text indicates sometimes for the sake of simplicity $\underline{\rho}$ as the steady-state distribution of $\mathbf{P}$.

or not [2], and most of them can be applied to an arbitrary SM. The matter is more delicate for the inverse eigenvalue problem (IEP), which consists in the reconstruction of a matrix from a given eigenvalue spectrum. In [3], the authors state the role of an IEP as "validating, determining, or estimating the unknown parameters of a system according to its observed or expected behavior". Since SMs are used to describe MCs, the IEP for a SM is tied to the problem of finding a valid transition matrix for a given eigenvalue spectrum — if no information is available on the physical mechanism driving the transitions from one state to the other. For instance, if the transition coefficients are unknown, an observed MC cannot be modeled numerically.

The trivial solution of the IEP for square matrices (a diagonal matrix holding all eigenvalues) does not work at all for SMs, which have an important constraint on the sum of all elements of their rows. This particular structure makes the *Inverse Stochastic Spectrum Problem* [4] (ISSP) a nonlinear problem, since the sum of two SMs is not a SM. ISSP is one amongst many existing *Structured IEP* [3], i.e. inverse eigenvalue problems applied to matrices having a given structure. Since it is possible to map through similarity transformations non-negative matrices to SMs [3], it is known that a solution to ISSP solves also IEP for non-negative matrices. However, to find the solution to these problems in the general case is an open problem [5], and answers are available only for small matrix size [6]. Algorithms exist to numerically construct the solutions of some structured problems [7, 8] and of the ISSP [4], using iterative methods that in some cases are unexpectedly slow to converge [4]. Indeed, iterative methods for the IEP typically have a rate of convergence that depends on the modulus of the second largest eigenvalue [9, 10], and a general algorithm that would cover all kinds of eigenvalue spectra is still to be found.

In this paper, we present a simple method to directly, i.e. without iterations nor numerical approximations, construct a SM from a given real eigenvalue spectrum through an elementary Markov state disaggregation method named state splitting and based on a matrix operator. After showing some results on how the state splitting operator influences the SM steady-state distribution, we show that the state splitting operator has a fundamental property: when applied to a SM $\mathbf{A}$ of size n-by-n, it yields a SM of size (n+1)-by-(n+1), whose eigenvalue spectrum is equal to that of $\mathbf{A}$, plus an additional eigenvalue, that belongs to a bounded interval. We use a construction method to prove that a SM matrix can be associated with any spectra made of real positive eigenvalues. Finally, we give a new sufficient condition for a set of real values to be the spectrum of a SM. Numerical experiments illustrating this condition conclude the paper.

**2. Splitting operators and state disaggregation.** Definition 2.1. *Given a real value $\lambda$, positive real $a$ such that $|\lambda| < a$ and a real probability factor* $\max\{h_{\min}, 0\} \leq r \leq \min\{h_{\max}, 1\}$*, where $h_{\min} = \lambda/(a+\lambda)$ and $h_{\max} = a/(a+\lambda)$ the scalar splitting operator $\hat{S}(\cdot, \lambda, r)$ is defined as:*

$$\hat{S}(a, \lambda, r) = \frac{a}{h_{\max}} \begin{pmatrix} r & h_{\max} - r \\ r - h_{\min} & 1 - r \end{pmatrix}, \tag{2.1}$$

*a $2 \times 2$ matrix whose second eigenvalue*[3] *is equal to $\lambda$.*

[3]For any square matrix the trace is equal to the sum of the eigenvalues and, by the Perron–Frobenius theorem, the principal (largest) eigenvalue of a SM is equal to 1. Since $(1/a)\hat{S}(a, \lambda, r)$ is stochastic (see Lemma 2.2), its trace being equal to $1/h_{\max} = 1 + \lambda/a$, its second eigenvalue must be equal to $\lambda/a$. The eigenvalues of $S(a, \lambda, r)$ are then $a$ and $\lambda$.

LEMMA 2.2. $\mathbf{H} := (1/a)\hat{S}(a, \lambda, r)$ *is stochastic.*

*Proof.* The sum of the elements of the first row of $\mathbf{H}$ is one. It is easy to prove that $h_{\max} + h_{\min} = 1$, so the sum of the elements of the second row is also one. All elements are non-negative by the definition of the range for $r$ leading to $\mathbf{H}$ stochastic. ∎

Markov state aggregation and disaggregation are powerful tools to investigate the dynamics of MCs [9, 11] and have been applied for a long time to the DEP [12]. The aggregation method is based on grouping together the Markov states using some criteria in order to simplify the SM holding the transition probabilities. Aggregation and disaggregation methods are based on the rearrangement of the evolution states of the MC; more precisely, they alter the partition [13] of states on which the MC is described.

The scalar 1 can be seen as a (degenerate) SM representing the simplest achievable transition matrix which is defined on the simplest possible partition $\mathfrak{A}$, limited to the certain event. It is a degenerate description because no dynamics is possible, because the certain event stays constant. The operator $\hat{S}(1, \lambda, r)$ splits the certain event into a couple of states, thus creating a new partition $\mathfrak{B}$ composed of those two states on which a system described by the new SM evolves. The splitting operator thus performs an elementary Markov state disaggregation of the certain event.

DEFINITION 2.3. *A block representation isolating the $k$-th column and $k$-th row from the rest of the matrix is obtained by writing the $k$-th row $(a_{k,1}, \cdots, a_{k,n})$ as the concatenation of three different row vectors $\underline{r}_{k1}, a_{k,k}, \underline{r}_{k2}$, where either $\underline{r}_{k1}$ or $\underline{r}_{k2}$ can be empty if $k = 1$ or $k = n$, respectively. Similar consideration for the $k$-th column leads to the definition of the row vectors*

$$\begin{aligned}
\underline{c}_{1k} &= (a_{1,k}, a_{2,k}, \cdots, a_{k-1,k}) \\
\underline{c}_{2k} &= (a_{k+1,k}, a_{k+2,k}, \cdots, a_{n,k}) \\
\underline{r}_{k1} &= (a_{k,1}, a_{k,2}, \cdots, a_{k,k-1}) \\
\underline{r}_{k2} &= (a_{k,k+1}, a_{k,k+2}, \cdots, a_{k,n}),
\end{aligned}$$

*which subsequently leads to the rewriting of matrix $\mathbf{A}$ into:*

$$\mathbf{A} = \left(\begin{array}{c|c|c}
\mathbf{A}_{11} & \underline{c}_{1k}^T & \mathbf{A}_{12} \\ \hline
\underline{r}_{k1} & a_{k,k} & \underline{r}_{k2} \\ \hline
\mathbf{A}_{21} & \underline{c}_{2,k}^T & \mathbf{A}_{22}
\end{array}\right). \tag{2.2}$$

DEFINITION 2.4. *Given a $n \times n$ SM $\mathbf{A}$, an index $1 \le k \le n$, a real value $\lambda$ such that $|\lambda| < a_{k,k}$ and a real probability factor $\max\{h_{\min}, 0\} \le r \le \min\{h_{\max}, 1\}$, where $h_{\min} = \lambda/(a_{k,k} + \lambda)$ and $h_{\max} = a_{k,k}/(a_{k,k} + \lambda)$, and writing $\mathbf{A}$ in the block representation of Definition 2.3, the state splitting operator $\hat{S}_M(\mathbf{A}, \lambda, r, k)$ is defined*

*as the operator that yields the* $(n+1)\times(n+1)$ *matrix (2.3):*

$$\hat{S}_M(\mathbf{A},\lambda,r,k) = \left(\begin{array}{c|cc|c} \mathbf{A}_{11} & r\underline{c}_{1k}^T & (1-r)\underline{c}_{1k}^T & \mathbf{A}_{12} \\ \hline \underline{r}_{k1} & & & \underline{r}_{k2} \\ & \hat{S}(a_{k,k},\lambda,r) & & \\ \underline{r}_{k1} & & & \underline{r}_{k2} \\ \hline \mathbf{A}_{21} & r\underline{c}_{2,k}^T & (1-r)\underline{c}_{2,k}^T & \mathbf{A}_{22} \end{array}\right) \tag{2.3}$$

LEMMA 2.5. $\mathbf{S} = \hat{S}_M(\mathbf{A},\lambda,r,k)$ *is stochastic.*

*Proof.* For the sake of simplicity, and with no loss of generality, let us consider the case $k=n$. One can write

$$\hat{S}_M(\mathbf{A},\lambda,r,k) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,k-1} & ra_{1,k} & (1-r)a_{1,k} \\ \vdots & \ddots & \cdots & \cdots & \cdots \\ a_{k-1,1} & \cdots & a_{k-1,k-1} & ra_{k-1,k} & (1-r)a_{k-1,k} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k}h_{1,1} & a_{k,k}h_{1,2} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k}h_{2,1} & a_{k,k}h_{2,2} \end{pmatrix}, \tag{2.4}$$

where $\mathbf{H} = (h_{i,j})_{1\le i,j\le 2} = (1/a_{k,k})\hat{S}(a_{k,k},\lambda,r)$.

The coefficients on the rows indexed by $\{1,\cdots,k-1\}$ sum to one, since they coincide with the sums of the rows of $\mathbf{A}$ with the same indices. Since $\mathbf{H}$ is stochastic (Theorem 2.2), $a_{k,k}(h_{i,1}+h_{i,2}) = a_{k,k}$ for $i=\{1,2\}$ and the sums of the coefficients on the rows indexed by $\{k,k+1\}$ are also equal to one. For the same reason, the elements in the rows indexed by $\{k,k+1\}$ are non-negative. Since $r$ and $(1-r)$ are non-negative, all elements of $\hat{S}_M(\mathbf{A},\lambda,r,k)$ are finally non-negative, and the matrix is stochastic. ∎

THEOREM 2.6. *The splitting operator* $\hat{S}_M(.,\lambda,r,k)$ *leaves unchanged the steady-state distribution* $\underline{\rho}$ *of the* $n\times n$ *SM* $\mathbf{A}$ *in all of its components except the* $k$*-th component* $\rho_k$*, which is split into two components* $f\rho_k$ *and* $(1-f)\rho_k$*, for* $0\le f\le 1$*, satisfying:*

$$f = (r + r\lambda - \lambda)/(1-\lambda). \tag{2.5}$$

*Proof.* Again, we consider for the sake of simplicity the case $k=n$ with no loss of generality. The unit eigenvalue equation $\underline{\rho} = \underline{\rho}\mathbf{A}$ for the SM $\mathbf{A}$ is a system of $k$ equations, with the $i$-th equation being

$$\rho_1 a_{1,i} + \cdots + \rho_k a_{k,i} + \cdots + \rho_n a_{n,i} = \rho_i. \tag{2.6}$$

Given $\mathbf{S} = \hat{S}_M(\mathbf{A},\lambda,r,k)$ described in (2.4), its unit eigenvalue equation $\underline{\rho'} = \underline{\rho'}\mathbf{S}$ yields $k+1$ equations:

$$\begin{array}{rcll} \rho'_1 a_{1,i} + \cdots + \rho'_{k-1}a_{k-1,i} + \rho'_k a_{k,i} + \rho'_{k+1}a_{k,i} & = & \rho'_i & \text{for} \quad i<k \\ \rho'_1 ra_{1,k} + \cdots + \rho'_{k-1}ra_{k-1,k} + \rho'_k a_{k,k}h_{1,1} + \rho'_{k+1}a_{k,k}h_{2,1} & = & \rho'_k & \text{for} \quad i=k \\ \rho'_1(1-r)a_{1,k} + \cdots + \rho'_{k-1}(1-r)a_{k-1,k} + & & & \\ \qquad + \rho'_k a_{k,k}h_{1,2} + \rho'_{k+1}a_{k,k}h_{2,2} & = & \rho'_{k+1} & \text{for} \quad i=k+1 \end{array} \tag{2.7}$$

A unique set of $k+1$ real, non-negative values $\rho'$ is the solution of the full system of $k+1$ equations, and Theorem 2.6 states that the steady-state distribution $\underline{\rho'}$ can be written in the form

$$(\rho_1, \cdots, \rho_{k-1}, f\rho_k, (1-f)\rho_k). \tag{2.8}$$

Let us define $\rho_{k0} = f\rho_k$ and $\rho_{k1} = (1-f)\rho_k$ to simplify the notation, and test if the vector described by (2.8) is the steady-state distribution that satisfies (2.7). Adding together $\rho_{k0}$ and $\rho_{k1}$, one finds immediately

$$\rho_{k0} + \rho_{k1} = \rho_k. \tag{2.9}$$

Let us suppose that $\underline{\rho'}$ is given by (2.8): the first set of equations in (2.7) can then be rewritten as

$$\begin{aligned} & \rho_1 a_{1,i} + \cdots + \rho_{k-1} a_{k-1,i} + \rho_{k0} a_{k,i} + \rho_{k1} a_{k,i} = \rho_i \\ \Leftrightarrow \quad & \rho_1 a_{1,i} + \cdots + \rho_{k-1} a_{k-1,i} + \rho_k a_{k,i} = \rho_i, \end{aligned}$$

which is true since $\underline{\rho} = \underline{\rho}\mathbf{A}$.

The $k$-th equation and $k+1$-th equation are not independent due to (2.9), so only the $k$-th equation needs to be considered:

$$\rho_1 r a_{1,k} + \cdots + \rho_{k-1} r a_{k-1,k} + \rho_{k0} a_{k,k} h_{1,1} + \rho_{k1} a_{k,k} h_{2,1} = \rho_{k0} \tag{2.10}$$

Adding $\rho_k r a_{k,k}$ on both sides and rearranging terms, the last equation becomes

$$\rho_k r = \rho_{k0} + \rho_k r a_{k,k} - \rho_{k0} a_{k,k} h_{1,1} - (\rho_k - \rho_{k0}) a_{k,k} h_{2,1}. \tag{2.11}$$

Rearranging terms one has

$$\rho_{k0}[1 + a_{k,k}(h_{2,1} - h_{1,1})] = \rho_k(r - r a_{k,k} + a_{k,k} h_{2,1}). \tag{2.12}$$

From Definition 2.1, $h_{2,1} = (r - h_{\min})/h_{\max} = [r(a_{k,k} + \lambda) - \lambda]/a_{k,k}$ and $h_{2,1} - h_{1,1} = -\lambda/a_{k,k}$. Substituting these expressions in (2.12 ), one gets

$$\rho_{k0}(1-\lambda) = \rho_k(r + r\lambda - \lambda). \tag{2.13}$$

In conclusion, if (2.13) is verified, the unit eigenvalue equation is satisfied by an eigenvector of the form (2.8).

From this we deduce that

$$f = (r + r\lambda - \lambda)/(1-\lambda). \tag{2.14}$$

Then since $|\lambda| < a_{k,k} \leq 1$, $0 \leq f \leq 1$ corresponds to $0 \leq r \leq 1/(1+\lambda)$ which is true. Indeed, if $\lambda \geq 0$ we get $r \leq a_{k,k}/(a_{k,k} + \lambda) \leq 1/(1+\lambda)$ since $a_{k,k} \leq 1$. On the contrary, when $\lambda \leq 0$, since $r \leq 1$ and $1/(1+\lambda) \geq 1$, we get the expected result. □

Let us now consider as an example the SM $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} 0.95 & 0.05 \\ 0.45 & 0.55 \end{pmatrix}.$$

Fig. 2.1, left, shows a finite-state machine representation of the MC associated with $\mathbf{P}$.

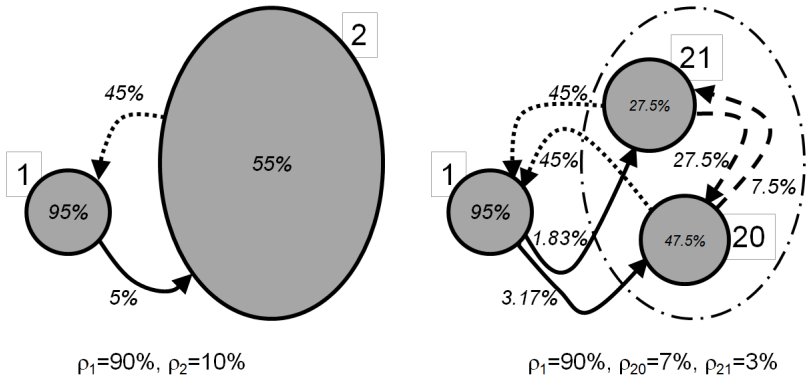


FIG. 2.1. *Finite-state machine representation of a MC having two states (left), and of a MC having three states (right). The two MCs are associated with matrices* $\mathbf{P}$ *and* $\mathbf{Q}$ *described in the text. The shape area for each state is not proportional to the state steady-state probability* $\rho$ *to find the system in that state.*

The steady-state distribution of this MC is $\underline{\rho} = (0.9, 0.1)$. The transition probabilities are shown near the corresponding arrows, and the probabilities for each state to evolve to the same state (self-transition probabilities) are written at the center of the states. Let us calculate $\mathbf{Q} = \hat{S}_M(\mathbf{P}, 0.2, 0.7, 2)$, whose three-digit approximation is

$$\tilde{\mathbf{Q}} = \begin{pmatrix} 0.95 & 0.0317 & 0.0183 \\ 0.45 & 0.475 & 0.075 \\ 0.45 & 0.275 & 0.275 \end{pmatrix}.$$

Note that by fixing $|\lambda| < p_{2,2}$, one determines an interval in which $r$ can be chosen. Fig. 2.1, right, shows the finite-state machine representation of the MC associated to the SM $\tilde{\mathbf{Q}}$, which has steady-state distribution $\underline{\rho'} = (0.9, 0.07, 0.03)$. States 20 and 21 not only have different steady-state probabilities, but also many differences in their transition probabilities.

Nevertheless, the total transition probability from state 1 to any other state is the same (and equal to 5%) across the two machines of Fig. 2.1, and vice versa the transition probability towards state 1 from the other states is 45%, no matter which of 20 or 21 is the starting state. Since state 1 also has the same self-transition probability, one can conclude that this state has exactly the same dynamics in the two machines, and its steady-state probability must be the same in the two MCs. If one wanted to consider the two states 20 and 21 of the three-state machine as a single state (dot-dashed line in Fig. 2.1, right), one would obtain exactly the machine of Fig. 2.1, left. The two MCs are indeed closely related; in particular, one can speculate that both machines describe *the same MC*, except that Fig. 2.1, left is a lower-resolution representation of the MC of Fig. 2.1, right, obtained if *no matter which of the states 20 or 21 are observed as a single state 2*. It is worth noting here that given a lower resolution for the MC, there exist infinitely many choices for the higher resolution extension each of them being related to a particular choice for $\lambda$ and $r$.

From a probabilistic point of view, let us consider a $n \times n$ SM $\mathbf{A}$ defined on a partition $\mathfrak{A}$ of $n$ states. The application of the splitting operator $\hat{S}_M(\cdot, \lambda, r, k)$ to $\mathbf{A}$ yields a $(n+1) \times (n+1)$ SM, that describes a MC on a partition $\mathfrak{B}$ of $n+1$ states. The splitting operator can be seen as an operator that performs an elementary state disaggregation of a MC, as already highlighted in the discussion of Fig. 2.1. More

generally, this operator transfers the description of a MC from a given event partition $\mathfrak{A}$ of $n$ states to a new event partition $\mathfrak{B}$, which is obtained from the former by splitting its $k$-th state into two sub-states $k0$ and $k1$. In this sense, the operator defined in Definition 2.4 is a *state splitting* operator.

**3. State splitting operator and eigenvalue spectrum.** The state splitting operator has a fundamental property with respect to the eigenvalue spectrum of the matrix to which it is applied.

THEOREM 3.1. *The eigenvalue spectrum of* $\mathbf{S} = \hat{S}_M(\mathbf{A}, \lambda, r, k)$ *denoted by* $\lambda(\mathbf{S})$ *is* $\{\lambda(\mathbf{A}), \lambda\}$*, where* $\lambda(\mathbf{A})$ *is the eigenvalue spectrum of* $\mathbf{A}$*.*

*Proof.* The characteristic polynomial of $\mathbf{S}$ in the determinant form is $\det(\mathbf{S} - xI)$. Let us suppose for the sake of simplicity that $k = n$, with no loss of generality. The explicit form of the determinant is:

$$\begin{vmatrix} a_{1,1} - x & \cdots & a_{1,k-1} & ra_{1,k} & (1-r)a_{1,k} \\ \vdots & \ddots & \cdots & \cdots & \cdots \\ a_{k-1,1} & \cdots & a_{k-1,k-1} - x & ra_{k-1,k} & (1-r)a_{k-1,k} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k}h_{1,1} - x & a_{k,k}h_{1,2} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k}h_{2,1} & a_{k,k}h_{2,2} - x \end{vmatrix} \tag{3.1}$$

to which a pair of linear transformations on columns and rows are applied, that leave the determinant unchanged by virtue of known properties of determinants [14]. The $k+1$-th column of $\mathbf{S} - xI$ is added to its $k$-th column, yielding

$$\begin{vmatrix} a_{1,1} - x & \cdots & a_{1,k-1} & a_{1,k} & (1-r)a_{1,k} \\ \vdots & \ddots & \cdots & \cdots & \cdots \\ a_{k-1,1} & \cdots & a_{k-1,k-1} - x & a_{k-1,k} & (1-r)a_{k-1,k} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k} - x & a_{k,k}h_{1,2} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k} - x & a_{k,k}h_{2,2} - x \end{vmatrix}. \tag{3.2}$$

Subtracting row $k$ to row $k+1$, one obtains

$$\begin{vmatrix} a_{1,1} - x & \cdots & a_{1,k-1} & a_{1,k} & (1-r)a_{1,k} \\ \vdots & \ddots & \cdots & \cdots & \cdots \\ a_{k-1,1} & \cdots & a_{k-1,k-1} - x & a_{k-1,k} & (1-r)a_{k-1,k} \\ a_{k,1} & \cdots & a_{k,k-1} & a_{k,k} - x & a_{k,k}h_{1,2} \\ 0 & \cdots & 0 & 0 & a_{k,k}(h_{2,2} - h_{1,2}) - x \end{vmatrix}, \tag{3.3}$$

and then, developing the determinant with respect to row $k+1$, that $\det(\mathbf{S} - xI) = \det(\mathbf{A})(a_{k,k}(h_{2,2} - h_{1,2}) - x)$. At the same time, using $a_{k,k}(h_{2,2} - h_{1,2}) = \lambda$, one finally gets

$$\det(\mathbf{S} - xI) = (\lambda - x)\det(\mathbf{A} - xI), \tag{3.4}$$

where the identity matrices on the two sides have different sizes. ∎

**4. Solution to the Inverse Eigenvalue Problem.** Theorem 3.1 allows in principle one to solve an IEP in a direct way, i.e. by *constructing* the solution matrix by repeated applications of the state splitting operator, which inserts one eigenvalue after the other in a matrix of growing size. This looks a very powerful method, which is extremely scalable to matrices of very large sizes. However, a successful construction

is possible only if the condition $|\lambda| < a_{i,i}$ in Definition 2.4 is satisfied at each step. After some splitting operations, the resulting matrix does not allow one to insert an arbitrary eigenvalue $\lambda$.

More generally, SMs obtained from the state splitting operator are a small subset of all possible SMs, since parameter $\lambda$ takes only real values, whereas SM may have complex eigenvalues. It is therefore interesting to determine under which conditions it is possible to construct the IEP solution using the state splitting operator and starting from a given set of $n$ real values (or real $n$-tuple). To know whether the construction is feasible is both of theoretical interest and practical importance, since it provides a new insight into eigenvalues of SM, as the next result shows.

THEOREM 4.1. *Any real and positive $n$-tuple $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ for which* $\max\{\lambda_i\} = 1$ *is the spectrum of infinitely many SMs.*

*Proof.* Let us suppose without loss of generality that the given values are in decreasing order, i.e. $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. From now on, the upper index inside parentheses will stand for a step index. Putting $\mathbf{A}^{(1)} = 1$, $\mathbf{A}^{(2)} = \hat{S}(1, \lambda_2, r^{(2)}, 2)$ can be constructed with a random $\lambda_2/(1+\lambda_2) < r^{(2)} < 1/(1+\lambda_2)$, the spectrum of $\mathbf{A}^{(2)}$ being $\{\lambda_1, \lambda_2\}$ by Definition 2.1.

The recursive law $\mathbf{A}^{(i)} = \hat{S}_M(\mathbf{A}^{(i-1)}, \lambda_i, r^{(i)}, i)$, with $\max\{0, h^{(i)}_{\min}\} \leq r^{(i)} \leq \min\{1, h^{(i)}_{\max}\}$ with $h^{(i)}_{\min} = \lambda_i/(a^{(i-1)}_{i-1,i-1} + \lambda_i)$ and $h^{(i)}_{\max} = a^{(i-1)}_{i-1,i-1}/(a^{(i-1)}_{i-1,i-1} + \lambda_i)$, defines for $i \geq 3$ a sequence of SMs. Indeed, the diagonal elements of $\mathbf{A}^{(2)}$ are $r^{(2)}(1+\lambda_2)$ and $(1 - r^{(2)})(1 + \lambda_2)$. Because $0 \leq \lambda_2 \leq 1$, we actually have $h^{(2)}_{\min} \leq r^{(2)} \leq h^{(2)}_{\max}$ so that we can write $r^{(2)} = (1-\epsilon)h^{(2)}_{\max} + \epsilon h^{(2)}_{\min}$ from which we deduce the diagonal coefficient of $\mathbf{A}^{(2)}$ are $1 - \epsilon + \epsilon\lambda_2$ and $(1-\epsilon)\lambda_2 + \epsilon$ both belonging to $[\lambda_2, 1]$. From this, we immediately deduce that $\lambda_3 \leq \lambda_2 \leq a^{(2)}_{2,2}$ and that the recursive law can be applied for $i = 3$.

Let us suppose that the construction has been successful for the first $q$ values of the $n$-tuple, i.e. one has $\lambda_i \leq a^{(i-1)}_{i-1,i-1}$ for $2 \leq i \leq q$. Then, the last coefficient on the diagonal of $\mathbf{A}^{(q)}$ is $a^{(q)}_{q,q} = (a^{(q-1)}_{q-1,q-1} + \lambda_q)(1 - r^{(q)}) = (1-\epsilon)\lambda_q + \epsilon a^{(q-1)}_{q-1,q-1}$, writing again $r^{(q)} = (1-\epsilon)h^{(q)}_{\max} + \epsilon h^{(q)}_{\min}$ ($\epsilon$ can be chosen with different value at each step, we just note $\epsilon$ for the sake of simplicity). So we deduce from this that $a^{(q)}_{q,q}$ belongs to $[\lambda_q, a^{(q-1)}_{q-1,q-1}]$ so that one has $\lambda_{q+1} \leq \lambda_q \leq a^{(q)}_{q,q}$. This finally implies that $\hat{S}_M(\mathbf{A}^{(q)}, \lambda_{q+1}, r^{(q+1)}, q+1)$ is stochastic by Lemma 2.5. By induction, one can therefore state that it is possible to construct $\mathbf{A}^{(n)}$, such that its spectrum is the given $n$-tuple. This result has $n$ degrees of freedom, since it holds independently on the chosen $n$-tuple $(r^{(1)}, \cdots, r^{(n)})$, provided each $r^{(i)}$ lies within its valid range. ∎

If one allows the values to be negative, not all real $n$-tuples may be the spectrum of a SM:

LEMMA 4.2. *A set of necessary conditions for an arbitrary real $n$-tuple $\{\lambda_i\} = \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ to be the spectrum of at least one SM is*

$$\max\{\lambda_i\} = \max\{|\lambda_i|\} = 1 \tag{4.1}$$

$$\sum_{i=1}^{n} \lambda_i \geq 0. \tag{4.2}$$

*Proof.* (4.1) is a known condition, tied to the fact that unity is the largest eigenvalue of SMs, after the Perron–Frobenius theorem. The left-hand side in (4.2) is the

trace of a SM with spectrum $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$. Since all elements in a SM are non-negative, and in particular the diagonal elements, the trace must be also non-negative, as required by (4.2). $\square$

One should note here that this necessary condition is very different from the sufficient condition found by Suleimanova in 1949 [3]. In this regard, we believe that condition to be far too restrictive. For instance, the following SM

$$\begin{pmatrix} 0.0025 & 0.9925 & 0.0025 & 0.0025 \\ 0.9925 & 0.0025 & 0.0025 & 0.0025 \\ 0.0025 & 0.0025 & 0.0075 & 0.9875 \\ 0.0025 & 0.0025 & 0.9875 & 0.0075 \end{pmatrix}$$

has an eigenvalue spectrum (1,0.99,-0.99,-0.98), even if it does not satisfy Suleimanova's condition $\sum_{i=p+1}^{n} |\lambda_i| \leq 1$, where the summation runs over all negative eigenvalues (supposing with no loss of generality that the first $p$ eigenvalues are positive). We now propose a new result that generalizes Suleimanova's:

THEOREM 4.3. *A real $n$-tuple $\underline{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ having $p$ positive values is the eigenvalue spectrum of at least one SM if it satisfies (4.1-4.2) and if the $(n-p)$ negative values can be grouped into $p$ groups (that can be empty), which can be mapped to the $p$ positive values of the $n$-tuple. The absolute value of the sum of the negative values of each group must be lesser than the corresponding positive value to which the group is mapped.*

Remark: Note that Suleimanova's result is a particular case of Theorem 4.3, where all negative eigenvalues are mapped to eigenvalue 1. Fig. 4 shows an example of the mapping required by Theorem 4.3.

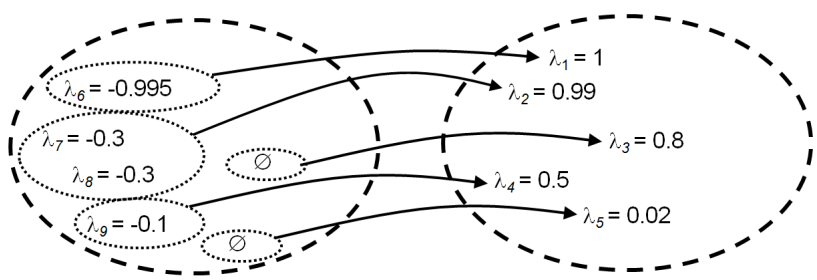


FIG. 4.1. *Arrows materialize the mapping between groups of negative values (dotted-line sets on the left) and isolated positive values (on the right). The set of all values shown in this example satisfies the conditions of Theorem 4.3.*

*Proof.* Let us suppose without loss of generality that $\underline{\lambda}$ is sorted in decreasing order and that the $p$ first values are positive, i.e. $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$. In what follows, an index inside parentheses will stand for a step index. Let us put $\mathbf{A}^{(1)} = 1$ and then define $h^{(2)}_{\min} = \lambda_2/(1+\lambda_2)$ and $h^{(2)}_{\max} = 1/(1+\lambda_2)$. In such a case, following Definition 2.1, one can define for any $h^{(2)}_{\min} \leq r^{(2)} \leq h^{(2)}_{\max}$, a SM matrix as $\mathbf{A}^{(2)} = \hat{S}(1, \lambda_2, r^{(2)}, 2)$. Putting $r^{(2)} = (1-\epsilon)h^{(2)}_{\max} + \epsilon h^{(2)}_{\min}$, one can rewrite $\mathbf{A}^{(2)}$ as:

$$\mathbf{A}^{(2)} = \begin{pmatrix} 1-\epsilon(1-\lambda_2) & \cdots \\ \cdots & \lambda_2 + \epsilon(1-\lambda_2) \end{pmatrix}.$$

Let us now define, for $3 \leq i \leq p$, $h^{(i)}_{\min} = \lambda_i/(a^{(i-1)}_{i-1,i-1} + \lambda_i)$ and $h^{(i)}_{\max} = a^{(i-1)}_{i-1,i-1}/(a^{(i-1)}_{i-1,i-1} + \lambda_i)$ for $i \leq p$, where $a^{(i-1)}_{i-1,i-1}$ is the $i-1$-th diagonal element of matrix $\mathbf{A}^{(i-1)}$ and

finally $r^{(i)} = (1-\epsilon)h^{(i)}_{\max} + \epsilon h^{(i)}_{\min}$, $\epsilon$ being the same at each step. We thus obtain a sequence of stochastic matrices as follows: $\mathbf{A}^{(i)} = \hat{S}_M(\mathbf{A}^{(i-1)}, \lambda_i, r^{(i)}, i)$ and a simple computation shows that:

$$(a^{(p)}_{1,1}, \cdots, a^{(p)}_{p,p}) = (1 + O(\epsilon), \lambda_2 + O(\epsilon), \cdots, \lambda_p + O(\epsilon)) \tag{4.3}$$

where $O(\epsilon)$ tends to 0 as $\epsilon$ tends to 0. Then, one defines $\mathbf{A}^{(p+1)} = \hat{S}_M(\mathbf{A}^{(p)}, \lambda_{p+1}, r^{(p+1)}, j)$, where $\lambda_{p+1}$ is the first negative value and where $\max\{0, h^{(p+1)}_{\min}\} \leq r^{(p+1)} \leq \min\{1, h^{(p+1)}_{\max}\}$, where $h^{(p+1)}_{\min}$ and $h^{(p+1)}_{\max}$ are defined the same way as for smaller step indices. We know that $\mathbf{A}^{(p+1)}$ is a SM provided $1 \leq j \leq p$ satisfies:

$$|\lambda_{p+1}| < a^{(p)}_{j,j}. \tag{4.4}$$

Note that since $\lambda_{p+1} < 0$, $0 \leq r^{(p+1)} \leq 1$, so that we can write $r^{(p+1)} = 1 - \epsilon$. The diagonal coefficients of matrix $\mathbf{A}^{(p+1)}$ which are all positive read:

$$(a^{(p)}_{1,1}, \cdots, a^{(p)}_{j-1,j-1}, (a^{(p)}_{j,j} + \lambda_{p+1})(1-\epsilon), (a^{(p)}_{j,j} + \lambda_{p+1})\epsilon, a^{(p)}_{j+1,j+1}, \cdots, a^{(p)}_{p,p}).$$

Let us now consider a second negative value, e.g. $\lambda_q$, with $p+1 < q \leq n$. If $a^{(p)}_{j,j} + \lambda_{p+1} + \lambda_q > 0$, then $|\lambda_q| < a^{(p+1)}_{j,j}$, and a new split operation $\hat{S}_M(\mathbf{A}^{(p+1)}, \lambda_q, 1-\epsilon, j)$ can be performed on the same $j$-th state. In general, if the $n$-tuple contains $l$ negative values $\{\lambda_{i_1}, \cdots, \lambda_{i_l}\}$ such that

$$a^{(p)}_{j,j} + \lambda_{i_1} + \cdots + \lambda_{i_l} > 0, \tag{4.5}$$

then one can perform a sequence of $l$ splitting operations on the same $j$-th state, because $a^{(p+m)}_{j,j}$ remains positive for $1 \leq m \leq l$. Then (4.3) shows that $a^{(p)}_{j,j}$ can be set arbitrarily close to $\lambda_j$ by taking a small enough $\epsilon$ which in turn means that (4.5) implies

$$|\lambda_{i_1} + \cdots + \lambda_{i_l}| < \lambda_j. \tag{4.6}$$

One can reorder the negative values in $p$ groups (that can be empty), denote with $l(j)$ the number of elements of the $j$-th group (with $j \in \{1, \cdots, p\}$ and $0 \leq l(j) \leq n-p$) and indicate as $i(j,k)$ the place in $\underline{\lambda}$ of the $k$-th element of the $j$-th group (if the group is non-empty). Such an arrangement of the negative values must satisfy the condition that all of them must be present once and only once[4] across all groups, i.e.

$$I = \bigcup_{\forall j \text{ s. t. } l(j)>0} \{i(j,1), \cdots, i(j,l(j))\} \tag{4.7}$$

must be a permutation without repetitions of $(p+1, \cdots, n)$.

If the elements are grouped in such a way that

$$\left| \sum_{k=1}^{l(j)} \lambda_{i(j,k)} \right| \leq \lambda_j \tag{4.8}$$

[4]Given $s > p$ and $t > p$ with $s \neq t$ ($s$ and $t$ are indexes to two negative values of $\underline{\lambda}$), if $\lambda_s = \lambda_t$, the two equal values are treated as two different items of the $n$-tuple.

for every group for which $l(j) > 0$, then it is possible to apply $l(j)$ times the state splitting operator to the $j$-th element of $\mathbf{A}^{(p)}$, with arguments $\lambda_{i(j,1)}, \cdots, \lambda_{i(j,l(j))}$. The result is a SM by virtue of Lemma 2.5, with the additional eigenvalues $\lambda_{i(j,1)}, \cdots, \lambda_{i(j,l(j))}$ by virtue of Theorem 3.1. Since this holds for any $j$, it is possible to apply the state splitting operator $n-p$ times, and obtain finally a SM $\mathbf{A}^{(n)}$ whose eigenvalue spectrum is the given $n$-tuple. $\square$

However, Theorem 4.3 is not a necessary condition since it is known the $n$-tuple (1,0.75,0.7,0.1,-0.75,-0.8,-0.8) is the eigenvalue spectrum of at least one SM [15]. The proposed splitting operator has been implemented numerically in MATLAB, and its properties have been verified within the numerical precision using the functions provided by the computing environment. We wrote an algorithm that, after applying the splitting operator with $\epsilon = 10^{-6}$ on the positive values taken in decreasing order, applies the splitting operator with $\epsilon = 0.5$ on the negative values. For simple groups of negative values, it is sufficient to apply the operator on the state that has the largest self-transition probability.

As an example of a $n$-tuple satisfying Theorem 4.3, but not Suleimanova's condition, one can form with the given algorithm a SM that has for eigenvalue spectrum the $n$-tuple $\underline{\lambda} = (1, 0.98, 0.6, 0.5, 0.45, 0.4, 0.3, -0.97, -0.99)$, for which the *eig* MATLAB function yields a vector whose largest difference with the true $\underline{\lambda}$ is $2 \times 10^{-15}$. We reproduce here the resulting matrix within a $10^{-4}$ precision:

$$\begin{pmatrix}
.0099 & .9901 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.9999 & .0001 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.0198 & .0002 & .0002 & .97 & .0065 & .0023 & .0007 & .0002 & 0 \\
.0198 & .0002 & .9702 & .0 & .0065 & .0023 & .0007 & .0002 & 0 \\
.0198 & .0002 & .3665 & .0037 & .6 & .0042 & .0014 & .0004 & .0001 \\
.0198 & .0002 & .3665 & .0037 & .1037 & .501 & .0036 & .0011 & .0003 \\
.0198 & .0002 & .3665 & .0037 & .1037 & .051 & .4505 & .0035 & .0011 \\
.0198 & .0002 & .3665 & .0037 & .1037 & .051 & .0505 & .4005 & .0040 \\
.0198 & .0002 & .3665 & .0037 & .1037 & .051 & .0505 & .1005 & .3040
\end{pmatrix}.$$

**5. Discussion and Conclusions.** In this paper, we have presented a simple method to directly construct a SM from a given eigenvalue spectrum through an elementary state disaggregation step named state splitting, described in terms of a matrix operator. We show that a splitting operation yields a SM having a steady-state distribution very close to the one of the original SM, so that the operator might be called a *state splitting* operator.

The state splitting operator has a fundamental property: when applied to a SM $\mathbf{A}$, it yields a matrix whose eigenvalue spectrum is equal to that of $\mathbf{A}$, plus an additional eigenvalue, which is *de facto* inserted by the operator. This property leads quite naturally to the solution of the IEP through step-by-step construction of a SM, inserting one by one the desired eigenvalues.

The state splitting operator has been used to prove quickly that all $n$-tuples of real and positive eigenvalues are realizable as eigenvalue spectra of infinitely many SM. A new, general condition has been found to prove the possibility for an arbitrary $n$-tuple with positive and negative values to be the eigenvalue spectrum of a SM.

REFERENCES

[1] S. Meyn and R. L. Tweedie, *Matrix Computations*, 2nd ed., Cambridge Univ. Press, Cambridge, UK, 2009.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., The Johns Hopkins University Press, Baltimore, MD, 1989.

[3] M. T. Chu and G. H. Golub, *Structured Inverse Eigenvalue Problems*, Acta Numer., 11 (2002), pp. 1-71.

[4] M. T. Chu and Q. Guo, *A Numerical Method for the Inverse Stochastic Spectrum Problem*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 1027-1039.

[5] J. L. Wu, *On Open Problems of Nonnegative Inverse Eigenvalues Proplem*, Adv. in Pure Math., 1 (2011), pp. 128-132.

[6] G. W. Soules, *Constructing Symmetric Nonnegative Matrices*, Linear Algebra Appl., 13 (1983), pp. 241-251.

[7] B. Z. Jian, *Numerical Methods for Inverse Eigenvalue Problems*, PhD dissertation, The Chinese University of Hong Kong, PRC, 2004.

[8] O. Rojo and H. Rojo, *Constructing Symmetric Nonnegative Matrices via the Fast Fourier Transform*, Comput. Math. Appl., 45 (2003), pp. 1655-1672.

[9] M. Haviv, *Aggregation/Disaggregation Methods for Computing the Stationary Distribution of a Markov Chain*, SIAM J. on Numer. Anal., 24 (1987), pp. 952-966.

[10] T. H. Haveliwala and S. D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Technical report 2003-20, Stanford Univ., CA, 2003.

[11] H. D. Kafeety, C. D. Meyer and W. J. Stewart, *A General Framework for Iterative Aggregation/Disaggregation Methods*, in Proceedings of the Fourth Copper Mountain Conference on Iterative Methods, 1992.

[12] F. Chatelin and L. W. Miranker, *Aggregation/Disaggregation for Eigenvalue Problems*, SIAM J. on Numer. Anal., 21 (1984), pp. 567-582.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill int., 1991.

[14] A. Jeffrey and H. H. Dai, *Handbook of Mathematical Formulas and Integrals*, 4th ed., Elsevier Press, Burlington, MA, USA, 2008.

[15] Y. Shang-jun and X. Chang-Qing, *Row Stochastic Inverse Eigenvalue Problem*, J. Inequal. Appl., 2011:24 (2011).