Research Article

Markov Chains of Molecular Processes of Biochemical Materials

Orchidea Maria Lecian*

Sapienza University of Rome, Rome, Italy

Check for updates

OPEN ACCESS

Abstract

Biochemical systems are analytically investigated after encoding the properties of the dynamics, which rule the time evolution of the transition properties, using some Markov models, such as the Hierarchical Markov-State Models. The present paper is aimed at analytically writing the (finite) Markov chain originating from the considered Markov models. Within this framework, the interaction with the environment is considered, and the ergodicity of the systems obtained from numerical simulation is controlled and compared with the qualities of the Markov chain. The (von Neumann) conditions to be imposed on the Bloch equations for the biomaterial structures to be described analytically in a consistent way are governed. The formalisms of the 'heat bath' and that of the control of the numerical errors ensure the good measure-theoretical framework and the ergodicity of the finite chain, respectively.

The finite Markov chains are investigated and the analytical expressions are presented, after which the Hierarchical Markov-State-Model provides the time evolution of the transition probabilities in biochemical systems.

The notion of heat bath is used to describe the interactions of the biomaterial with the environment and thus to control the uses of the projection operators in the Markovian processes where the appropriate measure is defined; the stochastic equations allow one to obtain the wanted measure from the probability spaces.

The cases in which a violation of the Markov property of the process occurs, i.e. in open systems, or dissipative processes are also considered. Furthermore, in complex molecules in biological systems, these features are investigated to be possibly even more dramatic. As far as molecular processes are concerned, this occurrence is associated with the appearance of chaotic effects with certain characteristics of potential surfaces: rather than the technique of isocommittors, the method of projectors in measure spaces is used for the Nakajima-Zwanzig paradigm for the density operator; this latter method complementary compares the time-convolution-less technique.

The finite Markov chains are finally proven to be ergodic after the control of the numerical errors which provide the Sinai-Markov partitions to be applied for the analysis of the measure space of the Markov chain, that is, one endowed with a Hilbert measure. The von Neumann conditions are therefore newly demonstrated to be apt to be applied to the Bloch equations for biomaterial structures after the use of the notion of heat bath, from which the measure space arises.

The qualities of the Hierarchical Markov-Sate Models which bring the analytical expression of the time evolution of probabilities of biomaterials are therefore newly analytically studied.

Introduction

The use of Hidden Markov Models in the approach to biology problems was analysed in a study [1] to be of use as far as the capability of encoding the statistical features of the sampled materials was concerned; in particular, the possibility to reproduce the patterns in time and space of the biological samples were delineated.

As from research [2], the Hierarchical Markov-State Models (MSM) of molecular processes are defined after the $p_i(t)$ probability of the system to occupy the state i at the time t: w_{ij} indicate the transition rate from i to j determined from the unrestrained simulations in the local MSM, while k_{ij} designate the transition rate from i to j such that each state is in a different local Markov state Bw n^* the MSM containing the state i, and S_n the partitions, i.e. the set of Markov states composing the Markov model. The evolution of probabilities $p_i(t)$ evolve in time as,

$$\frac{dp_i(t)}{dt} = \sum_{j \in S_n}\left[w_{ji}p_j(t) - w_{ij}p_i(t) \right] + \sum_{n \notin n^*}\sum_{j \in S_n}\left[k_{ji}p_j(t) - k_{ij}p_i(t) \right],$$

(1)

Where the Σwp accounts for transitions within the local MSM, while Σkp denotes the transition between local MSMs.

In another study [3], the different Markov Models are outlined, for the different biochemical systems, which are classified according to the different (bio-) molecular reactions, the Markov chain originating the processes are sketched for schematising the different biochemical systems according to the stochastic biochemical systems analyses.

As an example in a research work [4], the chosen MSM are shaped according to the experimental techniques for biological macromolecules: at small 'lag' times, an MSM is required to have more macrostates in order to make sure that each microstate is memoryless; differently, shorter 'lag' time is used to describe higher-resolution MSMs. i.e. such that more energy minima are described. As a result, at a lower-resolution MSM, only a few macrostates are separated after high-energy barriers. Several HMMs have been reviewed [5]; in particular, the profile HMMs, pair HMMs, and context-sensitive HMMs are considered. They are demonstrated to be of use in biochemical context and important in similarity search and classification.

The k_{ij} from Eq. (1) are issues after numerical simulation. Among the features of the simulation of Markov chains after Monte Carlo methods [6], the possibility to simulate the qualities of the Gibbs measure is outlined [7]. Markov chain Monte Carlo algorithms have been compared in a research work [8]. A study [9] presented the need for the convergence to the stationary distribution of Markov Chains Monte Carlo methods and recalled to be necessitated.

Stochastically exact simulations can be achieved from a few works [10,11] to shape the features of the biochemical network.

The measure space from the formalisms of the heat bath

The features of a system and a heat bath can be compared to the notion of an open quantum system (of which the latter is issued from [12]): the density operator is this way obtained, which obeys a time-evolution equation under the hypotheses that the generator be independent of time and that the von Neumann conditions be fulfilled. The hypothesis of the exponential decay of the correlation function is taken. The Bloch equations of the density matrix can thus be written, with respect to the von Neumann conditions.

The Markovian time evolution of quantum systems: The analysis of the Markovian time evolution of the quantum systems [13] can be schematised after the presence of a heat bath: the degrees of freedom pertinent to the latter are eliminated after the formulation of one-parameter groups generated after the infinitesimal generators, with respect to which the suitable projector operators on the corresponding Banach spaces are applied until the dependence on the projector is eliminated.

Alternatively, operators whose integral is bounded everywhere, and the integral is strongly continuous, can be used [14].

From a study [13], it can be stated that the time evolution of the transition probabilities is nevertheless formulated according to non-memory-less formalisms, where the memory-related parameter limit has to be examined in order for the realization of the Gibbs state to be achieved.

The Markovian equilibrium can be achieved in the free bath, as from a few works [13,15,16]; it can be discussed in comparison with the stochastic approach, where, from the probability space, the measure is obtained.

Given ρ an arbitrary trace class operator whose free evolution is given after the one-parameter group of isometries on the Banach space of the system (with formally-defined infinitesimal generator), a perturbation A can be introduced; the equilibrium state of the Markov processes is given as condition on the temperature: in the weak-coupling limit, the exponential-decay law is obtained [17,18]. The statistical approach is recovered in the Banach space B of the Markovian system coupled with the heat bath after the Banach-space evolution equation; for this purpose, with B_0 the Banach space of the Markovian system, the stochastic differential equations, a probability space can be defined

$$B = (\Omega, F, d\omega) \tag{2}$$

B is the space of the essentially bounded strongly F-measurable B_0-valued functions on Ω: this way, B_0 is identified as the constant functions on Ω [19]. As an alternative example, a dissipative operator Z can be considered, whose evolution on a Banach space is controlled after a Markov process [20]. From a study [19], the space of the Markovian system coupled with the heat bath is now successfully upgraded to a Hilbert space; given e^{Zt} from a unitary group on the Hilbert space, then iZ is a self-adjoint operator, and the orthogonal projector onto the null space of Z is defined. After A the perturbation of the Markovian system, then $A(\omega)$ is a 'random' operator-valued function: if iA is a self-adjoint operator, symmetric operators are proven to be obtained for the description of the Markov process.

The perturbative approach theory of (finite) Markov chains was studied [21].

Results on the Markov chain

Theorem: The perturbation formalism of finite Markov chains holds [22]. After a Markov chain containing a single irreducible set of states, derivatives of the stationary distributions are defined, and so is the fundamental matrix of the transition probabilities.

The following example is reconducted. Proof Let α be an N-state stationary Markov chain, endowed with a transition-probability matrix: the fundamental matrix always exists, and time averaged transition-probability matrix always exists.

Let α be a chain. Under the hypothesis that α contains only one subchain (i.e. only one irreducible set of states), then the solution of the equation of the stationary distribution always exists, as from [21].

Corollary: So do the qualities of the system β close to α.

Further qualities of the Markov chains

Under the hypothesis that the unperturbed system be geometrically ergodic (Foster-Ljapunov drift conditions), the perturbation is uniform in the weak sense on bounded time intervals, as from [23] after [24].

This opens the way to the construction of the Markov states [22].

The role of the Stochastic approaches and the taming of the non-Markovian properties: expressions of the corresponding Markov chain

In particular molecular processes, chaotic qualities are revealed, with certain characteristics of the potential surfaces of the Markov landscape. More in detail, thin high barriers or deep holes in exothermic reactions are described. It is therefore necessary that for complex molecules in biological systems such effects are understood and reconducted within the Markovian dynamics. This way, the possible apparent violations of Markovian property are tamed. The analysis of this aspect is here focused and framed within the used formalisms, and proven to introduce the formalisms of the density matrices.

In a study [25], it started with the analysis of an MSM with slow relaxation times. Within this scheme, trajectories occur, which are shorter than the 'slowest relaxation time' (i.e. for which some non-Markovian features would be expected). These trajectories are used de facto to reconstruct the transition probability matrix (i.e. also the associated Markov chain can be analytically written [26]). The method is to be applied to multiple pathways, and to 'poorly-relaxed' paths. From [27], the iso-committors method is proposed: iso-committor surfaces are those surfaces that are constructed from the set of all the phase-space points such that the committor they calculated is a constant. In the corresponding MSM, these surfaces are non-crossing surfaces situated in the Markov landscape between the reactant states and the product states, which are non-crossing.

In the present work, the iso-committor method is proven as poorly-grounded for the needed analyses. The method does not prove very efficient, since the committors are in general not orthogonal (i.e. they become orthogonal in the Galerkin

description). The methods of measure-theoretic analysis prove, differently, efficient, since they allow for the existence and uniqueness of projector operators (and their use).

From [28], some features of non-Markovian processes that can be reconducted to the Markovian scheme were studied. More in detail, the stochastic approach of the probability space has to be analysed as allowing for a definition of a probability space whose measure is determining the (measure) space in which the PDE solution is situated [29].

The non-Markovian memory properties are concerned with a method of coarse-graining to be compared with [30]; in this case, for the generalised Langevin equations, the memory term is schematised as dissipative forces. The definition of a 'memory kernel' allows one to apply the analysis also to samples of data with large statistical noise.

The presence of memory term within a measure-theoretical approach was started in [31] under the condition that the memory term be small with respect to the free term, among which the relation is cast through the proper projection operator.

Followingly, the way of the formalism of density matrices is introduced as complementary. More specifically, the projection-operator technique is employed also to describe those cases, in which the interaction of the system with the environment is not negligible. The Nakajima-Zwanzig projection-operator technique [32,33] is demonstrated as an alternative to the time-convolution-less technique [34]. The Nakajima-Zwanzig technique allows one to encode the properties of the time evolution of the transition probabilities into two different orthogonal subspaces, in which the density matrices corresponding to the two different analyses live. Differently, within the time-convolution-less technique, the projection operator technique is used to eliminate the memory kernel from the Nakajima-Zwanzig time equation of the probabilities evolution. The necessitated controls on the projection techniques are still under analysis [35].

Necessitated controls for biochemical materials

Controls on numerical simulations are necessitated [26].

Numerical simulations have to be performed about randomly-impulsed ODEs, about I'to SDEs, and about stochastic parabolic PDEs where white noise is approximated as Gaussian noise.

In the case of stochastic case PDEs (as those issued from [36]); in the geometrically ergodic case, the long-time weak convergence can therefore be proven: the perturbation theory is arising from numerical approximation.

Necessitated definitions: The density operator of an open quantum system is defined as the inverse of the dynamical map which governs the evolution of the density operator. The

quantum dissipation from the von Neumann conditions in the Bloch equation is to be set [26].

As far as the first born approximation Hilbert space is concerned, short-time approximation of the time evolution of probabilities and short-time approximation of the time evolution of probabilities must respect the von Neumann conditions [12].

The request [36] that the conditions needed to be imposed to prove the ergodicity of the perturbed chain must therefore be implemented [26].

The biochemical materials

From a study [37], the space-independent von Neumann Equation can be derived from the Bloch Equation.

The analytical expression of the density matrix in the quantum regime can be obtained; the quantum-mechanical formalism can be recovered for a pure state of electron spin.

The collapse of the wavefunction has to be taken as a measurement postulate.

Discussion

In a research work [38], the adeptness of Markov chains in simulating the qualities of biochemical networks is examined as far as the time evolution of the transition probabilities is concerned.

The features of the Markovian time evolution of the transition probabilities and the non-Markovian ones were compared in a study [39]. The further conditions on ergodicity must be controlled [40].

Conclusion

The present work is aimed at providing new analytical methods for the study of biochemical materials. In particular, the time evolution of the transition probabilities in molecular dynamics is analytically newly formulated. The scope of the research is to provide a new analytical expression of the finite ergodic Markov chain.

For these purposes, several tools are used. The time evolution of the transition probabilities in the molecular dynamics of biochemical materials is described within the framework of Hierarchical MSM; to this Hierarchical MSM there corresponds to the above-mentioned finite ergodic Markov chain. The work is devoted to the proof that the Markov chain is finite and ergodic.

The tools of measure theory are employed to analyse the interaction with the environment, which is schematised as a heat bath, which is analysed after the use of projector operators. The comparison of the stochastic models allows one to construct a triple (i.e. a probability space), in which the measure is defined. As a result, it is possible to pass from a

Banach space to a Hilbert space. The qualities of the Markov landscape are newly examined in great detail. More precisely, the effects which can lead to memory-property-like systems are scrutinised. These features are reconducted to Markovian dynamics through the definition of a density operator, on which the suitable projectors act: as a result, the Nakajima-Zwanzig method is juxtaposed to the time-convolution-less method as complementary.

The analysis is newly ready for the control of the definition of ergodicity after the means of the control of the errors in the numerical approximation.

The ergodicity is outlined as relevant for the sake of applying the Sinai-Markov partitions.

The finite ergodic Markov chain being newly defined, the von Neumann condition can be imposed on the Bloch equation. From this point of view, the quantum mechanical properties are newly demonstrated to lead to the postulate of the collapse of the wavefunction.

References

1. Lou XY. Hidden Markov model approaches for biological studies. Biom Biostat Int J. 2017; 5(4):132-144.

2. Wolfe DK, Persichetti JR, Sharma AK, Hudson PS, Woodcock HL, O'Brien EP. Hierarchical Markov State Model Building to Describe Molecular Processes. J Chem Theory Comput. 2020; 16: 1816-1826.

3. Ghosh P, Ghosh S, Basu K, Das SK. A markov model based analysis of stochastic biochemical systems. Comput Syst Bioinformatics Conf. 2007;6:121-32. PMID: 17951818.

4. Da LT, Sheong FK, Silva DA, Huang X. Application of Markov State Models to Simulate Long-Timescale Dynamics of Biological Macromolecules. In: Han Kl, Zhang X, Yang MJ. (eds) Protein Conformational Dynamics. Advances in Experimental Medicine and Biology. Springer, Cham. 80: 2014.

5. Yoon BJ. Hidden Markov Models and their Applications in Biological Sequence Analysis. Current Genomics. 2009; 10(6): 402-415.

6. Richey M. The Evolution of Markov Chain Monte Carlo Methods. The American Mathematical Monthly. 2010; 117383-413.

7. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6. 1984; 721-741.

8. Error Analysis for Markov Chain Data, Markov Chain Monte Carlo Simulations, and Their Statistical Analysis. 2004;196-235.

9. Toft N, Innocent G, Gettinby G, Reid SW. Assessing the convergence of Markov Chain Monte Carlo methods: An example from evaluation of diagnostic tests in absence of a gold standard. Preventive Veterinary Medicine. 2007; 79(2-4): 244-56.

10. Gillespie DT. A general method for numerically simulating the time evolution of coupled chemical reactions. J Comp Phys. 1976; 22: 403-434.

11. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977; 71(25): 2340-2361.

12. van Wonderen AJ, Lendi K. Quantum Theory of Dissipative Processes: The Markov Approximation Revisited. Journal of Statistical Physics. 1995; 80.

13. Davies EB. Markovian Master Equations, Commun Math Phys. 1974; 39: 91110.

14. Kato T. Perturbation theory for linear operators, Springer: BerlinHeidelberg-New York (1966).

15. Pule' JV, The Bloch equations, Communications in Mathematical Physics. 1974; 38: 241-256.

16. Balslev E, Verbeure A. States on Clifford algebras. Commun Math Phys. 1968; 7: 55-76.

17. Davies EB. Markovian Master Equations II. Math Ann. 1976; 219: 147-158.

18. Potts PP, Kalaee AAS, Wacker A. A thermodynamically consistent Markovian master equation beyond the secular approximation. New J Phys. 23: 123013.

19. Davies EB. Markovian master equations III, Annals of the Henri Poincar´e Institute, Section B, Calculation of probabilities and statistics. 1975; 11: 265-273.

20. Kurtz TG. A limit theorem for perturbed operator semigroups with applications to random evolutions. J Functional Anal. 1973; 12: 55-67.

21. Schweitzer PJ. Perturbation Theory and Finite Markov Chains. Journal of Applied Probability. 1968; 5: 401-413.

22. Lecian OM. Some properties of the Markov chains of the Markov Models of molecular processes, 4th International Conference on Biomaterials & Biodevices, 16 November 2023, Rome, Italy 2023.

23. Sinai YaG, Construction of Markov partitions. Functional Anal and Its Appl. 1968; 2:245-253.

24. Sinai YaG. Markov partitions and C-diffeomorphisms, Functional Anal and Its Appl. 1968; 2: 61-82.

25. Pan AC, Roux B. Building Markov state models along pathways to determine free energies and rates of transitions. J Chem Phys. 2008; 129: 064107.

26. Lecian OM. Markov chains reconducted from non-Markovian processes: the transition probabilities, in preparation.

27. Elber R, Bello-Rivas JM, Ma P, Cardenas AE, Fathizadeh A. Calculating Iso-Committor Surfaces as Optimal Reaction Coordinates with Milestoning. Entropy. 2017; 19(5): 219.

28. Bockius N. Model reduction techniques for the computation of extended Markov parameterizations for generalized Langevin equations. J Phys Condens Matter. 2021; 33: 214003.

29. Lecian OM. The existence and uniqueness of a measure of the Markov chain from the probability space of reconducted non-Markovian processes, in preparation.

30. Vanden Eijnden E, Venturoli M, Ciccotti G, Elber R. On the assumption underlying Milestoning. J Chem Phys 2008; 129.

31. Davies EB. Markovian Master Equations. Commun Math Phys. 1974; 39: 91110.

32. Nakajima S. On Quantum Theory of Transport Phenomena: Steady Diffusion, Progress of Theoretical Physics. 1958; 20: 948.

33. Zwanzig R. Ensemble Method in the Theory of Irreversibility. The Journal of Chemical Physics. 1960; 33: 1338 (1960).

34. Breuer HP, Petruccione F. The Theory of Open Quantum Systems, Oxford Univ. Press, Oxford, 2007.

35. Lecian OM. Projector techniques of memory kernels from exothermic reactions Markov landscapes, in preparation.

36. Shardlow T, Stuart AM. A Perturbation Theory for Ergodic Markov Chains and Application to Numerical Approximations, SIAM Journal on Numerical Analysis. 2000; 37: 1120-1137.

37. Wang LV. Derivation from Bloch Equation to von Neumann Equation to Schroedinger-Pauli Equation, Found Phys 2022; 52: 61.

38. Sandmann W. Discrete-time stochastic modeling and simulation of biochemical networks. Comput Biol Chem. 2008 Aug;32(4):292-7. doi: 10.1016/j.compbiolchem.2008.03.018. Epub 2008 Apr 10. PMID: 18499525.

39. Golatkar J. Markovian, Non Markovian process and Master equation. https://www.thphys.uni-heidelberg.de/w˜olschin/statsem21 5 s.pdf.

40. Chernov NI. Markov Approximations and Decay of Correlations for Anosov Flows, Annals of Mathematics Second Series. 1998; 147: 269-324.