ON THE CONVERGENCE OF A CLASS OF MULTILEVEL METHODS FOR LARGE, SPARSE MARKOV CHAINS*

PETER BUCHHOLZ† AND TUĞRUL DAYAR‡

Abstract. This paper investigates the theory behind the steady state analysis of large, sparse Markov chains (MCs) with a recently proposed class of multilevel (ML) methods using concepts from algebraic multigrid and iterative aggregation-disaggregation. The motivation is to better understand the convergence characteristics of the class of ML methods and to have a clearer formulation that will aid their implementation. In doing this, restriction (or aggregation) and prolongation (or disaggregation) operators of multigrid are used, and the Kronecker based approach for hierarchical Markovian models (HMMs) is employed, since it suggests a natural and compact definition of grids (or levels). However, the HMM formalism used to describe the class of ML methods for large, sparse MCs has no influence on the theoretical results derived.

Key words. Markov chains, multilevel methods, multigrid, aggregation-disaggregation, Kronecker based numerical techniques

AMS subject classifications. 60J27, 65F50, 65F10, 65B99, 65F15, 65F05, 15A72

1. Introduction. Markov chains (MCs) are a popular mathematical tool to describe real systems from various application areas like engineering, computer science or economics. For system analysis often one needs the steady state distribution of the MC to compute result measures for the modeled system. The problem in the continuous-time case is then to solve

$$\pi Q = 0 \quad \text{subject to} \quad \pi e = 1, \tag{1.1}$$

where Q is the infinitesimal generator or generator matrix (i.e., continuous-time Markov chain, CTMC) of order n underlying the modeled system, $\pi \geq 0$ is its (row) stationary probability vector, and e is the column vector of ones of appropriate length. We assume that the n states of Q are numbered starting from 0 and Q is irreducible, implying $\pi > 0$ and π is also the steady state vector. The nonnegative off-diagonal elements of Q represent exponential transition rates between different states and its diagonal elements are negated row sums of its off-diagonal elements. Hence, Q has row sums of zero (i.e., $Qe = 0$), is a singular matrix of rank $(n-1)$, and (1.1) represents a homogeneous linear system subject to a normalization condition, so that its solution vector π can be uniquely determined [25, Ch. 1]. At this level, states of the CTMC are numbered by consecutive integers. However, in almost all applications CTMCs result from some high level model like a stochastic automata network (SAN), a queueing network (QN) or a stochastic Petri net (SPN). In all these cases, the state space is multidimensional and is mapped for solution onto a set of consecutive integers. The multidimensional structure can be exploited in a compact representation of Q and can also be exploited to develop fast solvers for the computation of π.

Practical problems arise due to the state space size of MCs resulting from applications which often grows exponentially with the number of components in the specification. A popular way of dealing with this so called "state space explosion problem" is to employ Kronecker (or tensor) based representations of Q which remain compact even for considerably large state spaces. In the Kronecker based approach, the system of interest is modeled so that it is formed of smaller interacting components, and its larger underlying MC is neither generated nor stored but rather represented using Kronecker products of the smaller component matrices. This introduces significant storage savings at the expense of some overhead in the solution phase. In order to analyze large, structured Markovian models efficiently, various algorithms for vector-Kronecker product multiplication are devised [15, 16, 13] and used as kernels in iterative solution methods. The most effective solvers known for Kronecker representations of dimension four or larger are multilevel (ML) methods [11] and block successive over-relaxation (BSOR) preconditioned projection methods [12] as recently shown empirically by comparing different solvers on a large number of HMMs. Unfortunately, solvers using BSOR [10, 26] are sensitive to the ordering of components, the block partitionings chosen,

*Part of this work has been carried out at Dortmund University through a grant from the Alexander von Humboldt Foundation.

†Informatik IV, Universität Dortmund, D-44221 Dortmund, Germany; Tel: (+49) (231) 755-4746; Fax: (+49) (231) 755-4730 (peter.buchholz@cs.uni-dortmund.de).

‡Department of Computer Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey; Tel: (+90) (312) 290-1981; Fax: (+90) (312) 266-4047 (tugrul@cs.bilkent.edu.tr).

and the amount of fill-in in the factorized diagonal blocks so that a robust implementation for arbitrary models is difficult to achieve.

In this paper, we investigate the theory behind the steady state analysis of large, sparse MCs with the class of ML methods proposed in [11] using concepts from algebraic multigrid (AMG) [6, Ch. 8] and iterative aggregation-disaggregation (IAD) [25, Ch. 6]. Our motivation is to better understand the convergence characteristics of the class of ML methods and to have a clearer formulation that will aid their implementation. Convergence analysis of a two-level IAD method for MCs and its equivalence to AMG is provided in [19]. Another paper that investigates the convergence of a two-level IAD method for MCs using concepts from multigrid is [20]. Here we consider more than two levels, different types of smoothers, different types of cycles, and different orders of aggregation. In doing this, we use restriction (or aggregation) and prolongation (or disaggregation) operators of multigrid, and employ the Kronecker based approach for HMMs in [11]. This is due to three reasons. First, the hierarchy present in the HMM description suggests a natural definition of grids (or levels). This simplifies the description of the class of ML methods. Second, with the HMM description, one can store the aggregated MC at each level during implementation compactly in Kronecker form. It is not clear how the same effect can be achieved with a MC in sparse format (see [18]). Third, Kronecker operations to define large MCs underlying structured representations are natural for many application areas since complex systems are usually composed of interacting components. Almost all MCs resulting from applications can be represented as HMMs and this representation can be derived from the specification using an appropriate modeling tool [1]. Otherwise, the HMM formalism used in this paper to describe the class of ML methods for large, sparse MCs has no influence on the theoretical results derived.

The next section introduces the Kronecker based description of CTMCs underlying HMMs on a simple test case. The third section presents the proposed class of ML methods for HMMs with multiple macrostates and discusses how they work. The fourth section provides the convergence analysis. The fifth section illustrates one step of the ML method on the simple test case and the sixth section concludes the paper.

In what follows, calligraphic uppercase letters denote sets and lists, uppercase letters denote matrices, sets are defined using curly brackets, lists are defined using square brackets, matrices (and vectors) are defined using brackets, $|\cdot|$ denotes the cardinality of a set (list) when its argument is a set (list), $\emptyset$ denotes the empty set, $||\cdot||$ denotes the norm of a vector, $\cdot^T$ denotes the transpose operator, and $\mathrm{diag}(\cdot)$ represents a diagonal matrix having its vector argument along its diagonal.

2. Hierarchical Markovian Models. Hierarchical Markovian models (HMMs) are defined using the operations of Kronecker product and Kronecker sum [27]. First we introduce these operations.

DEFINITION 2.1. *The Kronecker product of two matrices $X \in \mathbb{R}^{r_X \times c_X}$ and $Y \in \mathbb{R}^{r_Y \times c_Y}$ is written as $X \otimes Y$ and yields the matrix $Z \in \mathbb{R}^{r_X r_Y \times c_X c_Y}$, whose elements satisfy $z_{i_X r_Y + i_Y, j_X c_Y + j_Y} = x_{i_X, j_X} y_{i_Y, j_Y}$. The Kronecker sum of two square matrices $U \in \mathbb{R}^{r_U \times r_U}$ and $V \in \mathbb{R}^{r_V \times r_V}$ is written as $U \oplus V$ and yields the matrix $S \in \mathbb{R}^{r_U r_V \times r_U r_V}$, which is defined in terms of two Kronecker products as $S = U \otimes I_{r_V} + I_{r_U} \otimes V$. Here I_{r_U} and I_{r_V} denote identity matrices of orders r_U and r_V, respectively. Both Kronecker product and Kronecker sum are associative and defined for more than two matrices.*

HMMs consist of multiple low level models (LLMs) which can be perceived as components, and a high level model (HLM) that defines how LLMs interact. The HLM is characterized by a single matrix, whereas each LLM is characterized by multiple matrices that define its interaction with other LLMs. The order of each LLM matrix is equal to the number of states of the particular component to which the matrix belongs. A formal definition of HMMs can be found in [8, pp. 387–390]. Here we extend the definition from [12] and introduce HMMs on a running example. We refer to the CTMC underlying an HMM as the matrix Q. We name the states of the HLM as *macrostates*, those of Q as *microstates*, and remark that macrostates define a partition of the microstates.

DEFINITION 2.2. *In a given HMM, let K be the number of LLMs, $\mathcal{S}^{(k)} = \{0, 1, \ldots, |\mathcal{S}^{(k)}| - 1\}$ be the state space of LLM k for $k = 1, 2 \ldots, K$, $\mathcal{S}^{(K+1)} = \{0, 1, \ldots, |\mathcal{S}^{(K+1)}| - 1\}$ be the state space of the HLM, $\mathcal{S}_j^{(k)}$ be the partition of states of LLM k mapped to macrostate $j \in \mathcal{S}^{(K+1)}$ so that $\cup_j \mathcal{S}_j^{(k)} = \mathcal{S}^{(k)}$ and $\mathcal{S}_i^{(k)} \cap \mathcal{S}_j^{(k)} = \emptyset$ when $i \neq j$, t_0 be a local transition (one per LLM), $\mathcal{T}_{i,j}$ be the set of LLM non-local transitions in element (i, j) of the HLM matrix, and D_j be the diagonal correction matrix that sums the rows of Q corresponding to macrostate j to zero. Then the diagonal block (j, j) of Q corresponding to*

element (j,j) of the HLM matrix is given by

$$Q(j,j) = \bigoplus_{k=1}^{K} Q_{t_0}^{(k)}(\mathcal{S}_j^{(k)}, \mathcal{S}_j^{(k)}) + \sum_{t_e \in \mathcal{T}_{j,j}} \bigotimes_{k=1}^{K} Q_{t_e}^{(k)}(\mathcal{S}_j^{(k)}, \mathcal{S}_j^{(k)}) + D_j, \tag{2.1}$$

and, when there are multiple macrostates, the off-diagonal block (i,j) of Q corresponding to element (i,j) of the HLM matrix is given by

$$Q(i,j) = \sum_{t_e \in \mathcal{T}_{i,j}} \bigotimes_{k=1}^{K} Q_{t_e}^{(k)}(\mathcal{S}_i^{(k)}, \mathcal{S}_j^{(k)}). \tag{2.2}$$

where $Q_{t_e}^{(k)}(\mathcal{S}_i^{(k)}, \mathcal{S}_j^{(k)})$ is a submatrix of order $(|\mathcal{S}_i^{(k)}| \times |\mathcal{S}_j^{(k)}|)$ including all transitions[1] *between states from $\mathcal{S}_i^{(k)}$ and $\mathcal{S}_j^{(k)}$ for LLM k under t_e.*

We remark that D_j can be expressed as a sum of Kronecker products:

PROPOSITION 2.3. *If D_j is the diagonal correction matrix that sums the rows of Q corresponding to macrostate j to zero, then*

$$D_j = -\bigoplus_{k=1}^{K} \operatorname{diag}(Q_{t_0}^{(k)}(\mathcal{S}_j^{(k)}, \mathcal{S}_j^{(k)})e) - \sum_{i \in \mathcal{S}^{(K+1)}} \sum_{t_e \in \mathcal{T}_{j,i}} \bigotimes_{k=1}^{K} \operatorname{diag}(Q_{t_e}^{(k)}(\mathcal{S}_j^{(k)}, \mathcal{S}_i^{(k)})e) \quad \text{for } j \in \mathcal{S}^{(K+1)}.$$

In order to enable the efficient implementation of numerical solvers, most of the time D_j is precomputed and stored explicitly as a vector. However, the off-diagonal part of Q is never stored explicitly, but represented in core through Definition 2.2 as sums of Kronecker products of small matrices, which are generally very sparse and therefore held in row sparse format [25, pp. 80–81].

For a definition of mapping used in the next proposition, see, for instance, [23, pp. 192–197].

PROPOSITION 2.4. *When the multidimensional states of Q are identified by the tuple $(s^{(1)}, s^{(2)}, \ldots, s^{(K)}, j)$, where $s^{(k)} \in \mathcal{S}^{(k)}$ is the state of LLM k for $k = 1, 2, \ldots, K$ and $j \in \mathcal{S}^{(K+1)}$ is the corresponding macrostate, the Kronecker product operation orders the state space of Q lexicographically, where each state is linearized through the one-to-one, onto mapping*

$$(s^{(1)}, s^{(2)}, \ldots, s^{(K)}, j) \longleftrightarrow \sum_{k=1}^{K} s^{(k)} \prod_{l=k+1}^{K} |\mathcal{S}_j^{(l)}| + \sum_{i=0}^{j-1} \prod_{k=1}^{K} |\mathcal{S}_i^{(k)}| \in \{0, 1, \ldots, n-1\}.$$

where $n = \sum_{j=0}^{|\mathcal{S}^{(K+1)}|-1} \prod_{k=1}^{K} |S_j^{(k)}|$.

The microstates corresponding to each macrostate result from the Cartesian (or cross) product [23, pp. 123–124] of the state space partitions of LLMs that are mapped to that particular macrostate. In contrast to other representations of CTMCs using Kronecker operators (e.g., [25, Ch. 9]), HMMs are generated in a way that only reachable states are considered [7, 8]. Note that each macrostate in an HLM may have a different number of microstates if LLMs have partitioned state spaces. When there are multiple macrostates, Q is effectively a block matrix having as many blocks in each dimension as $|\mathcal{S}^{(K+1)}|$. The diagonal and off-diagonal blocks of this partitioning are respectively the $Q_{j,j}$ and $Q_{i,j}$ matrices defined by (2.1) and (2.2). Due to the Kronecker structure suggested by Definitions 2.1 and 2.2, each of the blocks defined by the HLM matrix is also formed of blocks, and hence HMMs have nested block partitionings [10, 26].

Now, let us consider HMM *test* which gives rise to a (5×5) CTMC. The example is chosen deliberately to be very small since later we will be stepping through the ML method on this example.

Example 1. The HLM of 2 states describes the interaction among two LLMs (i.e., $K = 2$) each of which has 3 states. All states are numbered starting from 0. The mapping between LLM states and HLM states and the number of microstates are given in Table 2.1. In this example, Q has the following states in

[1] In this section, the concept of transition is used to refer to those that take place at the HMM level, except for this case where it is used to refer to nonzeros in a matrix at the state level.

TABLE 2.1
Mapping between LLM states and HLM states in test.

LLM 1	LLM 2	HLM	# of microstates
{0,1}	{0,1}	{0}	2 . 2 = 4
{2}	{2}	{1}	1 . 1 = 1

its rows and columns: $\{0,1\}\times\{0,1\}\times\{0\} \cup \{2\}\times\{2\}\times\{1\} = \{(0,0,0),(0,1,0),(1,0,0),(1,1,0),(2,2,1)\}$. One can think of these five states written in the given order as corresponding to the integers 0 through 4.

The values of the nonzeros in Q are determined by the rates of the transitions and their associated matrices. In Example 1, two transitions denoted by t_0 and t_1 take place and affect the LLMs. Transition t_0 covers all local transitions inside the LLMs, whereas transition t_1 is captured by the following (2×2) HLM matrix:

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left(\begin{array}{c|c} & t_1 \\ \hline t_1 & \end{array}\right) \end{array}. \tag{2.3}$$

To each transition in the HLM matrix corresponds a Kronecker product of two (i.e., number of LLMs, K) LLM matrices. The matrices associated with those LLMs that do not participate in a transition are identity. LLM 1 participates in t_1 with the matrix $Q_{t_1}^{(1)}$ and LLM 2 participates in t_1 with the matrix $Q_{t_1}^{(2)}$. In this example, the transition t_1 affects exactly two LLMs.

Other than Kronecker products due to the transitions in (2.3), there is a Kronecker sum implicitly associated with each diagonal element of the HLM matrix. Each Kronecker sum is formed of two (i.e., K) LLM matrices corresponding to *local transition* t_0. In the HLM matrix of *test* in (2.3), there does not exist any non-local transition along the diagonal. In general, this need not be so, as can be seen from Definition 2.2.

In our example, the second term in (2.1) is missing, and the matrices associated with t_0 and t_1 are given by

$$Q_{t_0}^{(1)} = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right),\ Q_{t_1}^{(1)} = \left(\begin{array}{cc|c} 0 & 0 & 2 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \end{array}\right), Q_{t_0}^{(2)} = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right),\ Q_{t_1}^{(2)} = \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 0 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array}\right).$$

Then the CTMC underlying HMM *test* can be obtained from

$$Q = \left(\begin{array}{c|c} Q_{t_0}^{(1)}(\{0,1\},\{0,1\}) \bigoplus Q_{t_0}^{(2)}(\{0,1\},\{0,1\}) & Q_{t_1}^{(1)}(\{0,1\},\{2\}) \bigotimes Q_{t_1}^{(2)}(\{0,1\},\{2\}) \\ \hline Q_{t_1}^{(1)}(\{2\},\{0,1\}) \bigotimes Q_{t_1}^{(2)}(\{2\},\{0,1\}) & Q_{t_0}^{(1)}(\{2\},\{2\}) \bigoplus Q_{t_0}^{(2)}(\{2\},\{2\}) \end{array}\right) + D, \tag{2.4}$$

where D is the diagonal correction matrix that sums the rows of Q to zero; hence,

$$Q = \left(\begin{array}{cccc|c} -4 & 1 & 1 & 0 & 2 \\ 1 & -2 & 0 & 1 & 0 \\ 1 & 0 & -3 & 1 & 1 \\ 0 & 1 & 1 & -2 & 0 \\ \hline 1 & 0 & 0 & 0 & -1 \end{array}\right). \tag{2.5}$$

The steady state vector of Q in four decimal digits of precision is

$$\pi = \left(\begin{array}{cccc|c} 0.1750 & 0.1500 & 0.1000 & 0.1250 & 0.4500 \end{array}\right).$$

If we neglect the diagonal of Q which is handled separately, from Definition 2.2 it follows that each nonzero element of the HLM matrix is essentially a sum of Kronecker products, since Kronecker sums can be expressed as sums of Kronecker products. This has a very nice implication on the choice of grids in the proposed ML method when LLM aggregation is used in forming the coarser grids. LLMs 1 through K

and the HLM define the least coarsest (in other words, the finest) grid. This grid is Q and in our example has five states. Regarding the intermediate grids, let us assume that LLMs are aggregated starting from 1 up to K. Thus LLMs 2 through K and the HLM define the first coarser grid when LLM 1 is aggregated. In our example, this grid has the states in $\{(0,0),(1,0),(2,1)\}$, where the first state in each tuple is an LLM 2 state and the second state in each tuple is the corresponding HLM state. The HLM and LLMs 3 through K define the second coarser grid when LLMs 1 and 2 are aggregated. In our example, this grid is the coarsest grid corresponding to the HLM and has the states $\{(0),(1)\}$. There are no other LLMs left to be aggregated in our example; otherwise aggregation continues with the next LLM.

Now, let us concentrate on the sizes of the grids defined by the LLMs and the HLM for the assumed order in which LLMs are aggregated. In Example 1, the grids defined in this way by LLMs 1-2 and the HLM, LLM 2 and the HLM, the HLM have respectively the sizes (5×5), (3×3), (2×2) (see Table 2.1 and (2.1)-(2.2)). Clearly, we are not limited to aggregating LLMs in the order 1 through K, and can consider other orderings. The number of possible orderings of LLMs equals $K!$

In the next section, we introduce the class of ML methods with the grid choices suggested by the Kronecker structure of HMMs and remark that, just like Q, none of the grids except the coarsest is explicitly generated.

3. A class of ML methods. The class of ML methods presented in this section for HMMs with multiple macrostates have the capability of using (V, W, F) cycles [28], (power, Jacobi over-relaxation—JOR, successive over-relaxation—SOR) methods as smoothers, and (fixed, cyclic, dynamic) orders in which LLMs can be aggregated in a cycle. These parameters are respectively denoted by C, S, and O. We remark that $C \in \{V, W, F\}$, $S \in \{POWER, JOR, SOR\}$, and $O \in \{FIXED, CYCLIC, DYNAMIC\}$. In a particular ML solver, C, S, and O are fixed at the beginning.

Algorithm 1 is the driver of the ML solver. It starts executing at the finest grid involving the LLMs and the HLM, and then invokes the recursive ML function in Algorithm 2 with the order of aggregation in the list $\mathcal{C}$. Each pass through the body of the repeat-until loop in Algorithm 1 corresponds to one cycle of the ML method. Observe that steps 3 through 8 in Algorithm 2 are almost identical to the statements between steps 3 and 4 in Algorithm 1.

The variable γ in the two algorithms determines the number of recursive calls to the ML function. It is initialized to 2 for a W- or an F-cycle and to 1 for a V-cycle before ML starts executing for the first time. After this point, there are two places where the value of γ changes, and these happen only for an F-cycle. Hence, for a V-cycle γ remains 1, and for a W-cycle it remains 2, meaning for V- and W-cycles 1 and 2 recursive calls are made to the ML function on the next coarser grid, respectively. On the other hand, for an F-cycle γ is set to 1 at the boundary case of the recursion (see step 2 in Algorithm 2). Hence, an F-cycle can be seen as a recursive call to a W-cycle followed by a recursive call to a V-cycle. After the F-cycle is over, γ is reset to 2 in step 4 of Algorithm 1 so as to be ready for a new ML cycle [28, pp. 174-175].

Each ML cycle starts and ends with some number of iterations using the smoother S. See respectively the two statements after step 3 and and before step 4 in Algorithm 1. The same is true for each execution of the recursive ML function at intermediate grids as can be seen in steps 3 and 8 of Algorithm 2. The first two arguments of the call to S in both algorithms represent respectively the grid to be used in the smoothing process and the vector to be smoothed. The user is given the flexibility to specify different numbers of pre- and post-smoothings in the two algorithms. Hence, we have the nonnegative integer pairs of parameters $(MIN\_OUT\_PRE, MAX\_OUT\_PRE)$, $(MIN\_OUT\_POST, MAX\_OUT\_POST)$ for the finest grid handled by Algorithm 1, and (MIN_IN_PRE, MAX_IN_PRE), $(MIN\_IN\_POST, MAX\_IN\_POST)$ for the coarser grids handled by Algorithm 2.

For each pair of parameters (MIN_*, MAX_*), S performs MAX_* smoothings when $MIN\_* \geq MAX\_*$. When $MIN\_* < MAX\_*$, S performs an adaptive number of smoothings using the two parameters ρ and RES_COUNT as follows. Upon entry to S, the residual norm of the current solution vector is computed and recorded. Then MIN_* smoothings are performed and the residual norm of the solution vector is recomputed. If the ratio of the two residual norms is less than ρ, then S stops executing; otherwise, smoothings continue till MAX_* iterations or the ratio of residual norms of two solution vectors RES_COUNT iterations apart are less than ρ. Note that the computation of the residual vector requires an extra implicit vector-grid multiply when $S = SOR$. However, this is performed only every RES_COUNT smoothings once S is beyond MIN_* smoothings. The parameter w in the call to S is the

relaxation parameter for JOR and SOR.

ALGORITHM 1. *ML Driver.*

main()
$\mathcal{D} = [1, 2, \ldots, K+1]$; $\tilde{Q}_{\mathcal{D}} = Q$; $x_{\mathcal{D}}$ = initial approximation; $it = 0$; $cyc = 0$; $stop = FALSE$; (step 1)
if ($C == W$ or $C == F$) then (step 2)
  $\gamma = 2$;
else
  $\gamma = 1$;
repeat (step 3)
  $x_{\mathcal{D}}^{'} = S(\tilde{Q}_{\mathcal{D}}, x_{\mathcal{D}}, w, MIN\_OUT\_PRE, MAX\_OUT\_PRE, \rho, RES\_COUNT, \nu_1)$;
  $\mathcal{C} = \mathcal{D} - [head(\mathcal{D})]$;
  $\tilde{Q}_{\mathcal{C}} = P_{x_{\mathcal{D}}^{'}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}$; $x_{\mathcal{C}} = x_{\mathcal{D}}^{'} R_{\mathcal{D}}$;
  if ($\gamma == 1$) then
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, \gamma)$;
  else
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, \gamma)$;
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, y_{\mathcal{C}}, \mathcal{C}, \gamma)$;
  $y_{\mathcal{D}} = y_{\mathcal{C}} P_{x_{\mathcal{D}}^{'}}$;
  $y_{\mathcal{D}}^{'} = S(\tilde{Q}_{\mathcal{D}}, y_{\mathcal{D}}, w, MIN\_OUT\_POST, MAX\_OUT\_POST, \rho, RES\_COUNT, \nu_2)$;
  if ($C == F$) then (step 4)
    $\gamma = 2$;
  $x_{\mathcal{D}} = y_{\mathcal{D}}^{'}$; $it = it + \nu_1 + \nu_2$; $cyc = cyc + 1$; (step 5)
  normalize($x_{\mathcal{D}}$); $r = -x_{\mathcal{D}} \tilde{Q}_{\mathcal{D}}$; (step 6)
  if ($it \geq MAX\_IT$ or $time \geq MAX\_TIME$ or $\|r\| \leq STOP\_TOL$) then (step 7)
    $stop = TRUE$;
  else if ($O == DYNAMIC$) then (step 8)
    sort LLM indices $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$ into increasing order of $\|r_k\|$,
    where r_k is the residual associated with LLM k and is computed from r;
  else if ($O == CYCLIC$) then
    circular_shift($\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$);
until($stop$);
take $x_{\mathcal{D}}$ as the steady state vector π of the HMM;

ALGORITHM 2. *Recursive ML Function on LLMs in $\mathcal{D}$.*

function ML($\tilde{Q}_{\mathcal{D}}, x_{\mathcal{D}}, \mathcal{D}, \gamma$)
if ($|\mathcal{D}| == 1$) then
  $y_{\mathcal{D}}^{'} = \mathrm{solve}(\tilde{Q}_{\mathcal{D}}, x_{\mathcal{D}})$ subject to $y_{\mathcal{D}}^{'} e = 1$; (step 1)
  if ($C == F$) then (step 2)
    $\gamma = 1$;
else
  $x_{\mathcal{D}}^{'} = S(\tilde{Q}_{\mathcal{D}}, x_{\mathcal{D}}, w, MIN\_IN\_PRE, MAX\_IN\_PRE, \rho, RES\_COUNT, \nu_1)$; (step 3)
  $\mathcal{C} = \mathcal{D} - [head(\mathcal{D})]$; (step 4)
  $\tilde{Q}_{\mathcal{C}} = P_{x_{\mathcal{D}}^{'}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}$; $x_{\mathcal{C}} = x_{\mathcal{D}}^{'} R_{\mathcal{D}}$; (step 5)
  if ($\gamma == 1$) then (step 6)
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, \gamma)$;
  else
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, \gamma)$;
    $y_{\mathcal{C}} = \mathrm{ML}(\tilde{Q}_{\mathcal{C}}, y_{\mathcal{C}}, \mathcal{C}, \gamma)$;
  $y_{\mathcal{D}} = y_{\mathcal{C}} P_{x_{\mathcal{D}}^{'}}$; (step 7)
  $y_{\mathcal{D}}^{'} = S(\tilde{Q}_{\mathcal{D}}, y_{\mathcal{D}}, w, MIN\_IN\_POST, MAX\_IN\_POST, \rho, RES\_COUNT, \nu_2)$; (step 8)
return($y_{\mathcal{D}}^{'}$);

The ML solver starts with $x_{\mathcal{D}}$ which is usually set to the uniform distribution and r as the corresponding residual vector. We remark that the smoothers of choice require two vectors of length n and two vectors (three in SOR) as long as the maximum number of microstates per macrostate in the HMM. One of the vectors of length n in SOR is required for the computation of residuals in the implementation of $DYNAMIC$ ordering of LLMs for aggregation. Furthermore, if one turns off the call(s) in Algorithm 1 to Algorithm 2, Algorithm 1 reduces to an iterative solver in which $(\nu_1 + \nu_2)$ iterations are performed on Q with the iterative method S at each cycle. This is a useful feature for debugging.

The order of aggregating LLMs in each ML cycle is determined by the list $\mathcal{D}$ defined in Algorithm 1. The elements of $\mathcal{D}$ from its head to its tail are denoted respectively by $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{K+1}$. The subscripts of these elements indicate their positions in $\mathcal{D}$. In each ML cycle, the HLM is always the last model to be handled due to its special position in the hierarchy. Hence, $\mathcal{D}_{K+1}$ is given the value $(K+1)$ and is associated with the HLM; the tail of $\mathcal{D}$ always has this value. Initially, LLM k is associated with element $\mathcal{D}_k$, which has the value k for $k = 1, 2, \ldots, K$ (see step 1 of Algorithm 1). In each ML cycle, LLMs are aggregated according to these values starting from the element at the head of the list (see the second statement in the repeat-until loop of Algorithm 1). Hence, LLM $\mathcal{D}_1$ is the first LLM to be aggregated.

In the $FIXED$ order of aggregating LLMs, the initial assignment of values to the elements of $\mathcal{D}$ does not change after the ML method starts executing; this is the default order. In the $CYCLIC$ order, at the end of each ML cycle a circular shift of elements $\mathcal{D}_1$ through $\mathcal{D}_K$ in the list is performed; this ensures some kind of fairness in aggregating LLMs in the next ML cycle. On the other hand, the $DYNAMIC$ order sorts the elements $\mathcal{D}_1$ through $\mathcal{D}_K$ according to the residual norms mapped (or restricted) to the corresponding LLM at the end of the ML cycle, and aggregates the LLMs in this sorted order in the next ML cycle (see step 8 of Algorithm 1). This ensures that LLMs which have smaller residual norms are aggregated earlier at finer grids. We expect small residual norms to be indicative of good approximations in those LLMs. Note that at each intermediate grid, the recursive ML function is invoked for the next coarser grid with the list of LLMs in $\mathcal{C}$, which is formed by removing the LLM at the head of the incoming list $\mathcal{D}$ (i.e., $head(\mathcal{D})$) by aggregation (see step 4 in Algorithm 2). Once the list of LLMs is exhausted, that is $(K+1)$ is the only value remaining in list $\mathcal{D}$, backtracking from recursion starts by solving a linear system as large as the HLM matrix (see step 1 in Algorithm 2).

Before we discuss the operation that computes the next coarser grid $\tilde{Q}_{\mathcal{C}}$ from the grid $\tilde{Q}_{\mathcal{D}}$ using the smoothed vector $x'_{\mathcal{D}}$ (see step 5 in Algorithm 2), let us define the state spaces of the grids used in the ML method for large, sparse MCs in terms of a mapping [23, pp. 192–197].

DEFINITION 3.1. *Let $\mathcal{S}_{\mathcal{D}}$ and $\mathcal{S}_{\mathcal{C}}$ respectively denote the state spaces of $\tilde{Q}_{\mathcal{D}}$ and $\tilde{Q}_{\mathcal{C}}$. Then the mapping $f_{\mathcal{D}} : \mathcal{S}_{\mathcal{D}} \longrightarrow \mathcal{S}_{\mathcal{C}}$ represents the transformation of states in $\mathcal{S}_{\mathcal{D}}$ to states in $\mathcal{S}_{\mathcal{C}}$; it is surjective (i.e., onto), it satisfies*

$$\exists s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}},\ f_{\mathcal{D}}(s_{\mathcal{D}}) = s_{\mathcal{C}} \quad \textit{for each} \quad s_{\mathcal{C}} \in \mathcal{S}_{\mathcal{C}},$$

and $|\mathcal{S}_{\mathcal{C}}| \leq |\mathcal{S}_{\mathcal{D}}|$. When $|\mathcal{S}_{\mathcal{C}}| = |\mathcal{S}_{\mathcal{D}}|$, the mapping becomes bijective (i.e., one-to-one onto).

From Definition 3.1 and [23, pp. 179], we have the next proposition.

PROPOSITION 3.2. *If $\tilde{f}_{\mathcal{D}}$ denotes the converse of $f_{\mathcal{D}}$, then $\tilde{f}_{\mathcal{D}}$ is a relation from $\mathcal{S}_{\mathcal{C}}$ to $\mathcal{S}_{\mathcal{D}}$, and will not be a mapping unless $|\mathcal{S}_{\mathcal{C}}| = |\mathcal{S}_{\mathcal{D}}|$ (i.e., $f_{\mathcal{D}}$ is bijective).*

Proposition 3.2 says that, if there is at least one state in $\mathcal{S}_{\mathcal{C}}$ to which multiple states from $\mathcal{S}_{\mathcal{D}}$ are mapped under $f_{\mathcal{D}}$ (i.e., $|\mathcal{S}_{\mathcal{C}}| < |\mathcal{S}_{\mathcal{D}}|$), then the converse of $f_{\mathcal{D}}$ cannot be a function; it is just a relation.

For HMMs, the Kronecker structure (see Definition 2.2 and Proposition 2.4) and the order of component aggregation determine $\mathcal{S}_{\mathcal{D}}$ and $\mathcal{S}_{\mathcal{C}}$ as in the next proposition.

PROPOSITION 3.3. *In Algorithms 1 and 2, the components in $\mathcal{D}$ and $\mathcal{C}$ respectively define $\mathcal{S}_{\mathcal{D}}$ and $\mathcal{S}_{\mathcal{C}}$ for HMMs, and*

$$\mathcal{S}_{\mathcal{D}} = \cup_{j \in \mathcal{S}^{(K+1)}} \times_{k=1}^{|\mathcal{D}|} \mathcal{S}_j^{(\mathcal{D}_k)} \quad \textit{and} \quad \mathcal{S}_{\mathcal{C}} = \cup_{j \in \mathcal{S}^{(K+1)}} \times_{k=1}^{|\mathcal{C}|} \mathcal{S}_j^{(\mathcal{C}_k)},$$

where $\times$ is the Cartesian product operator. Furthermore,

$$|\mathcal{S}_{\mathcal{D}}| = \sum_{j=0}^{|\mathcal{S}^{(K+1)}|-1} \prod_{k=1}^{|\mathcal{D}|} |\mathcal{S}_j^{(\mathcal{D}_k)}| \quad \textit{and} \quad |\mathcal{S}_{\mathcal{C}}| = \sum_{j=0}^{|\mathcal{S}^{(K+1)}|-1} \prod_{k=1}^{|\mathcal{C}|} |\mathcal{S}_j^{(\mathcal{C}_k)}|.$$

At the finest level in Algorithm 1, $|\mathcal{S}_\mathcal{D}| = n$.

Observe from Definition 2.2 that $\mathcal{S}_\mathcal{D}$ and $\mathcal{S}_\mathcal{C}$ for HMMs given in Proposition 3.3 satisfy the mapping $f_\mathcal{D} : \mathcal{S}_\mathcal{D} \longrightarrow \mathcal{S}_\mathcal{C}$ in Definition 3.1.

Now we return to the computation of the coarser grid and the coarser approximation. For each state $s_\mathcal{C} \in \mathcal{S}_\mathcal{C}$, the columns of the grid $\tilde{Q}_\mathcal{D}$ corresponding to the states in $\mathcal{S}_\mathcal{D}$ that get mapped to the same state $s_\mathcal{C}$ are summed. The aggregation on the columns of $\tilde{Q}_\mathcal{D}$ is also performed on the columns of the smoothed row vector $x'_\mathcal{D}$ yielding the vector $x_\mathcal{C}$ in step 5 of Algorithm 2. These are achieved by using the *restriction* [22] (or aggregation) operator defined next.

DEFINITION 3.4. *The* $(|\mathcal{S}_\mathcal{D}| \times |\mathcal{S}_\mathcal{C}|)$ *restriction operator* $R_\mathcal{D}$ *for the mapping* $f_\mathcal{D} : \mathcal{S}_\mathcal{D} \longrightarrow \mathcal{S}_\mathcal{C}$ *has its* $(s_\mathcal{D}, s_\mathcal{C})$*th element given by*

$$r_\mathcal{D}(s_\mathcal{D}, s_\mathcal{C}) = \begin{cases} 1 & \text{if } f_\mathcal{D}(s_\mathcal{D}) = s_\mathcal{C} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } s_\mathcal{D} \in \mathcal{S}_\mathcal{D} \text{ and } s_\mathcal{C} \in \mathcal{S}_\mathcal{C}.$$

PROPOSITION 3.5. *The restriction operator* $R_\mathcal{D}$ *is nonnegative (i.e.,* $R_\mathcal{D} \geq 0$*), has only a single nonzero with the value one in each row and therefore row sums of one (i.e.,* $R_\mathcal{D} e = e$*). Furthermore, since there is at least one nonzero in each column of* $R_\mathcal{D}$ *(i.e.,* $e^T R_\mathcal{D} > 0$*), it is also the case that* $rank(R_\mathcal{D}) = |\mathcal{S}_\mathcal{C}|$*. Thus the product* $\tilde{Q}_\mathcal{D} R_\mathcal{D}$ *yields a column aggregated grid whose row sums (i.e.,* $\tilde{Q}_\mathcal{D} R_\mathcal{D} e = \tilde{Q}_\mathcal{D} e$*) are zero if* $\tilde{Q}_\mathcal{D}$ *has row sums of zero (i.e.,* $\tilde{Q}_\mathcal{D} e = 0$*).*

For each state $s_\mathcal{C} \in \mathcal{S}_\mathcal{C}$, the rows of $\tilde{Q}_\mathcal{D} R_\mathcal{D}$ corresponding to the states in $\mathcal{S}_\mathcal{D}$ that are mapped to the same state $s_\mathcal{C}$ are multiplied with the corresponding normalized elements of the smoothed row vector $x'_\mathcal{D}$ and summed. This is achieved by using the *prolongation* [22] (or disaggregation) operator defined next.

DEFINITION 3.6. *The* $(|\mathcal{S}_\mathcal{C}| \times |\mathcal{S}_\mathcal{D}|)$ *prolongation operator* $P_{x'_\mathcal{D}}$ *for the mapping* $f_\mathcal{D} : \mathcal{S}_\mathcal{D} \longrightarrow \mathcal{S}_\mathcal{C}$ *has its* $(s_\mathcal{C}, s_\mathcal{D})$*th element given by*

$$p_{x'_\mathcal{D}}(s_\mathcal{C}, s_\mathcal{D}) = \begin{cases} x'_\mathcal{D}(s_\mathcal{D}) / \sum_{s_\mathcal{D} \in \mathcal{S}_\mathcal{D}, f_\mathcal{D}(s_\mathcal{D}) = s_\mathcal{C}} x'_\mathcal{D}(s_\mathcal{D}) & \text{if } f_\mathcal{D}(s_\mathcal{D}) = s_\mathcal{C} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } s_\mathcal{D} \in \mathcal{S}_\mathcal{D} \text{ and } s_\mathcal{C} \in \mathcal{S}_\mathcal{C}.$$

PROPOSITION 3.7. *If* $x'_\mathcal{D} > 0$*, the prolongation operator* $P_{x'_\mathcal{D}}$ *is nonnegative (i.e.,* $P_{x'_\mathcal{D}} \geq 0$*), has the same nonzero structure as the transpose of* $R_\mathcal{D}$ *(i.e.,* $R_\mathcal{D}^T$*), a single nonzero in each column (i.e.,* $e^T P_{x'_\mathcal{D}} > 0$*), and at least one nonzero in each row, implying* $rank(P_{x'_\mathcal{D}}) = |\mathcal{S}_\mathcal{C}|$*. Furthermore, when* $x'_\mathcal{D} > 0$*, each row of* $P_{x'_\mathcal{D}}$ *is a probability vector, implying* $P_{x'_\mathcal{D}}$ *has row sums of one (i.e.,* $P_{x'_\mathcal{D}} e = e$*) just like* $R_\mathcal{D}$*. Thus premultiplying* $\tilde{Q}_\mathcal{D} R_\mathcal{D}$ *by* $P_{x'_\mathcal{D}}$ *yields the* $(|\mathcal{S}_\mathcal{C}| \times |\mathcal{S}_\mathcal{C}|)$ *square grid* $\tilde{Q}_\mathcal{C}$*, which has row sums of zero regardless of the norm of* $x'_\mathcal{D}$*.*

The prolongation operator depends not only on $\mathcal{S}_\mathcal{D}$ and $\mathcal{S}_\mathcal{C}$, but also on the smoothed vector $x'_\mathcal{D}$, which is indicated by using the subscript $x'_\mathcal{D}$ rather than $\mathcal{D}$. This implies that the elements of $\tilde{Q}_\mathcal{C}$ depend on $x'_\mathcal{D}$ and will be different in each cycle of the ML solver.

LEMMA 3.8. *If* $x'_\mathcal{D} > 0$*, then* $P_{x'_\mathcal{D}} R_\mathcal{D} = I_\mathcal{C}$*, where* $I_\mathcal{C}$ *is the identity matrix of order* $|\mathcal{S}_\mathcal{C}|$*.*

Proof. The identity follows from Propositions 3.5 and 3.7 by the facts that $P_{x'_\mathcal{D}} \geq 0$, $R_\mathcal{D} \geq 0$, $P_{x'_\mathcal{D}}$ has the same nonzero structure as $R_\mathcal{D}^T$, $P_{x'_\mathcal{D}} e = e$, and $e^T R_\mathcal{D}^T = e^T$. □

When $x'_\mathcal{D} > 0$, we can state the next definition [21, p. 387] using $R_\mathcal{D}(P_{x'_\mathcal{D}} R_\mathcal{D}) P_{x'_\mathcal{D}} = R_\mathcal{D}(I_\mathcal{C}) P_{x'_\mathcal{D}} = R_\mathcal{D} P_{x'_\mathcal{D}}$ from Lemma 3.8, $R_\mathcal{D} \geq 0$, $R_\mathcal{D} e = e$ and $P_{x'_\mathcal{D}} \geq 0$, $P_{x'_\mathcal{D}} e = e$ from Propositions 3.5 and 3.7, respectively.

COROLLARY 3.9. *When* $x'_\mathcal{D} > 0$*, the* $(|\mathcal{S}_\mathcal{D}| \times |\mathcal{S}_\mathcal{D}|)$ *matrix*

$$H_{x'_\mathcal{D}} = R_\mathcal{D} P_{x'_\mathcal{D}}$$

defines a nonnegative projector (i.e., $H_{x'_\mathcal{D}} \geq 0$ *and* $H^2_{x'_\mathcal{D}} = H_{x'_\mathcal{D}}$*) which satisfies* $H_{x'_\mathcal{D}} e = e$*.*

LEMMA 3.10. *If* $x'_\mathcal{D} > 0$*, then* $x'_\mathcal{D} H_{x'_\mathcal{D}} = x'_\mathcal{D}$*.*

Proof. The identity follows from the definitions of restriction and prolongation operations (see Definitions 3.4 and 3.6) and the fact that the restricted and then prolonged row vector is $x'_\mathcal{D}$. □

The convergence analysis in section 4 is based on showing that the coarser grid $\tilde{Q}_{\mathcal{C}}$ is an irreducible CTMC and $x_{\mathcal{C}} > 0$ if the finer grid $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC and $x'_{\mathcal{D}} > 0$. This has been done for HMMs with one macro state in [9, p. 348]. In section 4, we show the results for the mapping $f : \mathcal{S}_{\mathcal{D}} \longrightarrow \mathcal{S}_{\mathcal{C}}$ in Definition 3.1.

Step 7 in Algorithm 2 corresponds to the opposite of what is done on $x'_{\mathcal{D}}$ in step 5; that is, it performs disaggregation using the newly computed vector $y_{\mathcal{C}}$ and the prolongation operator $P_{x'_{\mathcal{D}}}$ (which is based on the smoothed vector $x'_{\mathcal{D}}$) to obtain the vector $y_{\mathcal{D}}$. The next result follows from Proposition 3.7

PROPOSITION 3.11. *If $y_{\mathcal{C}} > 0$ and $x'_{\mathcal{D}} > 0$, then $y_{\mathcal{D}} = y_{\mathcal{C}} P_{x'_{\mathcal{D}}} > 0$, since $e^T P_{x'_{\mathcal{D}}} > 0$.*

Similar aggregation and disaggregation operations are performed in Algorithm 1 at the finest grid Q.

The Kronecker representation of $\tilde{Q}_{\mathcal{C}}$ for an HMM with one macrostate is given in [9, p. 347]. Here we extend it to multiple macrostates and show that $\tilde{Q}_{\mathcal{C}}$ can be expressed as a sum of Kronecker products as in Definition 2.2 using $\sum_{i,j \in \mathcal{S}^{(K+1)}} |\mathcal{T}_{i,j}|$ vectors each of length at most $\max_{j \in \mathcal{S}^{(K+1)}} (\prod_{k=2}^{|\mathcal{C}|} |\mathcal{S}_j^{(\mathcal{C}_k)}|)$ and the matrices corresponding to the components in $\mathcal{C}$ excluding $(K+1)$, which denotes the HLM (see Proposition 3.3). More specifically, we have the next definition.

DEFINITION 3.12. *If $h = \mathcal{D}_1$ is the index of the aggregated component, then the $s_{\mathcal{C}}$th element of the vector corresponding to the t_eth term in block (i,j) of the aggregated CTMC $\tilde{Q}_{\mathcal{C}}$ is defined as*

$$a_{(\mathcal{C},t_e),(i,j)}(s_{\mathcal{C}}) = \frac{\left(\sum_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}, f_{\mathcal{D}}(s_{\mathcal{D}}) = s_{\mathcal{C}}} x'_{\mathcal{D}}(s_{\mathcal{D}}) \; a_{(\mathcal{D},t_e),(i,j)}(s_{\mathcal{D}}) \; (e^T_{s_{\mathcal{D}}(h)} Q^{(h)}_{t_e}(\mathcal{S}^{(h)}_i, \mathcal{S}^{(h)}_j) e) \right)}{x_{\mathcal{C}}(s_{\mathcal{C}})}$$
$$\text{for } s_{\mathcal{C}} \in \mathcal{S}_{\mathcal{C}}, \; t_e \in \mathcal{T}_{i,j}, \; \text{and } i,j \in \mathcal{S}^{(K+1)},$$

where $a_{(\mathcal{D},t_e),(i,j)} = e$ if $\mathcal{D}$ corresponds to the finest level, $s_{\mathcal{D}}(h) \in \mathcal{S}^{(h)}$, and $e_{s_{\mathcal{D}}(h)}$ is the $s_{\mathcal{D}}(h)$th column of the identity matrix of order $|\mathcal{S}^{(h)}_i|$. Hence,

$$\begin{aligned}
\tilde{Q}_{\mathcal{C}}(j,j) = & \bigoplus_{k=1}^{|\mathcal{C}|-1} Q^{(\mathcal{C}_k)}_{t_0}(\mathcal{S}^{(\mathcal{C}_k)}_j, \mathcal{S}^{(\mathcal{C}_k)}_j) + \sum_{t_e \in \mathcal{T}_{j,j}} \bigotimes_{k=1}^{|\mathcal{C}|-1} \operatorname{diag}(a_{(\mathcal{C},t_e),(j,j)}) Q^{(\mathcal{C}_k)}_{t_e}(\mathcal{S}^{(\mathcal{C}_k)}_j, \mathcal{S}^{(\mathcal{C}_k)}_j) \\
& - \bigoplus_{k=1}^{|\mathcal{C}|-1} \operatorname{diag}(Q^{(\mathcal{C}_k)}_{t_0}(\mathcal{S}^{(\mathcal{C}_k)}_j, \mathcal{S}^{(\mathcal{C}_k)}_j) e) \\
& - \sum_{i \in \mathcal{S}^{(K+1)}} \sum_{t_e \in \mathcal{T}_{j,i}} \bigotimes_{k=1}^{|\mathcal{C}|-1} \operatorname{diag}(a_{(\mathcal{C},t_e),(j,i)}) \operatorname{diag}(Q^{(\mathcal{C}_k)}_{t_e}(\mathcal{S}^{(\mathcal{C}_k)}_j, \mathcal{S}^{(\mathcal{C}_k)}_i) e) \quad \text{for } j \in \mathcal{S}^{(K+1)}, \\
\tilde{Q}_{\mathcal{C}}(i,j) = & \sum_{t_e \in \mathcal{T}_{i,j}} \bigotimes_{k=1}^{|\mathcal{C}|-1} \operatorname{diag}(a_{(\mathcal{C},t_e),(i,j)}) Q^{(\mathcal{C}_k)}_{t_e}(\mathcal{S}^{(\mathcal{C}_k)}_i, \mathcal{S}^{(\mathcal{C}_k)}_j) \quad \text{for } i,j \in \mathcal{S}^{(K+1)}, i \neq j.
\end{aligned}$$

Observe from Proposition 2.3 that the last two terms of $\tilde{Q}_{\mathcal{C}}(j,j)$ return a diagonal matrix which sums the rows of $\tilde{Q}_{\mathcal{C}}(j,j)$ to zero. Furthermore, the vectors $a_{(\mathcal{D},t_e),(i,j)}$ for $t_e \in \mathcal{T}_{i,j}$ and $i,j \in \mathcal{S}^{(K+1)}$ at the finest level consist of all ones, and therefore need not be stored. When the recursion ends at the HLM, $\tilde{Q}_{\mathcal{C}}$ is a $(|\mathcal{S}^{(K+1)}| \times |\mathcal{S}^{(K+1)}|)$ CTMC, and therefore is generated and stored explicitly in sparse format so that it can be solved using a direct method. We remark that $a_{(\mathcal{C},t_e),(i,j)} = e$ for those t_e which have all $Q^{(\mathcal{C}_k)}_{t_e}(\mathcal{S}^{(\mathcal{C}_k)}_i, \mathcal{S}^{(\mathcal{C}_k)}_j)$ as diagonal matrices of size $(|\mathcal{S}^{(\mathcal{C}_k)}_i| \times |\mathcal{S}^{(\mathcal{C}_k)}_j|)$ with ones along their diagonal for $k = 1, 2, \ldots, |\mathcal{C}| - 1$ and $i,j \in \mathcal{S}^{(K+1)}$. Since component matrices forming $\tilde{Q}_{\mathcal{C}}(i,j)$ for $i,j \in \mathcal{S}^{(K+1)}, i \neq j$, can very well be rectangular, we refrain from using I, and remark that such vectors need not be stored either.

The next section presents the convergence analysis of the proposed class of ML methods for large, sparse Markov chains.

4. Convergence analysis. Let $\mathcal{D}$ represent the current level and $\mathcal{C}$ represent the next coarser level in the ML cycle as in Algorithms 1 and 2. Let $\mathcal{S}_{\mathcal{D}}$ and $\mathcal{S}_{\mathcal{C}}$ denote respectively the state spaces of $\tilde{Q}_{\mathcal{D}}$ and $\tilde{Q}_{\mathcal{C}}$, and assume that the mapping of states from $\mathcal{S}_{\mathcal{D}}$ to the states in $\mathcal{S}_{\mathcal{C}}$ is onto and satisfies $|\mathcal{S}_{\mathcal{C}}| \leq |\mathcal{S}_{\mathcal{D}}|$ as in Definition 3.1. The results that are presented in this section for Algorithms 1 and 2 are general in that the Kronecker representation of the grids particular to HMMs is not utilized.

4.1. Irreducibility of the coarser grids. Recall that $R_{\mathcal{D}} \geq 0$, $R_{\mathcal{D}}e = e$, $e^T R_{\mathcal{D}} > 0$ from Proposition 3.5, and if $x'_{\mathcal{D}} > 0$, then $P_{x'_{\mathcal{D}}} \geq 0$, $P_{x'_{\mathcal{D}}} e = e$, $e^T P_{x'_{\mathcal{D}}} > 0$ from Proposition 3.7. Now, consider the definition of irreducibility in [21, p. 209] and [25, p. 13].

DEFINITION 4.1. *Let $\mathcal{G}(A) = (\mathcal{V}, \mathcal{E})$ be the directed graph (digraph) associated with a square matrix A, where $\mathcal{V}$ is the set of vertices (or nodes) and $\mathcal{E}$ is the set of directed edges (or arcs). Then $\mathcal{V}$ has as many vertices as the order of A and $\mathcal{E}$ has a directed edge from vertex i to j if and only if $a_{i,j} \neq 0$. The digraph $\mathcal{G}(A)$ is said to be strongly connected, if for each pair of vertices (i, j), there is a sequence of directed edges leading from i to j. The matrix A is said to be irreducible if and only if $\mathcal{G}(A)$ is strongly connected. In MC terminology, the vertices in the graph correspond to states. A state is said to be reachable from another state if there is a path of transitions that lead to the state from the other state, and by definition a state is reachable from itself. A MC is irreducible if each state is reachable from every other state.*

PROPOSITION 4.2. *Irreducibility concerns transitions that appear in the off-diagonal part of a CTMC and the nonzeros corresponding to rates of these transitions have positive values.*

Now, we are in a position to state and prove a lemma for Algorithms 1 and 2 that will be used in the convergence analysis.

LEMMA 4.3. *The coarser grid $\tilde{Q}_{\mathcal{C}} = P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}$ is an irreducible CTMC and $x_{\mathcal{C}} = x'_{\mathcal{D}} R_{\mathcal{D}} > 0$ if the finer grid $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC and $x'_{\mathcal{D}} > 0$.*

Proof. First, we show that $\tilde{Q}_{\mathcal{C}} = P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}$ is an irreducible CTMC using Definition 4.1 and Proposition 4.2. Without losing generality, consider the pair of different states $s_{\mathcal{D}}, s'_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$. Through $f : \mathcal{S}_{\mathcal{D}} \longrightarrow \mathcal{S}_{\mathcal{C}}$ in Definition 3.1, these pair of states are mapped respectively to the states $s_{\mathcal{C}}, s'_{\mathcal{C}} \in \mathcal{S}_{\mathcal{C}}$ (i.e., $f(s_{\mathcal{D}}) = s_{\mathcal{C}}$ and $f(s'_{\mathcal{D}}) = s'_{\mathcal{C}}$). Since $\tilde{Q}_{\mathcal{D}}$ is irreducible, there exists a path of transitions from $s_{\mathcal{D}}$ to $s'_{\mathcal{D}}$ in $S_{\mathcal{D}}$ in the form $s_{\mathcal{D}} = s_1, s_2, \ldots, s_m = s'_{\mathcal{D}}$, where $m \leq |\mathcal{S}_{\mathcal{D}}|$, $s_k \in S_{\mathcal{D}}$, and $\tilde{q}_D(s_k, s_{k+1}) > 0$ for $k \in \{1, 2, \ldots, m-1\}$. Mapping this path onto $S_{\mathcal{C}}$ yields the path $s_{\mathcal{C}} = t_1, t_2, \ldots, t_m = s'_{\mathcal{C}}$, where $f(s_k) = t_k \in S_{\mathcal{C}}$. Now, let e_{t_k} denote the t_kth column of $I_{\mathcal{C}}$. Then, in the mapped path, we either have $t_k = t_{k+1}$ or $\tilde{q}_{\mathcal{C}}(t_k, t_{k+1}) > 0$, where the latter follows from

$$\tilde{q}_{\mathcal{C}}(t_k, t_{k+1}) = e_{t_k}^T \tilde{Q}_{\mathcal{C}} e_{t_{k+1}} = (e_{t_k}^T P_{x'_{\mathcal{D}}}) \tilde{Q}_{\mathcal{D}} (R_{\mathcal{D}} e_{t_{k+1}}) \geq p_{x'_{\mathcal{D}}}(t_k, s_k) \tilde{q}_{\mathcal{D}}(s_k, s_{k+1}) r_{\mathcal{D}}(s_{k+1}, t_{k+1}),$$

since $x_D(s_k) > 0$ (implying $p_{x'_{\mathcal{D}}}(t_k, s_k) > 0$ from Definition 3.6), $\tilde{q}_{\mathcal{D}}(s_k, s_{k+1}) > 0$, and $f(s_{k+1}) = t_{k+1}$ (implying $r_{\mathcal{D}}(s_{k+1}, t_{k+1}) = 1$ from Definition 3.4). Thus we conclude, $s'_{\mathcal{C}}$ is reachable from $s_{\mathcal{C}}$.

We have effectively shown that each state in $\tilde{Q}_{\mathcal{C}}$ is reachable from every other state. The question that arises at this point is whether a row of $\tilde{Q}_{\mathcal{C}}$ can become zero after the restriction. The answer is no, as long as $\mathcal{S}_{\mathcal{C}}$ has multiple states (i.e., $|\mathcal{S}_{\mathcal{C}}| > 1$), since all states in $\mathcal{S}_{\mathcal{D}}$ that are mapped to a particular state in $\mathcal{S}_{\mathcal{C}}$ cannot have all their transitions among themselves. This would imply that $\tilde{Q}_{\mathcal{D}}$ is reducible, which is a contradiction. Furthermore, since the row sums of $\tilde{Q}_{\mathcal{C}}$ are zero (i.e., $\tilde{Q}_{\mathcal{C}} e = (P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}) e = P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} (R_{\mathcal{D}} e) = P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} e = 0$ because $\tilde{Q}_{\mathcal{D}}$ is a CTMC and $\tilde{Q}_{\mathcal{D}} e = 0$), its diagonal must be equal to its negated off-diagonal row sums. Hence, $\tilde{Q}_{\mathcal{C}}$ is an irreducible CTMC.

Now, we show that $x_{\mathcal{C}} > 0$. Since $x_{\mathcal{C}} = x'_{\mathcal{D}} R_{\mathcal{D}}$, $x'_{\mathcal{D}} = e^T \text{diag}(x'_{\mathcal{D}})$ where $\text{diag}(x'_{\mathcal{D}})$ is the diagonal matrix with $x'_{\mathcal{D}}$ along its diagonal, $\text{diag}(x'_{\mathcal{D}}) R_{\mathcal{D}}$ has the same nonzero structure as $R_{\mathcal{D}}$, and $e^T R_{\mathcal{D}} > 0$, we have $x_{\mathcal{C}} = x'_{\mathcal{D}} R_{\mathcal{D}} = (e^T \text{diag}(x'_{\mathcal{D}})) R_{\mathcal{D}} = e^T (\text{diag}(x'_{\mathcal{D}}) R_{\mathcal{D}}) > 0$ when $x'_{\mathcal{D}} > 0$. ∎

COROLLARY 4.4. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC, $x'_{\mathcal{D}} > 0$, and $x'_{\mathcal{D}} \tilde{Q}_{\mathcal{D}} = 0$, then $x_{\mathcal{C}} \tilde{Q}_{\mathcal{C}} = 0$, where $\tilde{Q}_{\mathcal{C}} = P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}$ and $x_{\mathcal{C}} = x'_{\mathcal{D}} R_{\mathcal{D}}$.*

Proof. We have $x_{\mathcal{C}} \tilde{Q}_{\mathcal{C}} = (x'_{\mathcal{D}} R_{\mathcal{D}})(P_{x'_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}}) = (x'_{\mathcal{D}} R_{\mathcal{D}} P_{x'_{\mathcal{D}}}) \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}} = (x'_{\mathcal{D}} H_{x'_{\mathcal{D}}}) \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}} = (x'_{\mathcal{D}}) \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}} = (x'_{\mathcal{D}} \tilde{Q}_{\mathcal{D}}) R_{\mathcal{D}} = 0$, since $x'_{\mathcal{D}} H_{x'_{\mathcal{D}}} = x'_{\mathcal{D}}$ from Lemma 3.10 and $x'_{\mathcal{D}} \tilde{Q}_{\mathcal{D}} = 0$ by assumption. ∎

PROPOSITION 4.5. *If $\pi_{\mathcal{D}} = \pi > 0$ denotes the steady state vector of the irreducible grid $Q_{\mathcal{D}} = Q$ at the finest level $\mathcal{D}$, then the irreducible grid obtained by exact aggregation at the next coarser level $\mathcal{C}$ is $Q_{\mathcal{C}} = P_{\pi_{\mathcal{D}}} Q_{\mathcal{D}} R_{\mathcal{D}}$ and has the steady state vector $\pi_{\mathcal{C}} = \pi_{\mathcal{D}} R_{\mathcal{D}} > 0$. The result extends to all adjacent pairs of levels $\mathcal{D}$ and $\mathcal{C}$ as long as level $\mathcal{D}$ has the exact irreducible grid $Q_{\mathcal{D}}$ and its steady state vector $\pi_{\mathcal{D}}$ is used to compute the irreducible grid $Q_{\mathcal{C}}$ at the next coarser level $\mathcal{C}$.*

The proposition follows from $\pi_{\mathcal{C}} Q_{\mathcal{C}} = (\pi_D R_{\mathcal{D}})(P_{\pi_{\mathcal{D}}} Q_{\mathcal{D}} R_{\mathcal{D}}) = (\pi_D R_{\mathcal{D}} P_{\pi_{\mathcal{D}}}) Q_{\mathcal{D}} R_{\mathcal{D}} = (\pi_D H_{\pi_{\mathcal{D}}}) Q_{\mathcal{D}} R_{\mathcal{D}} = (\pi_D) Q_{\mathcal{D}} R_{\mathcal{D}} = (\pi_D Q_{\mathcal{D}}) R_{\mathcal{D}} = 0$ since $\pi_D H_{\pi_{\mathcal{D}}} = \pi_{\mathcal{D}}$ from Lemma 3.10 and $\pi_{\mathcal{D}} Q_{\mathcal{D}} = 0$ by assumption.

The next subsection specifies sufficient conditions for a converging smoother to provide improved solutions at each level.

4.2. Convergence of the smoothers. By definition at the finest level in Algorithm 1 and by construction at the coarser levels in Algorithm 2, the matrix $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC when $x'_{\mathcal{D}} > 0$ (see Lemma 4.3). Now, consider the nontransposed homogeneous singular linear system in the next definition (cf. (1.1)).

DEFINITION 4.6. *The problem at level $\mathcal{D}$ in the ML method is to solve*

$$\tilde{\pi}_{\mathcal{D}}\tilde{Q}_{\mathcal{D}} = 0 \quad \textit{subject to} \quad \tilde{\pi}_{\mathcal{D}}e = 1,$$

where $\tilde{\pi}_{\mathcal{D}} > 0$ is the steady state vector of the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$.

PROPOSITION 4.7. *At the finest level $\mathcal{D}$, the steady state vector of the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$ satisfies $\tilde{\pi}_{\mathcal{D}} = \pi$ since $\tilde{Q}_{\mathcal{D}} = Q$.*

Now, consider the splitting of $\tilde{Q}_{\mathcal{D}}$ in the next definition.

DEFINITION 4.8. *Let $\tilde{Q}_{\mathcal{D}}$ be split as*

$$\tilde{Q}_{\mathcal{D}} = D_{\mathcal{D}} - U_{\mathcal{D}} - L_{\mathcal{D}} = M_{\mathcal{D}} - N_{\mathcal{D}},$$

where $D_{\mathcal{D}}$, $U_{\mathcal{D}}$, and $L_{\mathcal{D}}$ are respectively the diagonal, negated strictly upper-triangular, and negated strictly lower-triangular parts of $\tilde{Q}_{\mathcal{D}}$, and $M_{\mathcal{D}}$ is nonsingular (i.e., $M_{\mathcal{D}}^{-1}$ exists).

PROPOSITION 4.9. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC, each of the terms $D_{\mathcal{D}}$, $U_{\mathcal{D}}$, and $L_{\mathcal{D}}$ in the splitting of $\tilde{Q}_{\mathcal{D}}$ is nonpositive (i.e., $D_{\mathcal{D}} \leq 0$, $U_{\mathcal{D}} \leq 0$, and $L_{\mathcal{D}} \leq 0$); furthermore, $\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}}) \neq 0$ for all $s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$, implying $D_{\mathcal{D}}^{-1}$ and $(D_{\mathcal{D}} - U_{\mathcal{D}})^{-1}$ exist.*

The next definition involving the iteration matrices of the $POWER$, JOR, and SOR smoothers follows from [25, Ch. 3].

DEFINITION 4.10. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC, then the POWER, JOR, and SOR smoothers are based on different splittings of $\tilde{Q}_{\mathcal{D}}$, where each yields an iteration matrix of the form*

$$T_{\mathcal{D}} = N_{\mathcal{D}}M_{\mathcal{D}}^{-1}$$

and the sequence of approximations

$$x_{\mathcal{D}}^{(m+1)} = x_{\mathcal{D}}^{(m)}T_{\mathcal{D}} \quad \textit{for} \quad m = 0, 1, \ldots.$$

The particular splittings corresponding to the three smoothers are

$$\begin{aligned} M_{\mathcal{D}}^{POWER} &= -\alpha I_{\mathcal{D}}, \quad N_{\mathcal{D}}^{POWER} = -\alpha(I_{\mathcal{D}} + \tilde{Q}_{\mathcal{D}}/\alpha), \\ M_{\mathcal{D}}^{JOR} &= D_{\mathcal{D}}/\omega, \quad N_{\mathcal{D}}^{JOR} = (1-\omega)D_{\mathcal{D}}/\omega + L_{\mathcal{D}} + U_{\mathcal{D}}, \\ M_{\mathcal{D}}^{SOR} &= D_{\mathcal{D}}/\omega - U_{\mathcal{D}}, \quad N_{\mathcal{D}}^{SOR} = (1-\omega)D_{\mathcal{D}}/\omega + L_{\mathcal{D}}, \end{aligned}$$

where $\alpha \in [max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}}|\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ is the uniformization parameter of POWER and $\omega \in (0, 2)$ is the relaxation parameter of JOR and SOR. The JOR and SOR splittings reduce to Jacobi and Gauss-Seidel (GS) splittings for $\omega = 1$. Hence, the iteration matrices corresponding to the three splittings are

$$T_{\mathcal{D}}^{POWER} = I_{\mathcal{D}} + \tilde{Q}_{\mathcal{D}}/\alpha,$$

$$T_{\mathcal{D}}^{JOR} = (1-\omega)I_{\mathcal{D}} + \omega(L_{\mathcal{D}} + U_{\mathcal{D}})D_{\mathcal{D}}^{-1} \quad \textit{and} \quad T_{\mathcal{D}}^{SOR} = ((1-\omega)D_{\mathcal{D}}/\omega + L_{\mathcal{D}})(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1}.$$

The next lemma specifies a fixed point for the iteration matrices of the $POWER$, JOR, and SOR smoothers.

LEMMA 4.11. *For the smoother $S = \{POWER, JOR, SOR\}$ at level $\mathcal{D}$, the steady state vector $\tilde{\pi}_{\mathcal{D}}$ of $\tilde{Q}_{\mathcal{D}}$ satisfies $\tilde{\pi}_{\mathcal{D}}T_{\mathcal{D}} = \tilde{\pi}_{\mathcal{D}}$, where $T_{\mathcal{D}}$ is the iteration matrix of the corresponding smoother.*

Proof. The proof rests on the particular form of the iteration matrices in Definition 4.10 and the fact that $\tilde{\pi}_{\mathcal{D}}\tilde{Q}_{\mathcal{D}} = 0$ in Definition 4.6. For $POWER$, we have

$$\tilde{\pi}_{\mathcal{D}}T_{\mathcal{D}}^{POWER} = \tilde{\pi}_{\mathcal{D}}I_{\mathcal{D}} + \tilde{\pi}_{\mathcal{D}}\tilde{Q}_{\mathcal{D}}/\alpha = \tilde{\pi}_{\mathcal{D}}.$$

For JOR, we have

$$\begin{aligned}\tilde{\pi}_{\mathcal{D}} T_{\mathcal{D}}^{JOR} &= \tilde{\pi}_{\mathcal{D}}(1-\omega) I_{\mathcal{D}} + \tilde{\pi}_{\mathcal{D}}\omega(L_{\mathcal{D}} + U_{\mathcal{D}}) D_{\mathcal{D}}^{-1} = (1-\omega)\tilde{\pi}_{\mathcal{D}} + \omega\tilde{\pi}_{\mathcal{D}}(D_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}}) D_{\mathcal{D}}^{-1} \\ &= (1-\omega)\tilde{\pi}_{\mathcal{D}} + \omega\tilde{\pi}_{\mathcal{D}}(I_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}} D_{\mathcal{D}}^{-1}) = (1-\omega)\tilde{\pi}_{\mathcal{D}} + \omega\tilde{\pi}_{\mathcal{D}} - \omega\tilde{\pi}_{\mathcal{D}}\tilde{Q}_{\mathcal{D}} D_{\mathcal{D}}^{-1} = \tilde{\pi}_{\mathcal{D}}\end{aligned}$$

since $L_{\mathcal{D}} + U_{\mathcal{D}} = D_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}}$ from Definition 4.8. For SOR, we have

$$\begin{aligned}\tilde{\pi}_{\mathcal{D}} T_{\mathcal{D}}^{SOR} &= \tilde{\pi}_{\mathcal{D}}((1-\omega) D_{\mathcal{D}}/\omega + L_{\mathcal{D}})(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1} = \tilde{\pi}_{\mathcal{D}}(D_{\mathcal{D}}/\omega - U_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}})(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1} \\ &= \tilde{\pi}_{\mathcal{D}}(I_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}}(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1}) = \tilde{\pi}_{\mathcal{D}} - \tilde{\pi}_{\mathcal{D}}\tilde{Q}_{\mathcal{D}}(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1} = \tilde{\pi}_{\mathcal{D}}\end{aligned}$$

since $(1-\omega) D_{\mathcal{D}}/\omega + L_{\mathcal{D}} = D_{\mathcal{D}}/\omega - U_{\mathcal{D}} - \tilde{Q}_{\mathcal{D}}$ from Definition 4.8. □

Before we state another lemma, we recall the definitions of primitivity and M-matrix from [25, p. 352 and p. 170] and remark that detailed information concerning M-matrices may be found in [4].

DEFINITION 4.12. *Let $\sigma(A)$ denote the set of eigenvalues (or spectrum) of the square matrix A (i.e., $\sigma(A) = \{\lambda \mid Av = \lambda v, v \neq 0\}$) and let $\rho(A)$ be the spectral radius of A (i.e., $\rho(A) = \{\max|\lambda| \mid \lambda \in \sigma(A)\}$). A nonnegative, irreducible matrix B is said to be primitive if it has a single eigenvalue with magnitude $\rho(B)$.*

DEFINITION 4.13. *Any square matrix A of the form $A = sI - B$ with $s > 0$ and $B \geq 0$ for which $s \geq \rho(B)$ is called an M-matrix.*

PROPOSITION 4.14. *The negated CTMC $-\tilde{Q}_{\mathcal{D}}$ is a singular M-matrix.*

The next proposition follows from [21, p. 640] and [25, p. 118].

PROPOSITION 4.15. *For the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$, the matrix $e\tilde{\pi}_{\mathcal{D}}$ has the steady vector of $\tilde{Q}_{\mathcal{D}}$ in each of its rows, and therefore is positive (i.e., $e\tilde{\pi}_{\mathcal{D}} > 0$), a probability matrix (i.e., $e\tilde{\pi}_{\mathcal{D}} e = e$), and of rank 1.*

COROLLARY 4.16. *When $\tilde{Q}_{\mathcal{D}}$ has a single state (i.e., $|\mathcal{S}_{\mathcal{D}}| = 1$), $\tilde{Q}_{\mathcal{D}} = 0$, and $\tilde{\pi}_{\mathcal{D}} = 1$.*

For HMMs, Corollary 4.16 applies at the coarsest level when the HLM has one macrostate.

Now, we are in a position to state and prove a lemma, which is essential in characterizing the convergence of the three smoothers.

LEMMA 4.17. *If the smoother $S \in \{POWER, JOR, SOR\}$ satisfies $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0,1)$, then the iteration matrix $T_{\mathcal{D}}$ associated with the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$ is nonnegative, irreducible, primitive, and has a spectral radius and an eigenvalue of one; furthermore, $T_{\mathcal{D}} = W_{\mathcal{D}} B_{\mathcal{D}} W_{\mathcal{D}}^{-1}$, where $B_{\mathcal{D}}$ is a probability matrix and $W_{\mathcal{D}}$ is a nonnegative, diagonal matrix having the right eigenvector of $T_{\mathcal{D}}$ corresponding to one along its diagonal, implying $\lim_{m\to\infty} T_{\mathcal{D}}^m = (W_{\mathcal{D}} e)\tilde{\pi}_{\mathcal{D}}/(\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} e) > 0$ and is of rank 1. When $POWER$ is the smoother, $W_{\mathcal{D}} = I_{\mathcal{D}}$ and $T_{\mathcal{D}}$ is a probability matrix, implying $\lim_{m\to\infty} T_{\mathcal{D}}^m = e\tilde{\pi}_{\mathcal{D}} > 0$.*

Proof. First, we show the nonnegativeness of the iteration matrices under the given assumptions using Definition 4.8, Proposition 4.9, and Definition 4.10. The nonnegativeness of $T_{\mathcal{D}}^{POWER}$ for $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ follows from $I_{\mathcal{D}} + D_{\mathcal{D}}/\alpha \geq 0$, $-(L_{\mathcal{D}} + U_{\mathcal{D}})/\alpha \geq 0$, and $T_{\mathcal{D}}^{POWER} = (I_{\mathcal{D}} + D_{\mathcal{D}}/\alpha) - (L_{\mathcal{D}} + U_{\mathcal{D}})/\alpha$. Similarly, the nonnegativeness of $T_{\mathcal{D}}^{JOR}$ for $\omega \in (0,1)$ follows from $(1-\omega) I_{\mathcal{D}} \geq 0$, $L_{\mathcal{D}} + U_{\mathcal{D}} \leq 0$, and $D_{\mathcal{D}}^{-1} \leq 0$. On the other hand, the nonnegativeness of $T_{\mathcal{D}}^{SOR}$ for $\omega \in (0,1)$ follows from $(1-\omega) D_{\mathcal{D}}/\omega + L_{\mathcal{D}} \leq 0$ and $(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1} \leq 0$ since $-(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})$ is a nonsingular M-matrix [21, p. 626].

The irreducibilities of $T_{\mathcal{D}}^{POWER}$ and $T_{\mathcal{D}}^{JOR}$ follow in a straightforward manner from the irreducibility of $\tilde{Q}_{\mathcal{D}}$ (see Proposition 4.2), since they have the same off-diagonal nonzero structure as $\tilde{Q}_{\mathcal{D}}$. On the other hand, $T_{\mathcal{D}}^{SOR}$ is the product of $(1-\omega) D_{\mathcal{D}}/\omega + L_{\mathcal{D}}$ and $(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})^{-1}$, both of which are negative with negative diagonals for $\omega \in (0,1)$. Note that this implies a positive diagonal in $T_{\mathcal{D}}^{SOR}$. Now, the nonzero structure of the former $T_{\mathcal{D}}^{SOR}$ factor below the diagonal is the same as that of $L_{\mathcal{D}}$ and the nonzero structure of the latter $T_{\mathcal{D}}^{SOR}$ factor is given by the transitive closure of $\mathcal{G}(D_{\mathcal{D}}/\omega - U_{\mathcal{D}})$ [17, p. 72]. Hence, the product of the two factors will yield at least as many nonzeros as there are in $\tilde{Q}_{\mathcal{D}}$, implying the irreducibility of $T_{\mathcal{D}}^{SOR}$. Furthermore, $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0,1)$ ensure positive diagonals in $T_{\mathcal{D}}^{POWER}$, $T_{\mathcal{D}}^{JOR}$, and $T_{\mathcal{D}}^{SOR}$, and hence, iteration matrices that are primitive (or aperiodic). Using the fact in Lemma 4.11 that $\tilde{\pi}_{\mathcal{D}} > 0$ is the left eigenvector corresponding to the iteration matrices $T_{\mathcal{D}}^{POWER}$, $T_{\mathcal{D}}^{JOR}$, and $T_{\mathcal{D}}^{SOR}$, we conclude that each of the nonnegative, irreducible, and primitive iteration matrices has a spectral radius and an eigenvalue of one from the Perron-Frobenius Theorem [21, p. 673]. Powers of such iteration matrices converge to constant, positive matrices [21, p. 674].

Now, recall that if $T \geq 0$ with $\rho(T) > 0$ and $v > 0$ such that $Tv = \rho(T)v$, then $T/\rho(T)$ is similar to a probability matrix [4, p. 49]. That is, for a nonnegative and irreducible matrix T with positive, right eigenvector v, there exists a diagonal matrix W with $w(i,i) = v(i)$ such that $B = W^{-1}TW$ is a probability matrix. Since $T_{\mathcal{D}}$ is nonnegative, is irreducible, and $\rho(T_{\mathcal{D}}) = 1$, after setting $T = T_{\mathcal{D}}$ we must have $T_{\mathcal{D}} = W_{\mathcal{D}} B_{\mathcal{D}} W_{\mathcal{D}}^{-1}$, where $B_{\mathcal{D}}$ is a probability matrix and $W_{\mathcal{D}}$ is a nonnegative, diagonal matrix having the positive, right eigenvector of $T_{\mathcal{D}}$ along its diagonal. Hence, $T_{\mathcal{D}}^m = W_{\mathcal{D}} B_{\mathcal{D}}^m W_{\mathcal{D}}^{-1}$ for $m \geq 0$. Now, since $\tilde{\pi}_{\mathcal{D}} > 0$ (see Definition 4.6) is a fixed point of $T_{\mathcal{D}}$ from Lemma 4.11, it must be unique because $T_{\mathcal{D}}$ is nonnegative and irreducible [21, p. 673]. Consequently, we have $\tilde{\pi}_{\mathcal{D}} T_{\mathcal{D}}^m = \tilde{\pi}_{\mathcal{D}}$, or $\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} B_{\mathcal{D}}^m W_{\mathcal{D}}^{-1} = \tilde{\pi}_{\mathcal{D}}$. But this can be written as $\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} B_{\mathcal{D}}^m = \tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}}$. Taking the limit, we obtain $\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} \lim_{m\to\infty} B_{\mathcal{D}}^m = \tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}}$, which implies $\lim_{m\to\infty} B_{\mathcal{D}}^m = e(\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}})/(\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} e)$, since $W_{\mathcal{D}}$ is a nonnegative, diagonal matrix and $\lim_{m\to\infty} B_{\mathcal{D}}^m$ must have the same probability vector in each of its rows. Then, $\lim_{m\to\infty} T_{\mathcal{D}}^m = W_{\mathcal{D}}(\lim_{m\to\infty} B_{\mathcal{D}}^m) W_{\mathcal{D}}^{-1} = (W_{\mathcal{D}} e)\tilde{\pi}_{\mathcal{D}}/(\tilde{\pi}_{\mathcal{D}} W_{\mathcal{D}} e)$. Since the numerator is an outer product and the denominator is a scalar, the limiting matrix is of rank 1. For the *POWER* smoother, we remark that $T_{\mathcal{D}}^{POWER} e = (I_{\mathcal{D}} + \tilde{Q}_{\mathcal{D}}/\alpha)e = e + (\tilde{Q}_{\mathcal{D}} e)/\alpha = e$ since $\tilde{Q}_{\mathcal{D}} e = 0$. Hence, $T_{\mathcal{D}}^{POWER}$ is a probability matrix, $W_{\mathcal{D}} = I_{\mathcal{D}}$, and powers of $T_{\mathcal{D}}^{POWER}$ converge to the probability matrix $e\tilde{\pi}_{\mathcal{D}}$ which has the steady state vector $\tilde{\pi}_{\mathcal{D}}$ of $\tilde{Q}_{\mathcal{D}}$ in its rows as shown in [25, p. 16]. □

Using Lemma 4.17, the next proposition expresses the pre- and post-smoothings at level $\mathcal{D}$ concisely.

PROPOSITION 4.18. *Given the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$, after ν_1 iterations of pre-smoothings at level $\mathcal{D}$ with the smoother S, the smoothed vector becomes*

$$x_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1} > 0;$$

after ν_2 iterations of post-smoothings at level $\mathcal{D}$ with S, the smoothed vector becomes

$$y_{\mathcal{D}}^{'} = y_{\mathcal{D}} T_{\mathcal{D}}^{\nu_2} > 0.$$

The next definition follows from Theorem 4.4 in [24, pp. 45–46] and is introduced to aid the characterization of the nonasymptotic convergence behavior of smoothings.

DEFINITION 4.19. *Let $S_{\mathcal{D}} \in \mathbb{R}^{|\mathcal{S}_{\mathcal{D}}| \times |\mathcal{S}_{\mathcal{D}}|}$ be nonsingular (i.e., $S_{\mathcal{D}}^{-1}$ exists). Then the function defined as*

$$\|w\|_{S_{\mathcal{D}}} = \|w S_{\mathcal{D}}\|_1 \quad \text{for } w \in \mathbb{R}^{1 \times |\mathcal{S}_{\mathcal{D}}|}$$

is a vector norm[2].

The next theorem characterizes the nonasymptotic convergence behavior of the smoothings through a lemma for positive probability matrices based on the discussion in [2, pp. 270–271] and proved in the appendix, and two results on nonnegative, irreducible matrices similar to positive matrices [5, p. 371 and p.375]. We remark that a similar theorem may be stated for the initial approximation $y_{\mathcal{D}}$.

THEOREM 4.20. *Given the initial approximation $x_{\mathcal{D}}^{(0)} = x_{\mathcal{D}} > 0$ for the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$ and the smoother $S \in \{POWER, JOR, SOR\}$ with iteration matrix $T_{\mathcal{D}}$ such that $x_{\mathcal{D}}^T \notin Range(I_{\mathcal{D}} - T_{\mathcal{D}}^T)$, if $T_{\mathcal{D}}^{\nu_1}$ is nonnegative (i.e., $T_{\mathcal{D}}^{\nu_1} \geq 0$), irreducible, and satisfies any of the three conditions:*

(i) $T_{\mathcal{D}}^{\nu_1}$ is positive (i.e., $T_{\mathcal{D}}^{\nu_1} > 0$),

(ii) $T_{\mathcal{D}}^{\nu_1}$ has a positive row $i_{\mathcal{D}}$ (i.e., $e_{i_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} > 0$) for some $i_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$ or a positive column $j_{\mathcal{D}}$ (i.e., $T_{\mathcal{D}}^{\nu_1} e_{j_{\mathcal{D}}} > 0$) for some $j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$,

(iii) $T_{\mathcal{D}}^{\nu_1}$ has a zero element in position $(i_{\mathcal{D}}, j_{\mathcal{D}})$ (i.e., $e_{i_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{j_{\mathcal{D}}} = 0$) for some $i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$,

(a) all other elements in row $i_{\mathcal{D}}$ are positive (i.e., $e_{i_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{k_{\mathcal{D}}} > 0$ for $k_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$, $k_{\mathcal{D}} \neq i_{\mathcal{D}}$) and $e_{i_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{i_{\mathcal{D}}} > e_{j_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{j_{\mathcal{D}}}$, or

(b) all other elements in column $j_{\mathcal{D}}$ are positive (i.e., $e_{k_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{j_{\mathcal{D}}} > 0$ for $k_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}$, $k_{\mathcal{D}} \neq j_{\mathcal{D}}$) and $e_{i_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{i_{\mathcal{D}}} < e_{j_{\mathcal{D}}}^T T_{\mathcal{D}}^{\nu_1} e_{j_{\mathcal{D}}}$,

then

$$\|a_{\mathcal{D}} x_{\mathcal{D}}^{'} - \tilde{\pi}_{\mathcal{D}}\|_{S_{\mathcal{D}}} \leq \left(1 - \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} g_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}})\right) \|a_{\mathcal{D}} x_{\mathcal{D}} - \tilde{\pi}_{\mathcal{D}}\|_{S_{\mathcal{D}}},$$

[2]This norm should not be confused with the elliptical norm [21, p. 288] defined as $\|w\|_{S_{\mathcal{D}}} = \|w S_{\mathcal{D}}\|_2$

where $x_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1}$, $G_{\mathcal{D}}$ *is a positive probability matrix (i.e.,* $G_{\mathcal{D}} > 0$, $G_{\mathcal{D}} e = e$*) defined as* $G_{\mathcal{D}} = S_{\mathcal{D}}^{-1} T_{\mathcal{D}}^{\nu_1} S_{\mathcal{D}}$ *for some* $S_{\mathcal{D}} \geq 0$ *such that* $0 < \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} g_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}}) \leq 1/|\mathcal{S}_{\mathcal{D}}|$, $\tilde{\pi}_{\mathcal{D}}$ *is the steady state vector of* $\tilde{Q}_{\mathcal{D}}$*, and* $a_{\mathcal{D}} = (\tilde{\pi}_{\mathcal{D}} S_{\mathcal{D}} e)/(x_{\mathcal{D}} S_{\mathcal{D}} e)$.

Proof. From Corollary 3 and Theorem 4 in [5], if $T_{\mathcal{D}}^{\nu_1}$ is nonnegative, is irreducible, and satisfies either of the conditions (ii) or (iii), then it is similar to a positive matrix, that is, $X_{\mathcal{D}}^{-1} T_{\mathcal{D}}^{\nu_1} X_{\mathcal{D}} = H_{\mathcal{D}} > 0$ for some $(|\mathcal{S}_{\mathcal{D}}| \times |\mathcal{S}_{\mathcal{D}}|)$ nonnegative matrix $X_{\mathcal{D}}$. Condition (i) is a special case for which $X_{\mathcal{D}} = I_{\mathcal{D}}$. Since these imply $\sigma(H_{\mathcal{D}}) = \sigma(T_{\mathcal{D}}^{\nu_1})$ and we have $\rho(T_{\mathcal{D}}^{\nu_1}) = 1$ from Lemma 4.17, $H_{\mathcal{D}} > 0$ must be similar to a positive probability matrix $G_{\mathcal{D}}$ as in $Y_{\mathcal{D}}^{-1} H_{\mathcal{D}} Y_{\mathcal{D}} = G_{\mathcal{D}} > 0$, where $Y_{\mathcal{D}}$ is a nonnegative, diagonal matrix having the positive, right eigenvector of $H_{\mathcal{D}}$ along its diagonal. Now, let $S_{\mathcal{D}} = X_{\mathcal{D}} Y_{\mathcal{D}}$ to obtain $T_{\mathcal{D}}^{\nu_1} = S_{\mathcal{D}} G_{\mathcal{D}} S_{\mathcal{D}}^{-1}$, where $S_{\mathcal{D}} \geq 0$, $G_{\mathcal{D}} > 0$, and $Ge = e$.

For a sequence of converging approximations, one needs to ensure for the initial approximation that $x_{\mathcal{D}}^T \notin Range(I_{\mathcal{D}} - T_{\mathcal{D}}^T)$ [3, pp. 26–28]; otherwise, there will be no improvement. Furthermore, since $\tilde{\pi}_{\mathcal{D}}$ is the unique, positive fixed point of $T_{\mathcal{D}}^{\nu_1}$ such that $\tilde{\pi}_{\mathcal{D}} e = 1$ from Lemma 4.11, the unique, positive fixed point of $G_{\mathcal{D}}$ with unit 1-norm must be $\psi_{\mathcal{D}} = (\tilde{\pi}_{\mathcal{D}} S_{\mathcal{D}})/(\tilde{\pi}_{\mathcal{D}} S_{\mathcal{D}} e)$. Now, rewrite $x_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1}$ using $T_{\mathcal{D}}^{\nu_1} = S_{\mathcal{D}} G_{\mathcal{D}} S_{\mathcal{D}}^{-1}$ to obtain $x_{\mathcal{D}}^{'} S_{\mathcal{D}} = x_{\mathcal{D}} S_{\mathcal{D}} (G_{\mathcal{D}})$. Since $x_{\mathcal{D}} > 0$, $S_{\mathcal{D}} \geq 0$, and $S_{\mathcal{D}}$ has full rank, we have $x_{\mathcal{D}}^{'} > 0$. Furthermore, note that $x_{\mathcal{D}}^{'} S_{\mathcal{D}} e = x_{\mathcal{D}} S_{\mathcal{D}} (G_{\mathcal{D}} e) = x_{\mathcal{D}} S_{\mathcal{D}} e$. Letting $\overline{x}_{\mathcal{D}}^{'} = (x_{\mathcal{D}}^{'} S_{\mathcal{D}})/(x_{\mathcal{D}} S_{\mathcal{D}} e)$ and $\overline{x}_{\mathcal{D}} = (x_{\mathcal{D}} S_{\mathcal{D}})/(x_{\mathcal{D}} S_{\mathcal{D}} e)$, we have from Lemma A.1 in the Appendix

$$\|\overline{x}_{\mathcal{D}}^{'} - \psi_{\mathcal{D}}\|_1 \leq \left(1 - \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} g_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}})\right) \|\overline{x}_{\mathcal{D}} - \psi_{\mathcal{D}}\|_1.$$

The result follows by taking each of $(\overline{x}_{\mathcal{D}}^{'} - \psi_{\mathcal{D}})$ and $(\overline{x}_{\mathcal{D}} - \psi_{\mathcal{D}})$ into $S_{\mathcal{D}}$ parantheses, multiplying bothsides of the inequality by $\tilde{\pi}_{\mathcal{D}} S_{\mathcal{D}} e$, letting $a_{\mathcal{D}} = (\tilde{\pi}_{\mathcal{D}} S_{\mathcal{D}} e)/(x_{\mathcal{D}} S_{\mathcal{D}} e)$, and using Definition 4.19. □

Theorem 4.20 indicates that the *normalized* solution vector, $a_{\mathcal{D}} x_{\mathcal{D}}$, improves with ν_1 pre-smoothings if $T_{\mathcal{D}}^{\nu_1}$ is positive or has a(n) (almost) positive row or column. Now, observe that the ordering of grids suggested by $O \in \{FIXED, CYCLIC, DYNAMIC\}$ has no effect on the assumptions of Theorem 4.20 and the parameters MIN_IN_PRE, MAX_IN_PRE, MIN_IN_POST, MAX_IN_POST, MIN_OUT_PRE, MAX_OUT_PRE, MIN_OUT_POST, MAX_OUT_POST, ρ, RES_COUNT in Algorithms 1 and 2 determine ν_1 and ν_2 at each level. Note from Lemma 4.17 that as ν_1 increases, $T_{\mathcal{D}}^{\nu_1}$ converges to a positive, rank 1 matrix. Hence, there is a value of $\nu_1 > 0$ for which the assumptions of Theorem 4.20 hold. We remark that $\tilde{Q}_{\mathcal{D}}$ is almost always sparse and the iteration matrices associated with the $POWER$ and JOR smoothers have the same off-diagonal nonzero structure as that of $\tilde{Q}_{\mathcal{D}}$. Hence, compared to $POWER$ and JOR, the SOR smoother has a higher chance of satisfying the conditions of Theorem 4.20 for a smaller value of ν_1, since its iteration matrix is likely to have a larger number of nonzeros as suggested in the proof of Lemma 4.17. Similar arguments are valid for post-smoothings. In summary, the smoothings can always be enforced to yield improved positive approximations at each level.

4.3. Convergence of the ML solver. Using the results in the previous subsections, we show that under certain conditions the devised class of ML methods provide converging iterations for different choices of the cycle parameter $C \in \{V, W, F\}$.

First, we define the ML iteration matrix at level $\mathcal{D}$ in Algorithms 1 and 2 using Propositions 3.5, 3.7, 4.15, and 4.18. Note that when there are only two levels, the W- and F-cycles are not defined, and the V-cycle yields an IAD solver. In order not to complicate the notation further, we refrain from introducing an index for the cycle number to the matrices and vectors at this point.

DEFINITION 4.21. *Let* $T_{\mathcal{D}}^{ML}$ *denote the ML iteration matrix that operates at level* $\mathcal{D}$ *on* $x_{\mathcal{D}} > 0$ *to give* $y_D > 0$ *at a particular cycle using the smoother* $S \in \{POWER, JOR, SOR\}$ *associated with the irreducible CTMC* $\tilde{Q}_{\mathcal{D}}$*, where* $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ *and* $\omega \in (0, 1)$*, and similarly let* $T_{\mathcal{C}}^{ML}$ *and* $T_{\mathcal{B}}^{ML}$ *denote the ML iteration matrices that operate at the next two coarser levels* $\mathcal{C}$ *and* $\mathcal{B}$*, respectively. Then*

$$y_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{ML},$$

where

$$T_{\mathcal{D}}^{ML} = \begin{cases} T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} T_{\mathcal{C}}^{ML} P_{x_{\mathcal{D}}^{'}} T_{\mathcal{D}}^{\nu_2} & \text{if} \quad C = V \\ T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} (T_{\mathcal{C}}^{ML})^2 P_{x_{\mathcal{D}}^{'}} T_{\mathcal{D}}^{\nu_2} & \text{if} \quad C = W \\ T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML^{'}} P_{x_{\mathcal{D}}^{'}} T_{\mathcal{D}}^{\nu_2} & \text{if} \quad C = F \end{cases},$$

$$T_{\mathcal{C}}^{ML'} = T_{\mathcal{C}}^{\nu_1} R_{\mathcal{C}} T_{\mathcal{B}}^{ML'} P_{x_{\mathcal{C}}'} T_{\mathcal{C}}^{\nu_2}, \quad x_{\mathcal{D}}' = x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1},$$

and when $\tilde{Q}_{\mathcal{C}}$ is the coarsest grid and solved exactly, $T_{\mathcal{C}}^{ML} = T_{\mathcal{C}}^{ML'} = (ey_{\mathcal{C}}')/(x_{\mathcal{C}}e) > 0$, where $y_{\mathcal{C}}' = \tilde{\pi}_{\mathcal{C}}$.

COROLLARY 4.22. *When POWER is the smoother and $x_{\mathcal{D}} > 0$ satisfies $x_{\mathcal{D}}e = 1$, the ML iteration matrix $T_{\mathcal{D}}^{ML}$ for $C \in \{V, W, F\}$ is a positive probability matrix (i.e., $T_{\mathcal{D}}^{ML} > 0$, $T_{\mathcal{D}}^{ML}e = e$) and therefore has a spectral radius of one (i.e., $\rho(T_{\mathcal{D}}^{ML}) = 1$).*

Proof. For the *POWER* smoother, at the coarsest level $\mathcal{C}$ we have $T_{\mathcal{C}}^{ML} = T_{\mathcal{C}}^{ML'} = e\tilde{\pi}_{\mathcal{C}}$ from Definition 4.21 when $x_{\mathcal{D}}e = 1$, implying a positive probability matrix, which has a spectral radius and an eigenvalue value of one. This forms the base case. Now, let us assume that the result is true for all levels from the coarsest up to an arbitrary level $\mathcal{C}$; this is the inductive hypothesis. We show that the result must be true for the next finer level $\mathcal{D}$. Noting that $R_{\mathcal{D}}e = e$ from Proposition 3.5, $(T_{\mathcal{C}}^{ML})e = e$ from the inductive hypothesis, $P_{x_{\mathcal{D}}'}e = e$ from Proposition 3.7, and $T_{\mathcal{D}}e = e$ from Lemma 4.17, we have $T_{\mathcal{D}}^{ML}e = T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} T_{\mathcal{C}}^{ML} P_{x_{\mathcal{D}}'} T_{\mathcal{D}}^{\nu_2} e = T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} T_{\mathcal{C}}^{ML} (P_{x_{\mathcal{D}}'} e) = T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}} (T_{\mathcal{C}}^{ML} e) = T_{\mathcal{D}}^{\nu_1} (R_{\mathcal{D}} e) = T_{\mathcal{D}}^{\nu_1} e = e$ for the V-cycle. The result follows similarly for W- and F-cycles. ∎

The interpretation of $T_{\mathcal{D}}^{ML}$ for V- and W-cycles is as follows. If the recursive call(s) to level $\mathcal{C}$ are turned off, then only $(\nu_1 + \nu_2)$ iterations are performed on $x_{\mathcal{D}}$ with the smoother S. Otherwise, the smoothed solution vector is restricted to level $\mathcal{C}$ (i.e., $x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1}$ is the smoothed solution vector and $x_{\mathcal{D}} T_{\mathcal{D}}^{\nu_1} R_{\mathcal{D}}$ is the restricted solution vector), the restricted solution vector is improved respectively one or two times with the iteration matrix $T_{\mathcal{C}}^{ML}$, the improved solution vector is projected back to level $\mathcal{D}$, and smoothed. The interpretation of $T_{\mathcal{D}}^{ML}$ for an F-cycle is similar to that for V- and W-cycles with the difference that the restricted solution vector is improved with the iteration matrix $T_{\mathcal{D}}^{ML}$ once followed by the iteration matrix of the V-cycle. This is exactly what is meant with a W-cycle followed by a V-cycle at each level.

The next lemma follows from Lemma 4.3, Lemma 4.17, and Definition 4.21.

LEMMA 4.23. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC, $x_{\mathcal{D}} > 0$, and the smoother $S \in \{POWER, JOR, SOR\}$ satisfies $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0, 1)$, then the ML iteration matrix $T_{\mathcal{D}}^{ML}$ for $C \in \{V, W, F\}$ is positive (i.e., $T_{\mathcal{D}}^{ML} > 0$).*

Proof. The proof is by induction. At the coarsest level $\mathcal{C}$, we have $T_{\mathcal{C}}^{ML} = T_{\mathcal{C}}^{ML'} > 0$ from Definition 4.21. This is the base case, and implies $(T_{\mathcal{C}}^{ML})^2 = T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML'} > 0$. Now, let us assume that the statement is true for all levels from the coarsest up to an arbitrary level $\mathcal{C}$. This is the inductive hypothesis. Now, we show that the statement must be true for the next finer level $\mathcal{D}$. Since $P_{x_{\mathcal{D}}'} \geq 0$ and each column of $P_{x_{\mathcal{D}}'}$ has one nonzero element from Proposition 3.7, the $(|\mathcal{S}_{\mathcal{C}}| \times |\mathcal{S}_{\mathcal{D}}|)$ matrices $T_{\mathcal{C}}^{ML} P_{x_{\mathcal{D}}'}$, $(T_{\mathcal{C}}^{ML})^2 P_{x_{\mathcal{D}}'}$, and $T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML'} P_{x_{\mathcal{D}}'}$ are positive. Furthermore, since $R_{\mathcal{D}} \geq 0$ and each row of $R_{\mathcal{D}}$ has one nonzero element from Proposition 3.5, the $(|\mathcal{S}_{\mathcal{D}}| \times |\mathcal{S}_{\mathcal{D}}|)$ matrices $R_{\mathcal{D}} T_{\mathcal{C}}^{ML} P_{x_{\mathcal{D}}'}$, $R_{\mathcal{D}} (T_{\mathcal{C}}^{ML})^2 P_{x_{\mathcal{D}}'}$, and $R_{\mathcal{D}} T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML'} P_{x_{\mathcal{D}}'}$ are also positive. Then the result follows from Lemma 4.17 by the fact that the iteration matrix associated with the smoother is nonnegative and irreducible, implying at least one nonzero in each row and column of $T_{\mathcal{D}}$ which pre- and post-multiplies the positive matrices $R_{\mathcal{D}} T_{\mathcal{C}}^{ML} P_{x_{\mathcal{D}}'}$, $R_{\mathcal{D}} (T_{\mathcal{C}}^{ML})^2 P_{x_{\mathcal{D}}'}$, and $R_{\mathcal{D}} T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML'} P_{x_{\mathcal{D}}'}$. ∎

The next result follows from Lemma 4.23 in that the positivity of $T_{\mathcal{D}}^{ML}$ implies its irreducibility and a positive diagonal, and hence its primitivity [4, p. 47].

COROLLARY 4.24. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC, $x_{\mathcal{D}} > 0$, and the smoother $S \in \{POWER, JOR, SOR\}$ satisfies $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0, 1)$, then the ML iteration matrix $T_{\mathcal{D}}^{ML}$ for $C \in \{V, W, F\}$ is irreducible and primitive.*

The next lemma shows that the steady state vector, $\pi_{\mathcal{D}}$, of the exactly aggregated grid, $Q_{\mathcal{D}}$, is the unique, positive, unit 1-norm fixed point of the ML iteration matrix, $T_{\mathcal{D}}^{ML}$, at level $\mathcal{D}$ upon convergence.

LEMMA 4.25. *If $\tilde{Q}_{\mathcal{D}}$ is an irreducible CTMC and equal to $Q_{\mathcal{D}}$, $x_{\mathcal{D}} = \pi_{\mathcal{D}}$, and the smoother $S \in \{POWER, JOR, SOR\}$ satisfies $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0, 1)$, then the ML iteration matrix $T_{\mathcal{D}}^{ML}$ for $C \in \{V, W, F\}$ has the unique, positive fixed point $\pi_{\mathcal{D}}$ (i.e., $\pi_{\mathcal{D}} T_{\mathcal{D}}^{ML} = \pi_{\mathcal{D}}$) such that $\pi_{\mathcal{D}}e = 1$; furthermore, $\rho(T_{\mathcal{D}}^{ML}) = 1$ and $y_{\mathcal{D}}' = \pi_{\mathcal{D}}$.*

Proof. The proof is by induction. At the coarsest level $\mathcal{C}$, we have $\tilde{Q}_{\mathcal{C}} = Q_{\mathcal{C}}$ and $x_{\mathcal{C}} = \pi_{\mathcal{C}} > 0$, implying $T_{\mathcal{C}}^{ML} = T_{\mathcal{C}}^{ML'} = e\pi_{\mathcal{C}} > 0$ from Definition 4.21. This positive matrix is stochastic and has the unique, positive fixed point $\pi_{\mathcal{C}}$ such that $\pi_{\mathcal{C}}e = 1$. Furthermore, it has a spectral radius of one and $y_{\mathcal{C}}' = x_{\mathcal{C}} T_{\mathcal{C}}^{ML} = \pi_{\mathcal{C}}(e\pi_{\mathcal{C}}) = (\pi_{\mathcal{C}}e)\pi_{\mathcal{C}} = \pi_{\mathcal{C}}$. This is the base case, and yields $(T_{\mathcal{C}}^{ML})^2 = T_{\mathcal{C}}^{ML} T_{\mathcal{C}}^{ML'} =$

$(e\pi_{\mathcal{C}})(e\pi_{\mathcal{C}}) = e(\pi_{\mathcal{C}}e)\pi_{\mathcal{C}} = e\pi_{\mathcal{C}} > 0$. Now, let us assume that the statement is true for all levels from the coarsest up to an arbitrary level $\mathcal{C}$. This is the inductive hypothesis. Now, we show that the statement must be true for the next finer level $\mathcal{D}$.

Since $x_{\mathcal{D}} = \pi_{\mathcal{D}} > 0$ is the fixed point of $T_{\mathcal{D}}$ from Lemma 4.11, $\pi_{\mathcal{D}}R_{\mathcal{D}} = \pi_{\mathcal{C}}$ from Definition 3.4, $\pi_{\mathcal{C}}T_{\mathcal{C}}^{ML} = \pi_{\mathcal{C}}$ by the inductive hypothesis, and $\pi_{\mathcal{C}}P_{\pi_{\mathcal{D}}} = \pi_{\mathcal{D}}$ from Definition 3.6, the result follows from Definition 4.21 for the V-cycle as $y_{\mathcal{D}}^{'} = \pi_{\mathcal{D}}T_{ML}^{D} = (\pi_{\mathcal{D}}T_{\mathcal{D}}^{\nu_1})R_{\mathcal{D}}T_{\mathcal{C}}^{ML}P_{\pi_{\mathcal{D}}}T_{\mathcal{D}}^{\nu_2} = (\pi_{\mathcal{D}}R_{\mathcal{D}})T_{\mathcal{C}}^{ML}P_{\pi_{\mathcal{D}}}T_{\mathcal{D}}^{\nu_2} = (\pi_{\mathcal{C}}T_{\mathcal{C}}^{ML})P_{\pi_{\mathcal{D}}}T_{\mathcal{D}}^{\nu_2} = (\pi_{\mathcal{C}}P_{\pi_{\mathcal{D}}})T_{\mathcal{D}}^{\nu_2} = \pi_{\mathcal{D}}T_{\mathcal{D}}^{\nu_2} = \pi_{\mathcal{D}}$. The result follows similarly for W- and F-cycles after interchanging $T_{\mathcal{C}}^{ML}$ respectively with $(T_{\mathcal{C}}^{ML})^2$ and $T_{\mathcal{C}}^{ML}T_{\mathcal{C}}^{ML'}$. The uniqueness and positiveness of the fixed point of $T_{\mathcal{D}}^{ML}$ follows from Lemma 4.23 by the fact that $T_{\mathcal{D}}^{ML}$ is positive [21, p. 666]. Clearly the spectral radius of $T_{\mathcal{D}}^{ML}$ is one. □

The next theorem characterizes the nonasymptotic convergence behavior of the ML solver with the initial approximation $x_{\mathcal{D}}$ by defining a unique, positive, unit 1-norm fixed point for the particular cycle.

THEOREM 4.26. *If $T_{\mathcal{D}}^{ML}$ is the ML iteration matrix that operates at level $\mathcal{D}$ on $x_{\mathcal{D}} > 0$, such that $x_{\mathcal{D}}^{T} \notin Range(I_{\mathcal{D}} - T_{\mathcal{D}}^{T})$, to give $y_{\mathcal{D}}^{'} > 0$ at a particular cycle using the smoother $S \in \{POWER, JOR, SOR\}$ associated with the irreducible CTMC $\tilde{Q}_{\mathcal{D}}$, where $\alpha \in (\max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|, \infty)$ and $\omega \in (0, 1)$, then $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ has a spectral radius of one and a vector $\phi_{\mathcal{D}}$ as its unique, positive fixed point (i.e., $\phi_{\mathcal{D}}(T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})) = \phi_{\mathcal{D}}$) such that $\phi_{\mathcal{D}}e = 1$. Furthermore, $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}) = Z_{\mathcal{D}}H_{\mathcal{D}}Z_{\mathcal{D}}^{-1}$, where $H_{\mathcal{D}}$ is a positive probability matrix (i.e., $H_{\mathcal{D}} > 0$, $H_{\mathcal{D}}e = e$) and $Z_{\mathcal{D}}$ is a nonnegative, diagonal matrix having the positive, right eigenvector of $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ along its diagonal. The unique, positive fixed point of $H_{\mathcal{D}}$ is given by $\psi_{\mathcal{D}} = (\phi_{\mathcal{D}}Z_{\mathcal{D}})/(\phi_{\mathcal{D}}Z_{\mathcal{D}}e)$ (i.e., $\psi_{\mathcal{D}}H_{\mathcal{D}} = \psi_{\mathcal{D}}$) such that $\psi_{\mathcal{D}}e = 1$. Finally,*

$$\|(b_{\mathcal{D}}/\rho(T_{\mathcal{D}}^{ML}))y_{\mathcal{D}}^{'} - \phi_{\mathcal{D}}\|_{Z_{\mathcal{D}}} \leq \left(1 - \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} h_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}})\right) \|b_{\mathcal{D}}x_{\mathcal{D}} - \phi_{\mathcal{D}}\|_{Z_{\mathcal{D}}},$$

where $b_{\mathcal{D}} = (\phi_{\mathcal{D}}Z_{\mathcal{D}}e)/(x_{\mathcal{D}}Z_{\mathcal{D}}e)$ and $0 < \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} h_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}}) \leq 1/|\mathcal{S}_{\mathcal{D}}|$. At the coarsest level, $\|(b_{\mathcal{D}}/\rho(T_{\mathcal{D}}^{ML}))y_{\mathcal{D}}^{'} - \phi_{\mathcal{D}}\|_{Z_{\mathcal{D}}} = 0$ if the system is solved directly. When $POWER$ is the smoother, $Z_{\mathcal{D}} = I_{\mathcal{D}}$, $H_{\mathcal{D}} = T_{\mathcal{D}}^{ML}$, $\rho(T_{\mathcal{D}}^{ML}) = 1$, $\psi_{\mathcal{D}} = \phi_{\mathcal{D}}$, and $b_{\mathcal{D}} = 1$.

Proof. Recall from Lemma 4.23 that $T_{\mathcal{D}}^{ML} > 0$. Since $\rho(T_{\mathcal{D}}^{ML}) > 0$ for $T_{\mathcal{D}}^{ML} \neq 0$, the matrix $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ is also positive, it satisfies $\sigma(T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})) = \{\lambda/\rho(T_{\mathcal{D}}^{ML}) \mid \lambda \in \sigma(T_{\mathcal{D}}^{ML})\}$, and therefore has a spectral radius of one. The uniqueness and positiveness of the fixed point $\phi_{\mathcal{D}}$ follows from Corollary 4.24. The row vector $\phi_{\mathcal{D}} > 0$ is assumed to be normalized so as to have unit 1-norm (i.e., $\phi_{\mathcal{D}}e = 1$).

To prove the second part, recall Corollary 4.24 and the result in [4, p. 49] which is also used in the proof of Lemma 4.17. These imply that $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ must have a positive, right eigenvector $\zeta_{\mathcal{D}}$ for which

$$T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}) = Z_{\mathcal{D}}H_{\mathcal{D}}Z_{\mathcal{D}}^{-1},$$

where $Z_{\mathcal{D}} = \mathrm{diag}(\zeta_{\mathcal{D}})$, $H_{\mathcal{D}} > 0$, and $H_{\mathcal{D}}e = e$. In other words,

$$H_{\mathcal{D}} = Z_{\mathcal{D}}^{-1}(T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}))Z_{\mathcal{D}}$$

is a probability matrix similar to $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ and its positivity follows from $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}) > 0$ and $\zeta_{\mathcal{D}} > 0$. Note that it does not matter whether $\zeta_{\mathcal{D}}$ is normalized or not, since $H_{\mathcal{D}}$ is defined in terms of $Z_{\mathcal{D}}$ and $Z_{\mathcal{D}}^{-1}$. The uniqueness and positiveness of the fixed point $\psi_{\mathcal{D}}$ follows from $H_{\mathcal{D}} > 0$. The row vector $\phi_{\mathcal{D}} > 0$ is assumed to be normalized so as to have unit 1-norm (i.e., $\phi_{\mathcal{D}}e = 1$) and it is given by $\psi_{\mathcal{D}} = \phi_{\mathcal{D}}Z_{\mathcal{D}}$ since $H_{\mathcal{D}}$ and $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML})$ are related by a similarity transformation, where the transformation matrix is $Z_{\mathcal{D}}$.

To prove the last part, rewrite

$$y_{\mathcal{D}}^{'} = \rho(T_{\mathcal{D}}^{ML})x_{\mathcal{D}}(T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}))$$

using $T_{\mathcal{D}}^{ML}/\rho(T_{\mathcal{D}}^{ML}) = Z_{\mathcal{D}}H_{\mathcal{D}}Z_{\mathcal{D}}^{-1} > 0$ as

$$(y_{\mathcal{D}}^{'}Z_{\mathcal{D}})/(\rho(T_{\mathcal{D}}^{ML})x_{\mathcal{D}}Z_{\mathcal{D}}e) = (x_{\mathcal{D}}Z_{\mathcal{D}})H_{\mathcal{D}}/(x_{\mathcal{D}}Z_{\mathcal{D}}e),$$

which is equivalent to $\overline{y}_{\mathcal{D}}^{'} = \overline{x}_{\mathcal{D}}H_{\mathcal{D}}$. Since $x_{\mathcal{D}} > 0$, $y_{\mathcal{D}}^{'} > 0$, $\rho(T_{\mathcal{D}}^{ML}) > 0$, and $\zeta_{\mathcal{D}} > 0$, we have $\overline{x}_{\mathcal{D}} = (x_{\mathcal{D}}Z_{\mathcal{D}})/(x_{\mathcal{D}}Z_{\mathcal{D}}e) > 0$, implying $\overline{x}_{\mathcal{D}}e = 1$, and $\overline{y}_{\mathcal{D}}^{'} = (y_{\mathcal{D}}^{'}Z_{\mathcal{D}})/(\rho(T_{\mathcal{D}}^{ML})x_{\mathcal{D}}Z_{\mathcal{D}}e) > 0$. Furthermore, since

$H_{\mathcal{D}} > 0$, $H_{\mathcal{D}}e = e$, and $\overline{x}_{\mathcal{D}}e = 1$, we obtain $\overline{y}'_{\mathcal{D}}e = 1$. Then, from Lemma A.1 we have

$$\|\overline{y}'_{\mathcal{D}} - \psi_{\mathcal{D}}\|_1 \leq \left(1 - \min_{i_{\mathcal{D}}, j_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} h_{\mathcal{D}}(i_{\mathcal{D}}, j_{\mathcal{D}})\right) \|\overline{x}_{\mathcal{D}} - \psi_{\mathcal{D}}\|_1.$$

The result follows by taking each of $(\overline{y}'_{\mathcal{D}} - \psi_{\mathcal{D}})$ and $(\overline{x}_{\mathcal{D}} - \psi_{\mathcal{D}})$ into $Z_{\mathcal{D}}$ parantheses, multiplying both sides of the inequality by $\phi_{\mathcal{D}} Z_{\mathcal{D}} e$, letting $b_{\mathcal{D}} = (\phi_{\mathcal{D}} Z_{\mathcal{D}} e)/(x_{\mathcal{D}} Z_{\mathcal{D}} e)$, and using Definition 4.19. The part for the coarsest level follows from Definition 4.21 by the fact that $T^{ML}_{\mathcal{D}} = (e y'_{\mathcal{D}})/(x_{\mathcal{D}} e)$ and $\rho(T^{ML}_{\mathcal{D}}) = 1/(x_{\mathcal{D}} e)$, implying $T^{ML}_{\mathcal{D}}/\rho(T^{ML}_{\mathcal{D}}) = H_{\mathcal{D}} = e y'_{\mathcal{D}}$ and $Z_{\mathcal{D}} = I_{\mathcal{D}}$. For the $POWER$ smoother, Corollary 4.22 implies $Z_{\mathcal{D}} = \mathrm{diag}(e) = I_{\mathcal{D}}$, and therefore, the respective results. □

The ML iteration matrix, $T^{ML}_{\mathcal{D}}$, changes at each cycle due to the dependence of $P_{x'_{\mathcal{D}}}$ on $x'_{\mathcal{D}}$, and therefore, the ML iteration is non-stationary. At the end of each cycle, the solution vector at the finest level $\mathcal{D}$, $y'_{\mathcal{D}}$, is normalized to be unit 1-norm and then assigned to $x_{\mathcal{D}}$ so as to start the next cycle. As long as $x'_{\mathcal{D}} \neq \pi_{\mathcal{D}}$, the aggregated CTMC $\tilde{Q}_{\mathcal{C}}$ at the next coarser level can only be approximative. Theorem 4.26 indicates that the *normalized* solution vector, $b_{\mathcal{D}} x_{\mathcal{D}}$, improves with respect to the fixed point $\phi_{\mathcal{D}}$ with a converging smoother as long as $x_{\mathcal{D}} > 0$ is not in the range of $(I - T^{ML}_{\mathcal{D}})^T$. For the solution to improve with respect to steady state vector $\tilde{\pi}_{\mathcal{D}}$ at each level, one requires sufficient conditions on the smoother as in Theorem 4.20. Then $x_{\mathcal{D}}$ at the finest level will improve from one cycle to the next, implying an improvement in the aggregated CTMC at each level, and thus an improved solution at each level. Then, recalling from Lemma 4.25 that $\tilde{Q}_{\mathcal{D}} = Q_{\mathcal{D}}$ and $\rho(T^{ML}_{\mathcal{D}}) = 1$ upon convergence, $\rho(T^{ML}_{\mathcal{D}})$ and $\phi_{\mathcal{D}}$ must be approaching one and $\pi_{\mathcal{D}}$, respectively, while the subdominant eigenvalue of $T^{ML}_{\mathcal{D}}$ in magnitude is approaching zero with an increasing number of cycles.

In [11], extensive numerical experiments have been conducted with the ML solver on HMMs. Therein, the values chosen for the parameters of the $POWER$, JOR, and SOR smoothers are $\alpha = \max_{s_{\mathcal{D}} \in \mathcal{S}_{\mathcal{D}}} |\tilde{q}_{\mathcal{D}}(s_{\mathcal{D}}, s_{\mathcal{D}})|/0.999$ and $\omega = 1$, and the initial approximation is the uniform distribution. Furthermore, at least one pre- and one post-smoothing is performed at each level and the coarsest system is solved using Gaussian elimination. Hence, $POWER$ is enforced to yield a converging smoother, and the JOR and SOR iteration matrices are nonnegative. Although, $\omega = 1$ does not guarantee converging JOR and SOR smoothers (see Lemma 4.10), the results indicate that convergence may still be achieved. Hence, we conclude that the conditions stated in Theorem 4.20 for the smoothers are sufficient for convergence, but not necessary.

The next section presents the results of an ML cycle.

5. A sample ML cycle. In this section, we walk through one ML cycle on the CTMC in Example 1 with the initial approximation

$$x^{(0)} = \left(\begin{array}{cccc|c} 0.2000 & 0.2000 & 0.2000 & 0.2000 & 0.2000 \end{array}\right)$$

using the parameters $C = V$ and $O = FIXED$. These parameters imply that V is the cycle type (i.e., $\gamma = 1$) and $FIXED$ is the ordering of components for aggregation. The computations are performed in IEEE double (i.e., about 16 decimal digits of) precision. For brevity, we present the results with four digits after the decimal point. Note that, from $r^{(0)} = -x^{(0)}Q$ and $o^{(0)} = x^{(0)} - \pi$, we respectively have the initial residual vector

$$r^{(0)} = \left(\begin{array}{cccc|c} 0.2000 & 0.0000 & 0.2000 & 0.0000 & -0.4000 \end{array}\right) \quad \text{with} \quad \|r^{(0)}\|_\infty = 0.4000$$

and the initial error vector

$$o^{(0)} = \left(\begin{array}{cccc|c} 0.0250 & 0.0500 & 0.1000 & 0.0750 & -0.2500 \end{array}\right) \quad \text{with} \quad \|o^{(0)}\|_\infty = 0.2500,$$

which satisfy $r^{(0)}e = 0$ and $o^{(0)}e = 0$. We consider two cases in the next two subsections, and for clarity we denote the three levels respectively by $\mathcal{D}$, $\mathcal{C}$, $\mathcal{B}$ as in Definition 4.21.

5.1. With smoothings. In this part, one pre- and one post-smoothing is performed at each level using $S = SOR$ and $\omega = 0.5$. That is, SOR is enforced to be a converging smoother and (MIN_IN_PRE, MAX_IN_PRE, MIN_IN_POST, MAX_IN_POST, MIN_OUT_PRE, MAX_OUT_PRE, MIN_OUT_POST, MAX_OUT_POST) $= (1, 1, 1, 1, 1, 1, 1, 1)$.

- *Finest level, going down:* The repeat-until loop in main() starts by smoothing $x_{\mathcal{D}} = x^{(0)}$ for $\mathcal{D} = [1, 2, 3]$ with one iteration of the SOR splitting

$$\tilde{Q}_{\mathcal{D}} = M_{\mathcal{D}} - N_{\mathcal{D}} = \left(\begin{array}{cccc|c} -8 & 1 & 1 & 0 & 2 \\ 0 & -4 & 0 & 1 & 0 \\ 0 & 0 & -6 & 1 & 1 \\ 0 & 0 & 0 & -4 & 0 \\ \hline 0 & 0 & 0 & 0 & -2 \end{array}\right) - \left(\begin{array}{cccc|c} -4 & 0 & 0 & 0 & 0 \\ -1 & -2 & 0 & 0 & 0 \\ -1 & 0 & -3 & 0 & 0 \\ 0 & -1 & -1 & -2 & 0 \\ \hline -1 & 0 & 0 & 0 & -1 \end{array}\right) \tag{5.1}$$

obtained by using $\omega = 0.5$ in Definition 4.10. Note that

$$T_{\mathcal{D}}^{SOR} = N_{\mathcal{D}} M_{\mathcal{D}}^{-1} = \left(\begin{array}{cccc|c} 0.5000 & 0.1250 & 0.0833 & 0.0521 & 0.5417 \\ 0.1250 & 0.5312 & 0.0208 & 0.1380 & 0.1354 \\ 0.1250 & 0.0312 & 0.5208 & 0.1380 & 0.3854 \\ 0.0000 & 0.2500 & 0.1667 & 0.6042 & 0.0833 \\ \hline 0.1250 & 0.0312 & 0.0208 & 0.0130 & 0.6354 \end{array}\right), \quad \text{yet} \quad \rho(T_{\mathcal{D}}^{SOR}) = 1.$$

Observe that $T_{\mathcal{D}}^{SOR}$ has at least one positive row; hence, Theorem 4.20 applies. Since $x_{\mathcal{D}} = x^{(0)} > 0$, from $x_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{SOR}$ we have

$$x_{\mathcal{D}}^{'} = \left(\begin{array}{cccc|c} 0.1750 & 0.1938 & 0.1625 & 0.1891 & 0.3563 \end{array}\right),$$

and $x_{\mathcal{D}}^{'} > 0$ is not a unit 1-norm vector. Aggregating LLM 1 gives $\mathcal{C} = [2, 3]$,

$$R_{\mathcal{D}} = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array}\right) \quad \text{and} \quad P_{x_{\mathcal{D}}^{'}} = \left(\begin{array}{cccc|c} 0.5185 & 0 & 0.4815 & 0 & 0 \\ 0 & 0.5061 & 0 & 0.4939 & 0 \\ \hline 0 & 0 & 0 & 0 & 1.0000 \end{array}\right).$$

Hence, the states $\{(0,0,0), (0,1,0), (1,0,0), (1,1,0), (2,2,1)\}$ defined by $\mathcal{D}$ are mapped onto the aggregated states $\{(0,0), (1,0), (2,1)\}$ defined by $\mathcal{C}$. The coarser grid is given by

$$\tilde{Q}_{\mathcal{C}} = P_{x_{\mathcal{D}}^{'}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}} = \left(\begin{array}{cc|c} -2.5185 & 1.0000 & 1.5185 \\ 1.0000 & -1.0000 & 0.0000 \\ \hline 1.0000 & 0.0000 & -1.0000 \end{array}\right).$$

Note that $\tilde{Q}_{\mathcal{C}} e = 0$ as expected. The restricted solution is obtained from $x_{\mathcal{C}} = x_{\mathcal{D}}^{'} R_{\mathcal{D}}$ as

$$x_{\mathcal{C}} = \left(\begin{array}{cc|c} 0.3375 & 0.3828 & 0.3563 \end{array}\right).$$

At this point the recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, 1)$ is made.

- *Coarser level, going down:* The recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, 1)$ starts by smoothing $x_{\mathcal{C}}$ for $\mathcal{C} = [2, 3]$ with one iteration of the SOR splitting

$$\tilde{Q}_{\mathcal{C}} = M_{\mathcal{C}} - N_{\mathcal{C}} = \left(\begin{array}{cc|c} -5.0370 & 1.0000 & 1.5185 \\ 0.0000 & -2.0000 & 0.0000 \\ \hline 0.0000 & 0.0000 & -2.0000 \end{array}\right) - \left(\begin{array}{cc|c} -2.5185 & 0.0000 & 0.0000 \\ -1.0000 & -1.0000 & 0.0000 \\ \hline -1.0000 & 0.0000 & -1.0000 \end{array}\right) \tag{5.2}$$

obtained by using $\omega = 0.5$ in Definition 4.10. Note that

$$T_{\mathcal{C}}^{SOR} = N_{\mathcal{C}} M_{\mathcal{C}}^{-1} = \left(\begin{array}{cc|c} 0.5000 & 0.2500 & 0.3796 \\ 0.1985 & 0.5993 & 0.1507 \\ \hline 0.1985 & 0.0993 & 0.6507 \end{array}\right), \quad \text{yet} \quad \rho(T_{\mathcal{C}}^{SOR}) = 1.$$

Observe that $T_{\mathcal{D}}^{SOR} > 0$; hence, Theorem 4.20 applies. Furthermore, from Lemma 4.11 $\tilde{\pi}_{\mathcal{C}} = (0.2842\ 0.2842 \mid 0.4316)$. Now, since $x_{\mathcal{C}} = (0.3375\ 0.3828 \mid 0.3563)$, from $x_{\mathcal{C}}^{'} = x_{\mathcal{C}} T_{\mathcal{C}}^{SOR}$ we have

$$x_{\mathcal{C}}^{'} = \left(\begin{array}{cc|c} 0.3155 & 0.3491 & 0.4177 \end{array}\right).$$

Aggregating LLM 2 gives $\mathcal{B} = [3]$,

$$R_{\mathcal{C}} = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \end{array}\right) \quad \text{and} \quad P_{x'_{\mathcal{C}}} = \left(\begin{array}{cc|c} 0.4747 & 0.5253 & 0 \\ \hline 0 & 0 & 1.0000 \end{array}\right).$$

Hence, the states in $\{(0,0),(1,0),(2,1)\}$ defined by $\mathcal{C}$ are mapped onto the aggregated states in $\{(0),(1)\}$ defined by $\mathcal{B}$, the HLM matrix in this case. The coarser grid is given by

$$\tilde{Q}_{\mathcal{B}} = P_{x'_{\mathcal{C}}} \tilde{Q}_{\mathcal{C}} R_{\mathcal{C}} = \left(\begin{array}{c|c} -0.7208 & 0.7208 \\ \hline 1.0000 & -1.0000 \end{array}\right).$$

Note that $\tilde{Q}_{\mathcal{B}} e = 0$ as expected. The restricted solution is obtained from $x_{\mathcal{B}} = x'_{\mathcal{C}} R_{\mathcal{C}}$ as

$$x_{\mathcal{B}} = \left(\begin{array}{c|c} 0.6646 & 0.4177 \end{array}\right).$$

At this point the recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{B}}, x_{\mathcal{B}}, \mathcal{B}, 1)$ is made.

- *Coarsest level:* The recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{B}}, x_{\mathcal{B}}, \mathcal{B}, 1)$ executes the then part of the if-statement since only the HLM is left in $\mathcal{B}$, and returns with the result of the call to solve$(\tilde{Q}_{\mathcal{B}}, x_{\mathcal{B}})$, where $x_{\mathcal{B}} = (0.6646 \mid 0.4177)$ and $x_{\mathcal{B}} e = 1.0823$. Since $\tilde{Q}_{\mathcal{B}}$ is (2×2), a direct method is employed, $x_{\mathcal{B}}$ is not used, and

$$y'_{\mathcal{B}} = \tilde{\pi}_{\mathcal{B}} = \left(\begin{array}{c|c} 0.5811 & 0.4189 \end{array}\right)$$

is obtained. An extra condition has to be enforced since $\tilde{Q}_{\mathcal{B}}$ is singular and the right-hand side is zero. This condition is $y'_{\mathcal{B}} e = 1$. Note that Definition 4.21 yields the matrix

$$T_{\mathcal{B}}^{ML} = e y'_{\mathcal{B}} / (x_{\mathcal{B}} e) = 0.9240 \left(\begin{array}{c|c} 0.5811 & 0.4189 \\ \hline 0.5811 & 0.4189 \end{array}\right),$$

which satisfies $y'_{\mathcal{B}} = x_{\mathcal{B}} T_{\mathcal{B}}^{ML} > 0$, $\sigma(T_{\mathcal{B}}^{ML}) = \{0,\ 0.9240\}$, $\rho(T_{\mathcal{B}}^{ML}) = 1/x_{\mathcal{B}} e = 0.9240$, and $\phi_{\mathcal{B}} = (b_{\mathcal{B}} / \rho(T_{\mathcal{B}}^{ML}))/y'_{\mathcal{B}}$ from Theorem 4.26.

- *Finer level, going up:* Once the call returns with $y_{\mathcal{B}} = (0.5811 \mid 0.4189)$, prolongation is performed following $y_{\mathcal{C}} = y_{\mathcal{B}} P_{x'_{\mathcal{C}}}$, and we have

$$y_{\mathcal{C}} = \left(\begin{array}{cc|c} 0.2758 & 0.3053 & 0.4189 \end{array}\right).$$

The recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{C}}, x_{\mathcal{C}}, \mathcal{C}, 1)$ ends by smoothing $y_{\mathcal{C}}$ for $\mathcal{C} = [2,3]$ with one iteration of the SOR splitting in (5.2). Since $y_{\mathcal{C}} = (0.2758\ 0.3053 \mid 0.4189)$, from $y'_{\mathcal{C}} = y_{\mathcal{C}} T_{\mathcal{C}}^{SOR}$ we have

$$y'_{\mathcal{C}} = \left(\begin{array}{cc|c} 0.2817 & 0.2935 & 0.4233 \end{array}\right).$$

Note that Definition 4.21 yields the ML iteration matrix

$$T_{\mathcal{C}}^{ML} = \left(\begin{array}{cc|c} 0.2940 & 0.3063 & 0.4418 \\ 0.2469 & 0.2572 & 0.3710 \\ \hline 0.2469 & 0.2572 & 0.3710 \end{array}\right),$$

which satisfies $y'_{\mathcal{C}} = x_{\mathcal{C}} T_{\mathcal{C}}^{ML}$, $\sigma(T_{\mathcal{C}}^{ML}) = \{0,\ 0.9222\}$, $\rho(T_{\mathcal{C}}^{ML}) = 0.9222$, and $\phi_{\mathcal{C}} = (b_{\mathcal{C}} / \rho(T_{\mathcal{C}}^{ML}))/y'_{\mathcal{C}}$ from Theorem 4.26.

- *Finest level, going up:* Once the call returns with $y_{\mathcal{C}} = (0.2817\ 0.2935 \mid 0.4233)$, prolongation is performed following $y_{\mathcal{D}} = y_{\mathcal{C}} P_{x'_{\mathcal{D}}}$, and we have

$$y_{\mathcal{D}} = \left(\begin{array}{ccc|c} 0.1461 & 0.1485 & 0.1356 & 0.1449 & 0.4233 \end{array}\right).$$

The recursive call $\mathrm{ML}(\tilde{Q}_{\mathcal{D}}, x_{\mathcal{D}}, \mathcal{D}, 1)$ ends by smoothing $y_{\mathcal{D}}$ for $\mathcal{D} = [1,2,3]$ with one iteration of the SOR splitting in (5.1). Since $y_{\mathcal{D}} = (0.1461\ 0.1485\ 0.1356\ 0.1449 \mid 0.4233)$, from $y'_{\mathcal{D}} = y_{\mathcal{D}} T_{\mathcal{D}}^{SOR}$ we have

$$y'_{\mathcal{D}} = \left(\begin{array}{ccc|c} 0.1615 & 0.1509 & 0.1189 & 0.1399 & 0.4326 \end{array}\right).$$

Note that Definition 4.21 yields the ML iteration matrix

$$T_{\mathcal{D}}^{ML} = \left(\begin{array}{cccc|c} 0.2000 & 0.1869 & 0.1473 & 0.1733 & 0.5359 \\ 0.1385 & 0.1294 & 0.1019 & 0.1200 & 0.3709 \\ 0.1873 & 0.1750 & 0.1379 & 0.1623 & 0.5019 \\ 0.1608 & 0.1502 & 0.1184 & 0.1393 & 0.4307 \\ \hline 0.1208 & 0.1128 & 0.0889 & 0.1046 & 0.3235 \end{array}\right),$$

which satisfies $y_{\mathcal{D}}^{'} = x_{\mathcal{D}} T_{\mathcal{D}}^{ML}$, $\sigma(T_{\mathcal{D}}^{ML}) = \{0,\ 0.9301\}$, $\rho(T_{\mathcal{D}}^{ML}) = 0.9301$, and $\phi_{\mathcal{D}} = (b_{\mathcal{D}}/\rho(T_{\mathcal{D}}^{ML}))y_{\mathcal{D}}^{'}$ from Theorem 4.26.

Hence, the improved solution is returned as $y_{\mathcal{D}}^{'}$ and copied to $x_{\mathcal{D}}$. Then $x_{\mathcal{D}}$ is normalized to be the unit 1-norm vector

$$x^{(1)} = \left(\begin{array}{cccc|c} 0.1609 & 0.1503 & 0.1184 & 0.1394 & 0.4310 \end{array}\right),$$

which yields from $r^{(1)} = -x^{(1)}Q$ the improved residual vector

$$r^{(1)} = \left(\begin{array}{cccc|c} -0.0562 & 0.0004 & 0.0551 & 0.0100 & -0.0092 \end{array}\right) \quad \text{with} \quad \|r^{(1)}\|_\infty = 0.0562 < \|r^{(0)}\|_\infty$$

and from $o^{(1)} = x^{(1)} - \pi$ the improved error vector

$$o^{(1)} = \left(\begin{array}{cccc|c} -0.0141 & 0.0003 & 0.0184 & 0.0144 & -0.0190 \end{array}\right) \quad \text{with} \quad \|o^{(1)}\|_\infty = 0.0190 < \|o^{(0)}\|_\infty.$$

Note that $r^{(1)}$ and $o^{(1)}$ are zero sum vectors as expected.

For this problem, the ML solver convergences to a tolerance of $STOP\_TOL = 10^{-8}$ in 14 cycles and the maximum norm of the residual vector r upon convergence is in the order 10^{-9} (i.e., $\|r\|_\infty \approx 10^{-9}$). If convergence had not taken place, the next cycle would start with the improved $x_{\mathcal{D}}$. On the other hand, When $POWER$ with $\alpha = 0.999$ and JOR with $\omega = 0.5$ are used as smoothers, the ML solver converges respectively in 12 and 14 cycles with $\|r\|_\infty \approx 10^{-9}$. In passing we remark that, although it does not satisfy the assumptions of Theorems 4.20 and 4.26, SOR with $\omega = 1$ (i.e., GS) as the smoother converges within one cycle to machine precision for this problem.

5.2. Without smoothings. Now, consider the same cycle without any smoothings. That is, all parameters are the same except (MIN_IN_PRE, MAX_IN_PRE, MIN_IN_POST, MAX_IN_POST, MIN_OUT_PRE, MAX_OUT_PRE, MIN_OUT_POST, MAX_OUT_POST) $= (0,0,0,0,0,0,0,0)$, and therefore $\nu_1 + \nu_2 = 0$. For this case, we observe the following.

- *Finest level, going down:* We have

$$R_{\mathcal{D}} = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array}\right) \quad \text{and} \quad P_{x_{\mathcal{D}}} = \left(\begin{array}{cccc|c} 0.5000 & 0 & 0.5000 & 0 & 0 \\ 0 & 0.5000 & 0 & 0.5000 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right),$$

implying the coarser grid

$$\tilde{Q}_{\mathcal{C}} = P_{x_{\mathcal{D}}} \tilde{Q}_{\mathcal{D}} R_{\mathcal{D}} = \left(\begin{array}{cc|c} -2.5000 & 1.0000 & 1.5000 \\ 1.0000 & -1.0000 & 0.0000 \\ \hline 1.0000 & 0.0000 & -1.0000 \end{array}\right)$$

and the restricted solution

$$x_{\mathcal{C}} = x_{\mathcal{D}} R_{\mathcal{D}} = \left(\begin{array}{cc|c} 0.4000 & 0.4000 & 0.2000 \end{array}\right).$$

- *Coarser level, going down:* We have

$$R_{\mathcal{C}} = \left(\begin{array}{c|c} 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \end{array}\right) \quad \text{and} \quad P_{x_{\mathcal{C}}} = \left(\begin{array}{cc|c} 0.5000 & 0.5000 & 0 \\ \hline 0 & 0 & 1.0000 \end{array}\right),$$

implying the coarser grid

$$\tilde{Q}_{\mathcal{B}} = P_{x_{\mathcal{C}}} \tilde{Q}_{\mathcal{C}} R_{\mathcal{C}} = \left(\begin{array}{r|r} -0.7500 & 0.7500 \\ \hline 1.0000 & -1.0000 \end{array} \right)$$

and the restricted solution

$$x_{\mathcal{B}} = x_{\mathcal{C}} R_{\mathcal{C}} = \left(\begin{array}{cc|c} 0.8000 & 0.2000 \end{array} \right).$$

- *Coarsest level:* A direct method is employed to give

$$y_{\mathcal{B}} = \left(\begin{array}{cc|c} 0.5714 & 0.4286 \end{array} \right).$$

- *Finer level, going up:* The call from the coarsest level returns with $y_{\mathcal{B}}$ to give

$$y_{\mathcal{C}} = y_{\mathcal{B}} P_{x_{\mathcal{C}}} = \left(\begin{array}{cc|c} 0.2857 & 0.2857 & 0.4286 \end{array} \right).$$

- *Finest level, going up:* The call from the coarser level returns with $y_{\mathcal{C}}$ to give

$$y_{\mathcal{D}} = y_{\mathcal{C}} P_{x_{\mathcal{D}}} = \left(\begin{array}{cccc|c} 0.1429 & 0.1429 & 0.1429 & 0.1429 & 0.4286 \end{array} \right).$$

For this cycle, the ML iteration matrix at the finest level can be computed to be $T_{\mathcal{D}}^{ML} = e y_{\mathcal{D}}$ from Definition 4.21. Note that the prolongation operator will not change in future cycles, which implies no improvement, and hence no convergence.

6. Conclusion. In this paper, the convergence of a class of multilevel (ML) methods for large, sparse Markov chains (MCs) is investigated. The particular class of ML methods are inspired by algebraic multigrid and iterative aggregation-disaggregation, and have the capability of using (V, W, F) cycles, (power, Jacobi over-relaxation—JOR, successive over-relaxation—SOR) methods as smoothers, and (fixed, cyclic, dynamic) orders in which coarser MCs can be formed by aggregation in a cycle. A detailed convergence analysis is carried out. The conditions sufficient for convergence are an irreducible MC, a positive initial approximation from an appropriate subspace, an onto mapping of states from a finer MC to a coarser MC at each level, a uniformization parameter larger than the minimum magnitude of the diagonal elements for the power method, a relaxation parameter less than one for JOR and SOR, a sufficient number of pre- and post-smoothings at each level so as to ensure a smoothing matrix which is positive or has a(n) (almost) positive row/column, and the accurate solution of the coarsest system at each cycle.

Appendix.

LEMMA A.1. *Let v be a probability vector (i.e., $v \geq 0$ and $ve = 1$), G be a positive probability matrix (i.e., $G > 0$, $Ge = e$) with state space $\mathcal{S}$, $z = vG$, and ψ be the unique, positive fixed point of G (i.e., $\psi > 0$, $\psi G = \psi$) such that $\psi e = 1$. Then $z > 0$, $ze = 1$, and*

$$\|z - \psi\|_1 \leq \left(1 - \min_{i,j \in \mathcal{S}} g(i,j)\right) \|v - \psi\|_1,$$

where $0 < \min_{i,j \in \mathcal{S}} g(i,j) \leq 1/|\mathcal{S}|$.

Proof. The positivity of z follows from $z = vG$ since $v \geq 0$ (with $v(j) > 0$ for at least one $j \in \mathcal{S}$) and $G > 0$, and its unit 1-norm follows from $ze = v(Ge) = ve = 1$ since $Ge = e$ and $ve = 1$. Furthermore from $z = vG$, we have

$$z(i) = \sum_{j \in \mathcal{S}} g(j,i) v(j) \quad \text{for } i \in \mathcal{S}.$$

Then using $\psi = \psi G$, we can write

$$z(i) - \psi(i) = \sum_{j \in \mathcal{S}} g(j,i)(v(j) - \psi(j)) \quad \text{for } i \in \mathcal{S}.$$

Since $v \geq 0$ and $\psi > 0$ such that $ve = \psi e = 1$, it is impossible to satisfy $v(j) \geq \psi(j)$ for all $j \in \mathcal{S}$ unless $v = \psi$. When $v = \psi$, we also have $z = \psi$, which proves the result trivially. Therefore, we consider the sets

$$\mathcal{X} = \{j \mid v(j) \geq \psi(j),\ j \in \mathcal{S}\} \quad \text{such that } \mathcal{X} \neq \emptyset \text{ and } \mathcal{X} \subset \mathcal{S}$$

and

$$\mathcal{Y} = \{j \mid z(j) \geq \psi(j),\ j \in \mathcal{S}\} \quad \text{such that } \mathcal{Y} \neq \emptyset \text{ and } \mathcal{Y} \subset \mathcal{S}.$$

Then, $\mathcal{S} - \mathcal{X} = \{j \mid v(j) < \psi(j),\ j \in \mathcal{S}\}$ and $\mathcal{S} - \mathcal{Y} = \{j \mid z(j) < \psi(j),\ j \in \mathcal{S}\}$. Now, for $i \in \mathcal{S}$ we clearly have

$$z(i) - \psi(i) = \sum_{j \in \mathcal{X}} g(j,i)(v(j) - \psi(j)) + \sum_{j \in (\mathcal{S}-\mathcal{X})} g(j,i)(v(j) - \psi(j)),$$

which implies

$$\sum_{j \in (\mathcal{S}-\mathcal{X})} g(j,i)(v(j) - \psi(j)) \leq z(i) - \psi(i) \leq \sum_{j \in \mathcal{X}} g(j,i)(v(j) - \psi(j))$$

due to the definition of $\mathcal{X}$. Observe that the summations on the left and right respectively evaluate to negative and positive values.

Since G is positive and has row sums of one, its minimum element is positive and is maximized if one is equally distributed across all rows. That is [2, p. 268],

$$0 < \min_{i,j \in \mathcal{S}} g(i,j) \leq 1/|\mathcal{S}|.$$

First, consider summing both sides of

$$z(i) - \psi(i) \leq \sum_{j \in \mathcal{X}} g(j,i)(v(j) - \psi(j))$$

over $i \in \mathcal{Y}$ so as to obtain

$$\begin{aligned}\sum_{i \in \mathcal{Y}} (z(i) - \psi(i)) &\leq \sum_{i \in \mathcal{Y}} \sum_{j \in \mathcal{X}} g(j,i)(v(j) - \psi(j)) \leq \sum_{j \in \mathcal{X}} \left(\sum_{i \in \mathcal{Y}} g(j,i) \right) (v(j) - \psi(j)) \\ &\leq \left(1 - \min_{i,j \in \mathcal{S}} g(i,j)\right) \sum_{j \in \mathcal{X}} (v(j) - \psi(j)),\end{aligned}$$

where $\sum_{i \in \mathcal{Y}} g(j,i) \leq 1 - \min_{i,j \in \mathcal{S}} g(i,j)$ follows from $0 < \min_{i,j \in \mathcal{S}} g(i,j) \leq g(i,j)$ by the fact that $|\mathcal{Y}| < |\mathcal{S}|$. Since each term in the final summations are positive and $1 - \min_{i,j \in \mathcal{S}} g(i,j) > 0$, this can be rewritten as

$$\sum_{i \in \mathcal{Y}} |z(i) - \psi(i)| \leq \left(1 - \min_{i,j \in \mathcal{S}} g(i,j)\right) \sum_{j \in \mathcal{X}} |v(j) - \psi(j)|.$$

due to the definitions of $\mathcal{X}$ and $\mathcal{Y}$.

Next, consider summing both sides of

$$\sum_{j \in (\mathcal{S}-\mathcal{X})} g(j,i)(v(j) - \psi(j)) \leq z(i) - \psi(i)$$

over $i \in (\mathcal{S} - \mathcal{Y})$ so as to obtain

$$\begin{aligned}\sum_{i \in (\mathcal{S}-\mathcal{Y})} \sum_{j \in (\mathcal{S}-\mathcal{X})} g(j,i)(v(j) - \psi(j)) &\leq \sum_{i \in (\mathcal{S}-\mathcal{Y})} (z(i) - \psi(i)) \\ \sum_{j \in (\mathcal{S}-\mathcal{X})} \left(\sum_{i \in (\mathcal{S}-\mathcal{Y})} g(j,i) \right) (v(j) - \psi(j)) &\leq \sum_{i \in (\mathcal{S}-\mathcal{Y})} (z(i) - \psi(i)) \\ \left(1 - \min_{i,j \in \mathcal{S}} g(i,j)\right) \sum_{j \in (\mathcal{S}-\mathcal{X})} (v(j) - \psi(j)) &\leq \sum_{i \in (\mathcal{S}-\mathcal{Y})} (z(i) - \psi(i)),\end{aligned}$$

where $\sum_{i\in(\mathcal{S}-\mathcal{Y})} g(j,i) \leq 1 - \min_{i,j\in\mathcal{S}} g(i,j)$ follows in a similar way as before; however, this time each term in the final summations are negative. Therefore, we can write

$$\sum_{i\in(\mathcal{S}-\mathcal{Y})} |z(i)-\psi(i)| \leq \left(1 - \min_{i,j\in\mathcal{S}} g(i,j)\right) \sum_{j\in(\mathcal{S}-\mathcal{X})} |v(j)-\psi(j)|$$

due to the definitions of $(\mathcal{S}-\mathcal{X})$ and $(\mathcal{S}-\mathcal{Y})$.

Combining the two inequalities, which involve absolute values, side by side, we obtain

$$\sum_{i\in\mathcal{S}} |z(i)-\psi(i)| \leq \left(1 - \min_{i,j\in\mathcal{S}} g(i,j)\right) \sum_{j\in\mathcal{S}} |v(j)-\psi(j)|.$$

But, this is equivalent to

$$\|z-\psi\|_1 \leq \left(1 - \min_{i,j\in\mathcal{S}} g(i,j)\right) \|v-\psi\|_1$$

and the result is proved. □

REFERENCES

[1] F. Bause, P. Buchholz, and P. Kemper, *A toolbox for functional and quantitative analysis of DEDS*, in Quantitative Evaluation of Computing and Communication Systems, Lecture Notes in Computer Science 1469, R. Puigjaner, N. N. Savino, and B. Serra, eds., Springer Verlag, 1998, pp. 356–359.

[2] R. Bellman, *Introduction to Matrix Analysis*, 2nd ed., SIAM Press, Philadelphia, Pennsylvania, 1997.

[3] M. Benzi and T. Dayar, *The arithmetic mean method for finding the stationary vector of Markov chains*, Parallel Algorithms and Applications, 6 (1995), pp. 25–37.

[4] A. Berman and R. J. Plemmons, *Nonnegative matrices in the Mathematical Sciences*, SIAM Press, Philadelphia, Pennsylvania, 1994.

[5] A. Borobia and J. Moro, *On nonnegative matrices similar to positive matrices*, Linear Algebra and Its Applications, 266 (1997), pp. 365–379.

[6] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd edition, SIAM Press, Philadelphia, Pennsylvania, 2000.

[7] P. Buchholz, *Hierarchical structuring of superposed GSPNs*, IEEE Transactions on Software Engineering, 25 (1999), pp. 166–181.

[8] P. Buchholz, *Structured analysis approaches for large Markov chains*, Applied Numerical Mathematics, 31 (1999), pp. 375–404.

[9] P. Buchholz, *Multilevel solutions for structured Markov chains*, SIAM Journal on Matrix Analysis and Applications, 22 (2000), pp. 342–357.

[10] P. Buchholz and T. Dayar, *Block SOR for Kronecker structured Markovian representations*, Linear Algebra and Its Applications, 386 (2004), pp. 83–109.

[11] P. Buchholz and T. Dayar, *Comparison of multilevel methods for Kronecker structured Markovian representations*, Computing, 73 (2004), pp. 349–371.

[12] P. Buchholz and T. Dayar, *Block SOR preconditioned projection methods for Kronecker structured Markovian representations*, SIAM Journal on Scientific Computing, 26 (2005), pp. 1289–1313.

[13] P. Buchholz, G. Ciardo, S. Donatelli, and P. Kemper, *Complexity of memory-efficient Kronecker operations with applications to the solution of Markov models*, INFORMS Journal on Computing, 12 (2000), pp. 203–222.

[14] P. Buchholz and P. Kemper, *On generating a hierarchy for GSPN analysis*, Performance Evaluation Review, 26 (1998), pp. 5–14.

[15] P. Fernandes, B. Plateau, and W. J. Stewart, *Efficient descriptor-vector multiplications in stochastic automata networks*, Journal of the ACM, 45 (1998), pp. 381–414.

[16] P. Fernandes, B. Plateau, and W. J. Stewart, *Optimizing tensor product computations in stochastic automata networks*, RAIRO Operations Research, 32 (1998), pp. 325–351.

[17] J. R. Gilbert, *Predicting structure in sparse matrix computations*, SIAM Journal on Matrix Analysis and Applications, 15 (1994), pp. 62–79.

[18] G. Horton and S. Leutenegger, *A multi-level solution algorithm for steady state Markov chains*, Performance Evaluation Review, 22 (1994), pp. 191–200.

[19] U. Krieger, *Numerical solution of large finite Markov chains by algebraic multigrid techniques*, in Computations with Markov Chains, W. J. Stewart, ed., Kluwer Academic Publishers, Boston, Massachusetts, 1995, pp. 403-424.

[20] I. Marek and P. Mayer, *Convergence analysis of an iterative aggregation/disaggregation method for computing stationary probability vectors of stochastic matrices*, Numerical Linear Algebra with Applications, 5 (1998), pp. 253–274.

[21] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM Press, Philadelphia, Pennsylvania, 2000.

[22] U. Rüde, *The Multigrid Workbench*. http://www.mgnet.org/mgnet/tutorials/xwb.html

[23] J.-P. Tremblay and R. Manohar, *Discrete Mathematical Structures with Applications to Computer Science*. McGraw-Hill, New York, 1975.

[24] G. W. Stewart, *Matrix Algorithms, Vol. I: Basic Decompositions*, SIAM Press, Philadelphia, Pennsylvania, 1998.

[25] W. J. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, New Jersey, 1994.

[26] E. Uysal and T. Dayar, *Iterative methods based on splittings for stochastic automata networks*, European Journal of Operational Research, 110 (1998), pp. 166–186.

[27] C. F. Van Loan, *The ubiquitous Kronecker product*, Journal of Computational and Applied Mathematics, 123 (2000), pp. 85–100.

[28] P. Wesseling, *An Introduction to Multigrid Methods*, John Wiley & Sons, Chichester, New York, 1992.