Markov Chain Aggregation with Error Bounds on Transient Distributions

Fabian Michel ✉[1 [0009−0005−7768−9111]] and Markus Siegle[1 [0000−0001−7639−2280]]

Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85579 Neubiberg, Germany
{fabian.michel,markus.siegle}@unibw.de

Abstract. We extend the existing theory of formal error bounds for the transient distribution of an aggregated (or lumped) Markov chain when compared to the transient distribution of the original chain, for both discrete- and continuous-time Markov chains. In the discrete-time setting, we bound the stepwise increment of the error, and in the continuous-time setting, we bound the rate at which the error grows. We then compare these error bounds with relevant concepts in the literature such as exact and ordinary lumpability as well as deflatability and aggregatability. These concepts define stricter than necessary conditions to identify settings in which the aggregation error is zero. We also consider possible algorithms for finding suitable aggregations for which the formal error bounds are low, and we analyse first experiments with these algorithms on different models.

Keywords: Markov chains · Aggregation · Formal error bounds.

1 Introduction

State aggregation in dynamic systems has been studied extensively since the 1960s (see [13]). Due to the curse of dimensionality, models with large state spaces are often computationally intractable without state space reduction, and one basic reduction technique is to aggregate multiple states into a single state in the aggregated model. Conditions under which an aggregated Markov chain is again a Markov chain are well known (see strong/weak lumpability in [8,12]), and various cases where exact transient or stationary probabilities of the original model can be derived from an aggregated model have been analysed (see, e.g. [3]).

However, formal error bounds for the approximation error when exact aggregation is not possible have only been studied rarely. [1] has presented error bounds for the transient distributions of discrete-time Markov chains, derived from an aggregated model. We extend the theory developed in [1] to support a more general way of disaggregation and to the continuous-time domain without falling back on uniformisation. Subsequently, we analyse the cases where the error bounds are zero, show optimality of the bounds, and compare them with lumpability concepts from [2,3,7,8]. We present two different algorithms, one based on [2] and one based on [4], with the goal to identify an aggregation resulting in low error bounds.

2 Preliminaries

2.1 Aggregation of Markov chains

We consider time-homogeneous discrete- and continuous-time Markov chains (DTMCs and CTMCs) on the state space $S = \{1, \dots, n\}$. The dynamics are given by the stochastic transition matrix $P \in \mathbb{R}^{n \times n}$ for DTMCs, where we have $P(r,s) = \mathbb{P}[X_{k+1} = s \mid X_k = r]$ if X_k denotes the state of the DTMC at time k. For CTMCs, the dynamics are defined via the generator matrix $Q \in \mathbb{R}^{n \times n}$, where $Q(r,s)$ is the transition rate from r to s, and $Q(r,r) = -\sum_{s \neq r} Q(r,s)$. Given an initial distribution $p_0 \in \mathbb{R}^n$, the transient distribution of a DTMC (respectively CTMC) is given by $p_k = p_0 P^k$ (respectively $p_t = p_0 e^{Qt}$), if we interpret p_k as a row vector.

An aggregation of the state space S consists of a set $\Omega = \{\Omega_1, \dots, \Omega_m\}$ of m aggregates, where Ω is a partition of S, i.e. $\rho \in \Omega$ is a subset of S which represents all states belonging to one aggregate. The aggregation function $\omega : S \to \Omega$ maps a state s to the aggregate to which s belongs, i.e. $s \in \omega(s)$. We approximate the dynamics of the original Markov chain by defining a stochastic transition matrix $\Pi \in \mathbb{R}^{m \times m}$ for DTMCs and a generator matrix $\Theta \in \mathbb{R}^{m \times m}$ for CTMCs on the aggregated state space. $\Pi(\rho, \sigma)$ for $\rho, \sigma \in \Omega$ should be an approximation of the probability to transition from one aggregate state into another, that is, an approximation of $\mathbb{P}[X_{k+1} \in \sigma \mid X_k \in \rho]$. Note that this probability may now depend on k, in contrast to the probability $\mathbb{P}[X_{k+1} = s \mid X_k = r]$ for $r, s \in S$, which is time-independent. However, we consider only time-independent approximations $\Pi(\rho, \sigma)$. Simlarly, for CTMCs, $\Theta(\rho, \sigma)$ should approximate the transition rate from aggregate ρ to aggregate σ.

The aggregation can also be applied to initial and transient distributions. We call $\pi_0 \in \mathbb{R}^m$, defined via $\pi_0(\sigma) = \sum_{s \in \sigma} p_0(s)$ for $\sigma \in \Omega$, the aggregated initial distribution and define aggregated transient distributions via $\pi_k = \pi_0 \Pi^k$ (discrete time) and $\pi_t = \pi_0 e^{\Theta t}$ (continuous time). In order to obtain an approximation of the transient probability for a given state in the original chain, we consider distributions $\alpha_\sigma \in \mathbb{R}^n$ with support on $\sigma \in \Omega$. As a shorthand, we write $\alpha(s) := \alpha_{\omega(s)}(s)$. The value $\alpha(s)$ should approximate the conditional probability of being in state s when we know that we are in aggregate $\omega(s)$, i.e. the probability $\mathbb{P}[X_k = s \mid X_k \in \omega(s)]$. Again, this probability is in general dependent on time, but we only consider time-independent approximations α. We then arrive at an approximated transient distribution $\widetilde{p}_k \in \mathbb{R}^n$ by the following formula: $\widetilde{p}_k(s) = \alpha(s)\pi_k(\omega(s))$, or, for continuous time, $\widetilde{p}_t(s) = \alpha(s)\pi_t(\omega(s))$.

We further define the aggregation matrix Λ and the disaggregation matrix A as follows:

$$\Lambda = \begin{pmatrix} | & & | \\ \mathbb{1}_{\Omega_1} & \dots & \mathbb{1}_{\Omega_m} \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times m}, \quad A = \begin{pmatrix} — \, \alpha_{\Omega_1} \, — \\ \vdots \\ — \, \alpha_{\Omega_m} \, — \end{pmatrix} \in \mathbb{R}^{m \times n} \qquad \text{(note: } A\Lambda = I\text{)}$$

We set $\widetilde{P} = \Lambda\Pi A$ and $\widetilde{Q} = \Lambda\Theta A$. This implies in particular that

$$\widetilde{p}_k\widetilde{P}^l = \underbrace{\widetilde{p}_k\Lambda}_{\pi_k}\Pi^l A = \pi_{k+l}A = \widetilde{p}_{k+l}$$

$$\widetilde{p}_t e^{\widetilde{Q}u} = \widetilde{p}_t \sum_{k=0}^{\infty} \frac{u^k}{k!}\widetilde{Q}^k = \underbrace{\widetilde{p}_t\Lambda}_{\pi_t}\left(\sum_{k=0}^{\infty}\frac{u^k}{k!}\Theta^k\right)A = \pi_t e^{\Theta u}A = \pi_{t+u}A = \widetilde{p}_{t+u}$$

$\widetilde{P}$ (which is stochastic) describes the approximated dynamics of the aggregated chain if we blow it up again to the original state space S. It holds that $\widetilde{P}(r,s) = \alpha(s)\Pi(\omega(r),\omega(s))$, so we approximate $P(r,s)$ by the transition probability from aggregate $\omega(r)$ to aggregate $\omega(s)$, weighted with the conditional probability $\alpha(s)$ of being in state s within aggregate $\omega(s)$. We also have $\widetilde{Q}(r,s) = \alpha(s)\Theta(\omega(r),\omega(s))$ and $\widetilde{Q}$ describes the approximated dynamics in a sense, via the equation $\widetilde{p}_t e^{\widetilde{Q}u} = \widetilde{p}_{t+u}$. However, $\widetilde{Q}$ is no longer a generator matrix. The row sums of $\widetilde{Q}$ are 0, but the negative entries are no longer confined to the diagonal.

2.2 Exact aggregation

Definition 1. *Given a partition Ω of the state space of a DTMC or CTMC, we call a probability distribution p on the state space S* ***compatible*** *with distributions α_σ with support on $\sigma \in \Omega$ if $p\Lambda A = p$.*

Compatibility of p and the distributions α means that

$$\alpha(s) = \frac{p(s)}{\sum_{s'\in\omega(s)} p(s')} \text{ for all } s \in S \text{ s.t. } \sum_{s'\in\omega(s)} p(s') > 0$$

Definition 2. *We call an aggregation Ω of the state space of a DTMC (respectively CTMC) with distributions $\alpha_{\Omega_1},\ldots,\alpha_{\Omega_m}$ and aggregated transition matrix Π (respectively Θ)* ***dynamic-exact*** *if $\Pi A = AP$ (respectively $\Theta A = AQ$).*

If the initial distribution p_0 is further compatible with the α distributions, i.e. if $\widetilde{p}_0 = p_0\Lambda A = p_0$, then we call the aggregation ***exact****.*

If Ω, $\alpha_{\Omega_1},\ldots,\alpha_{\Omega_m}$ and Π are an exact aggregation, then $\widetilde{p}_k = p_k$ for all k; if the aggregation is dynamic-exact, then it only holds that $\widetilde{p}_k = \widetilde{p}_0 P^k$ (this follows later from Corollary 1). Note the difference to $\widetilde{p}_k = \widetilde{p}_0\widetilde{P}^k$, which is always true.

The condition $\Pi A = AP$ has appeared in the literature before. Equation (4) on page 135 of [8] states that, if Π is set as in (1) below, and if the α distributions are compatible with the initial distribution p_0, then $\Pi A = AP$ implies weak lumpability of the DTMC. A DTMC is called weakly lumpable for Ω if there exists an initial distribution p_0 such that the process Y_k, defined by $Y_k = \sigma \in \Omega \iff X_k \in \sigma$, is a Markov chain. For such an initial distribution, the probabilities $\pi_k(\sigma)$ are equal to $\mathbb{P}[Y_k = \sigma] = \mathbb{P}[X_k \in \sigma]$. However, the concept of weak lumpability makes no statement about whether the probability $\mathbb{P}[X_k = s]$ for $s \in \sigma$ can be accurately derived from the knowledge of $\mathbb{P}[X_k \in \sigma]$.

[9, Definition 2.2] defined P to be A-lumpable if $\Pi A = AP$ and if Π is set as in (1). [9] then noted that, given π_k, an exact recovery of $p_k(s)$ is possible if the initial distribution p_0 is compatible with the α distributions. An exact aggregation is also called backward bisimulation of type 2 in [5, Definition 4.3].

2.3 Aggregated dynamics

Choosing the aggregates. We consider different possibilities for choosing the aggregates later. For now, assume Ω is fixed. We want to choose Π (or Θ) and α in a way which results in a good approximation of the original dynamics.

Aggregated transition and generator matrices. We consider the following aggregated transition probabilities and rates for $\rho, \sigma \in \Omega$:

$$\Pi(\rho,\sigma) = \sum_{r\in\rho} \alpha(r) \sum_{s\in\sigma} P(r,s) \qquad \Theta(\rho,\sigma) = \sum_{r\in\rho} \alpha(r) \sum_{s\in\sigma} Q(r,s) \tag{1}$$

The probability $\mathbb{P}[X_{k+1} \in \sigma \mid X_k \in \rho]$ is approximated via the weighted average (with weights $\alpha(r)$) of the probability to transition from a single state $r \in \rho$ into any of the states in σ, since we assume that if we are in aggregate ρ, the probability to be in state $r \in \rho$ is given (approximately) by $\alpha(r)$. In matrix notation, setting Π and Θ as in (1) corresponds to $\Pi = AP\Lambda$ and $\Theta = AQ\Lambda$. Note that Π is again stochastic, and Θ is again a generator matrix. Different choices for Π are discussed in [1], and the above choice yields good approximations in terms of the transient distribution in the experiments done in [1]. Only a so-called "median-based scheme" (see [1, equation (21)]) performs better in some settings. However, this scheme might result in a non-stochastic Π such that the aggregated chain can no longer be considered as a Markov chain.

Conditional distributions. We also need to choose the conditional distributions α_σ. For DTMCs, the following definitions provided good results, and are compatible with the aggregation techniques which will be considered later.

- The first possibility, called **proportional** α, is given by

$$\alpha(s) = \frac{\sum_{r\in S} P(r,s)}{\sum_{r\in S}\sum_{s'\in\omega(s)} P(r,s')} \tag{2}$$

  $\alpha(s)$ is the same as the probability of being in state s, conditioned on being in the aggregate of s, after the Markov chain took a single step, starting with a uniform distribution. Intuitively, the distributions α_σ should be approximations of this type of conditional probabilities, with the exception that we do not necessarily start with a uniform distribution. Note: proportional α is well-defined if the chain is irreducible.
- The second possibility, called **uniform** α, is given by $\alpha(s) = \frac{1}{|\omega(s)|}$.

For CTMCs, (2) cannot be used since the sum $\sum_{r\in S} Q(r,s)$ could be negative. We therefore use the following for CTMCs:

$$\alpha(s) = \left(\sum_{r\in S,\, r\notin\omega(s)} Q(r,s)\right)\left(\sum_{r\in S,\, r\notin\omega(s)} \sum_{s'\in\omega(s)} Q(r,s')\right)^{-1} \tag{3}$$

3 Bounding the approximation error

3.1 Error bounds for DTMCs

We follow [1] to derive error bounds for the difference between the transient distribution p_k of the DTMC and the approximation $\widetilde{p}_k$. Assume Ω, Π, and the distributions α_σ are given. We do not use the particular forms of Π and α given in (1) and (2); arbitrary choices are possible. [1] set $\alpha(s) = \frac{1}{|\omega(s)|}$ implicitly. We demonstrate that other choices of α do not significantly change the error bounds derived in [1]. Call $e_k = \widetilde{p}_k - p_k$ the error after step k. We want to bound $\|e_k\|_1 = \sum_{s=1}^{n} |e_k(s)|$ where $e_k(s)$ is the s-th entry of $e_k \in \mathbb{R}^n$. Note that

$$e_k = \widetilde{p}_{k-1} \cdot \left(\widetilde{P} - P + P\right) - p_{k-1} \cdot P = \widetilde{p}_{k-1} \cdot \left(\widetilde{P} - P\right) + \underbrace{(\widetilde{p}_{k-1} - p_{k-1})}_{e_{k-1}} \cdot P$$

$$\implies \|e_k\|_1 \leq \left\|\widetilde{p}_{k-1} \cdot \left(\widetilde{P} - P\right)\right\|_1 + \|e_{k-1} \cdot P\|_1 \tag{4}$$

Lemma 1. *Let $e \in \mathbb{R}^k$ be an arbitrary (row) vector and $P \in \mathbb{R}^{k \times k}$ be an arbitrary stochastic matrix. Then $\|e \cdot P\|_1 \leq \|e\|_1$.*

We omit the simple proof. As a consequence of Lemma 1 and (4), we have

$$\|e_k\|_1 \leq \underbrace{\|e_{k-1}\|_1}_{\text{previous error}} + \underbrace{\left\|\widetilde{p}_{k-1} \cdot \left(\widetilde{P} - P\right)\right\|_1}_{\text{error from using approximated transition probabilities}}$$

We bound the second term as follows (see [1, pages 15-17]):

$$\begin{aligned}
\left\|\widetilde{p}_{k-1} \cdot \left(\widetilde{P} - P\right)\right\|_1 &= \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \sum_{\rho \in \Omega} \sum_{r \in \rho} \widetilde{p}_{k-1}(r) \cdot \left(\widetilde{P}(r,s) - P(r,s)\right) \right| \\
&= \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \sum_{\rho \in \Omega} \sum_{r \in \rho} \alpha(r)\pi_{k-1}(\rho) \cdot \left(\alpha(s)\Pi(\rho,\sigma) - P(r,s)\right) \right| \\
&= \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \sum_{\rho \in \Omega} \pi_{k-1}(\rho) \cdot \left(\alpha(s)\Pi(\rho,\sigma) \sum_{r \in \rho} \alpha(r) - \sum_{r \in \rho} \alpha(r)P(r,s)\right) \right| \\
&= \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \sum_{\rho \in \Omega} \pi_{k-1}(\rho) \cdot \left(\alpha(s)\Pi(\rho,\sigma) - \sum_{r \in \rho} \alpha(r)P(r,s)\right) \right| \\
&\leq \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \sum_{\rho \in \Omega} \pi_{k-1}(\rho) \cdot \left| \alpha(s)\Pi(\rho,\sigma) - \sum_{r \in \rho} \alpha(r)P(r,s) \right| \\
&= \sum_{\rho \in \Omega} \pi_{k-1}(\rho) \cdot \underbrace{\sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \alpha(s)\Pi(\rho,\sigma) - \sum_{r \in \rho} \alpha(r)P(r,s) \right|}_{=: \tau(\rho)} =: \langle \pi_{k-1}, \tau \rangle
\end{aligned} \tag{5}$$

In particular, it follows that

$$\|e_k\|_1 \leq \|e_0\|_1 + \sum_{i=0}^{k-1} \underbrace{\sum_{\rho \in \Omega} \pi_i(\rho) \cdot \tau(\rho)}_{\leq \max_{\rho \in \Omega} \tau(\rho)} \leq \|e_0\|_1 + k \cdot \max_{\rho \in \Omega} \tau(\rho) \tag{6}$$

Calculating $\tau(\rho)$ for every $\rho \in \Omega$ once thus allows us to bound $\|e_k\|_1$ using just the error at time 0 and the aggregated transition probabilities π_i at times $i = 0, \ldots, k-1$. Even simpler, $\max_{\rho \in \Omega} \tau(\rho)$ gives an upper bound for the error growth in every step which gives an instant linear error bound on $\|e_k\|_1$.

3.2 Error bounds for CTMCs

We extend the above setting to continuous-time Markov chains. Set

$$\tau(\rho) := \sum_{\sigma \in \Omega} \underbrace{\sum_{s \in \sigma} \left| \alpha(s) \Theta(\rho, \sigma) - \sum_{r \in \rho} \alpha(r) Q(r, s) \right|}_{=: \tau(\rho, \sigma)} \tag{7}$$

which exactly matches the definition in the case of discrete time. [1] already proved the following: set Θ as in (1) and consider a uniformisation of the original CTMC with uniformisation rate q. Apply the same aggregation to the resulting DTMC and set Π as in (1). Then, the error growth in step $k \to k+1$ in the DTMC can be bounded by $\frac{1}{q} \cdot \sum_{\rho \in \Omega} \pi_k(\rho) \cdot \tau(\rho)$ where τ is derived for the aggregated CTMC as in (7). We now drop the detour via the uniformisation and claim that the factors τ can be interpreted as a rate of error growth for CTMCs.

Theorem 1. *Let e_t be the vector of component-wise error of the approximated transient distribution at time t, i.e. $e_t = \widetilde{p}_t - p_t$. Then:*

(i) We have the following bound for the error at time t:

$$\|e_t\|_1 \leq \|e_0\|_1 + \int_0^t \langle \pi_s, \tau \rangle \, \mathrm{d}s \qquad \text{where } \langle \pi_s, \tau \rangle = \sum_{\rho \in \Omega} \pi_s(\rho) \cdot \tau(\rho)$$

$$\text{and } \|e_t\|_1 \leq \|e_0\|_1 + t \cdot \max_{\rho \in \Omega} \tau(\rho)$$

(ii) $\|e_t\|_1$ is absolutely continuous, almost everywhere (a.e.) differentiable, and

$$\frac{\mathrm{d}}{\mathrm{d}t} \|e_t\|_1 \leq \langle \pi_t, \tau \rangle \text{ a.e.} \quad \text{and} \quad \limsup_{u \to 0} \frac{\|e_{t+u}\|_1 - \|e_t\|_1}{u} \leq \langle \pi_t, \tau \rangle \ \ \forall t \geq 0$$

Before being able to prove this, we need another lemma.

Lemma 2. *Assume that $f : \mathbb{R} \to \mathbb{R}$ is differentiable in 0, and that $f(0) = 0$.*

$$\text{Then: } \limsup_{u \to 0} \frac{|f(u)|}{u} = \lim_{\substack{u \to 0 \\ u > 0}} \frac{|f(u)|}{u} = |f'(0)|$$

We omit the simple proof and proceed to show Theorem 1.

Proof (of Theorem 1). As (ii) immediately implies (i), it suffices to prove (ii).

First, note the following: every component of e_t is continuously differentiable in t, as both p_t and $\widetilde{p}_t$ are continuously differentiable with respect to t. Indeed, calculating the derivative of all components of e_t simultaneously, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\widetilde{p}_t - p_t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\widetilde{p}_0 e^{\widetilde{Q}t} - p_0 e^{Qt}\right) = \widetilde{p}_0 e^{\widetilde{Q}t}\widetilde{Q} - p_0 e^{Qt}Q = \widetilde{p}_t\widetilde{Q} - p_t Q \tag{8}$$

$$\left\|\frac{\mathrm{d}}{\mathrm{d}t}(\widetilde{p}_t - p_t)\right\|_1 \leq \left\|\widetilde{p}_t\widetilde{Q}\right\|_1 + \|p_t Q\|_1 \leq |S| \cdot \left(\max_{r,s\in S}\left|\widetilde{Q}(r,s)\right| + \max_{r,s\in S}|Q(r,s)|\right)$$

As every component of e_t is continuously differentiable with bounded derivative, $\|e_t\|_1$ is absolutely continuous and differentiable a. e. (see [11, Section 5.4]).

$$\begin{aligned}
\|e_{t+u}\|_1 &= \|\widetilde{p}_{t+u} - p_{t+u}\|_1 = \left\|\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right) + (\widetilde{p}_t - p_t)\,e^{Qu}\right\|_1 \\
&\leq \left\|\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right\|_1 + \left\|(\widetilde{p}_t - p_t)\,e^{Qu}\right\|_1 \\
&\overset{\circledcirc}{\leq} \left\|\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right\|_1 + \|\widetilde{p}_t - p_t\|_1 = \left\|\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right\|_1 + \|e_t\|_1 \\
\implies \|e_{t+u}\|_1 - \|e_t\|_1 &\leq \left\|\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right\|_1 = \sum_{s\in S}\left|\left(\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right)(s)\right| \qquad (9)
\end{aligned}$$

where $\circledcirc$ follows from Lemma 1 since e^{Qu} is a stochastic matrix. We now take a closer look at the right hand side. In particular, we are interested in

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}u}\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right) = \widetilde{p}_t\left(e^{\widetilde{Q}u}\widetilde{Q} - e^{Qu}Q\right) \overset{\text{for } u=0}{=} \widetilde{p}_t\left(\widetilde{Q} - Q\right) \\
\implies &\left.\frac{\mathrm{d}}{\mathrm{d}u}\right|_{u=0}\left(\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right)(s) = \sum_{r\in S}\widetilde{p}_t(r)\left(\widetilde{Q}(r,s) - Q(r,s)\right)
\end{aligned}$$

Hence

$$\begin{aligned}
\limsup_{u\to 0}\frac{\|e_{t+u}\|_1 - \|e_t\|_1}{u} &\overset{(9)}{\leq} \sum_{s\in S}\limsup_{u\to 0}\frac{\left|\left(\widetilde{p}_t\left(e^{\widetilde{Q}u} - e^{Qu}\right)\right)(s)\right|}{u} \\
&\overset{\text{Lemma 2}}{=} \sum_{s\in S}\left|\sum_{r\in S}\widetilde{p}_t(r)\left(\widetilde{Q}(r,s) - Q(r,s)\right)\right| \overset{\circledast}{\leq} \sum_{\rho\in\Omega}\pi_t(\rho)\cdot\tau(\rho) = \langle\pi_t,\tau\rangle
\end{aligned}$$

where $\circledast$ follows from the same calculation as for DTMCs, see (5). □

3.3 When is the error bound 0?

In order to better understand the error bound which arises from the τ factors, this section first analyses in which cases the error bound is equal to 0, and then shows tightness of the error bound.

Lemma 3. *Given a DTMC, a partition Ω of its state space, arbitrary distributions α_σ with support on $\sigma \in \Omega$, and arbitrary Π, it holds that*

$$\forall \rho \in \Omega : \tau(\rho) = 0 \iff \Pi A = AP \overset{\text{Definition 2}}{\iff} \text{the aggregation is dynamic-exact}$$

The same holds for CTMCs with Π replaced by Θ and P replaced by Q.

Proof. Note: $\tau(\rho) = 0$ if, and only if, $\tau(\rho, \sigma) = 0$ (defined in (7)) for all σ. Hence,

$$\begin{aligned} &\forall \rho \in \Omega : \tau(\rho) = 0 \iff \forall \rho, \sigma \in \Omega : \tau(\rho, \sigma) = 0 \\ &\iff \forall \rho, \sigma \in \Omega : \sum_{s \in \sigma} \left| \alpha(s) \Pi(\rho, \sigma) - \sum_{r \in \rho} \alpha(r) P(r, s) \right| = 0 \\ &\iff \forall \rho, \sigma \in \Omega : \forall s \in \sigma : \ \ \alpha(s) \Pi(\rho, \sigma) = \sum_{r \in \rho} \alpha(r) P(r, s) \end{aligned} \tag{10}$$

This already proves Lemma 3 since

$$\alpha(s) \Pi(\rho, \sigma) = \underbrace{(\Pi A)}_{\in \mathbb{R}^{m \times n}}(\rho, s) \qquad \text{and} \qquad \sum_{r \in \rho} \alpha(r) P(r, s) = (AP)(\rho, s)$$

The same calculation is true with Π replaced by Θ and P replaced by Q. Also note that one can even show that $\max_{\rho \in \Omega} \tau(\rho) = \|\Pi A - AP\|_\infty$. □

Corollary 1. *Given a DTMC or CTMC, a partition Ω, arbitrary distributions α_σ with support on $\sigma \in \Omega$, and arbitrary Π (respectively Θ) such that $\Pi A = AP$ (respectively $\Theta A = AQ$, i.e. a dynamic-exact aggregation), it holds that*

$$\|\widetilde{p}_k - p_k\|_1 \leq \|\widetilde{p}_0 - p_0\|_1 \ \text{ or, for continuous time, } \ \|\widetilde{p}_t - p_t\|_1 \leq \|\widetilde{p}_0 - p_0\|_1$$

In particular, if an aggregation is exact, then $\|\widetilde{p}_k - p_k\|_1 = 0$ for all k (respectively $\|\widetilde{p}_t - p_t\|_1 = 0$ for all t).

This follows from Lemma 3 and (6) (respectively Theorem 1 for CTMCs). We next show tightness of the error bounds.

Theorem 2. *Given a DTMC or CTMC, a partition Ω, distributions α_σ with support on $\sigma \in \Omega$, and arbitrary Π or Θ, assume that $\tau(\rho) > 0$ for some ρ. Then, there exists an initial distribution p_0 which is compatible with the α distributions such that $\|\widetilde{p}_1 - p_1\|_1 = \langle \pi_0, \tau \rangle$ or, for CTMCs, $\lim_{t \to 0, t > 0} \frac{1}{t} \|\widetilde{p}_t - p_t\|_1 = \langle \pi_0, \tau \rangle$.*

Proof. There must be some aggregate $\rho \in \Omega$ with $\tau(\rho) > 0$. We choose $p_0 = \alpha_\rho$, which is clearly compatible with the α distributions (hence $\widetilde{p}_0 = p_0 = \alpha_\rho$). In the discrete-time case, we have $p_1 = \alpha_\rho P$ and $\widetilde{p}_1 = \alpha_\rho \Lambda \Pi A$. Hence

$$\begin{aligned} \|\widetilde{p}_1 - p_1\|_1 &= \sum_{s \in S} |\widetilde{p}_1(s) - p_1(s)| = \sum_{s \in S} \left| \alpha(s) \Pi(\rho, \omega(s)) - \sum_{r \in \rho} \alpha(r) P(r, s) \right| \\ &= \sum_{\sigma \in \Omega} \sum_{s \in \sigma} \left| \alpha(s) \Pi(\rho, \sigma) - \sum_{r \in \rho} \alpha(r) P(r, s) \right| = \tau(\rho) = \langle \pi_0, \tau \rangle \end{aligned}$$

since π_0 is the Dirac measure on ρ. We proceed simlarly for the continuous-time case, again setting $p_0 = \alpha_\rho$. Note that, as already established in (8), we have

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\widetilde{p}_t - p_t)(s) &= \left(\widetilde{p}_0\widetilde{Q} - p_0 Q\right)(s) \overset{\widetilde{p}_0 = p_0 = \alpha_\rho}{=} \left(\alpha_\rho\widetilde{Q} - \alpha_\rho Q\right)(s) \\ &= \left(\alpha_\rho \Lambda\Theta A - \alpha_\rho Q\right)(s)\end{aligned}$$

Hence, noting that $\|\widetilde{p}_0 - p_0\|_1 = 0$, we obtain

$$\begin{aligned}\lim_{\substack{t\to 0\\ t>0}} \frac{1}{t}\|\widetilde{p}_t - p_t\|_1 &= \sum_{s\in S} \lim_{\substack{t\to 0\\ t>0}} \frac{|(\widetilde{p}_t - p_t)(s)|}{t} \overset{\text{Lemma 2}}{=} \sum_{s\in S} \left|\left(\alpha_\rho \Lambda\Theta A - \alpha_\rho Q\right)(s)\right| \\ &= \ldots \text{ (as in the discrete-time case) } = \langle \pi_0, \tau\rangle \end{aligned}$$

□

4 Lumpability and aggregatability

The following definition was given in [3, Definition 1]:

Definition 3. *A partition $\Omega = \{\Omega_1, \ldots, \Omega_m\}$ of the state space of a DTMC is called **ordinarily lumpable** if*

$$\forall r, r' \in S \text{ s.t. } \omega(r) = \omega(r') : \forall \sigma \in \Omega : \qquad \sum_{s\in\sigma} P(r,s) = \sum_{s\in\sigma} P(r',s) \tag{11}$$

For CTMCs, Ω is called ordinarily lumpable if (11) *holds with P replaced by Q.*

For an ordinarily lumpable partition, we have that $\pi_k(\sigma) = \sum_{s\in\sigma} p_k(s)$ if Π is set as in (1), and for any initial distribution p_0 (and independently of the choice of α). See [3, Theorem 5]. The same holds in the continuous-time case. [3, Definition 1] also defines exact lumpability:

Definition 4. *A partition $\Omega = \{\Omega_1, \ldots, \Omega_m\}$ of the state space of a DTMC is called **exactly lumpable** if*

$$\forall s, s' \in S \text{ s.t. } \omega(s) = \omega(s') : \forall \rho \in \Omega : \qquad \sum_{r\in\rho} P(r,s) = \sum_{r\in\rho} P(r,s') \tag{12}$$

For CTMCs, Ω is called exactly lumpable if (12) *holds with P replaced by Q. Ω is further called **strictly lumpable** if it is both ordinarily and exactly lumpable.*

[2, Definition 2.1] also defines lumpability. Note that this definition of lumpability agrees with the definition of ordinary lumpability given above. [2, Definition 2.1] further defines deflatability and aggregatability:

Definition 5. *A partition Ω of the state space of a DTMC, together with distributions $\alpha_\sigma \in \mathbb{R}^n$ with support on $\sigma \in \Omega$, is called **deflatable** if*

$$\forall r \in S : \forall s \in S : \qquad P(r,s) = \alpha(s) \cdot \sum_{s'\in\omega(s)} P(r,s') \tag{13}$$

*The partition Ω, together with distributions α, is further called **aggregatable** if it is deflatable and if Ω is ordinarily lumpable.*

Note that aggregatability implies that $\widetilde{P} = P$ by [2, Proposition 2.6] (if Π is set as in (1)). Definition 5 cannot be extended to CTMCs easily.

Proposition 1. *Assume an irreducible DTMC or CTMC is exactly lumpable w.r.t. the partition Ω. When setting Π (respectively Θ) as in* (1) *and using proportional α as in* (2) *(respectively* (3)*)* or *uniform α, then it holds that*

$$\tau(\rho) = 0 \ \ \forall \rho \in \Omega \qquad \textit{and} \qquad \alpha(s) = \frac{1}{|\omega(s)|} \ \ \forall s \in S$$

Proof (Sketch). First, we look at DTMCs. Assume α is set as in (2), and $\omega(s) = \omega(s') = \sigma$. Then:

$$\alpha(s') \cdot \sum_{r \in S} \sum_{s'' \in \sigma} P(r,s'') = \sum_{r \in S} P(r,s') \overset{\circledast}{=} \sum_{r \in S} P(r,s) = \alpha(s) \cdot \underbrace{\sum_{r \in S} \sum_{s'' \in \sigma} P(r,s'')}_{>0 \text{ by irreducibility}}$$

where $\circledast$ holds by exact lumpability. Hence $\alpha(s) = \alpha(s')$. For CTMCs and α as in (3), the calculation is similar. Now, one way to proceed is to show that $\tau(\rho) = 0$ for all ρ is equivalent to:

$$\forall s, s' \text{ s.t. } \omega(s) = \omega(s') : \forall \rho \in \Omega : \frac{\sum_{r \in \rho} \alpha(r) P(r,s)}{\alpha(s)} = \frac{\sum_{r \in \rho} \alpha(r) P(r,s')}{\alpha(s')} \quad (14)$$

We skip the proof of equivalence here, but it relies only on simple calculations based on (10). More details are given in [10, Proposition 8] or in [7, Section 2 and 3], where the same equivalence is shown in a slightly different context (note that (14) is equal to [7, (Cond1) on page 771]). If we plug $\alpha(s) = \frac{1}{|\omega(s)|}$ into (14), we get the defininig equation (12) of exact lumpability, finishing the proof. □

Proposition 2. *Given an irreducible DTMC and a partition Ω, assume*

$$\forall r \in S : \forall s \in S : \qquad P(r,s) = c(s) \cdot \sum_{s' \in \omega(s)} P(r,s')$$

for constants $c(s) \in [0,1]$ which only depend on $s \in S$ (this is equivalent to the existence of distributions α s.t. Ω and α are deflatable). When setting Π as in (1) *and using proportional α as in* (2)*, it holds that*

$$\tau(\rho) = 0 \ \ \forall \rho \in \Omega \qquad \textit{and} \qquad \alpha(s) = c(s) \ \ \forall s \in S$$

We omit the proof of this proposition, it relies only on basic calculations and the equivalence of $\tau(\rho) = 0$ for all ρ to (14). We next show that none of the lumpability concepts above are necessary conditions for $\tau(\rho) = 0$ for all ρ. Except for [7], a large part of the literature has thus treated stricter than necessary conditions in order for dynamic-exact aggregation to be possible.

Proposition 3. *There are partitions Ω of the state space of a DTMC and probability distributions α_σ with support on $\sigma \in \Omega$ which are dynamic-exact (when Π is set as in* (1)*), but where Ω is neither ordinary lumpable, nor exactly lumpable, nor are Ω and the distributions α deflatable.*

Proof. Consider the state space $S = \{1, 2, 3\}$, the aggregation $\Omega = \{\{1\}, \{2, 3\}\}$ and $\alpha(1) = 1, \alpha(2) = \frac{1}{4}, \alpha(3) = \frac{3}{4}$ as well as the DTMC given by:

$$P = \begin{pmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{4}{9} & \frac{1}{18} & \frac{1}{2} \end{pmatrix} \quad \overset{(1)}{\Longrightarrow} \quad \Pi = \begin{pmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$\Pi A = AP$ holds, i.e. this aggregation is dynamic-exact and $\tau(\{1\}) = \tau(\{2, 3\}) = 0$. We show that none of the stated properties hold for Ω and α:

- ordinary lumpability: since $\omega(2) = \omega(3)$, by (11) in Definition 3, we would need $1 = P(2, 2) + P(2, 3) = P(3, 2) + P(3, 3) = \frac{5}{9}$ which is clearly not true.
- exact lumpability: since $\omega(2) = \omega(3)$, by (12) in Definition 4, we would need $\frac{5}{9} = P(2, 2) + P(3, 2) = P(2, 3) + P(3, 3) = 1$ which is clearly not true.
- deflatability: since $\omega(2) = \omega(3)$, by (13) in Definition 5, we would need

$$\frac{1}{2} = P(2, 2) = \alpha(2) \sum_{s \in \{2,3\}} P(2, s) = \alpha(2) \cdot 1 \implies \alpha(2) = \frac{1}{2}$$

$$\frac{1}{18} = P(3, 2) = \alpha(2) \sum_{s \in \{2,3\}} P(3, s) = \alpha(2) \cdot \frac{5}{9} \implies \alpha(2) = \frac{1}{10}$$

  so Ω and α are not deflatable, and there is no other deflatable choice for α.

In fact, there is no aggregation with $\Omega \neq \{S\}$ and $\Omega \neq \{\{1\}, \{2\}, \{3\}\}$ which is ordinary or exactly lumpable, or for which deflatable α distributions exist. □

For this example, neither using proportional α as in (2) nor uniform α delivers the choice of α which results in a dynamic-exact aggregation. We get $\alpha(1) = 1$, $\alpha(2) = \frac{29}{92} \approx 0.315, \alpha(3) = \frac{63}{92} \approx 0.685$ for proportional α. The proposed ways of calculating α can thus only be seen as an approximation of the optimal choice.

5 Choosing the aggregates

We want to choose a partition Ω such that the error bounds $\tau(\rho)$ are low, as this results in a good approximation $\widetilde{p}_k$ of p_k. In order to reduce the computational effort required to calculate $\widetilde{p}_k$, we would also like $|\Omega| = m \ll n = |S|$. An ideal algorithm would receive a parameter ε as input and determine the partition Ω with the fewest aggregates satisfying $\max_{\rho \in \Omega} \tau(\rho) < \varepsilon$. This would guarantee a stepwise error (or error growth rate) of at most ε (see (6) and Theorem 1 (i)). Solving this problem exactly will in general result in a runtime exceeding the time needed to compute p_k exactly for the original chain. We therefore consider different ways of choosing an Ω which is close to the optimal solution.

5.1 Almost aggregatability

We consider three algorithms based on [2] which identify almost aggregatable partitions for which the error bounds are low by Proposition 2. These algorithms use the singular value decomposition of P.

- **SVD sgn**: proposed as a simple algorithm in [2] with only limited practical applicability due to its numerical instability.
- **SVD seba**: proposed as a stable variant in [2] via a combination with [6].
- **SVD dir**: a new algorithm devised by us based on SVD sgn. The SVD algorithm from [2] basically assigns a (cropped) row vector of the right hand matrix V of the singular value decomposition $U\Sigma V^{\mathsf{T}}$ of P to every state in the Markov chain. SVD sgn then only analyses the sign structure of these vectors, while SVD dir exploits the fact that the vectors of two states in the same aggregate should point in approximately the same direction.

The three variants of the SVD algorithm can only be applied to DTMCs (aggregatable was only defined for DTMCs in Definition 5). In order to decide how coarse the aggregation should be, all three algorithms receive a parameter ε as input which is used to decide where to cut off the row vectors of V used for partitioning, i.e. ε is used to decide which dimension these vectors should have. Details are given in [10, Section 5.2].

5.2 ε-almost exact lumpability

By Proposition 1, if Ω is exactly lumpable, Π is set as in (1) and if proportional α as in (2) or uniform α is used, then the error bound is zero. In the general case, it is more likely that a partition exists which is close to being exactly lumpable. This motivates the following definition:

Definition 6. *We call a partition Ω ε-**almost exactly lumpable** if:*

$$\forall s, s' \in S \text{ s.t. } \omega(s) = \omega(s') : \qquad \sum_{\rho \in \Omega} \left| \sum_{r \in \rho} P(r,s) - \sum_{r \in \rho} P(r,s') \right| \leq \varepsilon$$

We now develop an algorithm (see Algorithm 1, similar to the ideas from [4, p. 269]) which finds an ε-almost exactly lumpable partition. It works for DTMCs as well as CTMCs. For a given ε, the algorithm should find a partition which is as coarse as possible and still satisfies ε-almost exact lumpability. The idea of the algorithm is as follows: we start with the initial partition $\Omega = \{S\}$, which is then successively refined. At every refinement step, for every aggregate $\sigma \in \Omega$ and for all states $s \in \sigma$, we construct vectors of incoming probabilities

$$\mathrm{inc}(s) = \left(\sum_{r \in \Omega_1} P(r,s)\,, \;\; \ldots\,, \;\; \sum_{r \in \Omega_m} P(r,s) \right) \in \mathbb{R}^m$$

where m is the current number of aggregates in Ω. By Definition 6, we have that the current partition Ω is ε-almost exactly lumpable if, and only if, we have that $\|\mathrm{inc}(s) - \mathrm{inc}(s')\|_1 \leq \varepsilon$ for all states s and s' belonging to the same aggregate σ. If this is not the case, the algorithm proceeds with the refinement by partitioning the states into smaller aggregates. This procedure stops when an ε-almost exactly lumpable partition is found (at the latest when every aggregate consists of a single state).

Algorithm 1 Calculating almost exactly lumpable partitions

Input: a Markov chain, defined via its transition matrix P on state space S, and the parameter ε (a generator matrix Q can be used instead of P)
Output: an aggregation function ω whose corresponding partition is ε-almost exactly lumpable

$\omega^{(1)} \leftarrow ((s \in S) \mapsto 1)$ ▷ aggregation function
$i \leftarrow 1$ ▷ iteration counter
$m \leftarrow 1$ ▷ number of aggregates
repeat
  $m_{\text{old}} \leftarrow m$ ▷ saves number of old aggregates
  $m \leftarrow 0$ ▷ counts number of new aggregates
  for all $j \in \{1, \dots, m_{\text{old}}\}$ **do** ▷ loop over old aggregates
    for all $s \in \{r \in S : \omega^{(i)}(r) = j\}$ **do** ▷ loop over states in same aggregate
      $\text{inc}(s) \leftarrow \mathbf{0} \in \mathbb{R}^{m_{\text{old}}}$
      for all $k \in \{1, \dots, m_{\text{old}}\}$ **do** ▷ loop over potential splitters
        $\text{inc}(s)_k \leftarrow \sum_{r \in S : \omega^{(i)}(r) = k} P(r, s)$ ▷ inc. prob. from agg. k to state s
      end for
    end for
    $C \leftarrow \text{cluster}(\{r \in S : \omega^{(i)}(r) = j\}, \text{inc}, \varepsilon)$ ▷ see below
    for all $\sigma \in C$ **do** ▷ loop over clusters
      for all $s \in \sigma$ **do**
        $\omega^{(i+1)}(s) \leftarrow m + 1$ ▷ states in σ are assigned to the same agg.
      end for
      $m \leftarrow m + 1$ ▷ increment aggregate number
    end for
  end for
  $i \leftarrow i + 1$
until $m_{\text{old}} = m$ ▷ stop when no aggregates were split
return $\omega^{(i)}$

cluster(T, f, ε) takes a subset of states $T \subseteq S$, a function $f : T \to \mathbb{R}^k$ and $\varepsilon > 0$ as input. The output is a partition C of T such that for any cluster $\sigma \in C$ and any two states $s, s' \in \sigma$, we have that $\|f(s) - f(s')\|_1 \leq \varepsilon$. We use Python and `scipy.cluster.hierarchy.fclusterdata` to calculate the clustering.

5.3 Experiments

We compare the performance of SVD sgn, SVD seba, SVD dir, and Algorithm 1 by comparing the error bounds given by the τ factors resulting from the aggregations returned by the different algorithms – the lower, the better. By default, we calculate the α distributions as in (2), and Π (or Θ) is set as in (1).

In Figure 1, we consider a setting for which the SVD algorithms were designed. We see that the SVD variants (except SVD seba) perform better than Algorithm 1 for almost aggregatable chains. The higher stability of SVD dir pays off in comparison to SVD sgn: we see a sharp drop in the error bounds around 20 aggregates (the number of aggregates in the almost aggregatable partition). For SVD sgn, the drop is more a gradual decrease. Algorithm 1 does not identify the

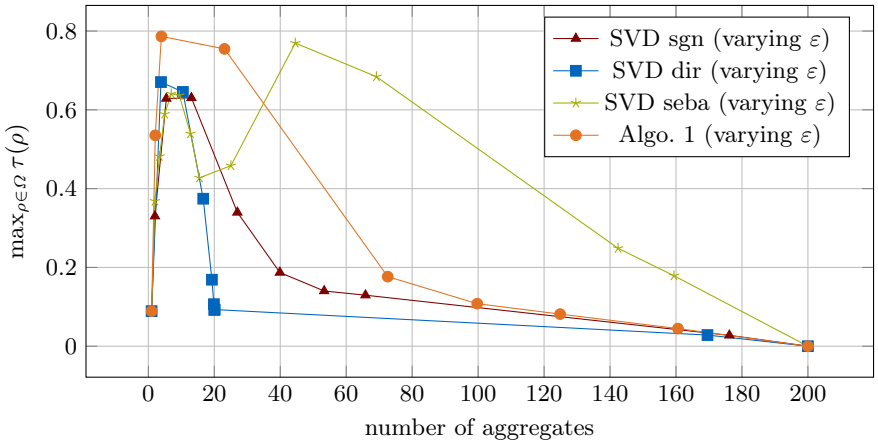

Fig. 1. SVD sgn, SVD dir, SVD seba, and Algorithm 1 executed on 100 randomly generated almost aggregatable DTMCs with 200 states, 20 aggregates and a probability of 0.5 to have no transition between a particular pair of aggregates. The almost aggregatable DTMCs were obtained by random perturbation of the transition matrix of an aggregatable DTMC with a magnitude of 0.002. Each plotted point is an average resulting from running the algorithms with a particular fixed input parameter ε on the 100 DTMCs. Multiple different parameters are used to obtain the different points.

almost aggregatable partition. SVD seba performs similarly to SVD dir for a low number of aggregates, but there is a sudden change around 20 aggregates when SVD seba starts to perform worse than all other algorithms. This is due to the fact that we limited the maximum number of iterations of the SEBA algorithm (see [6, Algorithm 3.1]) to 300 iterations because of its high runtime. Regardless of the number of maximum iterations, we could never observe SVD seba performing significantly better than SVD dir in all our experiments. The latter is therefore a good alternative. For details on implementation, refer to [10].

We also considered the compositional stochastic process algebra model RSVP from [14], comprising a lower and upper network channel with capacities for M and N calls, and a number of identical mobile nodes which request resources for calls. Due to the identical mobile nodes, lossless aggregation is possible. Comparing the different algorithms in Figure 2 for a uniformisation of this model, we see that only Algorithm 1 identifies the lossless aggregation (which is exactly lumpable). The SVD variants perform much worse. Figure 2 also compares the two different ways to calculate α: proportional α as in (2) and uniform α.

6 Conclusion & outlook

We extended the error bounds for the difference between the transient distribution of an aggregated chain and the original chain originally derived in [1] to a

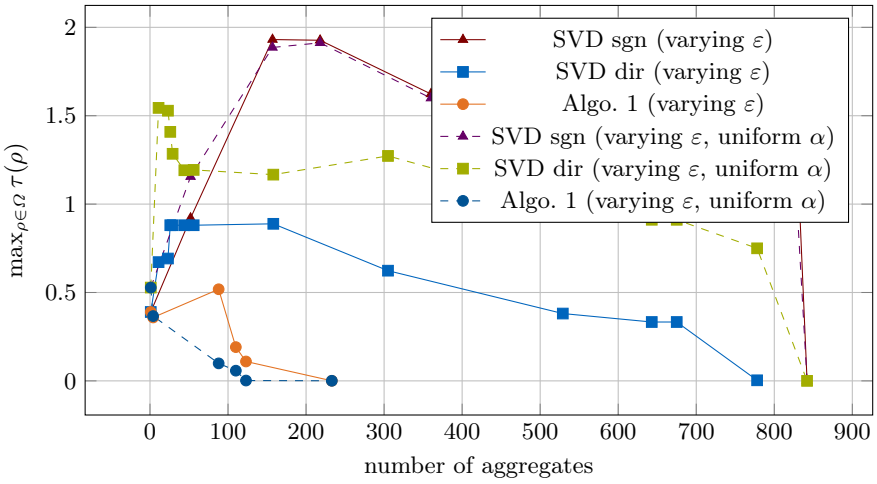

Fig. 2. SVD sgn, SVD dir, and Algorithm 1 executed on the uniformisation of the model from [14] with $M = 7$, $N = 5$ and 3 mobile nodes, resulting in a total of 842 states. By symmetry of the mobile nodes, a lossless reduction to 234 states is possible.

more general setting and proved that these bounds are tight. We also showed a relation of the error bounds to existing lumpability concepts, and we compared algorithms which identify different settings in which the error bounds are low. To obtain more reliable results, these algorithms have to be compared with other approaches in real-world applications in the future, and a combination with the adaptive aggregation from [1] should be investigated. Another interesting topic would be to develop an efficient algorithm which finds an approximate solution to $\Pi A = AP$, i.e. the most general case in which the error bounds are zero.

References

1. Abate, A., Andriushchenko, R., Češka, M., Kwiatkowska, M.: Adaptive formal approximations of Markov chains. Performance Evaluation **148**(102207) (2021). https://doi.org/10.1016/j.peva.2021.102207
2. Bittracher, A., Schütte, C.: A probabilistic algorithm for aggregating vastly undersampled large Markov chains. Physica D: Nonlinear Phenomena **416**(132799) (2021). https://doi.org/10.1016/j.physd.2020.132799
3. Buchholz, P.: Exact and ordinary lumpability in finite Markov chains. Journal of Applied Probability **31**(1), 59–75 (1994). https://doi.org/10.2307/3215235
4. Buchholz, P.: Exact performance equivalence: An equivalence relation for stochastic automata. Theoretical Computer Science **215**, 263–287 (1999). https://doi.org/10.1016/S0304-3975(98)00169-8
5. Buchholz, P.: Bisimulation relations for weighted automata. Theoretical Computer Science **393**, 109–123 (2008). https://doi.org/10.1016/j.tcs.2007.11.018

6. Froyland, G., Rock, C.P., Sakellariou, K.: Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification. Communications in Nonlinear Science and Numerical Simulation **77**, 81–107 (2019). https://doi.org/10.1016/j.cnsns.2019.04.012
7. Ganguly, A., Petrov, T., Koeppl, H.: Markov chain aggregation and its applications to combinatorial reaction networks. Journal of Mathematical Biology **69**(3), 767–797 (2014). https://doi.org/10.1007/s00285-013-0738-7
8. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Springer (1976), https://link.springer.com/book/9780387901923
9. Ledoux, J., Truffet, L.: Markovian bounds on functions of finite Markov chains. Advances in Applied Probability **33**(2), 505–519 (2001). https://doi.org/10.1017/S0001867800010910
10. Michel, F., Siegle, M.: Markov chain aggregation with error bounds on transient distributions. arXiv preprint (2024). https://doi.org/10.48550/arXiv.2403.07618
11. Royden, H.L.: Real Analysis. Collier Macmillan, 3rd edn. (1988)
12. Rubino, G., Sericola, B.: A finite characterization of weak lumpable Markov processes. Part II: The continuous time case. Stochastic Processes and their Applications **45**(1), 115–125 (1993). https://doi.org/10.1016/0304-4149(93)90063-A
13. Simon, H.A., Ando, A.: Aggregation of variables in dynamic systems. Econometric **29**(2), 111–138 (1961). https://doi.org/10.2307/1909285
14. Wang, H., Laurenson, D.I., Hillston, J.: Evaluation of RSVP and mobility-aware RSVP using performance evaluation process algebra. 2008 IEEE International Conference on Communications pp. 192–197 (2008). https://doi.org/10.1109/ICC.2008.43