Bounding the Lumping Error in Markov Chain Dynamics

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

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

[*]hoffmann@physik.tu-chemnitz.de

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

[†] salamon@math.sdsu.edu

February 5, 2008

Abstract

Forming lumped states in a Markov chain is a very useful device leading to a coarser level of description. The Markov chain on these lumped states is often taken as an approximation for the time evolution of the unlumped chain. In the present work we derive a bound on the error in this approximation.

Keywords: Markov chain, coarse graining error, state space dynamics,

1 Introduction

Markov chains are a standard modeling tool used in applications ranging from business [1,2] and sociology [3] to chemistry [4], physics [5,6], biology [7], computer science [8], and statistics [9]. The models in which such chains occur involve discrete stochastic dynamics represented by a matrix of transition probabilities. It is often convenient to partition the states of a chain into aggregates or lumps and to view the dynamics at a coarser level as the system of interest moves among the lumps. Examples of such aggregation include a weather model lumping the states "drizzle", "rain" and "snow" into one state called "precipitation" or a physical model aggregating the microstates of a physical system into so-called coarse grained "mesostates" each representing many microstates. Although the dynamics moving between the lumps is not even Markovian in general [10], there is a natural choice for a Markov chain model on the set of lumped states. This choice [10–13] matches the time evolution of the original unlumped chain started at equilibrium. In the present paper we bound the error of the dynamics predicted by this lumped chain considered as a model of the unlumped chain. Our goal in this letter is to analyze the accuracy of such coarse grained models as compared to the exact microscopic behavior, i.e., to bound the error in a coarse grained description.

2 Lumped Markov Chains

Let Ω be a finite set of states, and let p be a probability distribution on Ω viewed as a vector $p = (p_1, p_2, \ldots, p_N)$ with non-negative entries, where p_x is the probability that the system is in state $x \in \Omega$. The dynamics is given by

$$p(n+1) = p(n)G, \tag{1}$$

where G is the transition probability matrix and n denotes the (discrete) time. In the present work, we restrict our attention to Markov chains that are regular, time-homogeneous, and reversible. The time-homogeneous assumption says that G does not depend on n. Regular means that there exists an n_0 such that G^{n_0} has all positive entries and this in turn guarantees that G has a unique stationary distribution π such that

$$\pi = \pi G. \tag{2}$$

Reversible ensures that

$$\pi_x G_{xy} = \pi_y G_{yx} \qquad \forall x, y \in \Omega. \tag{3}$$

Our goal is to analyze the quantitative errors made by a lumped Markov chain model [10] compared with the unlumped chain. Let $\{L_j\}, j = 1, \ldots, M$ be a partition of Ω. We introduce the collecting matrix C with

$$C_{xj} = \begin{cases} 1 & \text{if} \quad x \in L_j \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

The matrix C serves to specify the lumped probability distribution $\hat{p} = pC$ on $\mathcal{O} = \{1, 2, \ldots, M\}$ corresponding to a distribution p on Ω. C collects the probability of the states belonging to each of the lumped states j. We also introduce the distributing matrix D with

$$D_{jx} = \begin{cases} \tilde{\pi}_x^j & \text{if} \quad x \in L_j \\ 0 & \text{otherwise} \end{cases} , \tag{5}$$

where $\tilde{\pi}_x^j = \pi_x / \sum_{x \in L_j} \pi_x$ is the stationary distribution restricted to L_j and renormalized such that its entries add to one. Note that for any probability distribution q in the lumped space, its image qD is locally equilibrated, i.e., qD restricted to any element L_j of our partition equals the stationary distribution π restricted to L_j and renormalized to add up to q_j. With $\hat{\pi} = \pi C$ we find

$$D_{jx} = C_{xj} \frac{\pi_x}{\hat{\pi}_j} \qquad \forall x \in \Omega, j \in \mathcal{O}. \tag{6}$$

Next we consider the local equilibration operator CD. The name is justified by considering its action. Starting from any distribution p on Ω, CD collects the probability in each L_j and then redistributes this much probability $(pC)_j$ among the states $x \in L_j$ as $(pC)_j \tilde{\pi}_x^j$. In particular,

$$\pi = \pi C D. \tag{7}$$

On the other hand $DC = I$, since C collects back exactly what D distributes.

Following [14], we now introduce the lumped dynamics with transition matrix $\hat{G} = DGC$. This choice of the transition matrix together with the choice of D insures that the distribution $\hat{\pi} = \pi C$ is the stationary distribution of $\hat{G}$ since multiplying (7) by GC we have

$$\pi C D G C = \pi G C \tag{8}$$

$$\pi C \hat{G} = \pi C. \tag{9}$$

We note that $\hat{G}$ is also reversible and thus $\hat{\pi}$ is its unique stationary distribution.

3 Bounding Coarse Graining Errors

Starting from a distribution p_0 on Ω, its time evolution among the aggregates L_j with the unlumped dynamics is $\hat{p}(t) = p_0 G^t C$ while its time evolution with the lumped dynamics is $q(t) = p_0 C \hat{G}^t$. The main question considered here is how different these two dynamics can be. For one time step the difference is

$$\begin{aligned} \|p_0 C \hat{G} - p_0 G C\| &= \|p_0 C D G C - p_0 G C\| \\ &= \|p_0 (C D G - G) C\|, \end{aligned} \tag{10}$$

where $\|\cdot\|$ is an as yet unspecified vector norm. The difference for n steps can be similarly rearranged to give

$$\begin{aligned} \|p_0 C \hat{G}^n - p_0 G^n C\| &= \|p_0 C (DGC)^n - p_0 G^n C\| \\ &= \|p_0 C D G C D \cdots G C D G C - p_0 G^n C\| \\ &= \|p_0 (CDG)^n C - p_0 G^n C\| \\ &= \|p_0 ((CDG)^n - G^n) C\|. \end{aligned} \tag{11}$$

We see from (10) and (11) the prominent role of the operator CDG. While many authors [12–14] have considered the lumped matrix $\hat{G}$, specifying the lumped dynamics on the aggregates $\{L_1, \ldots, L_M\}$, less attention has been paid to the matrix $CDG \equiv H$ which specifies the lumped dynamics on the original states Ω.

Our goal will be to bound the n-step difference $\|H^n - G^n\|$ in terms of the one step difference $\|H - G\|$. Note that $\|\cdot\|$ is here the induced matrix norm based on a vector norm to be specified. The general behavior of the n-step difference is that it can grow for a while, but must eventually decline to zero as, by construction, both the lumped and the unlumped chain converge to the same equilibrium distribution. Bounding how large the difference can get along the way in terms of the one step difference is the content of our

Theorem: Let G be a transition probability matrix, and let C and D be collecting and distributing matrices as introduced above. Define

$$\delta = \|CDG - G\|. \tag{12}$$

Then

$$\|(CDG)^n - G^n\| < K(n)\delta, \tag{13}$$

with $K(n)$ being a suitably chosen, bounded function of n, which will be specified in the proof below.

Proof: We start by noting that

$$\|H^n - G^n\| = \|(H - G)H^{n-1} + G(H^{n-1} - G^{n-1})\|. \tag{14}$$

Iterating, we find

$$\begin{aligned} \|H^n - G^n\| &= \|\sum_{k=0}^{n-1} G^k(H - G)H^{n-k-1}\| && (15)\\ &\leq \sum_{k=0}^{n-1} \|G^k(H - G)H^{n-k-1}\| && (16) \end{aligned}$$

To proceed further, we make use of the fact that both H and G have the common stationary distribution π. Accordingly, we define the projection $P_\pi = e^T\pi$, where e is the vector $e = (1, 1, \ldots, 1) \in I\!R^N$. The complimentary projection $P_\sigma = I - P_\pi$. P_σ projects any vector onto the subspace $\Sigma = \{v \mid ve^T = 0\}$. Σ is invariant under the action of any stochastic matrix. We end up with the following representation of G and H,

$$G = (P_\pi + P_\sigma)G = P_\pi + P_\sigma G, \tag{17}$$

and

$$H = (P_\pi + P_\sigma)H = P_\pi + P_\sigma H. \tag{18}$$

From (17) and (18) we obtain $H - G = P_\sigma H - P_\sigma G$, and $P_\pi P_\sigma G = P_\sigma G P_\pi = P_\pi P_\sigma H = P_\sigma H P_\pi = 0$. We note $\|CDG - G\| = \|P_\sigma H - P_\sigma G\| = \delta$. Then for integers k, n

$$\begin{aligned} & \|G^k(H - G)H^{n-k-1}\| && (19)\\ = \; & \|(P_\pi + P_\sigma G)^k(P_\sigma H - P_\sigma G)(P_\pi + P_\sigma H)^{n-1-k}\| && (20)\\ = \; & \|(P_\sigma G)^k(P_\sigma H - P_\sigma G)(P_\sigma H)^{n-1-k}\| && (21)\\ \leq \; & \|P_\sigma G\|^k\delta\|P_\sigma H\|^{n-k-1}. && (22) \end{aligned}$$

Combining (11), (16) and (22) we thus find

$$\begin{aligned} \|(CDG)^n - G^n\| &\leq \sum_{k=0}^{n-1} \|G^k(H - G)H^{n-k-1}\| \\ &\leq \sum_{k=0}^{n-1} \|P_\sigma G\|^k\delta\|P_\sigma H\|^{n-k-1}. \end{aligned} \tag{23}$$

So far we have not specified our choice of norm $\|\cdot\|$; our manipulations are valid for any submultiplicative matrix norm. We now choose $\|\cdot\|$ to be $\|\cdot\|_\pi$ with

$$\|v\|_\pi = \|vU_\pi\|_2 = \sqrt{\sum_x v_x^2/\pi_x}, \tag{24}$$

where U_π =diag$(1/\sqrt{\pi_1}, 1/\sqrt{\pi_2}, \ldots, 1/\sqrt{\pi_N})$ and $\|\cdot\|_2$ is the 2-norm. The corresponding operator norm is

$$\begin{aligned} \|A\|_\pi &= \max_{\|v\|_\pi=1} \|vA\|_\pi = \max_{\|vU_\pi\|_2=1} \|vU_\pi U_\pi^{-1} A U_\pi\|_2 & (25)\\ &= \max_{\|w\|_2=1} \|wU_\pi^{-1} A U_\pi\|_2 = \|U_\pi^{-1} A U_\pi\|_2 & (26) \end{aligned}$$

By reversibility, (3), $U_\pi^{-1} G U_\pi$ is a symmetric matrix. Therefore, its 2-norm is its dominant eigenvalue. In fact, as shown below, the matrix U_π also symmetrizes P_π, $P_\pi G$, $P_\sigma G$, and CD, and thus their norm also equals their dominant eigenvalue.

To start with, a direct calculation shows $\pi U_\pi = e U_\pi^{-1}$. Thus we have $U_\pi^{-1} P_\pi U_\pi = U_\pi^{-1} e^T \pi U_\pi = (eU_\pi^{-1})^T (\pi U_\pi) = (\pi U_\pi)^T (\pi U_\pi) = ((\pi U_\pi)^T (\pi U_\pi))^T = (U_\pi^{-1} P_\pi U_\pi)^T$ which proves that P_π is symmetrized by U_π.

So $U_\pi^{-1} G U_\pi$ and $U_\pi^{-1} P_\pi U_\pi$ are symmetric and commute with each other as G and P_π commute. Thus $U_\pi^{-1} P_\pi G U_\pi = U_\pi^{-1} P_\pi U_\pi U_\pi^{-1} G U_\pi$ is symmetric.

It follows that $U_\pi^{-1} P_\sigma G U_\pi = U_\pi^{-1}(I - P_\pi) G U_\pi$ is symmetric as well.

Finally, from (6) we have $D = U_{\hat\pi}^2 C^T U_\pi^{-2}$ or $U_{\hat\pi}^{-1} D U_\pi = U_{\hat\pi} C^T U_\pi^{-1}$ and thus

$$\begin{aligned} U_\pi^{-1} C D U_\pi &= (U_\pi^{-1} C U_{\hat\pi})(U_{\hat\pi}^{-1} D U_\pi) & (27)\\ &= (U_\pi^{-1} C U_{\hat\pi})(U_{\hat\pi} C^T U_\pi^{-1}) & (28)\\ &= (U_\pi^{-1} C U_{\hat\pi})(U_\pi^{-1} C U_{\hat\pi})^T. & (29) \end{aligned}$$

As the rhs is a matrix times its transpose, we conclude that CD is symmetrized by U_π.

Since $U_\pi^{-1} P_\sigma G U_\pi$ is symmetric and based on the definition of the norm we have

$$\|P_\sigma G\|_\pi = |\lambda_2|, \tag{30}$$

where λ_2 is the second largest eigenvalue of G. To bound $\|P_\sigma H\|_\pi$, we proceed as follows:

$$\begin{aligned} \|P_\sigma H\|_\pi &= \|CDG - P_\pi\|_\pi = \|CDG - CDP_\pi\|_\pi & (31)\\ &= \|CD(G - P_\pi)\|_\pi \le \|P_\sigma G\|_\pi \|CD\|_\pi = |\lambda_2| \|CD\|_\pi. & (32) \end{aligned}$$

Now noting that CD is a stochastic matrix and U_π symmetrizes CD, we conclude that $\|CD\|_\pi = 1$.

Thus our bound becomes

$$\|H^n - G^n\|_\pi \le n|\lambda_2|^{n-1}\delta \tag{33}$$

and we take $K(n) = n|\lambda_2|^{n-1}$. $K(n)$ can be bound from above by maximizing over n to give

$$K(n) \le \frac{-1}{\lambda e \ln(\lambda)}, \tag{34}$$

where e is the base of the natural logarithm, giving an n-independent bound for the deviation of our lumped dynamics and completing our proof. □

4 Conclusion

Above we presented the first quantitative bound on the errors made by using a lumped Markov chain instead of the unlumped chain. The bounds presented depend on the second largest eigenvalue of the transition probability matrix. The important finding is that indeed in a mesoscopic description the deviations from a

microscopic description can be uniformly bounded in terms of the deviation between the two descriptions in one time step. We find that the deviations decay as a power of the second largest eigenvalue of the transition matrix. This is to be expected for any asymptotic bound since the asymptotic convergence of the chain to its equilibrium must go as powers of the modulus of the second largest eigenvalue, known in this context as the coefficient of ergodicity [15]. The achievement of the present paper derives from the fact that our bounds hold for all time and are not just asymptotic.

The main technique making our bound possible concerned the use of a carefully chosen norm, which induced operator norms that are given by eigenvalues of our matrices. The technique depends on the fact that for reversible Markov chains the transition matrices are diagonally similar to symmetric matrices and thus we can find a norm that equals the dominant eigenvalue. While it is generally possible to choose matrix norms that come arbitrarily close to the dominant eigenvalue [15], it is not clear whether our arguments would carry through for Markov chains not satisfying the reversibility condition since different norms would be needed for the different operators in our expansion.

References

[1] G. George Yin and Qing Zhang. *Discrete-time Markov chains: two-time-scale methods and applications*, volume 55 of *Applications of mathematics*. Springer, New York, 2005.

[2] Frederick S. Hillier and Gerald J. Lieberman. *Introduction to Operations Research*. McGraw-Hill Higher Education, Boston, 8. edition, 2005.

[3] H. Peyton Young. Social dynamics: Theory and applications. In Leigh Tesfatsion and Kenneth L. Judd, editors, *Handbook of Computational Economics*, volume II, chapter 22, pages 1081–1108. North Holland, Amsterdam, 2006.

[4] B. L. Hammond, William A. Lester, Jr., and Peter J. Reynolds. *Monte Carlo Methods in Ab Initio Quantum Chemistry*. World Scientific Lecture and Course Notes in Chemistry. World Scientific, Singapore, 1994.

[5] Nicolaas G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam, 2. edition, 1992.

[6] Irwin Oppenheim, Kurt Egon Shuler, and George Herbert Weiss. *Stochastic Processes in Chemical Physics: The Master Equation*. MIT Press, Cambridge, Mass., 1977.

[7] Hal Caswell. *Matrix population models: construction, analysis, and interpretation*. Sinauer, Sunderland, Mass., 2. edition, 2001.

[8] Kishor Shridharbhai Trivedi. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[9] W. R. Gilks, S. Richardson, and David Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[10] John George Kemeny and James Laurie Snell, editors. *Finite Markov chains*. Van Nostrand, Princeton, NJ, 1960.

[11] B. Andresen, K.H. Hoffmann, K. Mosegaard, J. Nulton, J.M. Pedersen, and P. Salamon. On lumped models for thermodynamic properties of simulated annealing problems. *J. Phys. France*, 49:1485–1492, 1988.

[12] Konstantin E. Avrachenkov and Moshe Haviv. The first Laurent series coefficients for singularly perturbed stochastic matrices. *Linear Algebra Appl.*, 386:243–259, 2004.

[13] Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *The Annals of Applied Probability*, 14(4):1741–1765, 2004.

[14] C. J. Burke and M. Rosenblatt. A Markovian function of a Markov chain. *Annals of Mathematical Statistics*, 29(4):1112–1122, 1958.

[15] Uriel G. Rothblum and Choon Peng Tan. Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices. *Linear Algebra Appl.*, 66:45–86, 1985.