

Article

Sheaf Cohomology of Rectangular-Matrix Chains to Develop Deep-Machine-Learning Multiple Sequencing

Orchidea Maria Lecian 

Department of Clinical and Molecular Medicine, Sapienza University, 324-00185 Rome, Italy;
orchideamaria.lecian@uniroma1.it

Abstract: The sheaf cohomology techniques are newly used to include Morse simplicial complexes in a rectangular-matrix chain, whose singular values are compatible with those of a square matrix, which can be used for multiple sequencing. The equivalence with the simplices of the corresponding graph is proven, as well as that the filtration of the corresponding probability space. The new protocol eliminates the problem of stochastic stability of deep Markov models. The paradigm can be implemented to develop deep-machine-learning multiple sequencing. The construction of the deep Markov models for sequencing, starting from a profile Markov model, is analytically written. Applications can be found as an amino-acid sequencing model. As a result, the nucleotide-dependence of the positions on the alignments are fully modeled. The metrics of the manifolds are discussed. The instance of the application of the new paradigm to the Jukes–Cantor model is successfully controlled on nucleotide-substitution models.

Keywords: rectangular-matrix chains; sheaf cohomology; cochains; deep machine learning; multiple sequencing; amino-acids



Citation: Lecian, O.M. Sheaf Cohomology of Rectangular-Matrix Chains to Develop Deep-Machine-Learning Multiple Sequencing. *Int. J. Topol.* **2024**, *1*, 55–71. <https://doi.org/10.3390/ijt1010005>

Academic Editor: Michel Planat

Received: 5 November 2024

Revised: 13 December 2024

Accepted: 13 December 2024

Published: 16 December 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Topological Hidden Markov Model (THMM) from [1] is constructed from a mixture of normal distributions to be extended to a mixture in locally convex topological spaces in order to comprehend probability density functions in infinite-dimensional spaces. Differently, in the present work, the methods of sheaf cohomology will be used instead to construct the chain for deep-machine-learning multiple sequencing.

The present paper is aimed at proving how to construct a rectangular-matrix chain for multiple sequencing starting from a square two-dimensional-matrix Markov chain after adding Morse complexes (i.e., as from [2]).

For these purposes, the two-dimensional Markov models are considered. The more general context of rectangular chains is preferred, in which the compatibility with larger Markov models is ensured after the singular values. The corresponding probability spaces are constructed. A sheaf of constants is chosen in order to extend the two-state Markov models. The blocks to construct the (also, rectangular) new chain are added as Morse complexes, whose compatibility is proven with the simplices of the corresponding graph, i.e., with the filtration(s). The deep Markov models are constructed this way. The more general construction of rectangular-matrix Markov models is presented. The applications to develop sequencing are newly analytically modeled; this way, the dependence of the probabilities of the nucleotides in the sequences are explained.

In the present paper, the Jukes–Cantor model [3] is considered, according to which the amino-acids in a sequence have equal probability; a Poisson model was proposed in [4] and applied in [5] for the expression of the probability matrix of a Markov chain to be written.

As an application, the profile Markov model, which can be extracted from [6], for sequencing is newly analytically written; the insertion of new sequences, i.e., that of a single amino-acid and that of a sequence, is newly analytically written in order to extend

the profile Markov model to a Jukes–Cantor model [3] via the interrogation proposed in [5] (where the interrogation from [5] follows the long-standing problem after [4]). After the review of several two-dimensional square matrix Markov chain sequencing models, the limitations of the deep Markov models are outlined. The way to analytically manage insertions and deletions in the sequences is newly proven.

In more detail, the methods of sheaf cohomology are used in order to analytically add blocks to a matrix and to analytically delete blocks from a matrix, whose action corresponds to add insertions and to perform deletions to a sequence. The partial matchings that are to be managed are constructed starting from a chosen sequence. In particular, the Morse simplicial complexes that are treated are those that allow us to reconduct to the wanted filtration of the probability space.

The paradigm developed at the initial stage defined in the above can be of use to shape the profile Hidden Markov Models version treated in [7]. This item of the paradigm allows one to define the probability spaces of the two-state pair MM and of the profile HMM.

The applications to computation in biology are discussed in [8].

The measure of the probability space is that to which the proprieties of the Markov chain allow to upgrade starting from the sigma-algebra of the Borel (sub-)sets that constitute the state space of the models.

From the probability space, a metric space can be defined, therefore endowed with a metric, on which there lives the manifold on which the graphs associated with the chain are defined: it is one of the aims of the present paper to discuss and to prove the well-posedness of the concept of distance within the metric spaces, in the way such that the distance is well-posed if two chains are obtained from the initial two-items sequences, to which the opportune sheaf (or sheaves) are applied.

The possibility to construct rectangular-matrix chains is considered as well; this construction allows one to schematize the existence of latent states [9,10]. The sheaf-cohomology techniques here selected also apply to rectangular matrices; the analytical managing of the Morse simplicial complexes allows one to reconduct to the wanted filtration.

As a further advantage, it is possible to reproduce the site-dependence of the probability of finding the items in a sequence.

Therefore, the model is apt to be applied to the analysis of sequences of amino-acids, as outlined in [11].

In more detail, in the present paper, it is possible to analytically write the probabilities to find a certain number of items in a selected order. Accordingly, the new quality of the amino-acids to exhibit different probabilities to be found in a sequence, as very recently found in [12], is analytically modeled.

As an application of the methods developed here, two nucleotide-substitution models are compared: the method from [13] is proven to be obtained after the addition of the blocks of the sheaf needed to reproduce the wanted probability space, while the model from [14] is proven inconsistent with the presented protocol.

Therefore, the present paper contains answers to the interrogations delivered in [15] and in [16] as far as the comparison of methods of nucleotide substitutions is concerned. An illustrative example is contained in which two methods are compared: one is found to be an extension of the two-state Jukes–Cantor model, while the other one is proven as not being obtainable from the Jukes–Cantor model. The relevance of constructing sequences containing pairwise repetitions of the same pattern as far as timely requests are concerned about the evolution of species is outlined in [17].

The prospective studies to be envisaged are the recognition of patterns in the sequences and the outlining of the patterns from the experimental errors, as requested in [18]: the interrogation can be solved as indicated in [19] and applied as in [20]. The patterns are ibidem shown to be contained in rectangular-matrix constructions.

The implementation of the deep Markov models can therefore be argued to be apt for developing with the 'context-sensitive' Markov models, i.e., from [21], where the

rendering of the wanted states as ‘context-sensitive’, can be achieved using the substitution techniques developed in the present paper.

To apply the methods of [19], the singular-value decomposition is requested in order to control the filtration of the probability spaces.

The paper is organized as follows:

In Section 2, several models of sequencing are discussed, each of whose exhibits stringent peculiarities for application.

In Section 3, the ergodicity of HMMs with a ‘generalized observation structure’ is ensured.

In Section 4, the deep Markov models are reviewed and the limitations are outlined.

In Section 5, the probability spaces are constructed, whose features complete the limitations of the deep Markov models.

In Section 6, the filtration of the probability space is modeled according to the choice of a (constant) sheaf; furthermore, the way to add insertions is proven to be by adding Morse complexes, to which the compatibility with the simplices of the graph holds: the methods to perform deletions is, in this way, also delineated.

In Section 7, the outlook is proposed. In Appendix A, further complements of the cochain theory are studied in detail in order to illustrate the proof of the new Theorem 1.

In Appendix B, the method is demonstrated to be one of use to develop the multiple sequencing of amino-acids: the chains for sequencing are analytically constructed from the profile Markov model; the metric of the corresponding manifold is discussed. In Appendix C, two methods of amino-acid substitutions are compared: one of them is proven to be obtained from the Jukes–Cantor model after application of the Morse operator, while the other is demonstrated as consisting of different hypotheses.

2. Introductory Material

2.1. Profile-Hidden Markov Model

As described in [22], the construction of a Profile-hidden Markov Model (pHMM) starts with the consideration of a ‘multiple sequence alignment’.

Insertions and deletions are modeled.

The occupancy of a position and the probability to each position of the alignment are encoded in the pHMM.

In [23], the pHMM is commented to apply to large-scale sequence analyses.

Profile-Hidden Markov States models are introduced in [24]. The pHMM is constructed from an opportune HMM of an MSM. While the HMM consists of ‘training sets’ of unaligned sequences, the pHMM allows one to obtain the multiple alignment of all the training sequences.

In [8], the pHMM is analyzed to exhibit higher sensitivity and to ‘recall’ the rates compared to ‘pairwise alignment methods’.

The software implementation of pHMM is reviewed in [23].

From [7,22], a profile for an alignment is understood as a trivial HMM with one ‘match state’ for each column. The consecutive match states have consecutive probabilities.

The ‘output probabilities’ are defined as the probabilities given after the probabilities of finding one particular match state in the corresponding column.

‘Insertions’ are defined as ‘portions of sequences that do not overlap those found in the previous constructions’. They are represented after the state space S_{ij} , where the states modelize the inserts of the j -th column of the alignment. The ‘output probabilities’ of the insert states are requested to equal the background probabilities.

The probabilities P_{S_j} are given; such probabilities affect the pHMM such that the insertions can take place in different portions of the alignment. The P_{S_j} can be time-dependent as well: the ‘affine gap penalties’ are this way defined.

From [22], the construction of a pHMM starts with the consideration of a ‘multiple alignment’. Insertions and deletions can be modeled.

The occupancy of a position and the probability at each position are encoded in the ‘profile’.

Given the sample $\Sigma = \{b_1, \dots, b_n\}$, the wanted set of states $\mathcal{A} = \{1, \dots, k\}$ is chosen: the pertinent transition probabilities a_{ij} are specified under the initial condition as i ; the ‘emission probabilities’ are specified as $a_i\{b_k\}$: the notation in the latter formula employs the row–columns description of the entries of the matrix. The corresponding pair HMM relates the probabilities distribution over certain sequences of pairs of observations.

The $p_{s,j}$ can be time-dependent as well and are a-priori different from the probabilities P_{Sj} found in the previous calculations: the ‘affine gap penalties’ are this way defined.

2.2. The Context-Sensitive Hidden Markov Models

The context-sensitive Markov Models (csMMs) are presented in [21]. Context-sensitive HMMs are used to model long-range dependencies in symbol sequences—a case to which the study of amino-acids can be reconducted.

csMMs are apt to describe ‘nested dependences’ between symbols.

The csMMs are developed in order to study long-range correlations after ‘rendering’ some of the states in the model as ‘context-sensitive’.

2.3. Pair Hidden Markov Models

From [25], the Pair hidden Markov Models (pair HMMs) are constructed. The sample $\Sigma = \{b_1, \dots, b_k\}$ is given, with the wanted set of pairs of states $\tilde{S} = \{s_1, \dots, s_k\}$, the transition probabilities a_{ij} are written, specified under the initial conditions a_0 : the ‘emission probabilities’ of the pair hMM are different from those obtained in the profile hMM, and are obtained as $e_i(b_k)$. The pair HMM related the probability distributions a_{ij} over certain sequences of pairs of observations.

From [26], the use of conditional random fields is discussed in pair HMMs.

After the use of conditional random fields (CRFs) for time series data, linear chains are obtained [27].

2.4. Deep Markov Models of Sequential Data

The software implementation of featuring the deep Markov models of sequential data after [28] are written in [29].

From the analysis ibidem, the following developments are constructed.

The chain of latent variables is introduced: each latent variable from the chain is conditioned after the previous latent variables.

Potential highly non-linear dynamics are prospected.

Definition 1. *The wanted transition probabilities, which rule the dynamics of the latent variables, and the ‘emission probabilities’, which govern the observations, should be parameterized after non-linear neural networks.*

Example 1. *The sequence of observations is of length 3 as $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$; there exist a sequence of coherent random variables $\{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$.*

The implementation of the Markov properties of the model is to be studied.

3. About Ergodicity of HMMs with a ‘Generalized Observation Structure’

The conditions of ergodicity of HMMs with ‘generalized observation structure’ are addressed in [30].

Under the opportune set of hypotheses, which will be tested in the following for the verification of the extent to which the work [30] is useful in the present approach, the existence of a unique invariant measure of the pair (*filter – observation*) has to be verified.

The triple (*state – observation – filter*) is written within the asymptotic stability of probability.

The problem of ‘incorrectly initialized filters’ will be further addressed in the present work. The asymptotic properties of the filter and the state estimator are founded on both the observations and on the (knowledge of the) initial data.

A minimal invariant measure and a maximal invariant measure can be defined.

Let $\{ (x_n), (y_n) \}$ be a discrete time process of a probability space $(\Omega, \mathcal{F}, \mathcal{P})$: the kernel $P(x, dx')$ is the transition kernel.

The part (y_n) is called the ‘observation process’; its transition kernel is written as $P_1^{x'}(y, dy')$: it is parameterized after the current values of the hidden Markov process.

Let X^n be the set of states on a Borel subset endowed with a σ algebra $\sigma = \{ x_0, x_1, \dots, x_n \}$, and let $Y^n = \sigma \{ y_1, y_2, \dots, y_n \}$: here, it is necessary to remark that the implementation from $n = 2$ to $n = 3$ is non-trivial within the analysis of the ergodicity.

The transition probabilities are written as

$$P\{ x_{n+1} \in A \mid X^n, Y^n \} = P(x_n, A), \tag{1}$$

and

$$P\{ y_{n+1} \in B \mid X^n, Y^n \} = P_1^{x_{n+1}}(y_n, B) \tag{2}$$

It is crucial to remark in the present work that, from [30], for fixed $x, x', y, P(x, \cdot)$ and $P(y, \cdot)$ must be defined. In the case where they are defined, they are the probability measures of e_0 and of E , respectively.

It is further crucial to remark from [30] that, for a fixed Borel subset A of E_0 and, if the fixed Borel B of E exists, then the application that acts like sending

$$x \mapsto p(x, A), \tag{3}$$

and

$$(x', y) \mapsto P_1^{x'}(y, B) \tag{4}$$

has to be demonstrated as far as the new probabilities p exist to admit a measure $\mathcal{B}(E_0)$ and $\mathcal{B}(E)$ being \mathcal{B} the suitable Borel σ fields.

The hypothesis is made in [30] that, on the kernel P_1 , the following definition holds:

$$P_1^{x'}(y, B) = \int_B r(x', y, y') \eta dy' \tag{5}$$

with $\eta \in \mathcal{P}$, and \mathcal{P} from the probability space is the space of probability measures.

The observation is claimed to be admitted in [30] for the models

$$y_{n+1} = h(y_n, y_{n+1}, w_{n+1}) \tag{6}$$

being h C^1 diffeomorphisms on R^d .

It is one of the aims of the present work to analyze the construction.

4. About the Proposal of Deep Markov Models

In [28], the use of Gaussian state space models has been of use in ‘generative model of sequential data’ as a long-standing attribution.

Algorithms to learn a wide class of linear state space models and of non-linear state space models have been developed.

Developments were also implemented in which the emission distributions and the transition distributions were molded after neural networks.

The posterior distributions were obtained so as to be compared with the outcomes of a ‘learning algorithm’, in which both expand the properties of a ‘compiled inference network’ and the ‘generative model’, where room is left for parameterized variational approximation.

The scalability properties were exhibited.

The ‘structured approximation’ is shown to produce ‘posterior results’ with ‘significantly higher held-out distributions’. It is our aim to verify the statement analytically.

HMMs and Recurrent Neural Networks (RNNs) can be implemented with linear state space models or with Gaussian state space models.

Definition 2. A deep Markov model (DMM) is a ‘class of generative models’ in which the ‘classic’ linear emission and the classic transition distribution are substituted with complex multi-layer perceptrons (CMLPs).

There exists general state space model (GSSM’s) that are apt to keep the Markovian structure of the HMMs; in this case, however, the ‘representational power of the deep neural network to treat high-dimensional data is leveraged’.

If the DMM is ‘improved’ after increasing the latent states of the successive time step z_{t+1} , the DMM is *ibidem* hypothesized to be interpreted as a (possibly ‘more restrictive’) stochastic RNN [31], as far as variational RNNs are concerned [32].

The linear algorithm proposed *ibidem* is valid for executing ‘stochastic gradient ascent’ of the variational lower bound of the ‘likelihood’ [33].

A new family of ‘structured inference networks’ is assessed in [28]; it is parameterized after ‘recurrent neural networks’, and the following three possibilities are comprehended:

- (i) The ‘generative model’ is already known and assessed;
- (ii) The functional form of the parameters estimation is known;
- (iii) The learning deep Markov model is known.

5. The Probability Spaces

The proposed HMM is prepared from the complete model.

From the complete probability space $(\omega, \mathcal{F}, \mathcal{P})$ for the state space $\Omega \equiv \{ X_1, X_2, \dots, X_n \}$, the probability is defined

$$\mathcal{P}(X_{n+1} \in A \mid \mathcal{F})(\omega) \equiv p(X_n(\omega), A) \tag{7}$$

The HMM is defined as the probability space $(\Omega_1, \mathcal{F}_1, M_1)$ with the probability distribution M_1 and the measurable space $(\Omega_2, \mathcal{F}_2)$ with the probability distribution M_2 as

$$M_2 : \Omega_1 \times \mathcal{F}_2 \mapsto [0, 1] \tag{8}$$

The existence of the HMM is based on the verification that the map Equation (8) is still a measurable function $\forall A \in \mathcal{F}_2$. In the case where the requests after Equation (8) are verified, the section (product of σ fields) F defines a unique measure as

$$\mu(F) = \int_{\Omega_1} M_2(\omega_1, F(\omega_1)) M_1 d(\omega_1) \tag{9}$$

is still from.

The choice $\Omega = \{ X_1, X_2 \}$ does not allow for such a construction. It is verified after the verification of the measure. From this, it follows that the two models are on different manifolds.

6. Sheaf Cohomology of Models of Sequential Data

The two-dimensional matrices from [11,34,35] have an eigenvalue equal to 1, i.e., they are stochastic matrices.

The construction of the \hat{M} rectangular matrix in the thin singular-value decomposition is achieved as

$$\hat{M} = U_k \hat{\Sigma} V_k^T \tag{10}$$

where the apex T indicates the transpose. In the present case, the wanted vectors are such that

$$V^T \equiv U^T. \tag{11}$$

for the purposes of Appendix A. For the present purposes, it is even more conformable to investigate the thin compact singular-value decomposition

$$\hat{M} = U_r \hat{\Sigma}_r V_r \tag{12}$$

in which only the r non-vanishing singular values are considered.

The singular values σ_i of $\hat{\Sigma}$ allow for a decomposition

$$\hat{M} \sum_{i=1}^{i=r} \xi_i u_i v_i^T, \tag{13}$$

where \vec{u} and \vec{v} are from, in general, two sets of orthonormal bases.

The request that \vec{v} be unitary would lead to the definition of one corresponding normalized orthonormal basis.

It will be interesting to verify whether \hat{U} is a partial isometry.

From [2], the procedures are extended here in order to add at least one block $1 - \delta$ on the diagonal of $\hat{\Sigma}$ to discuss whether it is possible to pass from the two-state Markov models to the Deep Markov model of sequential data [28,29].

The wanted chain \hat{M} can be put on a manifold, on which a graph is constructed after the path of simplices implied after \hat{M} . A sheaf \mathcal{S} is fixed on a simplicial complex L .

The choice of the sheaf will be specified in the following as a sheaf of constants.

The following new definition is obtained starting from [2] Def 5.5:

Definition 3 (γ): \forall dimension $k \geq 0$, the vector space of k -cochains of L with \mathcal{R} coefficients is written as a product

$$C^k(L, \mathcal{R}) = \prod_{\dim \tau=k} \mathcal{R}(\tau) \tag{14}$$

Here, one newly chooses from [2] to develop the case in which the vertices of L are ordered. This way, it is possible to define the topological models (of which the 3-states model is that with the lowest dimensionality). This way, the following new definition is given:

Definition 4 (δ): For each $k > 0$, the k -th coboundary map of L with \mathcal{R} coefficients is the linear map

$$\partial_{\mathcal{R}}^k : C^k(L, \mathcal{R}) \rightarrow C^{k+1}(L, \mathcal{R}). \tag{15}$$

with $\partial_{\mathcal{R}}^k$ coboundary operator.

Remark 1. The arrow in Equation (15) is defined as a 'block action'.

For each pairs of simplices $\tau < \tau'$ such that $\dim \tau = k$ and such that $\dim \tau' = k + 1$, the component of $\partial_{\mathcal{R}}^k$

$$\mathcal{R}(\tau) - \mathcal{R}(\tau') \tag{16}$$

is written as

$$\partial_{\mathcal{R}}^k |_{\tau, \tau'} = [\tau : \tau'] \cdot \mathcal{R}(\tau \leq \tau'). \tag{17}$$

The sequence

$$0 \rightarrow C^0(L, \mathcal{R}) \xrightarrow{\partial_{\mathcal{R}}^0} C^1(L, \mathcal{R}) \xrightarrow{\partial_{\mathcal{R}}^1} \dots \xrightarrow{\partial_{\mathcal{R}}^{k-1}} C^k(L, \mathcal{R}) \xrightarrow{\partial_{\mathcal{R}}^k} C^{k+1}(L, \mathcal{R}) \xrightarrow{\partial_{\mathcal{R}}^{k+1}} \tag{18}$$

is forming a cochain complex over \mathbb{F} .

The composition

$$\partial_{\mathcal{S}}^k \circ \partial_{\mathcal{S}}^{k-1} = 0 \forall k \geq 1 \tag{19}$$

for every sheaf \mathcal{S} and, therefore, in particular, also for \mathcal{R} .

The Morse chain complex is now discussed.

Let K be a simplicial complex with ordered vertices.

Let $\zeta \in K$. Also, let $\tau \in K$ be any simplices.

Now, let $[\tau_1 : \zeta] \in \{0, \pm 1\}$: it defines the coefficient of ζ in the boundary of τ . The coefficient is non-vanishing iff $\zeta \triangleleft \tau$. Here, we are interested in the determination of $[\tau_1 : \zeta]$ and define the coefficient of ζ as 1.

In [2] page. 98, an acyclic partial matching is fixed: there, all acyclic partial matching is defined as one whose G -paths are all gradient.

The definition of ‘chain complex’ therefore follows Theorem 8.4 in ibidem.

From Definition 8.6 ibidem, which is aimed at associating an algebraic contribution to each G -path, the following new definition is given:

Definition 5 (ϵ): *The weight $w(\rho)$ on the G -path consisting of*

$$\zeta = (\zeta_1 \triangleleft \tau_1 \triangleright \zeta_2 \triangleleft \tau_2 \triangleright \dots \triangleright \zeta_m \triangleleft \tau_m) \tag{20}$$

is defined as the product

$$w(\zeta) = \frac{(-1)}{[\tau_1 : \zeta_1]} \cdot [\tau_2 \zeta_1] \cdot \frac{(-1)}{[\tau_2 : \zeta_2]} \cdot \dots \cdot [\tau_{m-1} : \zeta_{m-1}] \cdot \frac{(-1)}{[\tau_m : \zeta_m]}. \tag{21}$$

Differently, from [2], it is our aim to express the ‘collected’ version of the weight of the G -path, and to then choose the algebraic weight all equal to 1 and to study the k -th coboundary map.

The weights $w(\rho)$ are also written as

$$w(\rho) = (-1)^m \prod_{i=1}^{i=m-1} \frac{\tau_i : \zeta_i}{\prod_{i=1}^{i=m} [\tau_i : \zeta_i]}. \tag{22}$$

We need to specify Equation (22) for τ_1 as

$$w(\rho) = (-1)^m \prod_{i=1}^{i=m-1} \frac{\tau_1 : \zeta_i}{\prod_{i=1}^{i=m} [\tau_1 : \zeta_i]} \tag{23}$$

for the purposes of the sequencing procedures.

The constraints on $w(\rho)$ imply the looked-for entries of the singular-value decomposition form of the wanted vectors from $\hat{U}^T = \hat{V}^T$. To this aim, the following new definition is given:

Definition 6 (ζ): *For each $\dim k \geq 0$, the k -th Morse boundary operator is the linear map*

$$D_k^G : C_k^G(K) \rightarrow C_{k-1}^G(K) \tag{24}$$

which admits a matrix representation.

The entry of the matrix from Definition ζ are the columns of a critical k -simplex α , and the rows of a critical $(k - 1)$ -simplex ω are given as

$$[\alpha : \omega]_G = [\alpha : \omega] + \sum_{\rho} [\alpha \zeta_{\rho}] \cdot w(\rho : \omega). \tag{25}$$

We specify the choice of the orthonormal base (at least, \vec{v}) with all the weights of the G -path equal to 1.

The only paths that result in a non-zero contribution are those paths that can be written as

$$\alpha \triangleright (\zeta_1 \triangleleft \tau_1 \triangleright \zeta_2 \triangleleft \tau_2 \triangleright \dots \triangleright \zeta_m \triangleleft \tau_m) \triangleright \omega \tag{26}$$

where the role of τ_1 is outlined.

The choice of the path with weight equal to 1 is $\mathbb{F} = \mathbb{Z}/2$.

For the sake of the present purposes of sequencing, it is necessary to use Proposition 8.8 from [2].

It is necessary to control the number of the connecting G -paths.

The definition of a chain complex allows one to analyze the outcomes of the measurements of the sequencing resulting from the chain complex.

It is proven by induction that the wanted result is obtained when G consists of a single pair as $\zeta \triangleleft \tau$ of simplices in K .

Complements for the following matter are given in the Appendix.

The wanted set of critical simplices is defined as follows:

Theorem 1. *The set of critical simplices for sequencing is defined as*

$$C_G^{seq} \equiv K - \{ \zeta \triangleleft \tau \} \tag{27}$$

Proof. Equation (27) allows one to define the new boundary operator, which defines the chain complex. \square

It is crucial to comment the following:

Remark 2. *The chain complex is associated with an ‘acyclic partial matching’ G .*

The following is true:

Theorem 2. *The partial matching has the same homology as the boundary operator as the chain complex.*

Proof. From Proposition 8.10 in [2], the Morse chain complex is, in this case, chain-homotopy equivalent to the standard simplicial chain complex. \square

The wanted HMMs are, therefore, this way constructed.

Here, one needs to take a sheaf \mathcal{R} of constants as follows:

From Proposition 8.8 in [2], $C_\circ^\Sigma, D_\circ^\Sigma$ is a chain complex. From Proposition 8.10 in [2], the Morse chain complex $(C_\circ^\Sigma(K), \partial_\circ^\Sigma)$ is ‘chain homotopy equivalent’ to the standard simplicial complex $(C_\circ(K), \partial_\circ^k)$.

The two chain maps are described with a pair of chain homotopies.

To construct the chain homotopies, one processes the simplex-pairs $\sigma \triangleleft \tau$ in Σ one at a time. This means the block of the matrices can be added one at a time.

For this purpose, two more arbitrary simplices α and ω can be considered, as there is only one ‘path’ $\sigma \triangleleft \tau$.

The entries of the new blocks of the matrices to be inserted are this way found as

$$[\alpha : \omega] + [\alpha : \sigma] \cdot \frac{(-1)}{[\tau : \sigma]} [\tau : \omega]. \tag{28}$$

Construction of the Probability Space of the HMM

The chain homotopy is defined after the linear maps $\theta |_{\alpha,\beta}$ as

$$\theta |_{\alpha,\beta} = \frac{1}{[\tau : \sigma]}, \alpha = \sigma, \beta = \tau; \tag{29a}$$

$$\theta |_{\alpha,\beta} = 0 \text{ otherwise.} \tag{29b}$$

The difference between the ‘new’ entries and the previous ones are

$$\frac{[\tau' : \sigma] \cdot [\tau : \sigma']}{[\tau : \sigma]} \tag{30}$$

where the apex ‘ \prime ’ denotes the new entries.

Given a filtration of a simplicial complex, an acyclic ‘partial matching’ on Σ is compatible with the filtration if there exists a function b such that $b : k \rightarrow \mathbb{R}_+$ is associated with the map

$$\sigma \mapsto \{ t \geq 0 \mid \sigma \in F_t(K) \} \quad (31)$$

such that $b(\sigma) = b(\tau) \forall \sigma \triangleleft \tau \in \Sigma$.

The blocks can, therefore, be inserted one at time; nevertheless, the blocks are different from each other. It is nevertheless ensured that the effects in determining the sequencing are wanted. Indeed, a nested sequence of Morse complexes is found. From Theorem 8.25 in [2], there exist homeomorphisms of persistent homology groups ensuring that the inclusions of Morse complexes are equal.

7. Discussion

The present work is aimed at analyzing the DMM in deep machine learning as far as the dimensionality of the kernels is concerned for the comparison of the different models. It is proven by induction that the topological features of the HMM eventually descending from the latent states of a two-state MSM are not comparable with those of an N -state MSM with $N > 2$ because of the different properties of the manifold that the corresponding graphs stay on according to the analytical expressions of the eigenvalues; accordingly, not even a concept of distances between the models is well-posed.

In the present work, the dimensions of the possible matrices of the chains originating the transition probabilities of sequencing as ‘partial matching’ are achieved by added the blocks needed one at each time as nested sequence of Morse complexes; in more detail, the order at which the Morse complexes are added are proven irrelevant for the construction of the block matrix: indeed, not only one square matrix can be taken as originating the sequencing. In more detail, there exists an infinite set of rectangular matrices, discriminated only according to their singular values, which can serve for the scope of sequencing.

The definition of the probability space allows one to perform all of the requested measurements. In more detail, the measure of the probability spaces allows one to calculate the distances; the choice of the sheaf allows one to qualify the filtration of the partial matchings.

As a result, the new qualities of the DNA sequences [12] can be analytically modeled, according to which the probability to find an amino-acid in a sequence depends on the site; in more detail, further characterization is possible, according to which the probability to find a partial matching of amino-acids is dependent on the distances among other amino-acids in the complete sequence.

Accordingly, the construction of the complete most-general rectangular-matrix chain is possible. The construction of the chain is demonstrated from the proof of Theorem 1: the partial matching 12 is chosen from a profile Markov model, and the further items are added, which correspond to blocks of the matrix. The possibility to manage deletions analytically is also performed. The analytical paradigm here developed is now apt for several analytical applications.

As a first instance, it is possible to compare Markov models for sequencing that are aimed at generalizing the two-items sequence and that have the same dimensions: as a result, it is possible to prove which ones are originated from the Jukes–Cantor model and which are not: the proof is based on the analysis of the corresponding probability space.

As a further result, from the probability space, the metric space on which there lives the manifold of the graph associated with the Markov models can be determined; the determination of the metric space allows one to discern about the well-posedness of distances, i.e., as interrogated in [16]: the answer to the interrogation allows one to find the structures of the sequencing.

As an additional advantage, the search for hidden structures in the sequences is therefore possible; the relevance of the comparison with the Jukes–Cantor model is a long-standing interrogation, which is answered here after defining the suitable topology paradigm to be applied to implement the several two-state Markov models to a model with a higher number of states, i.e., starting from the three states for deep Markov models.

The data simulations for the papers analyzed in Appendix C can be described in a timely manner according to the newly introduced concepts of scalability and computational complexity, as introduced in [36,37]. A definition for scalability in this regard has been pointed out to be not complete yet. The notions are not used here, because all the computations performed are analytical.

The following considerations can be developed.

In [38], ‘transitions in the space of data’ are made to correspond to ‘transitions in the space of models’: in this context, the notion of ‘distance distribution functions’ are introduced. In the present paper, the well-posedness of distances in the metric spaces worked out of the probability spaces defined after the filtrations avoid this misconception according to our analytical derivations. In the case where the assistance of computer-based techniques is of interest, the pioneering work of [39] can be considered for further developments.

According to all the statements, the issue of inquiring about the stochastic stability of deep Markov models can be addressed. From [40], the deep Markov models from a multi-layer deep neural network are analyzed; it is observed that there exists a map between the transition probabilities and the emission probabilities: in the present work, the problem is overcome after the analytical definitions of the metric spaces, after which the ‘stability’ of the system is not stochastic.

Funding: This research received no external funding.

Data Availability Statement: The new calculations are analytical, and are written on this paper.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Complements of the Cochain Theory

Here, the cochain theory is recalled from [2] and the general theorems are specified for Theorem 1 in Section 6.

The boundary operator is built after the property.

Definition A1. A Morse boundary operator is the linear map

$$d_k^\Sigma : C_k^\Sigma(K) \rightarrow C_{k-1}^\Sigma(K). \tag{A1}$$

After the definition of a weight ρ over all the Σ paths, the Morse boundary operator d_k^Σ for Equation (A1) admits a matrix representation; it is constituted of entries such that the columns of a critical k simplex α and the row of a critical k -simplex ω are

$$[\alpha : \omega]_\Sigma = [\alpha : \omega] + \Sigma_\rho[\alpha : \sigma_\rho] \cdot \omega(\rho) \cdot [\tau_\rho : \omega] \tag{A2}$$

The following proposition holds:

Proposition A1. The pair $(C_\bullet^\Sigma, d_\bullet^\Sigma)$ is a chain complex.

The following example is worked out.

Theorem A1. Let Σ be an acyclic partial matching on a simplicial complex K and \mathbb{F} be a coefficient field. The cohomology group of $(C_\bullet^\Sigma(K, \mathbb{F}), d_\bullet^\Sigma)$ is isomorphic to those of K .

For the future purposes, we restrict to a special acyclic partial matching σ (i.e., which will be requested to be taken as that corresponding to the sequence 12).

Definition A2. $\forall \dim k \geq 0$, the k -th boundary operator is the linear map d^{sigma}_k

$$C_k^\sigma(K) \rightarrow C_{k-1}^\Sigma \tag{A3}$$

which admits a matrix representation.

Definition A2 must specified to σ in order to find the following:

Definition A3. $(C_{\bullet}^{\sigma}, D_{\bullet}^{\sigma})$ is the Morse chain complex associated to σ .

From Definition A3, we have the following:

Theorem A2. The Morse chain complex $(C_k^{\sigma}, D_k^{\sigma})$ is chain-homology equivalent to the standard simplicial chain complex $C_{\bullet}, D_{\bullet}^K$.

It is indeed our aim to specify the method to build a sequence starting form the partial sequence 12. For this, one notices the following:

Definition A4. A cochain complex $(C^{\bullet}, D^{\bullet})$ over \mathbb{F} is a sequence of vector spaces and of linear mappings d^k such that

$$d^{k-1} \otimes d^k = 0 \quad \forall k \geq 1. \tag{A4}$$

Appendix B. Construction of the Chain of Multiple Sequencing

As outlined in [20], the investigation of [5] is developed as follows. The fundamental matrix of the chain \hat{Q} is taken as

$$\hat{Q} = \hat{U} \hat{L} \hat{U}^{-1}, \tag{A5}$$

where \hat{L} is a diagonal matrix whose entries $l_{ii}, i = 1, 2, \dots, 20$ are written as $l_{ii} \equiv \lambda_i$, where λ_i are the eigenvalues of the fundamental matrix.

The probability matrix per unit time interval \hat{M} is set.

The matrix (\hat{U}) is formed from the eigenvectors of the matrix \hat{M} .

The probability matrix \hat{P} from the chain \hat{Q} is written from the entries u_{ij} of \hat{U} as

$$P_{ij} = \sum_k u_{ik} u_{kj}^{-1} e^{\lambda_k t}. \tag{A6}$$

The matrices \hat{R} is the matrix of relative rate of substitution.

The matrix \hat{A} is the mutation matrix.

A memory-less process can be implemented form these components after imposing the relation between the matrices \hat{A} and \hat{R} .

In the Markovian case, the matrix \hat{R} is the identity matrix \hat{I} .

In the Markovian case, the matrix \hat{M} is specified as

$$m_{ij} = \delta / 19, \quad i \neq j, \tag{A7a}$$

$$m_{ij} = 1 - \delta, \quad i = j. \tag{A7b}$$

The matrices \hat{R} and \hat{A} are related after the frequencies π_i^A from [41].

In [34], a general reversible process model of Markov process models of nucleotide substitution is provided according to a 4×4 fundamental matrix of a Markov chain proposed from [35]; in this case, the peculiarities of the fundamental matrix allow one to infer that one of the parameters Π_i (from which the parameters π_j in [5] are generalized) is redundant.

It is now demonstrated how to analytically construct and write the chain of multiple sequencing.

A rectangular-matrix chain can be chosen, whose singular values must be compatible with those of the rectangular matrix from [5]. This becomes a very special case of the singular-value decomposition Equation (10). In this very special case, the conditions

$$V^T \equiv U^T \tag{A8}$$

with U from Equation (A5).

The two-state Markov models can therefore be extended to be comprehended within a rectangular matrix (of the chain) after adding Morse complexes by adding ∂_S^k . Thus, the acyclic partial matching holds for every pair of simplices.

The metric of the manifold is dictated after the qualities of the vector \vec{U} .

Appendix B.1. The Jukes–Cantor Model

The Jukes–Cantor model [3] and its applications [42] are analyzed here.

The Jukes–Cantor model is one according to which the probability of finding an amino-acid in a sequence is requested to be independent of the amino-acids. The probability matrix \hat{P} is written in the original form as

$$\hat{P} \equiv \{p_{ij}\}$$

$$\hat{P} = \begin{bmatrix} 1 - \delta & \delta/(n - 1) & \delta/(n - 1) & \delta/(n - 1) \\ \delta/(n - 1) & 1 - \delta & \delta/(n - 1) & \delta/(n - 1) \\ \delta/(n - 1) & \delta/(n - 1) & 1 - \delta & \delta/(n - 1) \\ \delta/(n - 1) & \delta/(n - 1) & \delta/(n - 1) & 1 - \delta \end{bmatrix}$$

with $n = 4$.

The analysis from [20] is spelled according to the present analysis.

The Jukes–Cantor model is applied in [5] as follows: the sequences are requested to be described with ‘dependence among neighboring nucleotides’. A transition probability matrix is given in a general reversible Markov model. The probability matrix \hat{P} is rewritten as

$$\hat{P} = e^{\hat{Q}t} \tag{A9}$$

with

$$\hat{W} = \vec{U}\hat{M}U^{-1} \tag{A10}$$

for $n = 20$ and $\delta = 0.01$; the probabilities $p_{ij}(t)$ are spelled as

$$p_{ij}(t) = \sum_{k=1}^{k=20} u_{ik}u_{kj}^{-1}e^{t\lambda_k} \tag{A11}$$

with λ_k being the eigenvalues of \hat{M} .

Therefore, the system is newly, pertinently analyzed.

Here, the probabilities $p_{ij}(t)$ are newly remarked to be dependent on the 20 eigenvalues λ_k , $k = 1, 2, \dots, 20$: the probabilities $p_K \equiv p_{kk}$ are written as

$$p_K = (o(\delta))^n e^{\lambda_K t} \tag{A12}$$

i.e.,

$$\ln p_K = t\lambda_K + no(\delta). \tag{A13}$$

The probabilities p_K are here newly remarked not to depend on the neighboring entries but only on the phenomenological evidence δ and on the dimensionality n of the system.

It is our aim to start implementing the Jukes–Cantor model from the model [6]: ibidem, the sequence of two amino-acids is considered as consisting of sites independent of the quality of the amino-acid; therefore, here, the Jukes–Cantor mode is newly implemented in rewriting Equations (A9) and (A10) as $\mathcal{W} = \vec{u}\mathcal{M}\vec{v}^*$ with $\vec{u} = (u_1, u_2)$ and $\vec{v} = (v_1, v_2)$, where the two different vectors \vec{u} and \vec{v} are kept different for the discussion of the metric.

This way, sequence 12 has a probability

$$p_{12} = ((1 - \delta)v_1 + \delta v_2)u_1 \tag{A14}$$

while p_{21} is

$$p_{21} = +(\delta v_1 + (1 - \delta)v_2)u_2 \tag{A15}$$

therefore, we newly remark that

$$p_{12} \neq p_{21} \quad (\text{A16})$$

Furthermore, the eigenvalues of \mathcal{M} are $\lambda_1 = 1$ and $\lambda_2 = 1 - 2\delta$.

It is proven by induction that the model for $n \geq 3$ does not admit $\lambda_i = 1, i = 1, 2, \dots, n$.

The metric of the manifold on which the graph associated with the Markov process is staying can be discussed. As already anticipated, the metric depends on the vector \vec{u} : as an example, if one takes $u_1 \equiv 1/\sqrt{2}$ and $u_2 \equiv 1/\sqrt{2}$, the metric is Hilbert. The two vectors can now be taken as equal, if needed, i.e., for the sake of [5].

Appendix B.2. Probability to Find an Amino-Acid Alone from a Profile MM

A further example of profile MM is here compared with that of [6] as one from [43]; in [43], a profile MM is considered, from which entries are to be eliminated that correspond to pseudogenes: here, it is newly suggested that the block of the profile MM be eliminated after applying $\partial^{-1 i}(\delta)$ from the model constructed after [6], with $i = 1$ or $i = 2$; because the results of the elimination of blocks are dependent on the label i of the blocks, it is newly suggested that eliminating the blocks pertinent to the amino-acid i from a profile MM allows one to write the probability to find the amino-acid j alone. The two probabilities are a priori different.

Appendix B.3. Constructing Deep Markov Models of Sequential Data from a Profile MM

We outline how to construct a deep Markov model of sequential data starting from a profile MM or from a pair MM.

If a Jukes–Cantor model has to be constructed, starting for the sequence 12, the prescriptions outlined throughout the paper are to be followed.

In particular, to start with, the 3×3 matrix \mathcal{M} has to be built from the 2×2 matrix \mathcal{M} after the pertinent blocks. The blocks are appended after adding one Morse complex with ∂_S^3 , with S being the opportune phenomenological specification.

It is remarked that the hMM from [6] is a profile MM or a pair MM.

Appendix B.4. Substituting Blacks in Deep Markov Sequential Models

In [35], the procedure about how to substitute a nucleotide in a sequence with another nucleotide is studied; the reversible Markov model is analyzed to be more efficient than the method outlined in [44]. A nucleotide is substituted in a sequence according to the following paradigm: the entries i corresponding to the nucleotide a have to be substituted with new entries i corresponding to the nucleotide b . The entries i_a are removed after applying $\partial^{-1 i_a}(S_{i_a})$ of the opportune sheaf, and the entries corresponding to the nucleotide b in the same position i are added after $\partial^{i_b}(S_{i_b})$ of the newly selected sheaf.

In [11], the procedures about how to substitute a nucleotide in a sequence are studied, where the probability matrix is a 4×4 matrix: the procedures is found to be dependent on 6 parameters. An exhaustive comparison of nucleotide-substitution methods is found in Appendix C.

Appendix C. Comparison of Nucleotide-Sequencing Models

Several most common nucleotide-substitution methods are illustrated, i.e., in [16] Figure 16.12 ibidem.

In [15] Table 1 ibidem, the fundamental matrices of the Markov chain of several nucleotide-substitution methods are spelled out. In particular, it is shown that every (known) method of nucleotide substitution is descended from the (12 parameters) GH4 model. The choices are commented as far as the expected number of substitutions per site is concerned. Numerical examples are given ibidem. The differences in the probability spaces are appreciated also in [45].

Starting from this viewpoint, the aim of the present Appendix is to analyze how the steps are obtained, i.e., to discriminate about whether the several models can be connected via the Morse operations or not.

For the present purposes, it is our aim to compare two different models and explain how they differ as far as the construction of the blocks of the fundamental matrix of the Markov chain and how the difference is unavoidably observed in the filtration of the resulting probability space. The numerical simulations from the items of the bibliography are also commented.

The methods for comparing synonymous nucleotide substitutions and non-synonymous nucleotide substitutions [46] are summarized in [47,48].

In [17], the neighbor-dependent nucleotide substitution methods are described as based on repetitive elements.

The numerical methods of how to choose the sheaf of the third block starting from a two-amino-acid sequence are discussed in [45]; in this case, the examples are issued from the known data.

The choice of the blocks is reflected in the possibility to evaluate distances between the points of the graphs (on the suitable manifold); if two models are not obtained after the application of the Morse operators, the notion of distance is not well-posed.

In [11], two models are presented. The comparison is proposed, according to which one substitution mode is claimed to be equivalent to the other according to some identifications; it is the present aim to prove that the comparison does not hold. As the two models are constructed according to a different application of the Morse operator, the filtrations of the resulting probability spaces are different and it is not possible to identify the two models. Nevertheless, it was remarked in [11] that some of the details were not codified yet. The sequence of nucleotides *ATCG* is considered.

The four-parameter model of fundamental matrix \hat{Q}_{K1981} from [13] is spelled as

$$\hat{Q}_{K1981} \equiv \{q_{ij}\} \tag{A17}$$

$$\hat{Q}_{K1981} = \begin{bmatrix} -(\gamma + \theta\alpha + \alpha) & \gamma & \theta\alpha & \alpha \\ \gamma & -(\gamma + \theta\alpha + \alpha) & \alpha & \theta\alpha \\ \theta\beta & \beta & -(\gamma + \theta\beta + \beta) & \gamma \\ \beta & \theta\beta & \gamma & -(\gamma + \theta\beta + \beta) \end{bmatrix}$$

According to the analysis developed in the present paper, it is understood that the matrix \hat{Q}_{K1981} is built from the Jukes–Cantor model as far as the *AT* sequence is concerned; later on, the further blocks q_{ij} with $j = 3, 4$ are constructed according to the Morse operator after the choice of the opportune sheaf.

It is important to remark that the construction does not reflect the construction proposed after [6].

In [11], the six-parameter model is recalled from [14], whose fundamental Markov matrix \hat{Q}_{GNI} is spelled as

$$\hat{Q}_{GNI} \equiv \{q_{GNI\ ij}\} \tag{A18}$$

$$\hat{Q}_{GNI} = \begin{bmatrix} -(2\alpha + \alpha_1) & \alpha_1 & \alpha & \alpha \\ \beta_1 & -(2\alpha + \beta_1) & \alpha & \alpha \\ \beta & \beta & -(2\beta + \alpha_2) & \alpha_2 \\ \beta & \beta & \beta_2 & -(2\beta + \beta_2) \end{bmatrix}$$

The model from [14] is inequivalent to the Jukes–Cantor model, as the two-nucleotide sequence does not admit the eigenvalue 1 (with the opportune relabeling of the variables, as it is possible for \hat{Q}_{K1981}); moreover, the further blocks $q_{GNI\ ij}$ with $j = 3, 4$ are not added according to the prescription of the Morse operator. This way, it is understood that the probability space of the two-nucleotide *AT* model is not contained in the probability space of the four-nucleotide model. From the qualities of the filtration, it is also remarked that a comparison with [6] is not possible.

Therefore, the nucleotide substitution method proposed in [11] does not generalize the Jukes–Cantor model, even if the details of the comments in [11] are applied. The inconsistency is further pointed out in [49].

The system can be compared to the method exposed in [50], according to which phylogenetically independent pairwise sequences are used. The Jukes–Cantor model is therefore understood as the application of the choice of a constant sheaf apt to reproduce the same filtration.

References

1. Kashlak, A.B.; Loliencar, P.; Heo, G. Topological Hidden Markov Models. *J. Mach. Learn. Res.* **2023**, *24*, 1–49.
2. Nanda, V. Computational Algebraic Topology Lecture Notes, e-Print. Available online: <https://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf> (accessed on 4 November 2024).
3. Jukes, T.H.; Cantor, C.R. Evolution of protein molecules. In *Mammalian Protein Metabolism*; Munro, H.N., Ed.; Academic Press: New York, NY, USA, 1969; Volume III, pp. 21–132.
4. Adachi, J.; Hasegawa, M. *MOLPHY Version 2.3 Programs for Molecular Phylogenetics Based on Maximum Likelihood*; The Institute of Statistical Mathematics 4-6-7 Minami-Azabu: Tokyo, Japan, 1996. Available online: <https://stat.sys.i.kyoto-u.ac.jp/titech/class/doc/csm96.pdf> (accessed on 4 November 2024).
5. Adachi, J.; Hasegawa, M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* **1996**, *42*, 459. [[CrossRef](#)]
6. Henderson, J.; Salzberg, S.; Fasman, K.H. Finding genes in DNA with a Hidden Markov Model. *J. Comput. Biol.* **1997**, *4*, 127–141. [[CrossRef](#)]
7. Available online: <https://www.cs.princeton.edu/~mona/Lecture/HMM1.pdf> (accessed on 4 November 2024).
8. Helder, I.N. (Ed.) *Chapter 9 Rational Design of Profile Hidden Markov Models for Viral Classification and Discovery* Liliane Santana Oliveira and Arthur Gruber; Exon Publications: Brisbane, Australia, 2021.
9. Anderson, G.; Farcomeni, A.; Pittau, M.G.; Zelli, R. Rectangular latent Markov models for time-specific clustering, with an analysis of the well being of nations. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2019**, *68*, 603. [[CrossRef](#)]
10. Hoff, P.D. Model averaging and dimension selection for the singular value decomposition. *arXiv* **2007**. arXiv:math/0609042. [[CrossRef](#)]
11. Takashi Gojobori, T.; Ishii, K.; Nei, M. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J. Mol. Evol.* **1982**, *18*, 414. [[CrossRef](#)]
12. Fuellgrabe, J.; Gosal, W.S.; Creed, P.; Liu, S.; Lumby, C.K.; Morley, D.J.; Ost, T.W.; Vilella, A.J.; Yu, S.; Bignell, H.; et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. *Nat. Biotechnol.* **2023**, *41*, 1457. [[CrossRef](#)] [[PubMed](#)]
13. Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 454–458. [[CrossRef](#)] [[PubMed](#)]
14. Gojobori, T.; Nei, M.; Ishii, K. Mathematical model of nucleotide substitutions with unequal substitution rates. *Genetics* **1981**, *97*, s43.
15. Rodríguez, F.; Olivera, J.L.; Maran, A.; Medina, J.R. The General Stochastic Model of Nucleotide Substitution. *J. Theory Biol.* **1990**, *142*, 485–501. [[CrossRef](#)] [[PubMed](#)]
16. Saitou, N. Evolutionary Distances. In *Introduction to Evolutionary Genomics, Chapter 16. Computational Biology*; Springer: London, UK, 2013; Volume 17.
17. Arndt, P.F.; Hwa, T. Identification and measurement of neighbor-dependent nucleotide substitution processes. *Bioinformatics* **2005**, *21*, 2322–2328. [[CrossRef](#)] [[PubMed](#)]
18. Gojobori, T.; Li, W.H.; Graur, D. Patterns of Nucleotide Substitution in Pseudogenes and Functional Genes. *J. Mol. Evol.* **1982**, *18*, 360–369. [[CrossRef](#)]
19. Gong, T.; Zhang, W.; Chen, Y. Uncovering block structures in large rectangular matrices. *J. Multivar. Anal.* **2023**, *198*, 105211. [[CrossRef](#)]
20. Lecian, O.M. The chains of the Clusters of latent states in DNA sequencing. *J. Med. Care Health Rev.* **2004**, *1*, 1.
21. Yoon, B.-J.; Vaidyanathan, P.P. Context-Sensitive Hidden Markov Models for Modeling Long-Range Dependencies in Symbol Sequences. *IEEE Trans. Signal Process.* **2006**, *54*, 4169–4184. [[CrossRef](#)]
22. Available online: <https://www.ebi.ac.uk/training/online/courses/pfam-creating-protein-families/what-are-profile-hidden-markov-models-HMM's/> (accessed on 4 November 2024).
23. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [[CrossRef](#)] [[PubMed](#)]
24. Krogh, A.; Brown, M.; Mian, I.S.; Sjoelander, K.; Haussler, D. Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.* **1994**, *235*, 1501–1531. [[CrossRef](#)]
25. Available online: <https://web.stanford.edu/class/cs262/archives/notes/lecture8.pdf> (accessed on 4 November 2024).
26. Available online: <https://web.stanford.edu/class/cs262/presentations/lecture8.pdf> (accessed on 4 November 2024).
27. Available online: <https://www.cs.cmu.edu/~10715-f18/lectures/lecture2-crf.pdf> (accessed on 4 November 2024).

28. Krishnan, R.; Shalit, U.; Sontag, D. Structured Inference Networks for Nonlinear State Space Models. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), San Francisco, CA, USA, 4–9 February 2017.
29. Available online: <https://pyro.ai/examples/dmm.html> (accessed on 4 November 2024).
30. Masi, G.B.D.; Stettner, L. Ergodicity of hidden Markov models. *Math. Control. Signals Syst.* **2005**, *17*, 269–296. [[CrossRef](#)]
31. Bayer, J.; Osendorfer, C. Learning stochastic recurrent networks. *arXiv* **2014**. arXiv:1411.7610
32. Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A.; Bengio, Y. A recurrent latent variable model for sequential data. *Adv. Neural Inf. Process. Syst.* **2015**, *28*.
33. Hinton, G.E.; Dayan, P.; Frey, B.J.; Neal, R.M. The “wake-sleep” algorithm for unsupervised neural networks. *Science* **1995**, *268*, 1158–1161. [[CrossRef](#)]
34. Tavaré, S. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57.
35. Yang, Z. Estimating the Pattern of Nucleotide Substitution. *J. Mol. Evol.* **1994**, *39*, 105. [[CrossRef](#)] [[PubMed](#)]
36. Duboc, L.; Rosenblum, D.; Wicks, T. A framework for characterization and analysis of software system scalability. In Proceedings of the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT International Symposium on Foundations of Software Engineering, Dubrovnik, Croatia, 3–7 September 2007.
37. Arora, S.; Barak, B. Computational Complexity: A Modern Approach. In *Chapter: The Computational Complexity and Why It Does Not Matter*; Cambridge University Press: Cambridge, UK, 2009.
38. Torra, V.; Taha, M.; Navarro-Arribas, G. The space of models in machine learning: Using Markov chains to model transitions. *Prog. Artif. Intell.* **2021**, *10*, 321–332. [[CrossRef](#)]
39. Duboc, A.L.d.L. A Framework for the Characterization and Analysis of Software Systems Scalability. Ph.D. Thesis, University College London, London, UK, 2009.
40. Dragon'a, J.; Mukherjee, S.; Zhang, J.; Liu, F.; Halappanavar, M. On the Stochastic Stability of Deep Markov Models. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24033–24047.
41. Dayhoff, M.O.; Schwartz, R.M.; Orcutt, B.C. A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*; Dayhoff, M.O., Ed.; National Biomedical Research Foundation: Washington, DC, USA, 1978; pp. 345–352. Volume 5.
42. Lecture Notes: Markov models of sequence evolution Dannie Durand.
43. Porter, T.M.; Hajibabaei, M. Profile hidden Markov model sequence analysis can help remove putative pseudogenes from DNA barcoding and metabarcoding datasets. *BMC Bioinform.* **2021**, *22*, 256. [[CrossRef](#)]
44. Hasegawa, M.; Kishino, H.; Yano, T. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **1985**, *22*, 160–174. [[CrossRef](#)]
45. Andrey, Z. Estimation of Evolutionary Distances Between Nucleotide Sequences. *J. Mol. Evol.* **1994**, *39*, 315–329.
46. Ina, Y. New Methods for Estimating the Numbers of Synonymous and Nonsynonymous Substitutions. *J. Mol. Evol.* **1995**, *40*, 190–226. [[CrossRef](#)] [[PubMed](#)]
47. Miyata, T.; Yasunaga, T. Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* **1980**, *16*, 23–26. [[CrossRef](#)] [[PubMed](#)]
48. Li, W.-H.; Wu, C.-I.; Luo, C.-C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **1985**, *2*, 150–174. [[PubMed](#)]
49. Gojobori, T.; Moriyama, E.N.; Kimura, M. Statistical methods for estimating sequence divergence. *Methods Enzymol.* **1990**, *183*, 531–550.
50. Purvis, A.; Bromham, L. Estimating the transition/transversion ratio from independent pairwise comparisons with an assumed phylogeny. *J. Mol. Evol.* **1997**, *44*, 112. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.