Bounded state space truncation and censored Markov chains

Ana Bušić
INRIA - ENS
Paris, France
Email: ana.busic@ens.fr

Hilal Djafri
LSV, ENS Cachan
Cachan, France
Email: djafri@lsv.ens-cachan.fr

Jean-Michel Fourneau
PRiSM, Université de Versailles-St-Quentin
Versailles, France
Email: jmf@prism.fr

***Abstract*—Censored Markov chains (CMC) allow to represent the conditional behavior of a system within a subset of observed states. They provide a theoretical framework to study the truncation of a discrete-time Markov chain when the generation of the state-space is too hard or when the number of states is too large. However, the stochastic matrix of a CMC may be difficult to obtain. Dayar et al. (2006) have proposed an algorithm, called DPY, that computes a stochastic bounding matrix for a CMC with a smaller complexity with only a partial knowledge of the chain. We prove that this algorithm is optimal for the information they take into account. We also show how some additional knowledge on the chain can improve stochastic bounds for CMC.**

***Index Terms*—Censored Markov chains, stochastic bounds.**

I. INTRODUCTION

Since Plateau's seminal work on composition and compact tensor representation of Markov chains using Stochastic Automata Networks (SAN), we know how to model Markov systems with interacting components and large state space [1], [2], [3]. The main idea of the SAN approach is to decompose the system of interest into its components and to model each component separately. Once this is done, interactions and dependencies among components can be added to complete the model and obtain the transition probability (in discrete time) or transition rate matrix (in continuous time). The basic operations needed to build these matrices are Kronecker sum and product (sometimes denoted as tensor) and they are applied on local descriptions of the components and the actions.

The benefit of the tensor approach is twofold. First, each component can be modeled much easier compared to the global system. Second, the space required to store the description of components is in general much smaller than the explicit list of transitions, even in a sparse representation. The decomposition and tensor representation have been generalized to other modeling formalisms as well: Stochastic Petri nets [4], Stochastic Process Algebra [5]. So we now have several well-founded methods to model complex systems using Markov chains with huge state space.

However, it is often impossible to store in memory a probability vector indexed by the state space. Consider for instance the model of the highly available multicomponent systems studied by Muntz [6]. A typical system consists of several disks, CPUs and controllers with different types of failures. The system is operational if there are enough CPU, disks and controllers available. As the number of states grows exponentially with the number of different components, the state space is huge and the UP states are relatively rare. A typical system has more than $9 \cdot 10^{10}$ states and 10^{12} transitions. It is even not possible to store a probability vector of the size of the whole state space. Thus we have to truncate the state space to analyze a smaller chain (see for instance [7]). Few results have been published on this topic despite its considerable practical importance. A notable exception are the results obtained by Seneta [8, Chapter 7] on the approximation of an infinite Markov chain by a sequence of finite truncations. Here we investigate how one can study the truncation using the theory of Censored Markov Chains (CMC).

Besides for the truncation problem, CMC have also been used in other numerical approaches to solve Markov chains. Grassmann and Hayman have used censoring to deal with the block elimination for transition probability matrices of infinite Markov chains with repeating rows [9], [10]. Zhao, Li and Braun have used the censoring technique to study infinite state Markov chains whose transition matrices possess block-repeating entries [11]. They proved that a number of important probabilistic measures are invariant under censoring.

Let $\{X_t\}_{t\geq 0}$ be a Discrete Time Markov Chain (DTMC) with a state space $\mathcal{X}$. Let $\mathcal{E} \subset \mathcal{X}$ be the set of observed states, $\mathcal{X} = \mathcal{E} \cup \mathcal{E}^c$, $\mathcal{E} \cap \mathcal{E}^c = \emptyset$. For simplicity of presentation, we assume here that the state space is finite and denote by $n = |\mathcal{E}|$ and $m = |\mathcal{E}^c|$. Assume that successive visits of $\{X_t\}_{t\geq 0}$ to $\mathcal{E}$ take place at time epochs $0 \leq t_0 < t_1 < \ldots$ Then the chain $\{X_k^{\mathcal{E}}\}_{k\geq 0} = \{X_{t_k}\}_{k\geq 0}$ is called the censored chain with censoring set $\mathcal{E}$ [12]. Let P denote the transition probability matrix of chain $\{X_t\}_{t\geq 0}$. Consider the partition of the state space to obtain a block description of P:

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{1}$$

Blocks A, B, C, and D contain respectively transitions from $\mathcal{E}$ to $\mathcal{E}$, from $\mathcal{E}$ to $\mathcal{E}^c$, from $\mathcal{E}^c$ to $\mathcal{E}$, and from $\mathcal{E}^c$ to $\mathcal{E}^c$. The censored chain only observes the states in $\mathcal{E}$. It is out of the scope of this paper to study how one can find a good set $\mathcal{E}$ to get sufficiently accurate censoring process. In some special cases this can be done using stochastic bounds (an example is discussed in Section VI), or by analyzing the drift of a well chosen Lyapunov function (see for instance [13] for applications to Markov population models). We study here how we can find bounds of the chain once set $\mathcal{E}$ is fixed.

The tensor representation of the chain can be used to generate the blocks or a some elements of these blocks. Indeed, one can derive easily the set of successors or predecessors for a node taking into account the tensor representation. Such a property is the key idea for algorithm SAN2LIMSUB [14] which is based of the generation of a column of the stochastic matrix (*i.e.* the predecessor function) based on the tensor representation and allows the computation of a lumpable stochastic bound. Software tools dealing with Markov chains also support in general a successor function, provided by the user, which is used to build a row of the stochastic matrix. In some cases, a predecessor function is also supported by the tool.

We assume that $\mathcal{E}^c$ does not contain any reducible classes (so that the matrix $Id - D$ is regular). Then the transition probability matrix of the censored chain, often also called the stochastic complement of matrix A, is equal to [12]:

$$S_A = A + B(Id - D)^{-1}C = A + B\left(\sum_{i=0}^{\infty} D^i\right)C. \qquad (2)$$

The second term of the right-hand side represents the probabilities of paths that return to set $\mathcal{E}$ through states in $\mathcal{E}^c$.

In many problems, the size of the initial probability matrix P makes building of the four blocks extremely time and space consuming, or even impossible. In some cases, we are able to obtain block D but it is too difficult to compute $(Id - D)^{-1}$ to finally get S_A. Deriving bounds for S_A from block A of matrix P and from some information on the other blocks is thus an interesting alternative approach and several algorithms have been proposed in the literature. Truffet [15] considered the case when only block A is known. In that case, the stochastic bound is obtained by assuming that all the unknown returning transitions go to the last state of $\mathcal{E}$ (*i.e.* state n).

Dayar et al. [16] proposed an algorithm, called DPY, for the case when blocks A and C are known. We prove here that their algorithm is optimal when we do not have any information on blocks B (transitions between observed and non-observed states) and D (transitions between the non-observed states). We consider further how to improve bounds when some additional information is known on blocks B or D. We deal here with upper bounds but lower bounds may be computed as well. Once an upper bounding matrix is found, bounds on rewards, on steady-state and transient distributions, and on time to first visits may be derived as well (see [17] and [18] for some examples).

In Section II we introduce an alternative decomposition of stochastic complement S_A that turns out to be natural for deriving stochastic bounds. Section III contains some definitions and basic results on stochastic orders that will be used later in the paper. We also give an overview of existing results for censored Markov chains and we describe algorithm DPY. In Section IV we prove optimality of DPY using our new decomposition from Section II. Then in Section V we discuss how to use additional information on non-observed states to improve DPY bounds and give an illustrative example in Section VI. Section VII contains conclusions and final remarks.

II. Decomposition of stochastic complement

Let us first fix the notation used throughout the paper. For any $x \in \mathbb{R}$, $[x]^+ = \max\{x, 0\}$. Throughout the paper, the row (resp. column) vectors are denoted by small latin (resp. greek) letters and $\mathbf{0}$ (resp. $\mathbf{1}$) denotes the vector with all components equal to 0 (resp. 1). For a vector v, v^t denotes the transposed vector, and $\mathrm{diag}(v)$ is the matrix whose diagonal elements are given by vector v, *i.e.*

$$\mathrm{diag}(v)[i,j] = \begin{cases} v[i], & i = j, \\ 0, & i \neq j. \end{cases}$$

Furthermore, $\leq$ denotes the elementwise comparison of two vectors (or matrices), and $M[i, \cdot]$ is row i of matrix M.

Throughout the paper, we use the term positive vector (matrix) for a vector (matrix) whose all elements are non-negative. For any positive vector v we will denote by v^* the vector obtained from v by replacing all the zero elements by 1:

$$v^*[i] = \begin{cases} v[i], & v[i] > 0, \\ 1, & v[i] = 0. \end{cases}$$

For any positive matrix M, matrix $\mathrm{diag}(M\mathbf{1}^t)$ is a diagonal matrix whose diagonal elements are equal to the sum of rows of matrix M. If matrix M does not contain any zero row, matrix $\mathrm{diag}(M\mathbf{1}^t)$ is regular and $(\mathrm{diag}(M\mathbf{1}^t))^{-1}M$ is the renormalized matrix M (*i.e.* a matrix such that $\sum_j M[i,j] = 1, \forall i$. If matrix M contains a zero row, then $\mathrm{diag}(M\mathbf{1}^t)$ is singular so we use instead a modified matrix $\mathrm{diag}((M\mathbf{1}^t)^*)$ that is always regular. For any positive matrix M, we denote by $\mathrm{n}(M)$ the matrix:

$$\mathrm{n}(M) = \mathrm{diag}((M\mathbf{1}^t)^*)^{-1}M.$$

Clearly, for any positive matrix M, matrix $\mathrm{n}(M)$ is a positive matrix whose all rows are either stochastic or zero rows. We will use the following technical observation in our decomposition of stochastic complement:

Lemma 1. *For any positive matrix M:*

$$(\mathit{diag}((M\mathbf{1}^t)^*) - \mathit{diag}(M\mathbf{1}^t))\mathit{n}(M) = \mathbf{0}.$$

Proof: Denote by $Z = \mathrm{diag}((M\mathbf{1}^t)^*) - \mathrm{diag}(M\mathbf{1}^t)$. Matrix Z is a diagonal matrix with:

$$Z[i,i] = \begin{cases} 1, & M[i,\cdot] = \mathbf{0}, \\ 0, & \text{otherwise}. \end{cases}$$

The rows of matrix Z and the columns of M are orthogonal by the definition of matrix Z, so we have $ZM = \mathbf{0}$. This remains true for the renormalized matrix, thus $Zn(M) = \mathbf{0}$. ■

Let us now consider the block decomposition of matrix P given by (1). Denote by $W = \mathrm{diag}(C\mathbf{1}^t)$. Then clearly, for all $1 \leq i \leq m$: $W[i,i] = \sum_{j=1}^{n} C[i,j] = 1 - \sum_{k=1}^{m} D[i,k]$. If there is a state $i \in \mathcal{E}^c$ without any outgoing transition to set $\mathcal{E}$, then row i of W is equal to $\mathbf{0}$ and matrix W is singular. Finally, denote by $W^* = \text{diag}((C\mathbf{1}^t)^*)$.

The following proposition gives an alternative decomposition of stochastic complement. To the best of our knowledge such a representation was not previously stated even if it appears quite simple. Using this new representation we can derive new arguments to prove stochastic bounds based on comparison of stochastic vectors. Such an approach was harder with the usual representation in (2) as $(Id - D)^{-1}$ is a matrix of expectations.

Proposition 2 (Decomposition of stochastic complement)**.**

1) *Matrix $(W^*)^{-1}C$ has rows that are either stochastic or equal to $\mathbf{0}$.*
2) *Matrix $(Id - D)^{-1}W$ is stochastic.*
3) *Matrix S_A can be decomposed as:*

$$S_A = A + B(Id - D)^{-1}W(W^*)^{-1}C. \quad (3)$$

The proof is given in Appendix A.

As we now deal with stochastic matrices, we are able in the following to compute their stochastic bounds.

III. Stochastic bounds

We recall first the definition of strong stochastic ordering of random variables on a finite state space $\{1, \ldots, n\}$ (see [19] for more details on stochastic orders).

A. Some Fundamental Results on Stochastic Bounds

We will define operators r and v as in [20] and s for any positive $m \times n$ matrix M:

$$r(M)[i,j] = \sum_{k=j}^{n} M[i,k], \ \forall i,j, \quad (4)$$

$$v(M)[i,j] = \max_{k \le i}\{r(M)[k,j]\}, \ \forall i,j. \quad (5)$$

$$s(M)[j] = \max_{i}\{r(M)[i,j]\}, \ \forall j. \quad (6)$$

Let X and Y be two random variables with probability vectors p and q ($p[k] = P(X = k), q[k] = P(Y = k), \ \forall k$).

Definition 3. *$X \preceq_{st} Y$ if $\sum_{k=j}^{n} p[k] \le \sum_{k=j}^{n} q[k], \ \forall j$ (i.e. $r(p) \le r(q)$).*

Let $\{X_t\}_{t \ge 0}$ and $\{Y_t\}_{t \ge 0}$ be two DTMC with transition probability matrices P and Q. We say that $\{X_t\}_{t\ge 0} \preceq_{st} \{Y_t\}_{t \ge 0}$ if $X_t \preceq_{st} Y_t$ for all $t \ge 0$.

Definition 4. *For two probability matrices P and Q, $P \preceq_{st} Q$ if $r(P) \le r(Q)$.*

Definition 5. *A probability matrix P is $\preceq_{st}$-monotone if for any two probability vectors p and q:*

$$p \preceq_{st} q \ \Rightarrow \ pP \preceq_{st} qP.$$

We will use the following characterization of monotonicity (see [19] for the proof):

Proposition 6. *A probability matrix P is $\preceq_{st}$-monotone iff:*

$$P[i-1,\cdot] \preceq_{st} P[i,\cdot], \ \forall i > 1. \quad (7)$$

i.e. iff $v(P) = P$.

Sufficient conditions for comparison of two DTMC based on stochastic comparison and monotonicity can be found in [19]. These conditions can be easily checked algorithmically and it is also possible to construct a monotone upper bound for an arbitrary stochastic matrix P (see [20] for more details and proofs):

Proposition 7. *(Vincent's algorithm [20]) Let P be any stochastic matrix. Then the Vincent's bound is given by $Q = r^{-1}v(P)$, where r^{-1} denotes the inverse of r. Then Q is $\preceq_{st}$-monotone and $P \preceq_{st} Q$, therefore Q is a transition probability matrix of an upper bounding DTMC. Furthermore, if $P_1 \preceq_{st} P_2$, then $r^{-1}v(P_1) \preceq_{st} r^{-1}v(P_2)$.*

Corollary 8. *(Optimality [20, Lemma 3.1]) Let P be any stochastic matrix and $Q = r^{-1}v(P)$. Then Q is the smallest $\preceq_{st}$-monotone upper bound for P,* i.e. *if R is any stochastic matrix such that R is $\preceq_{st}$-monotone and $P \preceq_{st} R$, then $Q \preceq_{st} R$.*

B. Comparison of positive matrices

We can extend Definition 4 to positive (not necessarily square) matrices:

Definition 9. *Let M_1 and M_2 be any two positive matrices of the same size. We will say that $M_1 \preceq_{st} M_2$ if $r(M_1) \le r(M_2)$.*

We now state two simple properties that will be very useful later in Section V. The proofs are rather straightforward.

Lemma 10. *Let M_1 and M_2 be two positive matrices such that $M_1 \preceq_{st} M_2$, and Z any positive matrix. Then*

$$ZM_1 \preceq_{st} ZM_2.$$

Lemma 11. *Let Z be any positive matrix and M a positive matrix whose rows are either stochastic or equal to $\mathbf{0}$. Then:*

$$ZM \preceq_{st} \alpha r^{-1}(s(M)),$$

where $\alpha = ZM\mathbf{1}^t$.

C. Stochastic bounds for CMC

Now we can formally state the problems we consider:

1) Given only block A, compute a matrix Q such that $S_A \preceq_{st} Q$. Is there an optimal bound (in the sense of Definition 12), knowing only block A?
2) Given blocks A and C, compute a matrix Q such that $S_A \preceq_{st} Q$. Is this bound better than the one obtained knowing only block A? Is there an optimal bound knowing only blocks A and C?
3) Can some additional information on blocks B and D improve stochastic bounds for CMC?

The first question was already answered by Truffet [15]. Denote by $\beta = \mathbf{1}^t - A\mathbf{1}^t$ the column vector of probability slack for matrix A. Then the bound in Truffet [15] is given by:

$$T(A) = A + \beta(0, \ldots, 0, 1). \quad (8)$$

It is straightforward to see that $S_A \preceq_{st} T(A)$. Furthermore, this is the best bound one can obtain knowing only block A.

Definition 12. *Let $\mathcal{M}$ be a family of stochastic matrices. A stochastic matrix Q is an $\preceq_{st}$-upper bound for family $\mathcal{M}$ if:*

$$P \preceq_{st} Q, \ \forall P \in \mathcal{M}.$$

An $\preceq_{st}$-upper bound Q of $\mathcal{M}$ is optimal if:

$$Q \preceq_{st} R, \text{ for any } \preceq_{st}\text{-upper bound } R \text{ of } \mathcal{M}.$$

Let $\mathcal{R}$ be the set of all stochastic matrices such that $\mathcal{E}^c$ does not contain any reducible class (so that for any matrix $Z \in \mathcal{R}$, the stochastic complement S_Z is well defined by (2)). Then the Truffet's bound $T(A)$ in (8) is the optimal $\preceq_{st}$-upper bound for family:

$$\mathcal{M}(A) = \{S_Z \ : \ Z \in \mathcal{R}, \ Z_{\mathcal{E},\mathcal{E}} = A\}.$$

The proof is straightforward.

The second question was partially answered by Dayar et al. [16]: they derived an algorithm DPY that computes a stochastic bound for S_A when blocks A and C are known. In Algorithm 1 we give the algorithm DPY in its original form in [16]. We show in Section IV that DPY bound is optimal, which then fully answers the second question. The third question will be discussed in Section V.

Algorithm 1: DPY(A, C) [16]

```
Data: Blocks A and C.
Result: Matrix Q such that S_A ⪯_st Q.
begin
    β = 1^t − A1^t;
    for j = n downto 1 do
        H[j] = max_{k∈E^c} ( Σ_{l=j}^n C[k,l] / Σ_{l=1}^n C[k,l] );
        for i = 1 to n do
            F[i,j] = (β[i]H[j] − Σ_{l=j+1}^n F[i,l])^+;
    Q = A + F;
    return Q;
end
```

Using operators r and s as defined in relations (4) and (6), DPY can be rewritten as:

1) $H = s(n(C))$,
2) $F = \beta r^{-1}(H)$,
3) $Q = A + F$.

Thus:

$$DPY(A, C) = A + \beta r^{-1}(s(n(C))). \quad (9)$$

From the above equation it obviously follows that $DPY(A, C) \preceq_{st} T(A)$. The following example shows that DPY can improve the Truffet's bounds.

Example 1. *Let the blocks A, B, C, and D be respectively:*

$$\begin{bmatrix} 0.2 & 0.1 & 0.3 & 0 \\ 0.1 & 0 & 0.3 & 0 \\ 0 & 0.5 & 0 & 0.2 \\ 0.1 & 0 & 0.3 & 0.2 \end{bmatrix}, \begin{bmatrix} 0.1 & 0.1 & 0.2 & 0 & 0 & 0 \\ 0.2 & 0.2 & 0.2 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0 & 0 \end{bmatrix},$$
$$\begin{bmatrix} 0.2 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.2 & 0 \\ 0.2 & 0.2 & 0.1 & 0 \\ 0.2 & 0.1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0 & 0.2 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0.4 \\ 0.1 & 0 & 0.5 & 0 & 0 & 0.1 \end{bmatrix}.$$

Then the stochastic complement of A is equal to:

$$S_A = \begin{bmatrix} 0.3476 & 0.2257 & 0.3827 & 0.0440 \\ 0.3447 & 0.1900 & 0.4079 & 0.0574 \\ 0.1379 & 0.5960 & 0.0432 & 0.2229 \\ 0.2010 & 0.1227 & 0.4151 & 0.2612 \end{bmatrix}.$$

Truffet's bound $T(A)$ and DPY bound $DPY(A, C)$ are respectively:

$$\begin{bmatrix} 0.2 & 0.1 & 0.3 & 0.4 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0 & 0.5 & 0 & 0.5 \\ 0.1 & 0 & 0.3 & 0.6 \end{bmatrix}, \begin{bmatrix} 0.28 & 0.22 & 0.4 & 0.1 \\ 0.22 & 0.18 & 0.45 & 0.15 \\ 0.06 & 0.59 & 0.075 & 0.275 \\ 0.18 & 0.12 & 0.4 & 0.3 \end{bmatrix}.$$

Computing blocks A and C is based on the application of the predecessor function (*i.e.* the computation of a column of the stochastic matrix) for all the states in $\mathcal{E}$. Note that block D and B in Example 1 have typical properties of blocks generated by breadth first search visit algorithm based on the successor functions. Indeed, block D is block upper Hessenberg and the rightmost entries of B are 0.

Let us turn back to the algorithm and its theoretical properties. First, DPY Algorithm is even exact in some cases which have proved by Dayar and his co-authors in [16].

Proposition 13. *If block C has rank 1, then the bound provided by DPY is the true solution for the stochastic complement (see [16] for a proof).*

In the next section we prove that DPY is optimal when only the blocks A and C are known.

IV. OPTIMALITY OF DPY

We will show here the optimality of DPY for family

$$\mathcal{M}(A, C) = \{S_Z \ : \ Z \in \mathcal{R}, \ Z_{\mathcal{E},\mathcal{E}} = A, \ Z_{\mathcal{E}^c,\mathcal{E}} = C\}$$

of all transition probability matrices in $\mathcal{R}$ with given blocks A and C.

Theorem 14 (Optimality of DPY)**.** *Matrix $DPY(A, C)$ is the optimal $\preceq_{st}$-upper bound for family $\mathcal{M}(A, C)$.*

Proof: The proof that $DPY(A, C)$ is an $\preceq_{st}$-upper bound for family $\mathcal{M}(A, C)$ was already given in [16].

We prove here that $DPY(A, C)$ is the optimal bound for $\mathcal{M}(A, C)$. Consider any non-zero row j of matrix C (*i.e.* such that $s_j < 1$) and denote by B_j the matrix such that for all i, $B_j[i, k] = 0$, if $k \neq j$, and $B_j[i, j] = \beta[i]$. Let D_j be a matrix such that for all i, $D_j[i, k] = 0$, if $k \neq j$, and $D_j[i, j] = 1 - \sum_{k=1}^{n} C[i, k]$. Let Z_j be the matrix composed of blocks A, B_j, C, and D_j. Then clearly $Z_j \in \mathcal{M}(A, C)$. For Z_j all the returning paths from $\mathcal{E}^c$ to $\mathcal{E}$ go by state $j \in \mathcal{E}^c$. Denote $\widetilde{C} = n(C)$ to ease the notation. Any $\preceq_{st}$-upper bound R for family $\mathcal{M}(A, C)$ satisfies in particular:

$$S_{Z_j} = A + \beta \widetilde{C}[j, \cdot] \preceq_{st} R,$$

i.e. $r(A + \beta \widetilde{C}[j, \cdot]) \leq r(R)$. This is valid for all j such that $\widetilde{C}[j, \cdot] \neq 0$. Thus:

$$\max_j r(A + \beta \widetilde{C}[j, \cdot]) = \max_{\{j \,:\, \widetilde{C}[j,\cdot] \neq 0\}} r(A + \beta \widetilde{C}[j, \cdot]) \leq r(R).$$

And

$$\begin{aligned} \max_j r(A + \beta \widetilde{C}[j, \cdot]) &= r(A) + \beta s(\widetilde{C}) \\ &= r(A) + \beta H = r(A + F), \end{aligned}$$

so $DPY(A, C) = A + F \preceq_{st} R$. ∎

In order to obtain an upper bound for the chain $\{X_k^{\mathcal{E}}\}_{k \geq 0}$, we can now apply Proposition 7 and Corollary 8:

Corollary 15. *The smallest $\preceq_{st}$-monotone upper bound for $\{X_k^{\mathcal{E}}\}_{k \geq 0}$ is given by the transition probability matrix:*

$$r^{-1}(v(DPY(A, C))).$$

Remark 1 (Lower bounds)**.** *Similar algorithm to compute lower bounds can be obtained using the symmetry of $\preceq_{st}$ order.*

We proved the optimality of DPY for the case when only blocks A and C are known. In the following section we consider the case when we have some additional information about blocks B and D and how we can improve the bounds taking into account this new information.

V. USING ADDITIONAL INFORMATION

We consider in this section different assumptions on the (partial) knowledge of blocks of matrix P and we show how this can be used to improve bounds for the stochastic complement. In general, the bounds consist in two parts:

1) Find a deterministic part we can obtain from A, C and all the additional information on the model.
2) Then apply DPY to the unknown part.

Thus the optimality of DPY is not sufficient in general to imply the optimality of these bounds.

A. Known blocks A, B, and C

Let us first assume that we also know block B. Computing blocks A, B and C requires that we know both the predecessor function and the successor function. Using predecessor function we get the column of the stochastic matrix for all the states in $\mathcal{E}$ (*i.e.* blocks A and C) while the successor function gives rows of the states in $\mathcal{E}$ (*i.e.* blocks A and B).

Proposition 16. *Assume that A, B, and C are known. Then:*

$$S_A \preceq_{st} DPY(A + BC, C).$$

Proof: The proof is based on two steps. First we build a new expression for the stochastic complement associated with a new matrix. Then we prove that the matrix we have built is stochastic and we use DPY to obtain a bound of the stochastic complement of that matrix. Let us recall relation (3) and recall that as D does not contain any recurrent class we have:

$$(Id - D)^{-1} = \sum_{i=0}^{\infty} D^i = Id + D(Id - D)^{-1}.$$

After substitution we get:

$$S_A = A + B(Id + D(Id - D)^{-1}) W (W^*)^{-1} C.$$

After simplification we obtain:

$$S_A = A + BC + BD(Id - D)^{-1} W (W^*)^{-1} C.$$

Therefore we obtain S_A as the complement of matrix

$$\left(\begin{array}{c|c} A + BC & BD \\ \hline C & D \end{array} \right).$$

Simple algebraic manipulations allow to prove that this matrix is stochastic. Thus S_A is upper bounded by $DPY(A{+}BC, C)$. ∎

Example 2. *Using as example the same blocks A, B, and C already defined, we obtained a new upper bound of the stochastic complement S_A denoted as $H0$. Clearly the bound is better than $DPY(A, C)$ in Example 1.*

$$H0 = \begin{bmatrix} 0.298 & 0.217 & 0.3925 & 0.0925 \\ 0.252 & 0.178 & 0.435 & 0.135 \\ 0.074 & 0.591 & 0.0675 & 0.2675 \\ 0.188 & 0.112 & 0.4 & 0.3 \end{bmatrix}.$$

B. All blocks are known, but $(Id - D)^{-1}$ is difficult to compute

Now assume that we also know D, but we cannot compute $(Id - D)^{-1}$ because of the computational complexity. This assumption is similar to the one developed in [18] where graph theoretical arguments were used to obtain bounds.

Proposition 17. *For any $K \geq 0$, $S_A \preceq_{st} DPY(A + B \sum_{i=0}^{K} D^i C, C)$.*

The proof relies on the same technique as Proposition 16 and is omitted here for the sake of conciseness.

Example 3. *Let us turn back now to the example for the same blocks and for $K = 1$ (bound $H1$) and $K = 2$ (bound $H2$).*

$$H1 = \begin{bmatrix} 0.309 & 0.2205 & 0.39325 & 0.07725 \\ 0.2732 & 0.1818 & 0.4305 & 0.1145 \\ 0.0874 & 0.5911 & 0.06075 & 0.26075 \\ 0.1896 & 0.1184 & 0.412 & 0.28 \end{bmatrix},$$

$$H2 = \begin{bmatrix} 0.31878 & 0.22197 & 0.390825 & 0.068425 \\ 0.29052 & 0.18378 & 0.42485 & 0.10085 \\ 0.09834 & 0.59191 & 0.056475 & 0.253275 \\ 0.19408 & 0.12032 & 0.4136 & 0.272 \end{bmatrix}.$$

For the same blocks and for $K = 1$ we have also computed the bound obtained with an algorithm [18], based on breadth first search visit of the successors of the nodes in $\mathcal{E}$. The results are clearly less accurate than the bounds we obtain with Proposition 17:

$$FPY1 = \begin{bmatrix} 0.256 & 0.139 & 0.322 & 0.283 \\ 0.192 & 0.058 & 0.324 & 0.426 \\ 0.042 & 0.523 & 0.004 & 0.431 \\ 0.14 & 0.04 & 0.34 & 0.48 \end{bmatrix}.$$

C. Componentwise bounds on $(Id - D)^{-1}$

Here we assume that we know the blocks A, B, and C. Additionally, we know componentwise lower bounds for matrix $(Id - D)^{-1}$:

$$F \leq (Id - D)^{-1}.$$

The block D is either not completely known or $(Id - D)^{-1}$ is difficult to inverse as in the cases V-A and V-B. Recall that $(Id - D)^{-1} = \sum_{i=0}^{\infty} D^i$, so Proposition 17 can be seen as a special case for $F = \sum_{i=0}^{K} D^i$.

One way to obtain such a matrix F is to obtain first E a lower bound of D and then compute or approximate $(Id - E)^{-1}$. Notice that for non negative matrices E and D such that $E \leq D$ and $(Id - D)$ is regular, we have $(Id - E)^{-1} \leq (Id - D)^{-1}$.

Proposition 18. *Assume that we know the blocks A, B, C, and the matrix F such that* $F \leq (Id - D)^{-1}$. *Then:*

$$S_A \preceq_{st} A + BFC + \gamma r^{-1}(s(n(C))),$$

where $\gamma = \mathbf{1}^t - (A + BFC)\mathbf{1}^t$.

The proof is given in Appendix B.

Example 4. *Consider the same block decomposition as in Example 1. Assume that we know a element-wise lower bound of D denoted as D1 and which is equal to:*

$$\begin{bmatrix} 0.5 & 0.2 & 0. & 0 & 0 & 0 \\ 0.2 & 0.6 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that D1 is obtained by a breadth first search visit of the nodes in $\mathcal{E}^c$ *started with nodes of* $\mathcal{E}$ *and limited by a depth of two nodes. We easily compute* $(I - D1)^{-1}$*:*

$$\begin{bmatrix} 2.5 & 1.25 & 0.25 & 0 & 0 & 0 \\ 1.25 & 3.125 & 0.625 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Finally the bounds of S_A are :

$$\begin{bmatrix} 0.33825 & 0.218 & 0.371875 & 0.071875 \\ 0.3325 & 0.18 & 0.39375 & 0.09375 \\ 0.133 & 0.592 & 0.0375 & 0.2375 \\ 0.188 & 0.112 & 0.4 & 0.3 \end{bmatrix}.$$

If we do not know the block C, but only its elementwise lower bounds, then we can extract the known part and apply the Truffet's algorithm to the remaining part:

Proposition 19. *Assume we know the blocks A and B and the componentwise lower bounds of block C and matrix* $(Id - D)^{-1}$: $F \leq (Id - D)^{-1}$ *and* $H \leq C$. *Then:*

$$S_A \preceq_{st} T(A + BFH) = A + BFH + \delta(0, \ldots, 0, 1),$$

where $\delta = \mathbf{1}^t - (A + BFH)\mathbf{1}^t$.

D. Decomposition of the complement space

We assume now that the complement can be divided into two or more non-communicating subsets. Such an information may be provided by the tensor based representation of the DTMC (for some relations between tensor representation of the chain and its graph properties, see [21]).

We assume that D has the block diagonal form:

$$D = \text{diag}(D_{1,1}, D_{2,2}, \ldots, D_{K,K}).$$

Let m_k be the size of block k, $1 \leq k \leq K$. Blocks B and C can also be decomposed as:

$$B = \begin{bmatrix} B_1 & \ldots & B_K \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ \vdots \\ C_K \end{bmatrix}$$

where B_k is an $n \times m_k$ matrix and C_k an $m_k \times n$ matrix, for $1 \leq k \leq K$.

Proposition 20. *Let* $\beta_k = B_k \mathbf{1}^t$, $1 \leq k \leq K$. *We assume that blocks A and C are known and that we also know all* β_k, $1 \leq k \leq K$. *Then:*

$$S_A \preceq_{st} A + \sum_{k=1}^{K} \beta_k s_k,$$

where $s_k = r^{-1}(s(n(C_k)))$, $1 \leq k \leq K$.

Proof: We have: $S_A = A + \sum_{k=1}^{K} B_k (Id - D_{k,k})^{-1} W_k n(C_k)$, where $W_k = \text{diag}(C_k \mathbf{1}^t)$ and $n(C_k) = (W_k^*)^{-1} C_k$, $W_k^* = \text{diag}((C_k \mathbf{1}^t)^*)$, $1 \leq k \leq K$. Thus by Lemma 11,

$$B_k (Id - D_{k,k})^{-1} W_k n(C_k) \preceq_{st} B_k (Id - D_{k,k})^{-1} W_k \mathbf{1}^t s_k,$$

where $s_k = r^{-1}(s(n(C_k)))$, $1 \leq k \leq K$. Similarly as in the proof of Proposition 2, it can be shown that $(Id - D_{k,k})^{-1} W_k$ is a stochastic matrix. We obtain, $B_k (Id - D_{k,k})^{-1} W_k \mathbf{1}^t = B_k \mathbf{1}^t = \beta_k$ and $B_k (Id - D_{k,k})^{-1} W_k n(C_k) \preceq_{st} \beta_k s_k, 1 \leq k \leq K$. Thus, $S_A \preceq_{st} A + \sum_{k=1}^{K} \beta_k s_k$. ■

Example 5. *We now slightly modify Example 1. We consider the same state space and the same block decomposition except block D which is modified as follows:*

$$D2 = \left[\begin{array}{cc|cccc} 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0.3 & 0 & 0.2 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0.4 \\ 0 & 0 & 0.5 & 0 & 0.1 & 0.1 \end{array}\right].$$

The exact solution for the stochastic complement and the bound B2C based on Proposition 20 are respectively:

$$S_A = \begin{bmatrix} 0.384 & 0.228 & 0.358 & 0.030 \\ 0.417 & 0.195 & 0.358 & 0.030 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.201 & 0.123 & 0.415 & 0.261 \end{bmatrix},$$

$$B2C = \begin{bmatrix} 0.373 & 0.227 & 0.35 & 0.05 \\ 0.407 & 0.193 & 0.35 & 0.05 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.18 & 0.12 & 0.4 & 0.3 \end{bmatrix}.$$

The bound in Proposition 20 can be seen as a refinement of DPY bound (see Section III, equation (9)), using the decomposition of matrix D. Thus, as for DPY, we can similarly use additional knowledge on blocks B, C, and D, to improve the bounds in Proposition 20. For example, if we know blocks $B_1, \ldots, B_K$, then by combining the results of Section V-A with Proposition 20 we get:

$$S_A \preceq_{st} A + \sum_{k=1}^{K} (B_k C_k + \gamma_k s_k),$$

where $s_k = r^{-1}(s(n(C_k)))$ and $\gamma_k = \beta_k - B_k C_k \mathbf{1}^t$, $1 \leq k \leq K$. The proof uses similar arguments as the proofs of Proposition 16 and 20. We leave to the reader the details of the proof, as well as similar generalizations obtained by combining Proposition 20 with results in Sections V-B - V-C.

VI. Example

We illustrate the approach on a toy example of a repairable multicomponent system. The system consists of N identical components in the system that can have three different states: operational, in soft failure, and in hard failure. A soft (resp. hard) failure can occur at any operational component (resp. operational or in soft failure) and is distributed according to an exponential distribution with rate λ_S (resp. λ_H). Any component in a soft failure gets repaired after an exponential time with rate μ_S. As long as the whole system remains operational (i.e. there is at least one operational component), the hard failures are repaired one at a time and repair time is exponential with rate μ_H. When the system is not operational, all the components with hard failures are replaced by the new ones (as there is no risk of failures, the system can be repaired faster), and the time is exponential with rate μ). We are interested in the steady state availability of the system (i.e. the probability that the system is operational).

The state of the system is given by $n = (n_H, n_S)$, where n_S (resp. n_H) denotes the number of components in soft (resp. hard) failure. We consider the censoring set $\mathcal{E} = \{n : n_H \leq L\}$ for some threshold value L. In Table VI we give the upper bounds for the conditional probability $q_{\mathcal{E}}$ that the system is down (i.e. not operational) knowing $\mathcal{E}$, and the probability q that the system is down. Note that a simple lower bound on $P(\mathcal{E})$ (used to get an upper bound for q) can be obtained by assuming that the hard failure repairs are always done one by one at the rate $\min\{\mu_H, \mu\}$. The proof is straightforward using coupling arguments.

We consider different values for L and three different bounds. The first bound is obtained as DPY(A,C). Note that in this case matrix C has zero rows, so we have $DPY(A, C) = T(A)$. For the other two bounds, we use the complement space decomposition from Section V-D, together with Proposition 18. More precisely, bound B1 is obtained as: $T(A + B_1 G_1 C_1)$, and bound B2 as $T(A + B_1 G_1 (Id + D_{12} G_2 D_{21} G_1) C_1)$, where

		p_C	p
$L = 0$	DPY	9.9999e-006	0.0500
	B1	4.7337e-007	0.0500
	B2	4.3586e-008	0.0500
$L = 2$	DPY	5.5975e-007	1.2477e-004
	B1	5.5766e-007	1.2477e-004
	B2	5.5729e-007	1.2477e-004
$L = 5$	DPY	6.4621e-007	6.6137e-007
	B1	6.4382e-007	6.5897e-007
	B2	6.4339e-007	6.5854e-007

TABLE I
Upper bounds for the probability that the system is down ($K = 500$, $\lambda_S = 0.001$, $\lambda_H = 10^{-5}$, $\mu_S = 1$, $\mu_H = 0.1$, $\mu = 0.5$).

$G1 = (Id - D_{11})^{-1}$ and $G2 = (Id - D_{22})^{-1}$, and the decomposition of the complement space is made according to the number of hard failures. If we denote by S_i the set of states with $L + i$ hard failures, then $\mathcal{E}^c = \cup_{i=1}^{K-L}$, B_1 (resp. C_1) contain the transitions from $\mathcal{E}$ (resp. to $\mathcal{E}$), and $D_{i,j}$ the transitions from S_i to S_j.

VII. Conclusions

Our approach gives a theoretical framework for the partial generation of the state-space and the transition matrix of a really large Markov chain. Partial generation is often performed heuristically by software tools without any control on the accuracy of the results. If the chain is designed using an initial state and the successor function, when we stop the generation, we obtain blocks A and B. Similarly, using an initial state and the predecessor function we get blocks A and C when the partial generation is achieved. Tensor based representation [22] allows to build all blocks, but it is also possible to take advantage of a partial representation to reduce the complexity of the computational algorithms. Clearly, the more information (*i.e.* blocks) we put in the model, the more accurate are the bounds. Similarly, when we increase the number of steps to obtain a more accurate version of the blocks (*i.e.* parameter K in the visit-based algorithms in Section V-B), we also increase the tightness of the bounds.

References

[1] B. Plateau, "On the stochastic structure of parallelism and synchronization models for distributed algorithms," in *Proc. of the SIGMETRICS Conference*, Texas, 1985, pp. 147–154.

[2] B. Plateau, J. Fourneau, and K. Lee, "PEPS: a package for solving complex Markov models of parallel systems," in *Proceedings of the 4th Int. Conf. on Modelling Techniques and Tools for Computer Performance Evaluation, Spain*, 1988.

[3] B. Plateau and J.-M. Fourneau, "A methodology for solving Markov models of parallel systems," *J. Parallel Distrib. Comput.*, vol. 12, no. 4, pp. 370–387, 1991.

[4] S. Donatelli, "Superposed stochastic automata: A class of stochastic Petri nets with parallel solution and distributed state space," *Perform. Eval.*, vol. 18, no. 1, pp. 21–36, 1993.

[5] J. Hillston and L. Kloul., "An efficient Kronecker representation for PEPA models." in *PAPM-PROBMIV*, ser. Lecture Notes in Computer Science, L. de Alfaro and S. Gilmore., Eds., vol. 2165. Springer, 2001, pp. 120–135.

[6] R. R. Muntz, E. de Souza e Silva, and A. Goyal, "Bounding availability of repairable computer systems," *IEEE Trans. on Computers*, vol. 38, no. 12, pp. 1714–1723, 1989.

[7] E. de Souza e Silva and P. M. Ochoa, “State space exploration in Markov models,” *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 20, no. 1, pp. 152–166, 1992.
[8] E. Seneta, *Non-negative matrices and Markov chains*, ser. Springer Series in Statistics. New York: Springer, 2006, revised reprint of the second (1981) edition.
[9] W. K. Grassmann and D. P. Heyman, “Equilibrium distribution of block-structured markov chains with repeating rows,” *J. of Applied Probability*, vol. 27, pp. 557–576, 1990.
[10] ——, “Computation of steady-state probabilities for infinite-state markov chains with repeating rows,” *INFORMS Journal on Computing*, vol. 5, no. 3, pp. 292–303, 1993.
[11] Y. Zhao, W. Li, and W. Braun, “Censoring, factorizations, and spectral analysis for transition matrices with block-repeating entries,” *Methodology And Computing In Applied Probability*, vol. 5, no. 1, pp. 35–58, 2003.
[12] Y. Zhao and D. Liu, “The censored Markov chain and the best augmentation,” *Jour. of App. Prob.*, vol. 33, pp. 623–629, 1996.
[13] T. Dayar, H. Hermanns, D. Spieler, and V. Wolf, “Bounding the equilibrium distribution of Markov population models,” in *Proceedings of the Sixth International Workshop on the Numerical Solutions of Markov Chains (NSMC 2010)*, 2010, pp. 40–43.
[14] J.-M. Fourneau, B. Plateau, I. Sbeity, and W. J. Stewart, “SANs and lumpable stochastic bounds: Bounding availability,” in *Computer System, Network Performance and Quality of Service*. Imperial College Press, 2006.
[15] L. Truffet, “Near complete decomposability: Bounding the error by a stochastic comparison method,” *Ad. in App. Prob.*, vol. 29, pp. 830–855, 1997.
[16] T. Dayar, N. Pekergin, and S. Younès, “Conditional steady-state bounds for a subset of states in Markov chains,” in *SMCtools ’06*. ACM Press, 2006.
[17] A. Busic and J.-M. Fourneau, “Stochastic bounds for partially generated Markov chains: an algebraic approach,” in *Fith European Performance Engineering Workshop, EPEW 2008, Proceedings*, ser. Lecture Notes in Computer Science. Springer, 2008.
[18] J.-M. Fourneau, N. Pekergin, and S. Younès, “Censoring Markov chains and stochastic bounds,” in *EPEW 2007*, ser. LNCS 4748. Springer, 2007, pp. 213–227.
[19] A. Muller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*. New York, NY: Wiley, 2002.
[20] T. Dayar, J.-M. Fourneau, N. Pekergin, and J.-M. Vincent, “Polynomials of a stochastic matrix and strong stochastic bounds,” in *Markov Anniversary Meeting*. Boson Books, 2006, pp. 211–228.
[21] J.-M. Fourneau and F. Quessette, “Graphs and stochastic automata networks,” in *Proc. 2nd Int. Workshop on the Numerical Solution of Markov Chains*, Raleigh, NC, USA, 1995.
[22] P. Fernandes, B. Plateau, and W. J. Stewart, “Efficient descriptor-vector multiplications in Stochastic Automata Networks,” *J. ACM*, vol. 45, no. 3, pp. 381–414, 1998.

Note: A preliminary version of this work has been presented in NSMC’10 workshop, without proceedings. It can be found at: http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BDF-nsmc10.pdf.

Appendix

A. Proof of Proposition 2

Matrix W^* is always regular, so $(W^*)^{-1}$ is well defined.

1) We have $(W^*)^{-1}C = n(C)$ so the first statement is obvious.
2) We know that row i of matrix $(Id - D)^{-1}C$ is equal to the conditional probability vector of entering the set $\mathcal{E}$, knowing that we initially start in $i \in \mathcal{E}^c$. Let $G = (Id-D)^{-1}$. Therefore, for all i, $\sum_k (GC)[i,k] = 1$ and:
$$\sum_j (GW)[i,j] = \sum_j G[i,j] \sum_k C[j,k] = \sum_k (GC)[i,k] = 1.$$
Thus matrix $(Id - D)^{-1}W$ is stochastic.
3) Matrix S_A can be decomposed as:
$$S_A = A + B(Id - D)^{-1}W^*(W^*)^{-1}C. \qquad (10)$$
We have: $W = \text{diag}(C\mathbf{1}^t)$, $W^* = \text{diag}((C\mathbf{1}^t)^*)$, and $(W^*)^{-1}C = n(C)$. Thus by Lemma 1:
$$W^*(W^*)^{-1}C = W(W^*)^{-1}C. \qquad (11)$$
Relations (11) and (10) imply (3).

■

B. Proof of Proposition 18

We introduce first a new decomposition of stochastic complement (Proposition 21) that we need to prove bounds in Proposition 18. Matrix S_A can be decomposed as:
$$S_A = A + BFC + B((Id - D)^{-1} - F)W(W^*)^{-1}C.$$
Let $G = ((Id - D)^{-1} - F)W$, $V = \text{diag}(G\mathbf{1}^t)$, and $V^* = \text{diag}((G\mathbf{1}^t)^*)$. Then we have an additional decomposition of stochastic complement:

Proposition 21. *Matrix S_A can be decomposed as:*
$$S_A = A+BFC+BV(V^*)^{-1}((Id-D)^{-1}-F)W(W^*)^{-1}C. \qquad (12)$$
Matrices $(W^*)^{-1}C = n(C)$ and $(V^*)^{-1}((Id - D)^{-1} - F)W = n(G)$ have rows that are either stochastic or equal to $\mathbf{0}$.

Proof: Matrix S_A can be written as:
$$S_A = A+BFC+BV^*(V^*)^{-1}((Id-D)^{-1}-F)W(W^*)^{-1}C. \qquad (13)$$
Recall that $(V^*)^{-1}((Id - D)^{-1} - F)W = n(G)$, $V = \text{diag}(G\mathbf{1}^t)$, and $V^* = \text{diag}((G\mathbf{1}^t)^*)$. Thus, Lemma 1 implies that:
$$V^*(V^*)^{-1}((Id-D)^{-1}-F)W = V(V^*)^{-1}((Id-D)^{-1}-F)W. \qquad (14)$$
Relation (12) now follows from (13) and (14). ■

Proof of Proposition 18: By Proposition 21, we have $S_A = A + BFC + BVn(G)n(C)$, with $G = ((Id - D)^{-1} - F)W$. Lemma 11 for $Z = BVn(G)$ and $M = n(C)$ implies:
$$BVn(G)n(C) \preceq_{st} \alpha r^{-1}(s(n(C))),$$
for $\alpha = ZM\mathbf{1}^t = BVn(G)n(C)\mathbf{1}^t$. Therefore:
$$S_A \preceq_{st} A + BFC + \alpha r^{-1}(s(n(C))). \qquad (15)$$
After multiplying (12) by $\mathbf{1}^t$, we get: $\mathbf{1}^t = (A + BFC)\mathbf{1}^t + BVn(G)n(C)\mathbf{1}^t$, so: $\alpha = BVn(G)n(C)\mathbf{1}^t = \mathbf{1}^t - (A + BFC)\mathbf{1}^t = \gamma$. ■