Centrum voor Wiskunde en Informatica

A stochastic uncoupling process for graphs

S. van Dongen

Information Systems (INS)

INS-R0011 May 31, 2000

Report INS-R0011
ISSN 1386-3681

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

A Stochastic Uncoupling Process for Graphs

Stijn van Dongen
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT

A discrete stochastic uncoupling process for finite spaces is introduced, called the *Markov Cluster Process*. The process takes a stochastic matrix as input, and then alternates flow expansion and flow inflation, each step defining a stochastic matrix in terms of the previous one. Flow expansion corresponds with taking the k^{th} power of a stochastic matrix, where $k \in I\!N$. Flow inflation corresponds with a parametrized operator Γ_r, $r \geq 0$, which maps the set of (column) stochastic matrices onto itself. The image $\Gamma_r M$ is obtained by raising each entry in M to the r^{th} power and rescaling each column to have sum 1 again.

In practice the process converges very fast towards a limit which is idempotent under both matrix multiplication and inflation, with quadratic convergence around the limit points. The limit is in general extremely sparse and the number of components of its associated graph may be larger than the number associated with the input matrix. This uncoupling is a desired effect as it reveals structure in the input matrix.

The inflation operator Γ_r is shown to map the class of matrices which are diagonally similar to a symmetric matrix onto itself. The term *diagonally positive semi–definite* ($dpsd$) is used for matrices which are diagonally similar to a positive semi–definite matrix. It is shown that for $r \in I\!N$ and for M a stochastic $dpsd$ matrix, the image $\Gamma_r M$ is again $dpsd$. Determinantal inequalities satisfied by a $dpsd$ matrix M imply a natural ordering among the diagonal elements of M, generalizing a mapping of nonnegative column allowable idempotent matrices onto overlapping clusterings. The spectrum of $\Gamma_\infty M$, for $dpsd$ M, is of the form $\{0^{n-k}, 1^k\}$, where k is the number of endclasses of the ordering associated with M, and n is the dimension of M.

Reductions of $dpsd$ matrices are given, a connection with Hilbert's distance and the contraction ratio defined for nonnegative matrices is discussed, and several conjectures are made.

2000 Mathematics Subject Classification: 05B20, 15A48, 15A51, 62H30, 68R10, 68T10, 90C35.
Keywords and Phrases: Markov matrix, flow simulation, stochastic uncoupling, diagonal similarity, positive semi-definite matrices, circulant matrices, reinforced random walk.
Note: This report describes mathematical aspects of the MCL process. The process was introduced in [9] as a means for finding cluster structure in graphs. Cluster experiments are described in [11]. The work was carried out under project INS–3.2, Concept Building from Key–Phrases in Scientific Documents and Bottom Up Classification Methods in Mathematics.

1. Introduction

The subject of this report[1] is an algebraic process defined for stochastic[2] matrices, called the Markov Cluster Process (MCL process). The MCL process consists of alternation of flow expansion and flow inflation, where inflation means taking the Hadamard power of a stochastic matrix and subsequently scaling its columns to have sum 1 again. The process was first defined as a heuristic for a cluster algorithm for graphs, as described in [9]. The underlying idea is that a dense region in a graph corresponds with a node set S for which pairs of elements in S have the property that there are relatively many higher length paths between them, compared

[1] The report corresponds with Chapter 7 in the PhD thesis [10].
[2] Throughout the report stochastic matrices are assumed to be column stochastic.

to pairs of elements from different dense regions. By expansion (corresponding with the usual matrix power of a stochastic matrix) the higher step transition probabilities (*TPs*) are obtained; by inflation large *TPs* are promoted, and small *TPs* are demoted. It is to be expected that *TPs* which correspond with edges connecting different dense regions will suffer the most from the process of alternating expansion and inflation. Indeed, iteration of the two operators leads to a limit which is meaningful considering the original heuristic.

An *MCL* process (defined in Section 2) is characterized by an infinite row of pairs (e_i, r_i), where the e_i are integers greater than one, and the r_i are real numbers greater than zero. An input matrix M yields an infinite row of matrices $M_{(i)}$ by setting $M_1 = M$, defining the even–labeled iterands by $M_{2i} = {M_{2i-1}}^{e_i}$, and the odd–labeled iterands by $M_{2i+1} = \Gamma_{r_i} M_{2i}$ (See Definition 1 for the precise definition of the inflation operator Γ_r). For stochastic matrices diagonally similar to a symmetric matrix, the type of limit invariably found is that of a doubly idempotent matrix; idempotent under both matrix squaring and matrix inflation.

A nonnegative idempotent matrix L without zero columns induces an overlapping clustering on the column indices with the property that each cluster has at least one element not contained within any of the other clusters (Theorem 1 and Definition 5). The number of clusters, say k, is equal to the multiplicity of the eigenvalue 1 in the spectrum of L. The sets of unique elements in the clusters form the strongly connected components in the associated graph of L, the number of which is also k. Experiments yield that initiating an *MCL* process with an input matrix, the associated graph of which has only one strongly connected component, may in general give an idempotent limit with a larger number of connected components. An example showing this behaviour is given in Figure 2. Interpreting the limit according to Definition 5 yields a clustering whose distribution is invariably strongly related to the density characteristics of the input matrix. In this sense, the *MCL* process appears to be useful. The *MCL* process converges quadratically in the neighbourhood of doubly idempotent stochastic matrices for which all columns have precisely one nonzero entry. It converges quadratically on a macroscopic scale (in terms of block structure) for doubly idempotent stochastic matrices in general. The clustering associated with such a matrix is stable under perturbations of the *MCL* process (that is, it is essentially defined by the block structure), except for the phenomenon of overlap. This is covered extensively in the technical report [9].

The limit resulting from an *MCL* process is in general extremely sparse. Current evidence suggests that limit matrices which have columns with more than one nonzero entry imply the existence of a nontrivial automorphism for the underlying graph of the input matrix. Some examples illustrating this behaviour are given in [11]. The sparseness of the limit, and the sparseness in a 'weighted sense' of intermediate iterands have nice and important repercussions for the scalability of the cluster algorithm based on the *MCL* process. This is discussed in [9], and scaled experiments on randomly generated test graphs are described in [11].

The *MCL* process is interesting from a mathematical point of view, since it apparently has the power to 'inflate' the spectrum of a stochastic matrix, by pressing large eigenvalues towards 1. This effect is strong enough to overcome the effect of matrix exponentiation, which has the property of exponentiating the associated eigenvalues. The fundamental property established in this report is that Γ_r maps two nested classes of stochastic matrices with real spectra onto themselves (Theorem 3). The largest class is that of diagonally symmetrizable stochastic matrices, i.e. matrices which are diagonally similar to a symmetric matrix without further constraints. This class is mapped onto itself by Γ_r for arbitrary $r \in I\!R$. Defining *diagonally positive semi–definite* (*dpsd*) as the property of being diagonally similar to a positive semi–definite matrix, the second class is that of stochastic *dpsd* matrices. This class is mapped onto itself by Γ_r for $r \in I\!N$.

Using the property that minors of a *dpsd* matrix A are nonnegative, it is shown that the relation $\looparrowright$ defined on the nodes of its associated graph by $q \looparrowright p \equiv |A_{pq}| \geq |A_{qq}|$, for $p \neq q$, is a directed acyclic graph (*DAG*) if indices of identical[3] columns, resp. rows are lumped together (Theorem 5). This generalizes the mapping

[3]modulo multiplication by a scalar on the complex unit circle.

from nonnegative idempotent column allowable matrices onto overlapping clusterings (Definition 5), and it sheds some light on the tendency of the *MCL* process limits to have a larger number of strongly connected components than the input graph. It is then shown that applying Γ_∞ to a stochastic *dpsd* matrix M yields a matrix[4] which has spectrum of the form $\{0^{n-k}, 1^k\}$, where k, the multiplicity of the eigenvalue 1, equals the number of endclasses of the ordering of the columns of M provided by the associated DAG. It is not necessarily true that $\Gamma_\infty M$ is idempotent. However, the observation is confirmed that Γ_r tends to inflate the spectrum of M for $r > 1$, as $\Gamma_r M$ may be regarded as a function of varying r for fixed M, and as such is continuous.

The structure of this report is as follows. Section 2 consists of definitions and an example of an *MCL* process. It is meant to give the reader some intuition for the process and why it may be interesting. In Section 3 various lemmas and theorems formalizing the results described above are given. Section 4 introduces structure theory for *dpsd* matrices and gives properties of the inflation operator applied to stochastic *dpsd* matrices. Reductions of *dpsd* matrices are the subject of Section 5, and Section 6 is concerned with Hilbert's distance for positive vectors and a contraction ratio defined for nonnegative matrices defined in terms of the Hilbert distance. It is shown that inflation and expansion can both be described in this framework, which allows a simple description of the working of the *MCL* process on perturbed rank 1 stochastic matrices. Conclusions, further research, and related research make up the last section.

2. Definitions and an example of an *MCL* process

This section provides the basic definitions and concepts needed to describe the *MCL* process. Since the process was defined as a heuristic for clustering graphs, the relationship between nonnegative idempotent matrices and overlapping clusterings is established. Excerpts of an example *MCL* process are shown, including the idempotent limit resulting from it.

Submatrices of a matrix A are written $A[u|v]$, where u denotes a list of row indices, and v denotes a list of column indices. Let A be square of dimension n, and assume some ordering on the set of k–tuples with distinct indices. The k^{th} compound of A is the matrix of all minors of order k of A, and is written $\mathrm{Comp}_k(A)$. It has dimension $\binom{n}{k}$. Its pq entry is equal to $\det A[u_p|u_q]$, where u_i is the i^{th} k–tuple of distinct indices in the given ordering.

Following terminology used in [5] and [17], a nonnegative matrix is called column allowable if all its columns have at least one nonzero entry. With each square nonnegative matrix A of dimension n is a weighted graph G associated, defined on the set of indices $\{1, \dots, n\}$, where the weight of the arc going from l to k is defined as A_{kl}. If $A_{kl} = 0$, there is no arc going from l to k. A directed acyclic graph is abbreviated as DAG. Diagonal matrices are written as d_x, where x denotes the vector of diagonal elements. The Hadamard product between matrices (entrywise product) is denoted as $\circ$. The Hadamard power of a matrix A in which each element is raised to the power r, is written $A^{\circ r}$.

The inflation operator Γ_r is defined for arbitrary nonnegative matrices, in a columnwise manner. This implies that column stochastic matrices will be used rather than row stochastic matrices, which is merely a matter of preference and convention. There are no restrictions on the matrix dimensions to fit a square matrix, because this allows Γ_r to act on both matrices and vectors. There is no restriction that the input matrices be stochastic, since it is not strictly necessary, and the extended applicability is sometimes useful.

[4]The matrix $\Gamma_\infty M$ is defined as $\lim_{r\to\infty} \Gamma_r M$, which exists for all stochastic M. See Definition 9.

Definition 1 *Let r be a real nonnegative number, let $M \in \mathbb{R}_{\geq 0}{}^{m \times n}$ be nonnegative column allowable. The image of M under the parametrized operator Γ_r is defined by setting*

$$(\Gamma_r M)_{pq} \;=\; (M_{pq})^r \Big/ \sum_{i=1}^{m} (M_{iq})^r$$

The parameter r is assumed rather than required to be nonnegative. The reason is that in the setting of the *MCL* process nonnegative values r have a sensible interpretation attached to them. Values of r between 0 and 1 increase the homogeneity of the argument probability vector (matrix), whereas values of r between 1 and ∞ increase the inhomogeneity. In both cases, the ordering of the probabilities is not disturbed. Negative values of r invert the ordering, which does not seem to be of apparent use.

Lemma 1 *(Simple properties of Γ_r)*

i) If x is a nonnegative stochastic vector, then $x \prec \Gamma_r x$ for $r > 1$, and $\Gamma_r x \prec x$ for $r < 1$, where $\prec$ denotes the majorization *relationship.*

ii) If A and B are nonnegative matrices, then $\Gamma_r(A \otimes B) = (\Gamma_r A) \otimes (\Gamma_r B)$. where $\otimes$ denotes the usual Kronecker product.

iii) If r and s are real numbers, and A is a nonnegative matrix, then $\Gamma_r(\Gamma_s A) = \Gamma_{rs} A$.

□

These facts are easily verified. Statement i) is not evidently useful, since results from the theory of majorization of vectors do not carry over to matrices in such a straightforward way (i.e. the columns of one matrix majorizing the columns of another matrix). In [25] this issue is discussed at length. However, the statement clearly shows the inflationary or 'decontracting' effect of Γ_r, $r > 1$, as opposed to the contracting effect of multiplication of nonnegative matrices in terms of the so called *Hilbert distance* between positive vectors. This is given more thought in Section 6. Statement ii) is of use in studying the equilibrium states of the *MCL* process.

Definition 2 *Let S be some subset of the reals. Denote the operator which raises a square matrix A to the t^{th} power, $t \in S$, by Exp_t. Thus, $\mathrm{Exp}_t A = A^t$.*

This definition is put in such general terms because the class of diagonally *psd* matrices (to be introduced later) allows the introduction of fractional matrix powers in a well–defined way. Note that in general the identities $\mathrm{Exp}_r(A \otimes B) = \mathrm{Exp}_r(A) \otimes \mathrm{Exp}_r(B)$ and $\mathrm{Exp}_r(\mathrm{Exp}_s(A)) = \mathrm{Exp}_{rs}(A)$ hold. Thus the last two statements of Lemma 1 are valid for the operator Exp_r as well.

Definition 3 *An MCL* **process** *with input matrix M, where M is a stochastic matrix, is determined by M and two rows $e_{(i)}$, $r_{(i)}$, where $e_i \in \mathbb{N}, e_i > 1$, and $r_i \in \mathbb{R}, r_i \geq 0$. It is written*

$$(\ M\ , e_{(i)}, r_{(i)}) \tag{2.1}$$

Associated with an MCL process $(M, e_{(i)}, r_{(i)})$ is an infinite row of matrices $M_{(i)}$ where $M_1 = M$, $M_{2i} = \mathrm{Exp}_{e_i}(M_{2i-1})$, and $M_{2i+1} = \Gamma_{r_i}(M_{2i})$, $i = 1, \ldots, \infty$.

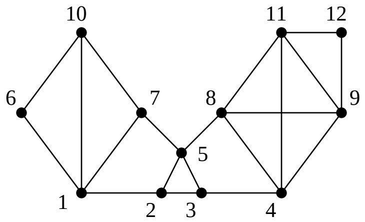

Figure 1: Graph H.

Figure 2 gives four excerpts of an example MCL process, namely the input matrix M, the iterand $M_3 = \Gamma_2 M^2$, the iterand $M_5 = \Gamma_2(\Gamma_2 M^2 \cdot \Gamma_2 M^2)$, and the stable limit denoted L_M. The process consists entirely of alternation of Exp_2 and Γ_2. The graph H associated with M is depicted in Figure 2. Every node in H has a loop; these are all left out in the figure. Weights are omitted as well. Note that there exists a diagonal matrix d such that Md is symmetric. This implies that $d^{-1/2}Md^{1/2}$ is symmetric and thus the spectrum of M is real. The clustering associated with L_M is $\{\{1,6,7,10\}, \{2,3,5\}, \{4,8,9,11,12\}\}$ (See Definition 5), which has good visual appeal. In order to achieve this appeal, it is necessary in this example to add loops to the nodes, in order to prevent a result reflecting the bipartite characteristics of H. Without adding loops, the MCL process limit yields the clustering $\{\{1,5,10\},\{2,6,7\},\{3,4,8,9,11,12\}\}$. This is in line with the heuristic underlying the process: The *TPs* which are initially boosted correspond with 2–step paths in H. The following theorem is preparatory to the mapping from nonnegative idempotent matrices to overlapping clusterings in Definition 5. Its proof is given in [9] and can also be derived from the decomposition of nonnegative idempotent matrices given in [2], page 65.

Theorem 1 *Let M be a nonnegative column allowable idempotent matrix of dimension n, let G be its associated graph. For s, t, nodes in G, write $s \to t$ if there is an arc in G from s to t. By definition, $s \to t \iff M_{ts} \neq 0$. Let α, β, γ be nodes in G. The following implications hold.*

$$(\alpha \to \beta) \wedge (\beta \to \gamma) \implies \alpha \to \gamma \tag{2.2}$$
$$(\alpha \to \alpha) \wedge (\alpha \to \beta) \implies \beta \to \alpha \tag{2.3}$$
$$\alpha \to \beta \implies \beta \to \beta \tag{2.4}$$

□

Definition 4 *Let G be the associated graph of a nonnegative column allowable idempotent matrix M of dimension n, with nodes labeled $1, \ldots, n$. The node α is called an* **attractor** *if $M_{\alpha\alpha} \neq 0$. If α is an attractor then the set of its neighbours is called an* **attractor system**.

By Theorem 1, each attractor system in G induces a weighted subgraph in G which is complete. These form the unique cores of the clustering associated with (nonnegative idempotent) M as stated below.

Definition 5 *Let M be a nonnegative column allowable idempotent matrix of dimension n, let G be its associated graph on the node set $V = \{1, \ldots, n\}$. Let $E_i, i = 1, \ldots, k$ be the different attractor systems of G. For $v \in V$ write $v \to E_i$ if there exists $e \in E_i$ with $v \to e$. Theorem 1 then implies that $v \to f$ for all $f \in E_i$. The (possibly) overlapping clustering $\mathcal{C} = \{C_1, \ldots, C_k\}$, associated with M, is defined by*

$$C_i = E_i \cup \{v \in V \mid v \to E_i\} \tag{2.5}$$

The example in Figure 2 indicates that the MCL process has remarkable convergence properties, regarding the structural properties of its iterands. Considering just this evidence, to some extent an analogy is suggested with the normal Markov process. Assuming that the associated graph of the input matrix M is strongly connected, it is known by Perron–Frobenius theory that 1 is the only eigenvalue of M of modulus 1. By considering the spectrum of the powers M^k it follows that the normal Markov process converges towards a rank–1 idempotent matrix, having spectrum $\{0^{n-1}, 1\}$. In the example shown here the process converges also

$$\begin{pmatrix}
0.2000 & 0.2500 & -- & -- & -- & 0.3333 & 0.2500 & -- & -- & 0.2500 & -- & -- \\
0.2000 & 0.2500 & 0.2500 & -- & 0.2000 & -- & -- & -- & -- & -- & -- & -- \\
-- & 0.2500 & 0.2500 & 0.2000 & 0.2000 & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & 0.2500 & 0.2000 & -- & -- & -- & 0.2000 & 0.2000 & -- & 0.2000 & -- \\
-- & 0.2500 & 0.2500 & -- & 0.2000 & -- & 0.2500 & 0.2000 & -- & -- & -- & -- \\
0.2000 & -- & -- & -- & -- & 0.3333 & -- & -- & -- & 0.2500 & -- & -- \\
0.2000 & -- & -- & -- & 0.2000 & -- & 0.2500 & -- & -- & 0.2500 & -- & -- \\
-- & -- & -- & 0.2000 & 0.2000 & -- & -- & 0.2000 & 0.2000 & -- & 0.2000 & -- \\
-- & -- & -- & 0.2000 & -- & -- & -- & 0.2000 & 0.2000 & -- & 0.2000 & 0.3333 \\
0.2000 & -- & -- & -- & -- & 0.3333 & 0.2500 & -- & -- & 0.2500 & -- & -- \\
-- & -- & -- & 0.2000 & -- & -- & -- & 0.2000 & 0.2000 & -- & 0.2000 & 0.3333 \\
-- & -- & -- & -- & -- & -- & -- & -- & 0.2000 & -- & 0.2000 & 0.3333
\end{pmatrix}$$

M

$$\begin{pmatrix}
0.3801 & 0.0867 & 0.0268 & -- & 0.0767 & 0.2945 & 0.2012 & -- & -- & 0.3195 & -- & -- \\
0.0467 & 0.3469 & 0.2099 & 0.0171 & 0.1503 & 0.0192 & 0.0657 & 0.0115 & -- & 0.0120 & -- & -- \\
0.0144 & 0.2099 & 0.3469 & 0.0555 & 0.1503 & -- & 0.0164 & 0.0460 & 0.0090 & -- & 0.0090 & -- \\
-- & 0.0268 & 0.0867 & 0.3021 & 0.0621 & -- & -- & 0.1839 & 0.1433 & -- & 0.1433 & 0.0828 \\
0.0577 & 0.2099 & 0.2099 & 0.0555 & 0.4057 & -- & 0.0832 & 0.0460 & 0.0090 & 0.0187 & 0.0090 & -- \\
0.1416 & 0.0171 & -- & -- & -- & 0.2945 & 0.0832 & -- & -- & 0.1836 & -- & -- \\
0.1131 & 0.0685 & 0.0171 & -- & 0.0621 & 0.0972 & 0.3326 & 0.0115 & -- & 0.1466 & -- & -- \\
-- & 0.0171 & 0.0685 & 0.1753 & 0.0491 & -- & 0.0164 & 0.2874 & 0.1433 & -- & 0.1433 & 0.0828 \\
-- & -- & 0.0171 & 0.1753 & 0.0123 & -- & -- & 0.1839 & 0.2876 & -- & 0.2876 & 0.2782 \\
0.2464 & 0.0171 & -- & -- & 0.0192 & 0.2945 & 0.2012 & -- & -- & 0.3195 & -- & -- \\
-- & -- & 0.0171 & 0.1753 & 0.0123 & -- & -- & 0.1839 & 0.2876 & -- & 0.2876 & 0.2782 \\
-- & -- & -- & 0.0438 & -- & -- & -- & 0.0460 & 0.1204 & -- & 0.1204 & 0.2782
\end{pmatrix}$$

$\Gamma_2 M^2$

$$\begin{pmatrix}
0.4478 & 0.0801 & 0.0226 & 0.0003 & 0.0681 & 0.4257 & 0.3593 & 0.0004 & 0.0000 & 0.4319 & 0.0000 & -- \\
0.0176 & 0.2849 & 0.2280 & 0.0070 & 0.1759 & 0.0056 & 0.0330 & 0.0049 & 0.0003 & 0.0068 & 0.0003 & 0.0000 \\
0.0048 & 0.2226 & 0.2895 & 0.0224 & 0.1726 & 0.0004 & 0.0101 & 0.0172 & 0.0030 & 0.0007 & 0.0030 & 0.0009 \\
0.0002 & 0.0180 & 0.0590 & 0.2217 & 0.0400 & 0.0000 & 0.0008 & 0.1870 & 0.1386 & 0.0000 & 0.1386 & 0.0990 \\
0.0265 & 0.3121 & 0.3136 & 0.0276 & 0.4389 & 0.0052 & 0.0539 & 0.0215 & 0.0033 & 0.0098 & 0.0033 & 0.0009 \\
0.1161 & 0.0069 & 0.0005 & 0.0000 & 0.0036 & 0.1574 & 0.0846 & 0.0000 & -- & 0.1308 & -- & -- \\
0.0963 & 0.0403 & 0.0127 & 0.0004 & 0.0371 & 0.0831 & 0.1968 & 0.0008 & 0.0000 & 0.1035 & 0.0000 & 0.0000 \\
0.0002 & 0.0115 & 0.0417 & 0.1723 & 0.0287 & 0.0000 & 0.0016 & 0.1982 & 0.1326 & 0.0001 & 0.1326 & 0.0964 \\
0.0000 & 0.0013 & 0.0147 & 0.2558 & 0.0088 & -- & 0.0001 & 0.2655 & 0.3264 & 0.0000 & 0.3264 & 0.3456 \\
0.2904 & 0.0209 & 0.0022 & 0.0000 & 0.0170 & 0.3225 & 0.2596 & 0.0001 & 0.0000 & 0.3164 & 0.0000 & -- \\
0.0000 & 0.0013 & 0.0147 & 0.2558 & 0.0088 & -- & 0.0001 & 0.2655 & 0.3264 & 0.0000 & 0.3264 & 0.3456 \\
-- & 0.0000 & 0.0008 & 0.0367 & 0.0005 & -- & 0.0000 & 0.0388 & 0.0694 & -- & 0.0694 & 0.1116
\end{pmatrix}$$

$\Gamma_2(\Gamma_2 M^2 \cdot \Gamma_2 M^2)$

$$\begin{pmatrix}
1.0000 & -- & -- & -- & -- & 1.0000 & 1.0000 & -- & -- & 1.0000 & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & 1.0000 & 1.0000 & -- & 1.0000 & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & 0.5000 & -- & -- & -- & 0.5000 & 0.5000 & -- & 0.5000 & 0.5000 \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- \\
-- & -- & -- & 0.5000 & -- & -- & -- & 0.5000 & 0.5000 & -- & 0.5000 & 0.5000 \\
-- & -- & -- & -- & -- & -- & -- & -- & -- & -- & -- & --
\end{pmatrix}$$

Limit L_M resulting from iterating $(\Gamma_2 \circ \mathrm{Exp}_2)$ infinitely many times with initial matrix M.

Figure 2: Iteration of $(\Gamma_2 \circ \mathrm{Exp}_2)$.

towards an idempotent limit. The multiplicity of its eigenvalue 1 equals 3 however, which is also the number of strongly connected components in the associated graph of the limit. The next section will give some insight in the phenomena that play a role in the *MCL* process, by focusing attention to two specific classes of stochastic matrices.

3. Properties of inflation and stochastic *dpsd* matrices

At first sight the inflation operator seems hard to get a grasp on mathematically. It is clear that describing its behaviour falls outside the scope of classical linear algebra, as it represents a scaling of matrices that is both non-linear, column-wise defined, and depends on the choice of basis. In general $\Gamma_r M$ can be described in terms of a Hadamard matrix power which is postmultiplied with a diagonal matrix. For a restricted class of matrices there is an even stronger connection with the Hadamard–Schur product. These are the class of stochastic diagonally symmetrizable matrices and a subclass of the latter, the class of stochastic diagonally positive semi-definite matrices.

Definition 6 *A square matrix A is called* **diagonally hermitian** *if it is diagonally similar to a hermitian matrix. If A is* real *then A is called* **diagonally symmetrizable** *if it is diagonally similar to a symmetric matrix. Given a hermitian matrix A, equivalent formulations for diagonal symmetrizability are:*

i) There exists a positive vector x such that ${d_x}^{-1}Ad_x$ is hermitian, or equivalently, such that $(x_l/x_k)A_{kl} = (x_k/x_l)\overline{A_{lk}}$. If A is real, Identity (3.1) holds.

$$ {d_x}^{-1}Ad_x = [A \circ A^T]^{\circ 1/2} \tag{3.1} $$

ii) There exists a positive vector y such that Ad_y is hermitian, or equivalently, such that $\overline{A_{kl}}/A_{lk} = y_k/y_l$. The vector y is related to x above via ${d_x}^2 = d_y$ or equivalently $y = x \circ x$. □

The fact that d_x can always be chosen with a positive real x depends on the following. Let d_u be a diagonal matrix, where each u_i is a complex number on the unit circle. Then the decomposition $A = d_x S {d_x}^{-1}$, where S is hermitian, can be rewritten as

$$ A = (d_x d_u)({d_u}^{-1} S d_u)(d_u d_x)^{-1} $$

with $S' = {d_u}^{-1} S d_u$ hermitian. This depends on the fact that on the unit circle the inverse of a complex number equals its conjugate.

Definition 7 *A square matrix is called* **diagonally positive-semi definite** *if it is diagonally similar to a positive semi-definite matrix, it is called* **diagonally positive definite** *if it is diagonally similar to a positive definite matrix. The phrases are respectively abbreviated as* dpsd *and* dpd. □

Remark. If M is diagonally symmetrizable stochastic, and y is such that Md_y is symmetric, then $My = y$; thus y represents the equilibrium distribution of M. In the theory of Markov chains, a stochastic diagonally symmetrizable matrix is called *time reversible* or said to satisfy the *detailed balance* condition (See e.g. [24, 36]). A slightly more general definition and different terminology was chosen here. The main reason is that the term 'time reversible' is coupled tightly with the idea of studying a stochastic chain via (powers of) its associated stochastic matrix, and is also used for continuous–time Markov chains. The *MCL* process does not have a straightforward stochastic interpretation, and the relationship between an input matrix and the subsequent iterands is much more complex. Moreover, it is natural to introduce the concepts of a matrix being diagonally similar to a positive (semi–) definite matrix; clinging to 'time reversible' in this abstract setting would be both contrived and unhelpful. The proposed phrases seem appropriate, since several properties of hermitian and *psd* matrices remain valid in the more general setting of diagonally hermitian and *dpsd* matrices. Lemma 2 lists the most important ones, which are easy to verify. Probably all these results are known.

Lemma 2 *Let A be diagonally hermitian of dimension n, let α be a list of distinct indices in the range $1 \ldots n$, let k and l be different indices in the range $1 \ldots n$. Let x be such that $S = {d_x}^{-1} A d_x$ is hermitian, and thus $A = d_x S {d_x}^{-1}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the decreasingly arranged eigenvalues of A (and S), let $a_1 \geq a_2 \geq \cdots \geq a_n$ be the decreasingly arranged diagonal entries of A.*

a) $A[\alpha|\alpha] = d_x[\alpha|\alpha] \;\; S[\alpha|\alpha] \;\; d_x[\alpha|\alpha]^{-1}$*, in particular, the diagonal entries of A equal the diagonal entries of S. This implies that the majorization relationship between eigenvalues and diagonal entries for hermitian matrices carry over to diagonally hermitian matrices: The spectrum of A majorizes the vector of diagonal entries of A:*

$$\sum_{i=1}^{k} \lambda_i \;\; \geq \;\; \sum_{i=1}^{k} a_i \qquad k = 1, \ldots, n$$

Together with the first equality this implies that diagonally hermitian matrices satisfy the same interlacing inequalities for bordered matrices as hermitian matrices do.

b) $\mathrm{Comp}_k(A) = \mathrm{Comp}_k(d_x) \; \mathrm{Comp}_k(S) \; \mathrm{Comp}_k({d_x}^{-1})$*, thus the compound of a diagonally hermitian matrix is diagonally hermitian. Moreover, the compound of a* dpd *(*dpsd*) matrix is again* dpd *(*dpsd*).*

c) $\det A[\alpha|\alpha] = \det S[\alpha|\alpha]$*, i.e. corresponding principal minors of A and S are equal. If A is* dpsd *then* $\det A[\alpha|\alpha] \geq 0$*, with strict inequality if A is* dpd.

d) If A is dpsd *and $A_{kk} = 0$ then the k^{th} row and the k^{th} column of A are zero. If A is* dpsd *and* $\det A[kl|kl] = 0$*, then row k and row l are proportional, and column k and column l are proportional.*

e) If A is dpsd*, then for each $k \in \mathbb{N}$, there exists a unique* dpsd *matrix B such that $B^k = A$. This matrix is defined by setting $B = d_x Q \Lambda^{1/k} Q^H {d_x}^{-1}$, where $Q \Lambda Q^H$ is a unitary diagonalization of S, Λ is the diagonal matrix of eigenvalues of S, and $\Lambda^{1/k}$ is the matrix Λ with each diagonal entry replaced by its real nonnegative k^{th} root. This implies that for* dpsd *A, the fractional power A^t, $t \in \mathbb{R}_{\geq 0}$, can be defined in a meaningful way.*

f) If A, B are both of dimension n and diagonally hermitian, dpsd, dpd*, then the Hadamard–Schur product $A \circ B$ is diagonally hermitian,* dpsd, dpd.

Proof. Most statements are easy to verify. For extensive discussion of the majorization relationship between diagonal entries and eigenvalues of hermitian matrices, as well as results on interlacing inequalities see [21]. Statement b) follows from the fact that the compound operator distributes over matrix multiplication, and the fact that the compound of a positive (semi–) definite matrix is again positive (semi–) definite. See [14] for an overview of results on compounds of matrices. c) follows from the fact that each term contributing to the principal minor in A is a product $\prod_i A_{k_i k_{i+1}}$ where each k_i occurs once as a row index and once as a column index, implying the equalities $\prod_i A_{k_i k_{i+1}} = \prod (x_{k_i}/x_{k_{i+1}}) A_{k_i k_{i+1}} = \prod S_{k_i k_{i+1}}$. Then it is a well known property of positive semi-definite matrices that the principal minors are nonnegative (see e.g. [21], page 404). The first statement in d) follows from the fact that principal minors (of dimension 2) are nonnegative. Also, if $\det A[kl|kl] = 0$, then the kl diagonal entry of $\mathrm{Comp}_2(A)$ is zero, and consequently the kl row and the kl column are also zero. Some calculations then confirm the second statement, which will be of use later on. For e) it is sufficient to use the fact that $Q \Lambda^{1/k} Q^H$ is the unique positive semi-definite k^{th} root of S (see [21], page 405). □

REMARK. The two most notable properties which do not generalize from hermitian matrices to diagonally hermitian matrices are the absence of an orthogonal basis of eigenvectors for the latter, and the fact that the sum of two diagonally hermitian matrices is in general not diagonally hermitian as well.

Statements c) and d) in Lemma 2 are used in associating a *DAG* with each *dpsd* matrix in Theorem 5. First the behaviour of the inflation operator on diagonally symmetrizable and *dpsd* matrices is described.

Theorem 2 *Let M be a column stochastic diagonally symmetrizable matrix of dimension n, let d_x be the diagonal matrix with positive diagonal such that $S = {d_x}^{-1} M d_x$ is symmetric, and let r be real. Define the positive vector z by setting $z_k = {x_k}^r (\sum_i {M_{ik}}^r)^{1/2}$, and the positive rank 1 symmetric matrix T by setting $T_{kl} = 1/(\sum_i {M_{ik}}^r)^{1/2} (\sum_i {M_{il}}^r)^{1/2}$. The following statement holds.*

$$ {d_z}^{-1} \, (\Gamma_r M) \; d_z \quad = \quad S^{\circ r} \circ T, \qquad \textit{which is symmetric.} $$

PROOF. Define the vector t by $t_k = \sum_i {M_{ik}}^r$. Then

$$
\begin{aligned}
\Gamma_r M &= M^{\circ r} \, {d_t}^{-1} \\
&= (d_x \, S \, {d_x}^{-1})^{\circ r} \, {d_t}^{-1} \\
&= {d_x}^{\circ r} \; S^{\circ r} \; ({d_x}^{\circ r})^{-1} \, {d_t}^{-1} \\
&= {d_t}^{1/2} {d_t}^{-1/2} \, {d_x}^{\circ r} \; S^{\circ r} \; ({d_x}^{\circ r})^{-1} {d_t}^{-1/2} {d_t}^{-1/2} \\
&= ({d_t}^{1/2} {d_x}^{\circ r}) \; ({d_t}^{-1/2} S^{\circ r} {d_t}^{-1/2}) \; ({d_t}^{1/2} {d_x}^{\circ r})^{-1}
\end{aligned}
$$

Since the matrix ${d_t}^{-1/2} S^{\circ r} {d_t}^{-1/2}$ equals $S^{\circ r} \circ T$, the lemma holds. □

Theorem 3 *Let M be square column stochastic diagonally symmetrizable, let z, S and T be as in Theorem 2.*

i) The matrix $\Gamma_r M$ is diagonally symmetrizable for all $r \in \mathbb{R}$.

ii) If M is dpsd *then $\Gamma_r M$ is* dpsd *for all $r \in \mathbb{N}$, if M is* dpd *then $\Gamma_r M$ is* dpd *for all $r \in \mathbb{N}$.*

PROOF. Statement i) follows immediately from Theorem 2. Statement ii) follows from the fact that the Hadamard–Schur product of matrices is positive (semi-) definite if each of the factors is positive (semi-) definite. Moreover, if at least one of the factors is positive definite, and none of the other factors has a zero diagonal entry, then the product is positive definite (see e.g. [22], page 309). These are basic results in the theory of Hadamard–Schur products, an area which is now covered by a vast body of literature. An excellent exposition on the subject is found in [22]. It should be noted that $r \in \mathbb{N}$ is in general a necessary condition ([22], page 453). The above result is pleasant in the sense that it gives both the inflation operator and the *MCL* process mathematical footing.

Theorem 4 *Let M be diagonally symmetric stochastic, and consider the MCL process $(M, e_{(i)}, r_{(i)})$.*

i) All iterands of this process have real spectrum.

ii) If $r_i = 2$ eventually, and $e_i = 2$ eventually, then the iterands of the process $(M, e_{(i)}, r_{(i)})$ are dpsd *eventually.*

These statements[5] follow from the fact that Exp_2 maps diagonally symmetric matrices onto *dpsd* matrices and from Theorem 2 ii). □

Theorem 4 represents a qualitative result on the *MCL* process. Under fairly basic assumptions the spectra of the iterands are real and nonnegative. In the [9] it was furthermore proven that the *MCL* process converges quadratically in the neighbourhood of nonnegative doubly idempotent matrices These combined facts indicate that the *MCL* process has a sound mathematical foundation. The fact remains however that much less can be said about the connection between successive iterands than in the case of the usual Markov process. Clearly, the process has something to do with mixing properties of different subsets of nodes. If there are relatively few paths between two subsets, or if the combined capacity of all paths is low, then flow tends to evaporate in the long run between the two subsets. It was actually this observation which originally led to the formulation of the *MCL* process as the basic ingredient of a cluster algorithm for graphs.

The question now rises whether the *MCL* process can be further studied aiming at quantitative results. It was seen that $\Gamma_r M$, $r \in I\!N$, can be described in terms of a Hadamard–Schur product of positive semi-definite matrices relating the symmetric matrices associated with M and $\Gamma_r M$ (in Theorem 3). There are many results on the spectra of such products. These are generically of the form

$$\sum_{i=1}^{k} \sigma_i(A \circ B) \;\leq\; \sum_{i=1}^{k} f_i(A)\sigma_i(B), \qquad k = 1, \ldots, n.$$

Here $\sigma_i()$ denotes the i–largest singular value, and $f_i(A)$ may stand (among others) for the i–largest singular value of A, the i–largest diagonal entry of A, the i–largest Euclidean column length, or the i–largest Euclidean row length (see [22]). Unfortunately such inequalities go the wrong way in a sense. Since the inflation operator has apparently the ability to press several large eigenvalues towards 1, what is needed are inequalities of the type

$$\sum_{i=1}^{k} \sigma_i(A \circ B) \;\geq\; ??? \quad .$$

However, the number of eigenvalues pressed towards 1 by Γ_r depends on the density characteristics of the argument matrix, and it could be zero (noting that one eigenvalue 1 is always present). Moreover, Γ_r has also the ability to press small eigenvalues towards zero. Clearly, one cannot expect to find inequalities of the '$\geq$' type without assuming anything on the density characteristics of M. It is shown in the next section that the classic majorization relation formulated in Lemma 2 a) between the eigenvalues and diagonal entries of a *dpsd* matrix, plus a classification of the diagonal entries of a *dpsd* matrix, gives useful information on the relationship between eigenvalues of a stochastic *dpsd* matrix and its image under Γ_r.

A second area of related research is found in the field of rapidly mixing Markov chains. A good reference is [36]. The focus is also on mixing properties of node subsets of Markov graphs, and the Markov graphs used are generally of the time reversible kind, i.e. correspond with diagonally symmetrizable matrices. Transfer of results is not likely however. The derived theorems depend crucially on the fact that a Markov process is considered which corresponds with the row of powers of a given Markov matrix. Bounds securing a minimal amount of mixing are sought in terms of the second–largest eigenvalue of a Markov matrix, and in terms of the notion of conductance, which depends on the equilibrium distribution of the matrix.

4. Structure in *dpsd* matrices

The main objective for this section is to establish structure theory for the class of *dpsd* matrices, and study the behaviour of Γ_∞ using these results. It will be shown that for stochastic *dpsd* M the spectrum of the

[5] Clearly the condition under ii) can be weakened; it is only necessary that e_i is at least one time even for an index $i = k$ such that $r_i \in I\!N$ for $i \geq k$. However, the assumptions under ii) can be viewed as a standard way of enforcing convergence in a setting genuinely differing from the usual Markov process.

matrix Γ_∞ is of the form $\{0^{n-k}, 1^k\}$, where k is related to a structural property of M. Throughout this section two symbols are used which are associated with a *dpsd* matrix A, namely the symbol $\looparrowright$ which denotes an arc relation defined on the indices of A, and the symbol $\sim$ which denotes an equivalence relation on the indices of A. It should be clear from the context which matrix they refer to. All results in this section are stated in terms of columns; the analogous statements in terms of rows hold as well.

Definition 8 *Let A be* dpsd *of dimension n, let k and l be different indices in the range $1\ldots n$.*

i) Define the equivalence relation $\sim$ on the set of indices $\{1,\ldots,n\}$ by $k \sim l \equiv$ columns k and l of A are scalar multiples of each other via scalars on the complex unit circle.

ii) Define the arc relation $\looparrowright$ on the set of indices $\{1,\ldots,n\}$, for $p \neq q$, by $q \looparrowright p \equiv |A_{pq}| \geq |A_{qq}|$.

iii) Let E and F be different equivalence classes in $\{1,\ldots,n\}/\sim$. Extend the definition of $\looparrowright$ by setting $F \looparrowright E \equiv \exists e \in E, \exists f \in F[f \looparrowright e]$. By definition of $\looparrowright$ and $\sim$ the latter implies $\forall e' \in E, \forall f' \in F[f' \looparrowright e']$. □

Lemma 3 *Let A be* dpsd *of dimension n, let k and l be distinct indices in the range $1\ldots n$. Then*

$l \looparrowright k \wedge k \looparrowright l$ implies $k \sim l$.

□

This follows from Lemma 2 d) and the fact that the premise implies $\det A[kl|kl] = 0$. Lemma 3 can be generalized towards the following statement.

Theorem 5 *Let A be* dpsd *of dimension n.*

The arc $\looparrowright$ defines a directed acyclic graph (DAG) on $\{1,\ldots,n\}/\sim$.

Note that the theorem is stated in a column-wise manner. The analogous statement for rows is of course also true. The proof of this theorem follows from Lemma 4. □

Lemma 4 *Let A be* dpsd *of dimension n, suppose there exist k distinct indices $p_i, i = 1,\ldots,k, k > 1$, such that $p_1 \looparrowright p_2 \looparrowright \cdots \looparrowright p_k \looparrowright p_1$. Then $p_1 \sim p_2 \sim \cdots \sim p_k$, and thus all $p_i, i = 1,\ldots,k$ are contained in the same equivalence class in $\{1,\ldots,n\}/\sim$. Furthermore, if A is real nonnegative then each of the subcolumns $A[p_1\ldots p_k|p_i]$ is a scalar multiple of the all-one vector of length k.*

Proof. Without loss of generality assume $1 \looparrowright 2 \looparrowright \cdots \looparrowright k \looparrowright 1$. The following inequalities hold, where the left–hand side inequalities follow from the inequalities implied by $\det A[i\, i{+}1] \geq 0$ and $i \looparrowright i+1$.

$$\begin{array}{ccccc} |A_{i\,i+1}| & \leq & |A_{i+1\,i+1}| & \leq & |A_{i+2\,i+1}| \\ |A_{k-1\,k}| & \leq & |A_{kk}| & \leq & |A_{1k}| \\ |A_{k1}| & \leq & |A_{11}| & \leq & |A_{21}| \end{array}$$

Now let x be positive such that $x_q A_{pq} = x_p \overline{A_{qp}}$. On the one hand, $|A_{kk}| \leq |A_{1k}|$. On the other hand,

$$\begin{aligned}
|A_{kk}| &\geq |A_{k-1\,k}| \\
&= \tfrac{x_{k-1}}{x_k}|A_{k\,k-1}| \\
&\geq \tfrac{x_{k-1}}{x_k}|A_{k-2\,k-1}| \\
&= \tfrac{x_{k-1}}{x_k}\tfrac{x_{k-2}}{x_{k-1}}|A_{k-1\,k-2}| \\
&\dots \\
&\geq \tfrac{x_{k-1}}{x_k}\tfrac{x_{k-2}}{x_{k-1}}\cdots\tfrac{x_1}{x_2}|A_{k1}| \\
&= \tfrac{x_1}{x_k}|A_{k1}| \\
&= |A_{1k}|
\end{aligned}$$

This implies that $|A_{k-1\,k}| = |A_{kk}| = |A_{1k}|$ and the identities $|A_{i-1\,i}| = |A_{ii}| = |A_{i+1\,i}|$ are established by abstracting from the index k. From this it follows that $\det A[i,i+1|i,i+1] = 0$, and consequently $i \sim i+1$ for $i = 1,\dots,k-1$ by Lemma 3. The identities $|A_{i-1\,i}| = |A_{ii}| = |A_{i+1\,i}|$ also imply the last statement of the lemma. □

Definition 9 *Define Γ_∞ by $\Gamma_\infty M = \lim_{r\to\infty}\Gamma_r M$.* □

This definition is meaningful, and it is easy to derive the structure of $\Gamma_\infty M$. Each column q of $\Gamma_\infty M$ has k nonzero entries equal to $1/k$, (k depending on q), where k is the number of elements which equal $\max_p M_{pq}$, and the positions of the nonzero entries in $\Gamma_\infty M[1\dots n|q]$ correspond with the positions of the maximal entries in $M[1\dots n|q]$.

Theorem 6 *Let M be stochastic* dpsd *of dimension n. Let D_M be the directed graph defined on $\{1,\dots,n\}/\sim$ according to Definition 8, which is acyclic according to Theorem 5. Let k be the number of nodes in $\{1,\dots,n\}/\sim$ which do not have an outgoing arc in D_M. These nodes correspond with (groups of) indices p for which M_{pp} is maximal in column p.*

The spectrum of $\Gamma_\infty M$ equals $\{0^{n-k}, 1^k\}$.

Proof. For the duration of this proof, write S_A for the symmetric matrix to which a diagonally symmetrizable matrix A is similar. Consider the identity

$$S_{(\Gamma_r M)} = [\Gamma_r M \circ (\Gamma_r M)^T]^{\circ 1/2}$$

mentioned in Definition 6 i). The matrices $\Gamma_r M$ and $S_{\Gamma_r M}$ have the same spectrum. Now, let r approach infinity. The identity is in the limit not meaningful, since $\Gamma_\infty M$ is not necessarily diagonalizable, and thus the left–hand side may not exist in the sense that there is no symmetric matrix to which $\Gamma_\infty M$ is similar. However, the identity 'spectrum of $\Gamma_\infty M$ = spectrum of $[\Gamma_\infty M \circ (\Gamma_\infty M)^T]^{\circ 1/2}$' *does* remain true, since the spectrum depends continuously on matrix entries (see e.g. [21], page 540), and both limits exist. Thus, it is sufficient to compute the spectrum of S_∞, which is defined as

$$S_\infty = [\Gamma_\infty M \circ (\Gamma_\infty M)^T]^{\circ 1/2}$$

Note that the nonzero entries of $\Gamma_\infty M$ correspond with the entries of M which are maximal in their column. Whenever $[\Gamma_\infty M]_{kl} \neq 0$ *and* $[\Gamma_\infty M]_{lk} \neq 0$, it is true that $k \not\rightarrow l$ and $l \not\rightarrow k$. Now consider a column q in S_∞, and assume that $S_{\infty p_i q} \neq 0$, for $i = 1, \ldots, t$. It follows that $q \not\rightarrow p_i \wedge p_i \not\rightarrow q$ for all i, thus $q \sim p_i$ for all i, and $S_\infty[p_1 \ldots p_t | p_1 \ldots p_t]$ is a positive submatrix equal to $1/tJ_t$, where J_t denotes the all one matrix of dimension t. This implies that S_∞ is block diagonal (after permutation), with each block corresponding with an equivalence class in $\{1, \ldots, n\}/\sim$ which has no outgoing arc in the $\not\rightarrow$ arc relation. Each block contributes an eigenvalue 1 to the spectrum of S_∞. Since the spectrum of S_∞ equals the spectrum of $\Gamma_\infty M$, and there are assumed to be k equivalence classes with the stated properties, this proves the theorem. □

Observation. It was shown that the inflation operator has a decoupling effect on *dpsd* matrices by considering its most extreme parametrization. This result connects the uncoupling properties of the *MCL* process to the effect of the inflation operator on the spectrum of its operand, and it generalizes the mapping of nonnegative column allowable idempotent matrices onto overlapping clusterings towards a mapping of column allowable *dpsd* matrices onto directed acyclic graphs. This generalization is most elegantly described by considering a *dpsd* stochastic matrix M and the matrix $D = \Gamma_\infty M$. From the proof given above it follows that D is a matrix for which some power D^t is idempotent. The overlapping clustering associated with D by taking as clusters all endclasses and the nodes that reach them, is exactly the overlapping clustering resulting from applying Definition 5 on page 5 to D^t. In both cases the clustering is obtained by taking as clusterings the weakly connected components of the graph.

5. Reductions of *dpsd* matrices

The following simple reductions of *dpsd* matrices have not yet been of immediate use in the analysis of the *MCL* process, but knowledge of their existence surely will not harm. The first part of Theorem 7 below is a decomposition of a *dpsd* matrix into mutually orthogonal rank 1 idempotents. This decomposition is in general possible for matrices which are (not necessarily diagonally) similar to a hermitian matrix, but is still of particular interest. It is extensively used in the analysis of rapidly mixing Markov chains, where the relationship between the diagonal matrix transforming a matrix to symmetric form and the stationary distribution is of crucial importance. The second part is a decomposition of a *dpsd* matrix into rank 1 matrices with a particular bordered 0/1 structure. For this decomposition, diagonal similarity is responsible for preserving the bordered structure.

Theorem 7 *Let A be* dpsd *of dimension n, such that $A = {d_t}^{-1} S d_t$. Then A can be written in the forms*

i) $A = \sum_{i=1}^{n} \lambda_i(A) E_i$, *where the E_i are a set of mutually orthogonal rank 1 idempotents.*
ii) $A = {d_t}^{-1}(\sum_{i=1}^{n} x_i {x_i}^*) d_t$, *where the last $i-1$ entries of x_i are zero. If A is real, the vectors x_i can be chosen real.*

The *reduction* aspect of this statement is that all partial sums $\sum_{i=1}^{k} x_i {x_i}^*$ are positive semi-definite as well (by the property that all hermitian forms are nonnegative), so that all partial sums ${d_t}^{-1}(\sum_{i=1}^{k} x_i {x_i}^*) d_t$ are *dpsd*.

Proof. Let u_i be a set of orthonormal eigenvectors of S. Then S can be written as the sum of weighted idempotents $U_i = \lambda_i(S) u_i {u_i}^*$. Statement i) now follows from setting $E_i = {d_t}^{-1} U_i d_t$. Statement ii) is adapted from a similar theorem by FitzGerald and Horn [15] for hermitian matrices. The proof of ii) follows from their argument for the hermitian case, given in the lemma below. □

Lemma 5 *[15] Let B be positive definite of dimension n. If $B_{nn} > 0$ write b_n for the n^{th} column of B, and let x be the vector b_n scaled by a factor ${B_{nn}}^{-1/2}$, so that $[xx^*]_{nn} = B_{nn}$. If $B_{nn} = 0$ let x be the null vector of dimension n. In either case, the matrix $B - xx^*$ is positive semi-definite and all entries in the last row and column are zero.*

PROOF. The latter statement is obvious. For the first part, if $B_{nn} = 0$ (and thus $x = 0$) the proof is trivial, as $B - xx^*$ then equals C which is positive semi-definite because it is a principal submatrix of B. Otherwise consider the hermitian form $u^* B u$ and partition B and u conformally as

$$B = \begin{pmatrix} C & y \\ y^* & B_{nn} \end{pmatrix} \qquad u = \begin{pmatrix} v \\ z \end{pmatrix}$$

where u and v are complex vectors of dimension n and $n-1$ respectively. Expand the hermitian form $u^* B u$ as $v^* B v + \overline{z} y^* v + v^* y z + \overline{z} B_{nn} z$. This expression is greater than or equal to zero for any choice of u. For arbitrary v fix z in terms of v as $-(y^* v)/B_{nn}$. In the further expansion of the hermitian form $u^* B u$ two terms cancel, and the remaining parts are $v^* C v - (y v^* y^* v)/B_{nn}$ which can be rewritten as $v^*(C - (yy^*)/B_{nn})v$. Because this expression is greater than or equal to zero for arbitrary v, and because $C - (y^* y)/B_{nn}$ equals $B - xx^*$, the lemma follows.

A positive semi-definite matrix B has a decomposition $B = \sum_{i=1}^{n} x_i {x_i}^*$ by repeated application of Lemma 5, yielding the theorem of FitzGerald and Horn. Using diagonal similarity, this decomposition translates to the decomposition given in Theorem 7 for *dpsd* matrices. □

The following theorem provides a reduction of diagonally symmetric and *dpsd* stochastic matrices to smaller dimensional counterparts, by taking two states together. This theorem may be of use for the proof of existence or non-existence of stochastic *dpsd* matrices (e.g. with respect to the associated DAGs).

Definition 10 *Let M be a diagonally symmetric stochastic matrix of dimension n, and let π be its stationary distribution, so that ${d_\pi}^{-1/2} M {d_\pi}^{1/2}$ is symmetric. Let k and l be two states corresponding with columns of M. The stochastic contraction M' of M with respect to k and l is the diagonally symmetric matrix in which the states k and l are contracted into a new state $\{kl\}$ as follows.*

$$\begin{aligned} M'_{a\{kl\}} &= \tfrac{M_{ak}\pi_k + M_{al}\pi_l}{\pi_k + \pi_l} \\ M'_{\{kl\}a} &= M_{ka} + M_{la} \\ M'_{\{kl\}\{kl\}} &= \tfrac{(M_{lk} + M_{kk})\pi_k + (M_{ll} + M_{kl})\pi_l}{\pi_k + \pi_l} \end{aligned}$$

□

Theorem 8 *Taking the stochastic contraction commutes with taking powers of M. If M is* dpsd *then so is its stochastic contraction M'.*

PROOF. Without loss of generality, assume that 1 and 2 are the two states being contracted. Identify the new state $\{1,2\}$ with the second column and second row. It is easily verified that the equilibrium distribution π' of M' equals $(0, \pi_1 + \pi_2, \pi_3, \ldots, \pi_n)^T$ and that $M d_{\pi'}$ is symmetric. Let S_M and $S_{M'}$ be the matrices to which M and M' are respectively diagonally similar. Then $[S_{M'}]_{2a} = \sqrt{\pi_a/(\pi_1 + \pi_2)}(M_{1a} + M_{2a})$, and all entries of $S_{M'}$ corresponding with column and row indices greater than two are identical to the entries of S_M. This establishes that $S_{M'}$ can be factorized as below (remembering that $[S_M]_{kl}$ equals $\sqrt{(\pi_l/\pi_k)} M_{kl}$).

$$S_{M'} = \begin{pmatrix} 0 & 0 & & & \\ \frac{\pi_1}{\sqrt{\pi_1+\pi_2}} & \frac{\pi_2}{\sqrt{\pi_1+\pi_2}} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \quad S_M \quad \begin{pmatrix} 0 & \frac{\pi_1}{\sqrt{\pi_1+\pi_2}} & & & \\ 0 & \frac{\pi_2}{\sqrt{\pi_1+\pi_2}} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

This factorization establishes both the commuting part of the theorem and the fact that contraction preserves $dpsd$-ness. The latter follows from considering a symmetric form $xS_{M'}x$; using the factorization it is reduced to a particular symmetric form yS_My □.

6. Hilbert's projective metric

For convenience, all vectors and matrices in this section are assumed to be positive. This is not strictly necessary, see e.g. [5]. Hilbert's projective metric d for two positive vectors x and y both of dimension n is defined as

$$d(x,y) \;=\; \ln\left[(\max_i \frac{x_i}{y_i})(\max_j \frac{y_j}{x_j})\right] \;=\; \max_{i,j} \ln\left(\frac{x_i y_j}{x_j y_i}\right)$$

It can be defined in the more general setting of a Banach space [4]. Hilbert's metric is a genuine metric distance on the unit sphere in $\mathbb{R}^n$, with respect to any vector norm (see [4]). For a positive matrix A define the contraction ratio τ and the cross-ratio number ϕ by

$$\tau A \;=\; \sup_{x,y} \frac{d(Ax,Ay)}{d(x,y)} \qquad\qquad \phi A \;=\; \min_{i,j,k,l} \frac{A_{ik}A_{jl}}{A_{jk}A_{il}}$$

These are related to each other via

$$\tau A \;=\; \frac{1-\sqrt{\phi A}}{1+\sqrt{\phi A}} \tag{6.1}$$

For proofs see [3, 17, 34]. The quantity τ is used to measure the deviation of large products of nonnegative matrices from the set of rank 1 matrices (see e.g. [4, 5, 17, 34]). There is a straightforward connection between Γ_r and ϕ. For M nonnegative stochastic,

$$\phi(\Gamma_r A) \;=\; (\phi A)^r \tag{6.2}$$

It follows immediately from the definition of τ, that for A and B nonnegative,

$$\tau(AB) \;\le\; \tau(A)\tau(B) \qquad \text{c.q.} \qquad \tau(A^k) \;\le\; \tau(A)^k \tag{6.3}$$

Observation. Equations (6.2) and (6.3) supply the means for a simple[6] proof of the fact that the MCL process converges quadratically around the class of rank 1 stochastic matrices. Suppose that M is a rank 1 column stochastic matrix (so that $\phi M = 1$), and that $M' = M + E$ is a perturbation of M such that $\phi M' = 1 - \epsilon$. Equation 6.1 yields that $\tau M'$ is of order $\epsilon/4$. Then $\phi\Gamma_2 M'$ is of order $1 - 2\epsilon$ and $\tau\Gamma_2 M'$ is of order $\epsilon/2$. So for small perturbations inflation has a linear effect on the contraction ratio, whereas quadratic expansion (i.e. Exp_2) squares the contraction ratio. It follows immediately that the MCL process applied to M' will result in a limit that has rank equal to one. □

Experiments with the MCL process suggest that if the process converges towards a doubly idempotent limit, then appropriately chosen submatrices of the iterands have the property that their contraction ratio approaches zero. For example, consider an MCL process converging towards the limit

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

[6] For simplicity it is still assumed that all vectors under consideration are positive.

partition the iterands M_i as

$$M_i \;=\; \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

with each matrix $A, \ldots, D$ of dimension 2×2. The observation is that all four quantities $\tau(A|B)$, $\tau(B^T|D^T)$, $\tau(C|D)$, and $\tau(A^T|C^T)$ tend to zero as i goes to infinity. This presumption is not of crucial importance for the matter of convergence, since it is already known that the *MCL* process converges quadratically (in one of the usual matrix norms) in the neighbourhood of doubly idempotent matrices. However, the connection between Γ_r and ϕ may just lead to new insights in the *MCL* process. What is needed is results on the square of the matrix M_i above (not assuming anything on the dimension of M_i), in term of the inherited partition. Thus, bounds are sought for $\tau(A^2 + BC|AB + BD)$, $\tau((AB + BD)^T|(CB + D^2)^T)$, $\tau(CA + DC|CB + D^2)$, and $\tau((A^2+BC)^T|(CA+DC)^T)$. For this it may be interesting to investigate a notion like 'mutual contraction ratio', e.g. the quantity τ' defined as

$$\tau'(A, B) \;=\; \sup_{x,y} \frac{d(Ax, By)}{d(x, y)}$$

It is difficult to assess the potential of this line of research, but it is interesting to see that inflation and expansion can be described in the same framework.

7. Discussion and conjectures

Theorem 5 and 6 shed light on the structure and the spectral properties of the iterands of the *MCL* process. Theorem 5 also gives the means to associate an overlapping clustering with each *dpsd* iterand of an *MCL* process, simply by defining the end nodes of the associated *DAG* as the unique cores of the clustering, and adding to each core all nodes which reach it.

There is a further contrasting analogy with the usual Markov process. Consider a Markov process with *dpsd* input matrix M. Then the difference $M^k - M^l$, $k < l$, is again *dpsd* (they have the same symmetrizing diagonal matrix, and the spectrum of $M^k - M^l$ is nonnegative). From this it follows that all rows of diagonal entries ${M^{(k)}}_{ii}$, for fixed diagonal position ii, are non-increasing. Given a stochastic *dpsd* matrix M, the Γ_r operator, $r > 1$, (in the setting of *dpsd* matrices) always increases some diagonal entries (at least one). The sum of the increased diagonal entries, of which there are at least k if k is the number of endnodes of the *DAG* associated with both M and $\Gamma_r M$, is a lower bound for the combined mass of the k largest eigenvalues of $\Gamma_r M$ (Lemma 2 a)).

In [9] circulant matrices were introduced for which inflation and expansion act as each other's inverse, that is, matrices M satisfying $\Gamma_2(M^2) = M$. An example of such a matrix (of dimension 3) is $\frac{1}{6}J_3+\frac{1}{2}I_3$, where J_n denotes the matrix of dimensions $n \times n$ which has a one in every position. It was shown in [9] that for each dimension n there is a matrix F_n of the form $\frac{a}{n}J_n+(1-a)I_n$ such that $\Gamma_2({F_n}^2) = F_n$. It was also shown that the *MCL* process with default parameters is unstable around the cyclic limit F_n. For the study of flip–flop equilibrium states the many results on circulant matrices are likely to be valuable, for example the monograph [7], and the work on group majorization in the setting of circulant matrices in [16]. It may also be fruitful to investigate the relationship with *Hilbert's distance* and the *contraction ratio* for positive matrices, introduced in Section 6.

The *MCL* process converges quadratically in the neighbourhood of the doubly idempotent matrices. Proving (near–) global convergence seems to be a difficult task. I do believe however that a strong result will hold.

Conjecture 1 *All MCL processes $(M, e_{(i)}, r_{(i)})$, with $e_i = 2, r_i = 2$ eventually, converge towards a doubly idempotent limit, provided M is irreducible,* dpsd*, and cannot be decomposed as a Kronecker product of matrices in which one of the terms is a flip–flop equilibrium state.*

It is a worthy long standing goal to prove or disprove this conjecture. Subordinate objectives are:

i) For a fixed *MCL* process $(\,\cdot\,, e_{(i)}, r_{(i)})$, what can be said about the basins of attraction of the *MCL* process. Are they connected?

ii) What can be said about the union of all basins of attraction for all limits which correspond with the same overlapping clustering (i.e. differing only in the distribution of attractors)?

iii) Can the set of limits reachable from a fixed nonnegative matrix M for all *MCL* processes $(M, e_{(i)}, r_{(i)})$ be characterized? Can it be related to a structural property of M?

iv) Given a node set $I = \{1, \ldots, n\}$, and two directed acyclic graphs D_1 and D_2 defined on I, under what conditions on D_1 and D_2 does there exist a *dpsd* matrix A such that the *DAG*s associated with A according to Theorem 5, via respectively rows and columns, equals D_1 and D_2? What if A is also required to be column stochastic?

v) Under what conditions do the clusters in the cluster interpretation of the limit of a convergent *MCL* process $(M, e_{(i)}, r_{(i)})$ correspond with connected subgraphs in the associated graph of M?

vi) For A *dpsd*, in which ways can the *DAG* associated with A^2 be related to the *DAG* associated with M?

vii) Is it possible to specify a subclass $\mathcal{S}$ of the stochastic *dpsd* matrices and a subset R' of the reals larger than $I\!N$, such that $\Gamma_r M$ is in $\mathcal{S}$ if $r \in R'$ and $M \in \mathcal{S}$?

REMARK. There is no obvious non-trivial hypothesis regarding item vi), unless such a hypothesis takes quantitative properties of M into account. This is due to the fact that the equilibrium state corresponding with a connected component of M corresponds with a *DAG* which has precisely one endclass. The breaking up of connected components which can be witnessed in the *MCL* process is thus always reversible in a sense. With respect to v), I conjecture the following.

Conjecture 2 *The clustering associated with a limit of an MCL process with* dpsd *input matrix M, corresponds with subsets of the node set of the associated graph G of M which induce subgraphs in G that are connected.*

There are several lines of research which may inspire answers to the questions posed here. However, for none of them the connection seems so strong that existing theorems can immediately be applied. The main challenge is to further develop the framework in which the interplay of Γ_r and Exp_s can be studied. Hadamard–Schur theory was discussed in Section 3. Perron–Frobenius theory, graph partitioning by eigenvectors (e.g. [32, 33]), and work regarding the second largest eigenvalue of a graph (e.g. [1, 6]), form a natural source of inspiration. The theory of Perron complementation and stochastic complementation as introduced by Meyer may offer conceptual support in its focus on uncoupling Markov chains [26, 27]. There are also papers which address the topic of matrix structure when the subdominant eigenvalue is close to the dominant eigenvalue [18, 31]. The literature on the subject of diagonal similarity does not seem to be of immediate further use, as it is often focussed on scaling problems (e.g. [12, 20]).

Regarding flip–flop states, several interesting questions are open:

i) For the *MCL* process with both parameter rows constant equal to 2, are there orbits of length greater than 2 in the class of *dpsd* matrices?

ii) Must an indecomposable *dpsd* (in terms of the Kronecker product) flip–flop state necessarily be a symmetric circulant? It seems obvious that this must be the case.

iii) For flip–flop states which are symmetric circulants, how close is is Exp_2 to $\Gamma_{1/2}$? Note that both operators have a contracting effect on positive matrices.

iv) For each dimension n, does there exist a flip–flop state which is the circulant of a vector

$$
\begin{aligned}
(p_1, p_2, p_3, \ldots, p_{k-1}, p_k, p_{k-1}, \ldots, p_2), \qquad & n = 2k-2 \\
(p_1, p_2, p_3, \ldots, p_{k-1}, p_k, p_k, p_{k-1}, \ldots, p_2), \qquad & n = 2k-1,
\end{aligned}
$$

where all p_i are different, $i = 1, \ldots, k$?

v) For which $r > 1$ and $s > 1$ do there exist nonnegative *dpsd* matrices A such that $\Gamma_r(A^s) = A$, where A^s is defined according to Lemma 2e)?

Conjecture 3 *For every* dpsd *flip–flop equilibrium state which is indecomposable in terms of the Kronecker product, there is no trajectory leading to this state other than the state itself.*

References

1. N. Alon and V.D. Milman. λ_1, Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38:73–88, 1985.
2. Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices In The Mathematical Sciences.* Number 9 in Classics in Applied Mathematics. SIAM, 1994. Corrected and extended republication of the 1979 book.
3. Garret Birkhoff. *Lattice Theory.* Number 25 in AMS Colloquium publications. American Mathematical Society, third edition, 1967. Chapter XVI, Positive Linear Operators, pp. 380–396.
4. P.J. Bushell. Hilbert's metric and positive contraction mappings in a Banach space. *Archive for Rational Mechanics and Analysis*, 52:330–338, 1973.
5. Joel E. Cohen. Contractive inhomogeneous products of non–negative matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 86:351–364, 1979.
6. Dragoš Cvetković and Slobodan Simić. The second largest eigenvalue of a graph (a survey). *Filomat*, 9(3):449–472, 1995.
7. Philip J. Davis. *Circulant Matrices.* John Wiley & Sons, 1979.
8. Stijn van Dongen. A new cluster algorithm for graphs. Technical Report INS–R9814, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, December 1998.
9. Stijn van Dongen. A cluster algorithm for graphs. Technical report, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, 2000. Revised version of [8]. To appear.
10. Stijn van Dongen. *Graph Clustering by Flow Simulation.* PhD thesis, University of Utrecht, May 2000.
11. Stijn van Dongen. Performance criteria for graph clustering and Markov cluster experiments. Technical report, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, 2000. To appear.
12. Tommy Elfving. On some methods for entropy maximization and matrix scaling. *Linear Algebra & Applications*, 34:321–339, 1980.
13. Brian S. Everitt. *Cluster Analysis.* Hodder & Stoughton, third edition, 1993.
14. Miroslav Fiedler. *Special matrices and their applications in numerical mathematics.* Martinus Nijhoff Publishers, 1986.

15. Carl H. Fitzgerald and Roger A. Horn. On fractional Hadamard powers of positive definite matrices. *Journal of Mathematical Analysis and Applications*, 61:633–342, 1977.

16. Alessandra Giovagnoli and Henry P. Wynn. Cyclic majorization and smoothing operators. *Linear Algebra & Applications*, 239:215–225, 1996.

17. J. Hajnal. On products of non–negative matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 79:521–530, 1976.

18. D.J. Hartfiel and Carl D. Meyer. On the structure of stochastic matrices with a subdominant eigenvalue near 1. *Linear Algebra & Applications*, 1272:193–203, 1998.

19. Michiel Hazewinkel. Tree–tree matrices and other combinatorial problems from taxonomy. Technical Report AM–R9507, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, April 1995.

20. Daniel Hershkowitz et al. Minimization of norms and the spectral radius of a sum of nonnegative matrices under diagonal equivalence. *Linear Algebra & Applications*, 241–243:431–453, 1996.

21. Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, 1990.

22. Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

23. Nicholas Jardine and Robin Sibson. *Mathematical Taxonomy.* Wiley Series In Probabilistic And Mathematical Statistics. John Wiley & Sons, 1971.

24. L. Lovász. Random walks on graphs: A survey. In Miklos et al. [28], pages 353–397.

25. Albert W. Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and Its Applications.* Number 143 in Mathematics In Science And Engineering. Academic Press, 1979.

26. Carl D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.

27. Carl D. Meyer. Uncoupling the Perron eigenvector problem. *Linear Algebra & Applications*, 114/115:69–74, 1989.

28. D. Miklos et al., editors. *Combinatorics, Paul Erdős is eighty*, volume II. Janos Bolyai Mathematical Society, 1996.

29. Henryk Minc. *Nonnegative Matrices.* Wiley Interscience Series In Discrete Mathematics And Optimization. John Wiley & Sons, 1988.

30. Boris Mirkin. *Mathematical Classification and Clustering.* Kluwer Academic Publishers, 1996.

31. B.N. Parlett. Invariant subspaces for tightly clustered eigenvalues of tridiagonals. *BIT*, 36(3):542–562, 1996.

32. Alex Pothen, Horst D. Simon, and Kang-Pu Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11(3):430–452, July 1990.

33. David L. Powers. Structure of a matrix according to its second eigenvector. In Uhlig and Grone [37], pages 261–265.

34. E. Seneta. *Non–negative matrices and Markov chains.* Springer, second edition, 1981.

35. Oved Shisha, editor. *Inequalities.* Academic Press, 1965. Wright–Patterson Air Force Base, Ohio, August 19–27.

36. Alistair Sinclair. *Algorithms for Random Generation and Counting, A Markov Chain Approach.* Number 7 in Progress in Theoretical Computer Science. Birkhäuser, 1993.

37. Frank Uhlig and Robert Grone, editors. *Current trends in matrix theory. Proceedings of the Third Auburn Matrix Theory Conference, Auburn University, Auburn, Alabama, U.S.A., March 19–22, 1986.* Elsevier Science Publishing Co., 1987.