International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 11 Issue 09 September 2023, Page no. 3746-3754

Index Copernicus ICV: 57.55, Impact Factor: 7.362

DOI: 10.47191/ijmcr/v11i9.08

Analytical results from the two-states Markovv-states model and applications to validation of molecular dynamics

Orchidea Maria Lecian
Sapienza University of Rome, Rome, Italy.

Abstract

The two-states Markovv-states-model of molecular dynamics is newly analytically studied. The total reward of the path integral of the reaction within a crisp Markovv landscape is proven to be expressed as a Laplace integral (kernel) after the opportune Radon measure. The evolution of the eigenvalues is newly exactly analytical calculated; the corresponding relative error is newly analytically calculated. The problem of an m-states model is established within this framework.

Key-words

Markovv-states model; two-states systems; calculation of eigenvalues; molecular dynamics; protein folding.

1 Introduction

The *Markovv-states models* (MSM) is a chain-of states which are a sample over large free-energy barriers.

The *Markovv hierarchies* depict ensembles of pathways and the kinetic networks [19].

Within the *independent Markovv decomposition* a system is decomposed into separable subsystems; the MSM for each subsystem can be later coupled to reproduce the behaviour of the global system. Different decomposition strategies allow one to describe complex systems [2].

In the *Hidden Markovv states* (HMM), the prescription is realazed, that states correspond to a discrete partition of the states space . Accordingly, jump processes over a finite number of states are allowed [3].

In [11], the problem of the calculation of the evolution of the eigenvalues and the relative error of a two-states Markovv-states model of molecular dynamics is envisaged.

The present paper is devoted to analysisng the problem form an analytical point of view; from the analysis, after the explanation of the mathematical tools needed, the requested evolution of the eigenvalue is newly calculated analytically, and the relative error is newly calculated as well, as requested from [11].

In particular, the total reward of the path integral is proven to be reconducted to a Laplace integral from the Radon measure when the dynamics of the crisp Markovv landscape is taken into account.

The evolution of the chosen eigenvalue is newly calculated analytically; the relative error is calcualted analytically, rather than estimated as frm the inequality [11]. The manuscript is organised as follows.

In Section 1, The Markovv models are introduced.

In Section 2, Markovv-states model is analysed; the two-states model is revised within the framework of a crisp Markovv landscape.

In Section 3, the corresponding dynamics is revised.

In Section 4, the problem of the minimisation of the partition error is presented.

In Section 5, the path integral within the framework of the Markovv landscapes is defined from a Radon measure; the corresponding total reward is set. Within a crisp Markovv landscape, the total reward of the path integral is proven to be written as a Laplace integral of the two-states model. In Section 5, the time evolution of the chosen eigenvalue is newly analytically calculated fro the resulting kernel; the relative error is newly calculated analytically.

In Section 8, the prospective studies are envisaged; more in detail, the problem of a many-states model is framed.

2 Introductory material on Markovv models

The dynamical systems under the paradigm of a Markovv models is characterised after the $n \times n$ square matrices named *transition-probability matrices.* The state probability (of a conditional pairwise probability)encode the items of kinetic information, i.e. the enumeration of possible paths between states.

2.1 Dynamical-systems characterization of Markovv-states models

The dynamical-systems characterization within the framework of Markovv-states models is achieved via the investigation of thermodynamical equilibrium, symmetry with respect to the equilibrium distribution, ergotic propoerty, and a-periodicity: under these hypotheses, the transfer operators admit eigenfunctions (for further description, the example of [4]) is here followed.

Under the hypothesis of the detailed balance, from [11] the following two-states model is described.

In the case of a ’crisp’ partitioning of the state space, it is the aim to find *eigen-functions approximated by step functions, which are constant within the discrete states.*
The crisp Markov landscape is defined after a partitioning of the states space into n sets $\mathcal{S}_n$; the functions $\chi_i(\vec{x})$ express the probability of a point $\vec{x}$ to be found in the $i-th$ set; the functions χ_i are explicitely written as

$$\chi_i^{crisp}(\vec{x}) = 1, \quad \vec{x} \in \mathcal{S}_i \tag{1}$$

and

$$\chi_i^{crisp}(\vec{x}) = 0, \quad \vec{x} \notin \mathcal{S}_i \tag{2}$$

The *kinetics* is described as the evolution according to the intrinsic *time scales* of systems (measurable with spectroscopic techniques),

folding and activation free energies [5],

ultrarapid-mixing continuous-flow method: trapping misfolded structures [6];

upper limit for the speed of formation of the first side-chain contacts (i.e. during protein folding) [7];

constraints from small systems: requirements for *high time resolution* and *high spatial resolution* have to be matched simultaneously [8].

2.2 Long-time dynamics

An ergotic Markovv process is denoted as $\vec{x}_t$, within the phase space Ω; the phase space Ω reversible, endowed with a positive stationarity density $\mu(\vec{x})$ of measure $d\mu(\vec{x})$.
By means of these tools, at time t, the ensemble is described after probability distribution $p_t(\vec{x})$: the equilibrium-weighted probability density $u_t(\vec{x})$ is deifnes as

$$u_t(\vec{x}) \equiv p_t(\vec{x})\mu^{-1}(\vec{x}) \tag{3}$$

The evolution of $p_t(\vec{x})$ after a time interval is therefore well-posed; it is obtained after the *transfer operators* $\mathcal{T}(\tau)$, which evolve the system of a time interval τ: in particular, $\mathcal{T}(T)$ lets the system evolve form the state $u_t(\vec{x})$ to $u_{t+T}(\vec{x})$: because $\mathcal{T}$ bounded, and because $\mathcal{T}$ is self-adjoint, a scalar product can be defined as

$$< f, g > 0 \int_\Omega f(\vec{x})g(\vec{x})\mu(\vec{x})d\vec{x}. \tag{4}$$

The evolution of the probability density according to the eigenfunction decomposition with the eigenvalues λ_i of the equilibrium-weighted probability density

$u_t(\vec{x})$ is

$$u_{t+n\tau} = \mathcal{T}^n(\tau)u_t(\vec{x}) \equiv \sum_{i=1}^{i=N} \lambda_i^n < \psi_i, u_t(x) > \psi_i(\vec{x}), \tag{5}$$

with N as needed, also $N = \infty$, and the chosen ordered eigenvalues with $\lambda_1 = 1$, $\lambda_1 < \lambda_2 < ... < \lambda_N$ (where this technique finds also applications in *biomolecular dynamics*) [9].

3 About the dynamics

The hypothesys of the detailed balance is taken.
Be $x(t) \in \Omega$ a dynamical process; let $x(t)$ be a discrete process in the full Ω, with instantaneous (continuous) change.
The time evolution of an ensemble density is studied.
The transition probability density $p(\vec{x}, \vec{y}; \tau)$ is the change undergone by the system at a time τ and is calculated from the Radon-Stieltjes integration [26].
The operator $\mathcal{Q}(\tau)$ modifies the probability density as

$$p_{t+\tau}(\vec{y}) = \mathcal{Q}(\tau)p_t(\vec{y}). \tag{6}$$

$u(t)$ is defined from the probability density as modified after the measure as

$$u_t(\vec{x}) = \mu^{-1}(\vec{x})p_t(\vec{x}) \tag{7}$$

so that

$$\mathcal{T}(\tau)u_t = u(t+\tau) \tag{8}$$

and

$$\mu u_{t+\tau} = p_{t+\tau} \tag{9}$$

One there defines the generators $\mathcal{L}$ of a continuous basis of rate matrices.
The composition law of iterations of $\mathcal{Q}$ is defined as

$$p_{t+k\tau} =\mid \mathcal{Q}(\tau) \mid^k p_t(\tau). \tag{10}$$

The composition law of iterations of $\mathcal{T}$ is defined as

$$u_{t+k\tau} =\mid \mathcal{T}(\tau) \mid^k u_t(\tau). \tag{11}$$

$\mathcal{T}(\tau)$ can be approximated by a reversible transition matrix on a discrete state space; its eigenfunctions are approximated by the eigenvectors; int he case of m eigenvectors, m eigenvalues λ_i, $i = 1, ..., m$ are considered.
In general, a continuous spectrum of eigenvalues is present.
In the Galerkin approximation, the generators $\mathcal{L}$ are chosen int he case of a long-time $\tau > 0$ as

$$\mathcal{T}(\tau) = e^{\mathcal{L}\tau}, \tag{12}$$

so that the eigenvalues read

$$\lambda_{i,\tau} = e^{\Lambda_{i,\tau}} \tag{13}$$

with Λ_i the eigenvalues of $\mathcal{L}$.
Approximation is possible in terms of density propagation.
From [10]

4 Minimisation of the partition error

Within the framework of high-metastability partions, the the trace of the transition matrix $\mathcal{T}(t)$ is calculated; if the system remains in each partition for sufficiently long time in order to approximately lose memory, the discretized dynamics must be approximately Markovv.
The discretisation error is minimised by the most metastable partition: let m be metastable sets, with

$$\lambda_m >> \lambda_{m+1}; \tag{14}$$

then the most metastable partition into $n = m$ sets *minimises the discretization error.* the model evolves with the Markovv transfer operators $\mathcal{T}(\tau)$.
The focus of these topics is analysed in [11] after expanding the tools of the techniques developped in [12]. The techinques are presented as in Subsection 4.1.

From [12], given a two-states system, the relative error $E_{rel}(\tau, \delta)$ with respect to the eigenvalue $\hat{\lambda}$ of the discrete-time process reads

$$E_{rel}(\tau, \delta) = \frac{\mid \lambda_{1,\tau} - \hat{\lambda}_{1,\tau} \mid}{\lambda_{1,\tau}}; \tag{15}$$

more in particular, $\hat{\lambda}$ is an eigenvalue of QTQ, where once $\tilde{n}$ sets are chosen in one-to-one correspondence with the choice of a basis of an $\tilde{n}$-dimensional subspace D, the transition matrix of the Markovv-states-model, with T the transfer operator of the original Markov process, and Q is the orthogonal projection on-to the subspace D, as analysed in [14]. Application of off-equilibrium simulations are presented in [13]

4.1 Analytical expression of the minimisation of the Markovv-model error

It was proposed it is necessary to look for an **exploration of the extrema of the transition probability density for the diffusion process** of a small partition of the phase space Ω after the **choice of an opportune time interval** in order to

find a new implementation of Markovv-chain Monte-Carlo sampling of transition matrices which extremise the error δ in [11].

Given t_2 *the slowest relaxation time of the system,* **it satisfies the error**

$$\frac{\mid \lambda_j(\tau) - \mathcal{T}_\delta \lambda_j(\tau) \mid}{\lambda_j(\tau)} \leq \bar{\delta}^2. \tag{16}$$

It is the aim of the next Section to evaluate $\bar{\delta}$ analytically instead.

5 Analytical demonstrations

A kernel is Markovv when it is an (evolutionary) map which can therefore be explained as a transition matrix in a finite state space [18].

From [9], given a measurable space, a measurable transformation $\mathcal{T}$ of the measurable space can be defined, for which there associates a rpobability distribution ρ on the measurable space, and a positive kernel Q such that, $\forall f, g$ positive-measurable functions, the following evolution holds

$$\rho(f \odot \mathcal{T} g) = \rho(fQg) : \tag{17}$$

Q can be a Frobenius-Perron operator, ρ is $Q-$invariant, and ρ is $\mathcal{T}-$invariant when the kernel is Markovv. From [19], the transition operator which qualifies transfer operators in ergotic context is studied in metric compact space.
Given the (transition) operators P, the series of its iterates is demonstrated to be convergent itn hte Markovv case; furthermore, in the Markovv case, the corresponding central limit theorem for the Markovv chain is proven to hold. More in particular, the operators are transition operators of Markovv chains, and they are transfer operators int eh case of ergodic theory. The coding of the dynamical partitions is explained in [20]. The decay of correlations is studied in [21]. In its formulation, the continuous-time Markovv chain $X(t)$ is taken

within a partition of integers $S = \{0, 1, 2, ...\}$; A is a subset of S: the path integral Γ is defined as

$$\Gamma = \int_0^\tau f_{X(t)} dt, \tag{18}$$

with f an application which sends A to the interval $[0, \infty)$, and τ the first exit time of A. If a Banach space is given, f qualifies a Radon measure, where a Borel subset is obtained [9].
The path-integral Γ is therefore the ’total reward’ over the time the system spends on A; (some further specific applications are proposed in [22], [23].)

From [10], be $\mathcal{Q}$ the matrix q_i such that

$$q_i = \sum_{i \neq j} q_{ij} < \infty \tag{19}$$

(under the assumption that A contains no ’absorbing’ states), i.e. such that

$$\sum_{i \in S} q_{ij} z_j = \theta f_i z_i, \tag{20}$$

with $0 < z_i < 1$.
For given $i \in A$ and the total time the system spends in A, the Laplace transform $E_i(e^{-\theta\Gamma})$ of the distribution of path integral Eq. (18) is written as

$$E_i(e^{-\theta\Gamma}) = \int_0^\infty \sum_{k\neq i} e^{-\theta f_i u} E_k(e^{-\theta\Gamma}) \frac{q_{ik}}{q_i} q_i e^{-q_i u} du \qquad (21)$$

6 Implementation to the two-states model: analytical calculations

The evolution of the eigenvalue $\hat{\lambda}$ is calculating after specifying the expression of the 'reward' Eq. (21) to the two-states systems as the Laplace integral

$$\tilde{\lambda}_{1,\tau} = \int_0^\infty e^{-\theta\Lambda(t+\tau)} e^{-\theta\tilde{\delta}\Lambda} d\theta = \frac{1}{\Lambda(t+\tau) + \tilde{\delta}\Lambda} \qquad (22)$$

from Eq. (13), which descends form Eq. (12), where the (auxiliary) time variable θ does not correspond to any exit time.
The time evlotoin of the chosen eigenvalue is therefore caluclated analytically in the Garlenkin model.

7 Analytical calculation of the relative error

The error Eq. (16) is therefore newly exactly calculated anaytically within the Galerkin model.
As a result, the relative error $E_{rel\quad Gal}$ is newly analytically calculated as

$$E_{rel\quad Gal}(\tau, \delta) = \frac{\mid \lambda_{1,\tau} - \frac{1}{\Lambda(t+\tau)+\tilde{\delta}\Lambda} \mid}{\lambda_{1,\tau}} \qquad (23)$$

8 Outlook and Perspectives

The aim of the present work is to analyse some features of the two-states Markovv model for a crisp Markovv landscape. The Markovv model is characterised after a path integral on the crisp Markovv landscape, which is proven to be rewritten ans the Laplace integral (kernel). The evolution of the chosen eigenvalue is newly analytically calculated. The corresponding relative error is newly analytically calculated in the Garlenkin model.

It is now possible to extend the results about the two-states system from [11] to an n-states system without the problems if issuing the states.
It can be achieved after defining the opportune Markovv transfer operator $\mathcal{T}_\delta(\tau)$.
On therefore is provided with **the new parameter(s) that controls the discretization error**

$max_{j=1,...,m} \mid \lambda_j(\tau) - \mathcal{T}_\delta \lambda_j(\tau) \mid \leq (m-1)\lambda_2(\tau)\delta^2$
of the m eigenfunctions.

From the obtained results, a comparison with the analysis of [25] will be possible.

References

[1] D. K. Wolfe et al., Hierarchical Markov State Model Building to Describe Molecular Processes, J. Chem. Theory Comput. 16, 1816 (2020).

[2] T. Hempel, Independent Markov decomposition: Toward modeling kinetics of biomolecular complexes, PNAS 118, 2021 (2021).

[3] C. R. Schwantes, R. T. McGibbon, V. S. Pande, Perspective: Markov models for long-timescale biomolecular dynamics, J. Chem. Phys. 141, 090901 (2014).

[4] B. E. Husic, V. S. Pande, Markov State Models: From an Art to a Science, J. Am. Chem. Soc. 2018, 140, 2386 (2018).

[5] M. Jaeger, H. Nguyen, J. C. Crane, J. W. Kelly, M. Gruebele, The folding mechanism of a beta-sheet: the WW domain, J. Mol. Biol. 311, 373 (2001).

[6] C.-K. Chan, Y. Hu, S. Takahashi, D. L. Rousseau, W. A. Eaton, J. Hofrichter, Submillisecond protein folding kinetics studied by ultrarapid mixing, Proc. Natl. Acad. Sci. U.S.A. 94, 1779 (1997).

[7] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, T. Kiefhaber, The speed limit for protein folding measured by triplet–triplet energy transfer, Proc. Natl. Acad. Sci. U.S.A. 96, 9597 (1999).

[8] H. Neuweiler, S. Doose, M. Sauer, A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate, Proc. Natl. Acad. Sci. U.S.A. 102, 16650 (2005).

[9] C. R. Schwantes, R. T. McGibbon, V. S. Pande, Perspective: Markov models for long-timescale biomolecular dynamics, J. Chem. Phys. 141, 090901 (2014).

[10] P. K. Pollett, V. T. Stefanov, Path Integrals for Continuous-Time Markov Chains, Journal of Applied Probability 39, 901 (2002).

[11] J. H. Prinz et al., Markov models of molecular kinetics: Generation and validation, The Journal of Chemical Physics 134, 174105 (2011).

[12] N. Djurdjevac, M. Sarich, C. Schuette, Estimating the Eigenvalue Error of Markov State Models, Multiscale Modeling & Simulation 10, 48 (2012).

[13] Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, PNAS 106, 19011 (2009).

[14] M. Sarich, C. Schuette, Approximating selected non-dominant timescales by Markov state models, Communications in Mathematical Sciences 10, 1001 (2012).

[15] J. D. Chodera, F. Noe', Probability distributions of molecular observables computed from Markov models. II. Uncertainties in observables and their time-evolution, J. Chem. Phys. 133, 105102 (2010).

[16] J. D. Chodera, W. C. Swope, J. W. Pitera, K. A. Dill, Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations, Multiscale Model. Simul. 5, 1214 (2006).

[17] P. Metzner, F. Noe', C. Schuette, Estimating the sampling error: Distribution of transition matrices and functions of transition matrices for given trajectory data, Phys. Rev. E 80, 021106 (2009).

[18] R. D. Reiss, A course on point processes, Springer series in statistics, Springer, New York, USA (1993).

Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, Lecture Notes in Mathematics (LNM, volume 1766), Chapter IX Stochastic Properties Of Dynamical Systems Theorems, Springer, Berlin, Heidelberg (2001).

[19] J. P. Conze, A. Raugi, Convergence of iterates of a transfer operator, application to dynaical systems and Markovv chains, ESAIM: Probability and Statistics 7, 115 (2003).

[20] Ya. G. Sinai, Gibbs measure in ergodic theory, Russ. Math. Surv. 27, 21 1972.

[21] V. Baladi, Advanced Series in Nonlinear Dynamics: Volume 16: Positive Transfer Operators and Decay of Correlations, World Scientific (2000).

[22] G. Rubino, B. Sericola, Sojourn times in finite Markov processes, J. Appl. Prob. 27, 744 (1989).

[23] R. Syski, Passage Times for Markov Chains, IOS Press, Amsterdam, The Netherlands (1992).

[24] VEDI P. Collet, S. Isola, On the Essential Spectrum of the Transfer Operator for Expanding Markov Maps, Communications in Mathematical Physics 139, 551 (1991).

[25] B. Trendelkamp-Schroer, F. Noe', Efficient Bayesian estimation of Markov model transition matrices with given stationary distribution, J. Chem. Phys. 138, 164113 (2013).

[26] J. Ito, Transactions of the American Mathematical Society 110, 152 (1964).