ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Accuracy of coarse grained Markovian dynamics

Karl Heinz Hoffmann [a,*], Peter Salamon [b]

[a] Institut für Physik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany

[b] Department of Mathematical Sciences, San Diego State University, CA 92182-7720, San Diego, USA

ARTICLE INFO

Article history:
Received 23 December 2010
Received in revised form 19 April 2011
Available online xxxx

Keywords:
Irreversible thermodynamics
Markov processes
Markov chain
Lumping techniques
Coarse grained Markovian dynamics

ABSTRACT

Markov chain models on a mesoscopic level are a widely used description for complex systems. They are based on the assumption that certain sets of microstates can be coarse grained as their internal dynamics is faster than the time scales considered in the modeling. Here we analyze quantitatively the errors made by using lumping techniques and present the first rigorous proof for bounds on such errors. Our bounds express the deviations from a full microscopic description for all subsequent time steps in terms of the deviations in the first time step.

1. Introduction

In physics, Markov chains provide a paradigm dynamics for irreversible processes. The applications of Markov chains are typically via the master equation [1,2], with the transition probability matrix containing the specifics of the system under investigation. But Markov chains also occur in other fields of science. They are an often used modeling tool with applications ranging from business [3,4] and sociology [5] to chemistry [6], biology [7], and computer science [8].

Most applications of Markov chains model the states of the chain already as some kind of average over many factors which cannot be individually controlled. In physical theories based on the master equation, this comes about through coarse graining often implicitly by defining the states of the Markov chain to be aggregates or, in the mathematical language, lumps of microstates. Our goal is to analyze the accuracy of such coarse grained models as compared to the exact microscopic behavior, i.e., to bound the error in the dynamics as predicted by a coarse grained description. Apart from the mathematical facts [9] we present application areas as well as examples showing the strength of our bounds.

We present the first rigorous proof of a fact that mesoscopic models take for granted. In these approaches, certain sets of microstates are not resolved in the modeling, but are rather treated as mesostates. One assumes that the mesoscopic dynamics appropriately models the behavior of the original system, at least on time scales that are large compared to the time scale of relaxation within one mesostate. While this assumption seems reasonable, it remains unclear to what extent the dynamics quantitatively mirrors the underlying system dynamics. The rationale behind the mesoscopic approach is to facilitate calculations with a simplified model as a full microscopic approach is usually beyond technical feasibility.

Our theorem below bounds the error that results from the mesoscopic modeling in terms of the error in one step. It thereby gives a precise criterion for when and to what extent such an approach is justified.

* Corresponding author.
E-mail address: hoffmann@physik.tu-chemnitz.de (K.H. Hoffmann).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.04.027

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

2. Relaxation dynamics in complex systems

As an example, consider a complex system characterized by an energy landscape with many local minima [10]. At sufficiently low temperatures, the relaxation within the basin of one local minimum is fast compared to the time scale on which the states within this basin communicate with states outside this basin. However, due to energetic barriers, it can take a very long time to reach the completely relaxed Boltzmann distribution. For such examples, frequent use is made of dynamics in which the states within one basin are replaced by a single mesostate. Such an approach leads to levels of complexity which allow an explicit treatment. Typical examples include cluster dynamics [11], the aging behavior of spin glasses [12,13], proteins [14], and solid state chemistry [15].

In these problems the dynamics is described by the vector of probabilities $p = (p_1, p_2, \ldots, p_l)^T$, whose components $p_i(t)$ represent the probability that the system is in state i at time t. The dynamics is given in terms of a discrete-time master equation

$$p(t+1) = Gp(t) \tag{1}$$

where G is the transition probability matrix. In the following we assume that the matrix G satisfies detailed balance

$$G_{ij}s_j = G_{ji}s_i \tag{2}$$

with the Boltzmann distribution s as its unique stationary distribution. Such an assumption is often used (see for instance [16] and references therein) and guarantees that G can be symmetrized, a property needed in the derivation below. We here refrain from discussing possible generalizations.

To analyze the quantitative errors made by a mesoscopic model, we compare its results with the results of a full microscopic description. Technically the mesoscopic model will also rely on a master equation dynamics, albeit in a much lower dimensional space. The transition from the microscopic to the mesoscopic is achieved by coarse graining the sets of states which are close to their internal equilibrium distribution. Often such sets of states can be found empirically by studying the time evolution of the probability distribution [17,18]. In the mathematical literature the coarse graining procedure is also known as lumping [19], aggregation [20], or projection [21]. The formal procedure is briefly described in the next section.

3. Coarse graining

Coarse graining the state space is performed by lumping together a number of states. Let Ω be the state space of size I, and let $L \subset \Omega$ be the subset of size m which is to be lumped—the lumped set. Here we consider only lumping a single set to facilitate the exposition. The general case has more complicated formulas, but our bounds are equally valid also for this case. Without loss of generality we number the states such that the elements of L are the first m elements of Ω.

When we coarse grain the elements of L into a single mesostate $\mathcal{L}$ we consider transitions into $\mathcal{L}$ to have occurred whenever the underlying dynamics makes a transition to any state in L. Thus the transition probability from any state to $\mathcal{L}$ is the sum of the transition probabilities from that state to the various states of L. Technically, this is achieved by applying the (collecting) matrix C with

$$C_{ij} = \begin{cases} 1 & \text{if} \quad i = 1,\ j \in L \\ 1 & \text{if} \quad i > 1,\ j = i + m \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

As an illustration, consider a four state system with $L = \{1, 2\}$. For that example,

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

The matrix C also serves to specify the lumped counterpart $\hat{p} = Cp$ of any probability distribution p.

The transition probabilities out of $\mathcal{L}$ need a bit more care. Recalling that our approach is predicated on the probabilities of states in L being approximately locally equilibrated, we average the exiting transition probabilities to any state outside L using a weighted average over transitions out of states in L. The weights in the average are given by the stationary distribution restricted to L. This is achieved with the (distributing) matrix D with

$$D_{ij} = \begin{cases} \tilde{s}_i & \text{if} \quad i \in L,\ j = 1 \\ 1 & \text{if} \quad i > m,\ i = j + m \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\tilde{s}_i = s_i / \sum_{j=1}^{m} s_j$ is the stationary distribution renormalized such that its entries add to one in the lumped set. For our four state example,

$$D = \begin{pmatrix} \tilde{s}_1 & 0 & 0 \\ \tilde{s}_2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6}$$

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

The lumped dynamics then proceeds with transition matrix [22] $\hat{G} = CGD$ which acts on a probability distribution $q(t)$ on the lumped states according to the lumped master equation

$$q(t+1) = \hat{G}q(t). \tag{7}$$

Note that in general $q(t) \neq \hat{p}(t)$. This master equation has the distribution $\hat{s} = Cs$ as its stationary distribution [22,23]. In order to prove this we find by direct calculation that

$$DC = \begin{pmatrix} \tilde{s}_1 & \tilde{s}_1 & 0 & 0 \\ \tilde{s}_2 & \tilde{s}_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{8}$$

Thus the action of DC on the stationary distribution is given by

$$DCs = s. \tag{9}$$

Multiplying (9) by CG we have

$$CGDCs = CGs \tag{10}$$

$$\hat{G}Cs = Cs \tag{11}$$

where we used the definition of $\hat{G}$ on the left side and the fact that s is the stationary distribution of G on the right hand side. Thus $\hat{s} = Cs$ is the stationary distribution of $\hat{G}$.

We assume that G is indecomposable, else we could do two (or more) separate problems. This assumption along with detailed balance guarantees that s is the unique stationary distribution of G and it then follows that $\hat{s}$ is the unique stationary distribution of $\hat{G}$, which also satisfies detailed balance [24].

The t step dynamics for the lumped master equation does however not mimic the true dynamics of the system, i.e., $\hat{G}^t = (CGD)^t$ is not close in general to CG^tD. The underlying reason is of course that the elements in L are not sufficiently equilibrated. As discussed later giving the system more time to establish a partial equilibrium is helpful in that respect, and thus we introduce the long time transitions matrix

$$A = G^N \tag{12}$$

and rescale our time to proceed in units of N so

$$p(n+1) = Ap(n). \tag{13}$$

4. Operator norms

We seek conditions which guarantee that the coarse graining of the state space does not introduce overly large errors in the subsequent dynamics. To measure the error in a way that applies to arbitrary initial states, we use an operator norm $\|\cdot\|$ applied directly to the time propagation operators. The operator norm we choose is induced by vector norms for probability distributions, both lumped and unlumped. Before introducing our norms, we review some relevant facts.

Fact 1. Given a vector space V and a norm $\|\cdot\|_a$ on V, the induced operator norm for operators $f : V \to V$ is defined by

$$\|f\|_a = \max_{v\in V, \|v\|_a \leq 1} \|fv\|_a. \tag{14}$$

The implication of this definition is that the induced norm $\|\cdot\|_a$ on operators is submultiplicative, i.e., it satisfies

$$\|fv\|_a \leq \|f\|_a \|v\|_a \tag{15}$$

and thus also

$$\|f_1 f_2\|_a \leq \|f_1\|_a \|f_2\|_a \tag{16}$$

where f_1 and f_2 are any two operators $V \to V$.

Fact 2. Given two vector spaces V and W with norms $\|\cdot\|_a$ and $\|\cdot\|_b$ respectively, the induced operator norm for operators $g : V \to W$ is defined by

$$\|g\|_{ab} = \max_{v\in V, \|v\|_a \leq 1} \|gv\|_b. \tag{17}$$

The implication of this definition is that the induced norm $\|\cdot\|_{ab}$ on operators is submultiplicative, i.e., it satisfies

$$\|gv\|_b \leq \|g\|_{ab} \|v\|_a \tag{18}$$

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

and thus also

$$\|gf\|_{ab} \leq \|g\|_{ab}\|f\|_a \tag{19}$$

where f and g are any two operators $V \to V$ and $V \to W$ respectively.

The norm $\|\cdot\|_s$ we choose to use on the space of (unlumped) probability distributions is defined by

$$\|p\|_s = \|Up\|_2 = \sqrt{\sum_i p_i^2/s_i}, \tag{20}$$

where $U = \mathrm{diag}(s_1^{1/2}, \ldots, s_l^{1/2})$. U is in fact the change of basis matrix which makes our transition matrix symmetric. The representation of our Markov chain in this basis and the subsequent use of this norm is rather frequent in various physical contexts [1,20]. The consequences of using this norm for probability distributions is that the induced operator norm becomes

$$\|A\|_s = \max_{\|v\|_s \leq 1} \|Av\|_s = \max_{\|Uv\|_2 \leq 1} \|UAU^{-1}Uv\|_2 \tag{21}$$

$$= \max_{\|w\|_2 \leq 1} \|UAU^{-1}w\|_2 = \|UAU^{-1}\|_2. \tag{22}$$

Thus for any operator A sending (unlumped) probability distributions to (unlumped) probability distributions with UAU^{-1} being symmetric,

$$\|A\|_s = \max_{\lambda \text{ an eigenvalue of } A} |\lambda|. \tag{23}$$

For our space of lumped distributions, we use the ordinary 2 norm. The resulting mixed norm $\|\cdot\|_{2s}$ satisfies

$$\|f\|_{2s} = \max_{\|v\|_s \leq 1} \|fv\|_2 \leq \max_{\|v\|_2 \leq 1} \|fv\|_2 = \|f\|_2, \tag{24}$$

where we have made use of the fact that $\|v\|_s \geq \|v\|_2$ and thus $\|v\|_s \leq 1$ implies that $\|v\|_2 \leq 1$.

Thus for example for the matrix product CA^3p, we have

$$\|CA^3p\| \leq \|C\|_{2s}\|A\|_s^3\|p\|_s \leq \|C\|_2\|A\|_s^3\|p\|_s. \tag{25}$$

5. Error bounds

Our theorem bounds the error in an n step process in terms of the error in a single step. Consider an initial distribution p_0. One step with the non-lumped dynamics is just Ap_0, but this lives in a higher dimensional space than $CADCp_0$, which is a single step with the lumped dynamics, applied to the lumped initial state Cp_0. In order to compare both we bring Ap_0 to the lumped space by using the collecting matrix C. Thus the error of one step in the lumped space is $\|CADCp_0 - CAp_0\| = \|C(ADC - A)p_0\|$. By omitting the last collecting step a comparison in the non-lumped space is obtained and thus in terms of the operator norm we introduce

$$\delta = \|ADC - A\|_s \equiv \|B - A\|_s \tag{26}$$

as the single-step lumping-error. We see that the matrix $ADC - A$ is a measure of the deviation between the dynamics represented by A and the dynamics which results by collecting and redistributing the total probability within L before applying the dynamic matrix A. Since the matrix $B \equiv ADC$ plays a prominent role in the derivation below, it is convenient to give it a name: the locally equilibrated transition probability matrix.

The error in the lumped space for an n-step process is

$$\|(CAD)^nCp_0 - CA^np_0\| = \|CADCAD\cdots DCp_0 - CA^np_0\|$$

$$= \|C((ADC)^n - A^n)p_0\|, \tag{27}$$

or equivalently in the non-lumped space and in terms of the matrix norm $\|(ADC)^n - A^n\|_s$, which is addressed as the n-step lumping-error. Technically our aim is now to prove

$$\|(ADC)^n - A^n\|_s \leq K(n)\delta \tag{28}$$

for some appropriately chosen $K(n)$.

To proceed further, we examine the general form of the difference in the nth power of two stochastic matrices A and B. The relevance to the calculation above is for $B = ADC$.

$$\|B^n - A^n\| \tag{29}$$

$$= \|(B - A)B^{n-1} + A(B^{n-1} - A^{n-1})\| \tag{30}$$

$$= \|(B - A)B^{n-1} + A((B - A)B^{n-2} + A(B^{n-2} - A^{n-2}))\| \tag{31}$$

$$= \|(B - A)B^{n-1} + A(B - A)B^{n-2} + A^2(B^{n-2} - A^{n-2})\|. \tag{32}$$

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

Iterating, we find

$$\|B^n - A^n\| = \left\| \sum_{k=0}^{n-1} A^k (B-A) B^{n-k-1} \right\| \tag{33}$$

$$\leq \sum_{k=0}^{n-1} \|A^k (B-A) B^{n-k-1}\| \tag{34}$$

$$= b_0(n). \tag{35}$$

While not really useful from a practical point of view, the bound b_0 is a good starting point for the more useful bounds we present below.

We now note that $Bs = s$ follows directly from (9) and $As = s$, and thus A and B share a common stationary distribution. The key factor for our proof is the representation for the matrices A and B which follows from that fact. To obtain this representation, consider the underlying probability space. This space can be decomposed into a one dimensional subspace spanned by the stationary distribution s and a complementary subspace

$$U = \left\{ x \;\middle|\; \sum_{i=1}^{I} x_i = (1, 1, \ldots, 1)x = 0 \right\}. \tag{36}$$

U is an invariant subspace of A. This follows from the fact that the columns of A sum to one, which can be interpreted as saying that the row vector $(1, 1, \ldots, 1)$ is a left eigenvector of A with eigenvalue one, i.e.,

$$(1, 1, \ldots, 1) = (1, 1, \ldots, 1)A. \tag{37}$$

Multiplying on the right by a vector $x \in U$, we see that

$$0 = (1, 1, \ldots, 1)x = (1, 1, \ldots, 1)Ax \tag{38}$$

and thus Ax is in U.

It follows that relative to a basis $\{s, v_2, \ldots, v_I\}$, where v_i are chosen from U, the matrix of A takes the form

$$A = K^{-1} \begin{pmatrix} 1 & 0 \\ 0 & A_\perp \end{pmatrix} K, \tag{39}$$

where K is the matrix facilitating the basis change, and $A_\perp$ is a matrix acting on the space spanned by $\{v_2, \ldots, v_I\}$. Since we used only the fact that A is a stochastic matrix with unique stationary distribution s, the same basis also puts B into the form

$$B = K^{-1} \begin{pmatrix} 1 & 0 \\ 0 & B_\perp \end{pmatrix} K. \tag{40}$$

Let

$$\bar{A}_\perp = K^{-1} \begin{pmatrix} 0 & 0 \\ 0 & A_\perp \end{pmatrix} K, \quad \bar{B}_\perp = K^{-1} \begin{pmatrix} 0 & 0 \\ 0 & B_\perp \end{pmatrix} K. \tag{41}$$

Matrix calculation then leads directly to

$$A^k (B-A) B^{n-k-1} \tag{42}$$

$$= K^{-1} \begin{pmatrix} 1 & 0 \\ 0 & A_\perp^k \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & B_\perp - A_\perp \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & B_\perp^{n-1-k} \end{pmatrix} K \tag{43}$$

$$= K^{-1} \begin{pmatrix} 0 & 0 \\ 0 & A_\perp^k (B_\perp - A_\perp) B_\perp^{n-1-k} \end{pmatrix} K \tag{44}$$

$$= \bar{A}_\perp^k (\bar{B}_\perp - \bar{A}_\perp) \bar{B}_\perp^{n-k-1}. \tag{45}$$

Note that due to (23) $\|\bar{A}_\perp\|_s = |\lambda_2|$, where λ_2 is the eigenvalue of A with second largest modulus. This is not so for $\|\bar{B}_\perp\|_s$, as $\bar{B}_\perp$ need not be symmetric. We can however bound $\|\bar{B}_\perp\|_s$ by $|\lambda_2|$ as follows. First we note that from $B - A = \bar{B}_\perp - \bar{A}_\perp$ follows $\bar{B}_\perp = B - (A - \bar{A}_\perp)$; furthermore, as $(A - \bar{A}_\perp)$ projects onto the subspace spanned by s, we have $(A - \bar{A}_\perp) = (A - \bar{A}_\perp)DC$. Then

$$\|\bar{B}_\perp\|_s = \|ADC - (A - \bar{A}_\perp)\|_s \tag{46}$$

$$= \|ADC - (A - \bar{A}_\perp)DC\|_s \tag{47}$$

$$= \|(A - A + \bar{A}_\perp)DC\|_s \tag{48}$$

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

$$= \|\bar{A}_\perp DC\|_s \tag{49}$$

$$\leq \|\bar{A}_\perp\|_s \|DC\|_s \tag{50}$$

$$= |\lambda_2|, \tag{51}$$

where we used that $UDCU^{-1}$ is symmetric with largest eigenvalue 1.

It follows that

$$\|A^k(B-A)B^{n-k-1}\|_s = \|\bar{A}_\perp^k(\bar{B}_\perp - \bar{A}_\perp)\bar{B}_\perp^{n-1-k}\|_s \leq |\lambda_2|^{n-1}\delta. \tag{52}$$

Using this and summing we have

$$b_0(n) \leq \sum_{k=0}^{n-1} \|\bar{A}_\perp\|_s^k \|(\bar{B}_\perp - \bar{A}_\perp)\|_s \|\bar{B}_\perp\|_s^{n-1-k} \tag{53}$$

$$= \sum_{k=0}^{n-1} \|\bar{A}_\perp\|_s^k \|\bar{B}_\perp\|_s^{n-1-k}\, \delta \equiv b_1(n) \tag{54}$$

$$\leq |\lambda_2|^{n-1} n\, \delta \equiv b_2(n). \tag{55}$$

Finally we note that one can obtain an n-independent bound from (53) by calculating the maximum of (55) as a function of n:

$$b_3 = \frac{-1}{|\lambda_2| e \ln(|\lambda_2|)}\, \delta, \tag{56}$$

where e is the base of the natural logarithm.

From a general point of view we expect that the above results should be transferable to continuous-time Markov processes, and indeed results similar in spirit have been already obtained [25].

6. Applications

The strength of the bounds introduced rests with the ability to find a small δ in the first place. While a thorough discussion would go beyond the scope of this presentation, we note that based on (26) $\delta = \|ADC - A\|_s = \|G^N DC - G^N\|_s$ can be made arbitrarily small by taking N sufficiently large. This is due to the fact that $G^N DC$ and G^N have the same stationary distribution. In many applications there are however subsets in state space which are close to local equilibrium at a much earlier time. Coarse graining these will lead to small δ's. A careful study of the eigenvectors of G would provide the necessary information needed, but the art of clever coarse graining might as well rest on a physical intuition of the relaxation process. For instance, in the relaxation dynamics on energy landscapes, the basins around local minima are good candidates for coarse graining.

6.1. Example: energy landscapes

To illustrate coarse graining and the resulting errors in the dynamics we present a caricature example of the thermal relaxation of a complex system. The relaxation dynamics is described by a Markov process with Metropolis transition rates. The energy landscape on which the relaxation takes place consists of only 4 states, which are connected as a line. States 1 and 2 are one local minimum, and state 4 another. These state space regions (basins) are dynamically separated by state 3, which represents an energy barrier. Its height is 1 in units of $k_B T$ viewed from state 4 and 2 viewed from state 2. This becomes apparent in the transition probability matrix

$$G = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5(1-e^{-2}) & 0.5 & 0 \\ 0 & 0.5e^{-2} & 0 & 0.5e^{-1} \\ 0 & 0 & 0.5 & 1-0.5e^{-1} \end{pmatrix}. \tag{57}$$

Matrix G has eigenvalues $(1., 0.8989, -0.2078, 0.0573)$ in decreasing order of their absolute size. In this example we expect that due to the energy barrier at state 3 states $(1, 2)$ will equilibrate faster than any other pair of states. This is made evident in Table 1, which shows the initial δ for different sets L and for $A = G^5$ and for $A = G$. Consequently, we lump states $(1, 2)$.

The resulting collecting matrix is the one given in (4). The stationary distribution of G is $s = (0.3995, 0.3995, 0.0541, 0.1470)^T$, which leads to the distributing matrix

$$D = \begin{pmatrix} 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{58}$$

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

Table 1
Our bound for the error in coarse grained dynamics is proportional to the initial error δ. The table shows values of δ for different choices of the set L to be lumped and for different time scales. The times scales are set by using the dynamical matrix $A = G^N$. The effect of choosing $N = 1$ and $N = 5$ is shown.

set L	{1, 2}	{1, 3}	{1, 4}	{2, 3}	{2, 4}	{3, 4}
$G^5\delta$	0.0286	0.200	0.539	0.186	0.518	0.145
$G^1\delta$	0.1390	0.333	0.830	0.340	0.794	0.276

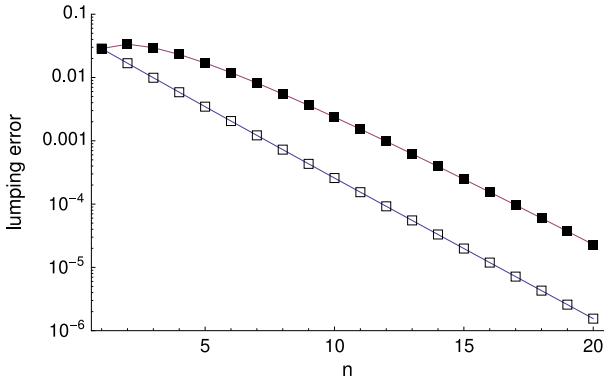

Fig. 1. The n-step lumping-error (open squares), i.e. the deviation between full and coarse grained descriptions, as well as bound b_2 (full squares) are shown as a function of time. Note that the bound increases only in the second step but decreases thereafter. The two curves are asymptotically parallel indicating that the bound follows the deviation with a fixed ratio.

Using matrices C and D we can lump $A = G$

$$CAD = \begin{pmatrix} 0.9662 & 0.5 & 0 \\ 0.0338 & 0 & 0.1839 \\ 0 & 0.5 & 0.8161 \end{pmatrix} \tag{59}$$

which has eigenvalues $(1., 0.8967, -0.1145)$. We also determine the locally equilibrated transition probability matrix

$$B = ADC = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.4662 & 0.4662 & 0.5 & 0 \\ 0.0338 & 0.0338 & 0 & 0.1839 \\ 0 & 0 & 0.5 & 0.8161 \end{pmatrix} \tag{60}$$

which has the eigenvalues $(1., 0.8967, -0.1145, 0)$.

While for $A = G$ the corresponding $\delta = 0.139$ is quite large, it is much smaller for $A = G^5$ as explained above. Thus in the following we use $A = G^5$ in our example system and we compute the quantity $\|(A^n - B^n)\|_s$ as well as its bound $b_2(n)$. Fig. 1 shows these as a function of time n.

For longer times, bound $b_2(n)$ decreases with n, even though the initial increase leads to some loss in bounding efficiency. In Fig. 1 one sees that both bounds asymptotically decay with the second largest eigenvalue of A. For the initial error of 0.0286 the bounds are highly efficient. Already at the second step bound $b_2(2)$ uniformly limits the maximum subsequent error to less than 0.035. We note that $b_3 = \delta * 1.1762 = 0.0337$, thus the errors are limited to about a 20% increase against the initial error.

6.2. Example: random matrices

In order to illustrate the strength of our bounds we present an application to random transition matrices, for which we analyzed the error due to coarse graining.

We created random matrices of dimension 20 with elements chosen to be either zero with probability 1/2 or a random number from a uniform distribution on [0, 1]. These matrices were symmetrized by adding their respective transpose, and thereafter they were column-normalized. For these matrices the first five states were lumped. We studied ensembles of 50 matrices and determined the mean value of $\|A^n - B^n\|$ as well as of the bounds and the respective standard deviations. Fig. 2 shows the deviation between lumped and non-lumped description and bound $b_2(n)$ as a function of time n. The surprising result is that already for $n = 2$ the deviation is smaller than the initial value, which shows the strength of the bound obtained. Table 2 gives the error bars on the average values shown in Fig. 2.

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

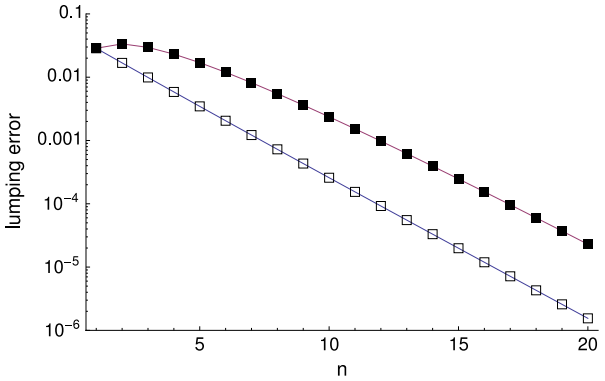

Fig. 2. This figure shows the results for our computation of random 20×20 transition probability matrices. The deviation between full and coarse grained descriptions (n-step lumping-error, open squares) as well as bound b_2 (full squares) are displayed as a function of time. We lump 5 random states out of the 20 states. Interestingly, the bound decreases monotonically and follows the deviation between full and coarse grained dynamics closely.

Table 2
Mean values and standard deviations of the data shown in Fig. 2.

n	$\|A^n - B^n\|_s$	$b_2(n)$
2	0.0936 ± 0.0184	0.199 ± 0.0311
4	0.0142 ± 0.00483	0.0583 ± 0.0167
6	0.00227 ± 0.00116	0.0131 ± 0.00569
8	0.000369 ± 0.000253	0.00267 ± 0.00159

7. Conclusions

We conclude by noting the implications of our calculations for the original problem with stochastic matrix G. Since $A = G^N$, the larger we take N, the smaller our initial deviation δ. This is further helped by the fact that

$$\lambda_2(A) = \lambda_2(G)^N. \tag{61}$$

Thus the larger we choose N, the smaller the value of both δ and $\lambda_2 = \lambda_2(A)$ in (55) and (56). Thus we have shown that on sufficiently large time scales, the error in coarse graining a master equation can be made arbitrarily small. Although our discussion was entirely in terms of a single lump of states, the calculation is valid *mutatis mutandis* for simultaneously lumping many groups of states. This merely makes the notation in the definitions of the collecting and distributing operators C and D more complicated.

While the above examples provide already considerable insight in the strength of the bounds discussed, we note that under certain circumstances it can happen, that A and ADC have the same second largest eigenvalue and thus $ADC - A$ lies in an even smaller subspace such that the spectral norms of A and ADC restricted to this subspace become even smaller.

Of course, the question remains open, whether even more stringent bounds can be obtained. It seems to us that especially the problem under which circumstances small initial errors δ can be expected is an important next step, as is the analysis of how to select good candidates for subsets to be lumped.

References

[1] V.N.G Kampen, Stochastic Processes in Physics and Chemistry, 3rd ed., North-Holland, North-Holland Personal Library, 2007, E-Book; URL http://www.sciencedirect.com/science/book/9780444529657.
[2] I. Oppenheim, K.E. Shuler, G.H. Weiss, Stochastic Processes in Chemical Physics: The Master Equation, MIT Press, Cambridge, Mass, 1977.
[3] G.G. Yin, Q. Zhang, Discrete-time Markov Chains: Two-time-scale Mehtods and Applications, in: Applications of Mathematics, vol. 55, Springer, New York, 2005.
[4] F.S. Hillier, G.J. Liebermann, Introduction to Operations Research, 8th ed., McGraw-Hill Higher Eduction, Boston, 2005.
[5] H.P. Young, Social dynamics: theory and applications, in: L. Tesfatsion (Ed.), Handbook of Computationsl Economics, vol. II, North-Holland, Amsterdam, 2006, pp. 1081–1108. chap. 22.
[6] B.L. Hammond, W.A.J. Lester, P.J. Reynolds, Monte Carlo Methods in ab initio Quantum Chemistry, in: World Scientific Lecture and Course Notes in Chemistry, World Scientific, Singapore, 1994.
[7] H. Caswell, Matrix Population Models: Construction, Analysis, and Interpretation, 2nd ed., Sinauer Associates, Sunderland, MA, 2001.
[8] K.S. Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, Prentice-Hall, Englewood Cliffs, NJ, 1982.
[9] K.H. Hoffmann, P. Salamon, Bounding the lumping error in Markov chain dynamics, Appl. Math. Lett. 22 (2009) 1471–1475. doi:10.1016/j.aml.2009.03.016.
[10] K.H. Hoffmann, P. Sibani, Diffusion in hierarchies, Phys. Rev. A 38 (8) (1988) 4261–4270.
[11] K.D. Ball, R.S. Berry, R.E. Kunz, F.Y. Li, A. Proykova, D.J. Wales, From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters, Science 271 (5251) (1996) 963–966.
[12] P. Sibani, K.H. Hoffmann, Hierarchical models for aging and relaxation of spin glasses, Phys. Rev. Lett. 63 (26) (1989) 2853–2856.
[13] K.H. Hoffmann, P. Sibani, Relaxation and aging in spin glasses and other complex systems, Z. Phys. B 80 (1990) 429–438.

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027

[14] H. Frauenfelder, S.G. Sligar, P.G. Wolynes, The energy landscapes and motions of proteins, Science 254 (5038) (1991) 1598–1603.
[15] J.C. Schön, P. Sibani, Properties of the energy landscape of network models for covalent glasses, J. Phys. A: Math. Gen. 31 (1998) 8165–8178.
[16] A.Y. Mitrophanov, The spectral gap and perturbation bounds for reversible continuous-time Markov chains, J. Appl. Probab. 41 (2004) 1219–1222.
[17] P. Sibani, J.C. Schön, P. Salamon, J.O. Andersson, Emergent hierarchical structures in complex system dynamics. A phase space analysis of the travelling salesman problem, Europhys. Lett. 22 (1993) 479–485.
[18] C. Uhlig, K.H. Hoffmann, P. Sibani, Relaxation in self similar hierarchies, Z. Phys. B 96 (1995) 409–416.
[19] J.G. Kemeney, J.L. Snell, Finite Markov Chains, Springer Verlag, New York, 1976.
[20] G. Ruppeiner, Riemannian geometry in thermodynamic fluctuation theory, Rev. Modern Phys. 67 (3) (1995) 605–659. doi:10.1103/RevModPhys.67.605.
[21] M. Jerrum, J.B. Son, P. Tetali, E. Vigoda, Elementary bonds on Poincaré and log-Sobolev constants for decomposable Markov chains, Ann. Appl. Probab. 14 (4) (2004) 1741–1765.
[22] C.J. Burke, M. Rosenblatt, A markovian function of a Markov chain, Ann. Math. Statist. 29 (4) (1958) 1112–1122.
[23] B. Andresen, K.H. Hoffmann, K. Mosegaard, J. Nulton, J.M. Pedersen, P. Salamon, On lumped models for thermodynamic properties of simulated annealing problems, J. Phys. 49 (1988) 1485–1492.
[24] R. Grone, K.H. Hoffmann, P. Salamon, An interlacing theorem for reversible markov chains, J. Phys. A: Math. Gen. 41 (2008) 1–7.
[25] A.Y. Mitrophanov, A. Lomsadze, M. Borodosky, Sensivitvity of hidden Markov models, J. Appl. Probab. 42 (2005) 632–642.

Please cite this article in press as: K.H. Hoffmann, P. Salamon, Accuracy of coarse grained Markovian dynamics, Physica A (2011), doi:10.1016/j.physa.2011.04.027