Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

FIETE HAACK

Properties of Reduced Reversible Markov Chains

ZIB-Report 08-42 (October 2008)

Properties of Reduced Reversible Markov Chains

Fiete Haack

November 13, 2008

Abstract

The enormous time lag between fast atomic motion and complex protein folding events makes it almost impossible to compute molecular dynamics on a high resolution. A common way to tackle this problem is to model the system dynamics as a Markov process. Yet for large molecular systems the resulting Markov chains can hardly be handled due to the curse of dimensionality. Coarse graining methods can be used to reduce the dimension of a Markov chain, but it is still unclear how far the coarse grained Markov chain resembles the original system. In order to answer this question, two different coarse-graining methods were analysed and compared: a classical set-based reduction method and an alternative subspace-based approach, which is based on membership vectors instead of sets. On the basis of a small toy system, it could be shown, that in contrast to the subset-based approach, the subspace-based reduction method preserves the Markov property as well as the essential dynamics of the original system.

AMS MSC 2000: 62-07, 60J22, 65C40

Keywords: Markov chain, PCCA+, Conformation dynamics

Contents

1 Introduction

Atomic motion is extremely fast. Thermal vibrations like rotation, oscillation or fluctuation have a timescale of only a few femtoseconds. On the contrary, most essential dynamics, like complex protein-folding or protein-ligand binding processes, have a timescale of several seconds. The resulting time lag makes it almost impossible to simulate molecular motion with high resolution. Therefore a model is needed, which allows the simulation of molecular motion on a coarse level without loosing the essential dynamics.

1.1 Molecular dynamics as stochastic process

Most molecules have metastable conformations, i.e. on large scales the molecule has the same geometric structure, whereas on small scales the system may rotate, oscillate or fluctuate. The essential dynamic of most molecules can therefore be described as a jump process with the molecule staying in one conformation for long periods of time and rare switches between these conformations. A physical explanation for these conformations can be given by the free energy landscape, which consists of deep wells, representing local minima. In general, each well is surrounded by large barriers, which separate each well from another. Due to these high energy barriers, transitions between different local minima are rare events [2].
Consider now the dynamic of a molecular system in equilibrium, with the conformational space decomposed into a set of N disjoint but contiguous states. By observing a trajectory of this system at discrete time steps $t = 0, 1, .., i$ the trajectory can be seen as discrete stochastic process with the system having a specific state $q^{(0)}$, $q^{(1)}$, ..., $q^{(i)} \in \Omega$ at each of these discrete timesteps [1; 6]. This stochastic process defines a Markov chain, if the probability of the current state $X_i = q^{(i)}$ only depends on its previous state $X_{i-1} = q^{(i-1)}$:

$$P(X_i = q^{(i)} | X_{i-1} = q^{(i-1)}, .., X_0 = q^{(0)}) = P(X_i = q^{(i)} | X_{i-1} = q^{(i-1)}),$$

i.e. it fulfiles the Markov property. For a Markov chain with finite state space $|\Omega| = N$, the transition matrix P contains the conditional probabilities $P(a, b)$ for each pair of states $q^{(i-1)} = a$ and $q^{(i)} = b$. P is a stochastic matrix with non-negative elements and row sum 1.
However, for large molecular systems this simplified model still suffers from the curse of dimensionality, due to the enormous amount of local minima located on the free energy landscape. Therefore one has to find a way to further reduce the dimension of the Markov chain. This can be done by coarse graining.

1.2 Coarse graining

In order to reduce a given Markov chain, the easiest and most intuitive way would be a set-based approach. The standard practice contains the following steps:

- identify all metastable sets
- assign all states uniquely to one set
- calculate the transition frequencies between all pairs of different sets