Proceedings of the 31st Conference
on Decision and Control
Tucson, Arizona • December 1992

Identification of Coupled Markov Chain Model with Application

Rodney A. Kennedy
Department of Systems Engineering
Research School of Physical Sciences and Engineering
Australian National University
GPO Box 4
Canberra ACT 2601 Australia.

Shin-Ho Chung
Protein Dynamics Unit
Faculty of Chemistry
Australian National University
GPO Box 4
Canberra ACT 2601 Australia.

Supported by Clive and Vera Ramaciotti Foundations, NH & MRC of Australia and the Australian Research Council.

Abstract

A parameterized Markov chain model is developed to model the action of a biological ion channel measured in noise. In detail, the model takes the form of a set of binary chains which are interdependent according to a coupling parameter. When varied this parameter realises the range of behaviors from tight coupling to complete independence. Other parameters describe the intrinsic characteristics of the binary chains as well as their behaviour when fully coupled. An identification procedure for the model parameters is developed based on hidden Markov modelling ideas but incorporating a novel parameter estimation step (due to the nonlinear form of the dependency of the model on the parameters). The model and identification methods are tested on real data.

1 Introduction

This paper presents an identification problem motivated from a specific application. We open with a brief account of the physical system under study and then outline the contents of the paper.

A current dual problem in biology is: (i) performing difficult low level measurements of ion currents which occur across membranes of neurons and (ii) the subsequent explanation of the observed phenomena [1, 2, 3]. The ultimate goal is, naturally, to explain on physical grounds the mechanisms observed. As a preliminary step towards such a physical model a simplified engineering model is developed in this paper. The first part of the paper is devoted towards developing such a model which takes the form of a set of partially coupled Markov chains. In explaining the flow of ions, the conceptual model is that of a (biological) channel composed of a number of pores each of which may open and close. With the action of each pore a simple binary Markov chain is associated and these are coupled on account of their close physical proximity, *i.e.*, if one pore closes then neighbouring pores have a greater tendency to close. There are a number of precedents to the use of Markov chains in the study of such physical systems recently in the biological literature, see [4, 5, 6].

Of greater interest to the engineer is the explicit model development and the subsequent identification of the model parameters. The model is treated in section 2, and the identification in section 3. Processing of real data is treated in section 4.

2 Problem Formulation

Introductory Remarks

This section develops an abstract model. In the context of the application outlined in section 1 it is useful to connect some of the abstract objects to follow with features from the physical channel under study.

We model a single channel as a collection of L *pores* in close vicinity. Each pore will be modelled by a simple binary (open and closed) Markov chain, denoted $s_k^{(r)}$ where $1 \leq r \leq L$ and k denotes discrete time. When open the conductance level of a single pore will be denoted q_1 and when closed q_2. The collection of L pores can then take any configuration corresponding to all variations of open and closed pores. The *state* of the process at time k will be denoted S_k and may be thought of as a binary vector (labelled with a corresponding decimal number in the range $(0, 1, \cdots, 2^L - 1)$ which we call the index).

In terms of the observations which are made of the total conductance of the channel (measured in noise), *i.e.*, the sum of the conductances of the L pores, it is possible to develop a second Markov process which we called an aggregated Markov chain and is central to our work. The details to follow will be punctuated with interpretations based on the multi-pore channel model described above.

Coupled Markov Chain Model

Binary Vector Chain Process

We begin with a mathematical formulation of the signal model on interest. Consider L identical discrete-time, binary, homogeneous Markov processes denoted $s_k^{(1)}, \cdots, s_k^{(L)}$, such that, at each time k, $s_k^{(r)}$, $r = 1, \cdots, L$, is a binary random variable taking on one of two states in the set $\mathbf{q} \triangleq \{q_1, q_2\}$. These two states represent open and closed states. Define the vector process (which describes the emsemble of the L scalar processes)

$$S_k \triangleq \left(s_k^{(1)}, s_k^{(2)}, \cdots, s_k^{(L)}\right)$$

with state space

$$\mathbf{q}^L \triangleq \mathbf{q} \times \mathbf{q} \times \cdots \times \mathbf{q},$$

i.e., $\mathbf{q}^L$ is the L times Cartesian product of $\mathbf{q}$. Note that the set $\mathbf{q}^L$ has $N = 2^L$ elements. We will index each of these states according to a binary ordering:

Index	State
0	$(q_1, q_1, \cdots, q_1)$
1	$(q_2, q_1, \cdots, q_1)$
2	$(q_1, q_2, \cdots, q_1)$
3	$(q_2, q_2, \cdots, q_1)$
$\vdots$	$\vdots$
$N-1=2^L-1$	$(q_2, q_2, \cdots, q_2)$

Each of the identical binary processes will be modelled as a Markov chain, with transition probabilities

$$v_{ij} \triangleq P\Big(s_{k+1}^{(r)} = q_j \Big|\ s_k^{(r)} = q_i\Big),\ i,j \in \{1,2\},$$

for all $1 \le r \le L$. The corresponding 2×2 transition probability matrix will be explicitly represented

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \triangleq \begin{bmatrix} \zeta & 1-\zeta \\ 1-\rho & \rho \end{bmatrix} \tag{2.1}$$

and is sufficiently parameterized, as indicated, by two parameters ζ and ρ.

Let $\mathbf{P}^{(G)} = (p_{mn}^{(G)})$ denote a generic transition probability matrix of the vector binary process S_k, *i.e.*, for $Q_m \triangleq (q_{i_1}, \cdots, q_{i_L})$ and $Q_n \triangleq (q_{j_1}, \cdots, q_{j_L})$,

$$\begin{aligned} p_{mn}^{(G)} &\triangleq P\Big(S_{k+1} = Q_n \Big|\ S_k = Q_m\Big) \\ &= P\Big(s_{k+1}^{(1)} = q_{j_1}, \cdots, s_{k+1}^{(L)} = q_{j_L} \Big| \\ &\qquad s_k^{(1)} = q_{i_1}, \cdots, s_k^{(L)} = q_{i_L}\Big). \end{aligned} \tag{2.2}$$

Then $\mathbf{P}^{(G)}$ is a square matrix with 4^L elements.

Uncoupled Case

We can determine the specific transition probability matrix when the constituent components of the vector process are *independent*, as a special case of (2.2). That is, $\mathbf{P}^{(I)} = (p_{mn}^{(I)})$ is defined as the L times tensor (Kronecker) product of $\mathbf{V}$ (2.1):

$$\mathbf{P}^{(I)} \triangleq \underbrace{\mathbf{V} \otimes \mathbf{V} \otimes \cdots \otimes \mathbf{V}}_{L \text{ times}} \tag{2.3}$$

where $\otimes$ denotes the tensor product. Alternatively, this means the components of (2.3) can be factored as

$$p_{mn}^{(I)} = \prod_{r=1}^{L} P\Big(s_{k+1}^{(r)} = q_{j_r} \Big|\ s_k^{(r)} = q_{i_r}\Big).$$

As an example, with $L = 2$ and $\mathbf{V}$ as in (2.1) then $\mathbf{V} \otimes \mathbf{V}$ is given by

$$\begin{bmatrix} \zeta^2 & \zeta(1-\zeta) & (1-\zeta)\zeta & (1-\zeta)^2 \\ \zeta(1-\rho) & \zeta\rho & (1-\zeta)(1-\rho) & (1-\zeta)\rho \\ (1-\rho)\zeta & (1-\rho)(1-\zeta) & \rho\zeta & \rho(1-\zeta) \\ (1-\rho)^2 & (1-\rho)\rho & \rho(1-\rho) & \rho^2 \end{bmatrix}. \tag{2.4}$$

In the context of the multi-pore channel model this means individual pores open and close without interacting with neighbouring pores. An example of a real channel composed of 2 pores where such an uncoupled model is identified was found to be evident in real data (see section 4).

Coupled Case

A second special case of (2.2), which we denote as $\mathbf{P}^{(C)}$, plays an important role. This (vector) chain represents the situation where the L binary chains are tightly coupled,

$$\mathbf{P}^{(C)} \triangleq \begin{bmatrix} \xi & 0 & \cdots & 0 & 1-\xi \\ 1-\delta & 0 & \cdots & 0 & \delta \\ \vdots & \vdots & & \vdots & \vdots \\ 1-\delta & 0 & \cdots & 0 & \delta \\ 1-\eta & 0 & \cdots & 0 & \eta \end{bmatrix}. \tag{2.5}$$

This is described by three parameters, ξ, η and δ. The only recurrent vector states are $(q_1, \cdots, q_1)$ (corresponding to all pores open) and $(q_2, \cdots, q_2)$ (corresponding to all pores closed). Therefore, after any transient, the vector chain mimicks the 2 state chain (2.1) except with parameters ξ and η replacing ζ and ρ. Finally, the parameter δ (also a probability) determines the behaviour of the (at most one step) transient in a natural way.

Partial Coupling

In the most useful model that we consider the constituent chains need not be independent (2.3) nor fully coupled (2.5), leading to the transition probability matrix

$$\mathbf{P} \triangleq (1-\kappa)\mathbf{P}^{(I)} + \kappa\mathbf{P}^{(C)}, \qquad 0 \le \kappa \le 1. \tag{2.6}$$

where κ (a probability) is the coupling factor. Thus $\mathbf{P}$ can be parameterized by the set

$$\Theta \triangleq \Big\{\zeta, \rho, \xi, \eta, \delta, \kappa\Big\}. \tag{2.7}$$

(The number of channels, L, also implicitly parameterizes the model.) Despite the relatively low number of parameters we will see that a very rich range of qualitative behaviours is possible.

Aggregated Markov Chain

In this section we will see that because of the nature of the observation process (to be introduced) it is possible to simplify (particularly in terms of level of computation required) the above Markov chain model to a significantly lower order one via an aggregation procedure. That is, we group together the binary vector states into aggregated states in a natural way.

Motivated by our application we define a scalar finite-state Markov process corresponding to a noiseless observation of the form

$$z_k \triangleq \mathbf{1}^T S_k = \sum_{r=1}^{L} s_k^{(r)}, \tag{2.8}$$

i.e., the inner product of the vector S_k and the vector with all elements equal to one. The state space of z_k then consists of $L+1$ distinct values in the range

$$u_0 \triangleq Lq_1$$

$$u_1 \triangleq (L-1)q_1 + q_2$$
$$u_2 \triangleq (L-2)q_1 + 2q_2,$$
$$\vdots \quad \vdots$$
$$u_L \triangleq Lq_2$$

This output process (2.8) adds the conductances of the L pores. Recall also that, in our previous modelling, all the L pores are assumed identical. The marriage of these two facts makes it plain that on the basis of measurements of the (potentially noisy) observations of z_k it will be impossible to distinguish between channel configurations with the same number of open and closed pores (For example, if a maximum likelihood identification were done, then each of the configurations would have identical likelihood functions.) This leads to the notion of Markov chain formed by aggregating states (although at this stage it is not clear that such an aggregation preserves the important Markov property—this is established in Appendix A).

A separate motivation for considering the aggregation is that the dimensionality of the problem is reduced considerably. A naive application of (2.6) would have a transition probability matrix involving 4^L elements; whereas if we define z_k (2.8) as the state at time k the transition probability matrix has only $(L+1)^2$ elements. For the range of L values that we wish to consider the latter dimension is easily handled.

In the light of the output equation (2.8), which defines the process z_k, we define aggregated states:

> **Aggregated State:** A state S_k belongs to **aggregated state** i if there are i pores which are closed, *i.e.*, exactly i pores with conductance q_2 and $L-i$ pores with conductance q_1. As such, there are $\binom{L}{i}$ (binary vector) states in aggregated state i.

In other words, states which are equivalent up to permutations of the binary components belong to the same aggregated state (equivalence class). The aggregated states are indexed by $i \in \{0, 1, 2, \cdots, L\}$. Aggregated state 0 corresponds to $(q_1, q_1, \cdots, q_1)$ and aggregated state L corresponds to $(q_2, q_2, \cdots, q_2)$. As a further explicit example, if $L = 3$ then we aggregate the $2^L = 8$ S_k states into the following $L+1 = 4$ aggregated states:

i	Set of States
0	$\{(q_1, q_1, q_1)\}$
1	$\{(q_2, q_1, q_1), (q_1, q_2, q_1), (q_1, q_1, q_2)\}$
2	$\{(q_2, q_2, q_1), (q_2, q_1, q_2), (q_1, q_2, q_2)\}$
3	$\{(q_2, q_2, q_2)\}$

Finally in terms of the output measurement process we have the simple observation

$$z_k = u_i \iff S_k \in \text{aggregated state } i \iff \text{exactly } i \text{ pores closed.} \quad (2.9)$$

It is possible to express the transition matrix of the aggregated system in terms of the transition matrix of the nonaggregated system (2.6). The $(L+1) \times (L+1)$ aggregated probability transition matrix $\mathbf{A}$ corresponding to (2.6) takes the form:

$$\mathbf{A} \triangleq \mathbf{LPR} = (1-\kappa)\underbrace{\mathbf{LP}^{(I)}\mathbf{R}}_{\mathbf{A}^{(I)}} + \kappa\underbrace{\mathbf{LP}^{(C)}\mathbf{R}}_{\mathbf{A}^{(C)}} \quad (2.10)$$

where $\mathbf{L}$ is an $(L+1) \times 2^L$ matrix with components

$$l_{ij} \triangleq \begin{cases} \binom{L}{i}^{-1} & \text{if state } j \in \text{aggregated state } i \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbf{R}$ is a $2^L \times (L+1)$ matrix with components

$$r_{ji} \triangleq \begin{cases} 1 & \text{if state } j \in \text{aggregated state } i \\ 0 & \text{otherwise.} \end{cases}$$

(This matrix, $\mathbf{L}$, is nonunique, *i.e.*, there exists an infinite number of matrices $\tilde{\mathbf{L}}$ such that $\tilde{\mathbf{L}}\mathbf{PR} = \mathbf{LPR}$. However, the above choice is convenient.) Here we note that under aggregation $\mathbf{LP}^{(C)}\mathbf{R}$ maintains the form of (2.5) save that the dimensions are $(L+1) \times (L+1)$ rather than $2^L \times 2^L$. However $\mathbf{LP}^{(I)}\mathbf{R}$ differs in form from (2.3), *e.g.*, if $L = 2$ then $\mathbf{V} \otimes \mathbf{V}$ is given by (2.4) whereas

$$\mathbf{A}^{(I)} \triangleq \mathbf{L}\Big(\mathbf{V} \otimes \mathbf{V}\Big)\mathbf{R} = \quad (2.11)$$

$$\begin{bmatrix} \zeta^2 & 2\zeta(1-\zeta) & (1-\zeta)^2 \\ \zeta(1-\rho) & \zeta\rho + (1-\zeta)(1-\rho) & (1-\zeta)\rho \\ (1-\rho)^2 & 2(1-\rho)\rho & \rho^2 \end{bmatrix}.$$

At this point we note that evaluation of the components of $\mathbf{A}$ according to (2.10) is somewhat impractical for large L because of the computation of the uncoupled component $\mathbf{A}^{(I)}$ which involves the Kronecker product expression. In Appendix B we show how to compute the coefficients with a simple recursion which is important for software development.

Measurement Noise

Next we assume that the chain z_k is hidden, that is, indirectly observed by measurements y_k of the form

$$y_k = z_k + w_k. \quad (2.12)$$

where w_k is zero mean gaussian noise of variance σ_w^2. Define the vector of conditional probability functions $\mathbf{b}(\sigma_w^2, \mathbf{q}, y_k) = (b_i(\sigma_w^2, \mathbf{q}, y_k))$ where

$$b_i(\sigma_w^2, \mathbf{q}, y_k) \triangleq P(y_k | z_k = u_i) = \frac{1}{\sqrt{2\pi}\,\sigma_w} e^{-\frac{(y_k - u_i)^2}{2\sigma_w^2}} \quad (2.13)$$

for $0 \le i \le L$. We denote the sequence of observations $\{y_1, y_2, \cdots, y_k\}$ by Y_k. The problem takes the form of a hidden Markov model (HMM, [7]) because w_k is white, leading to the independence property

$$P(y_k |\, z_k = u_i, z_{k-1} = u_j, Y_{k-1}) = P(y_k |\, z_k = u_i), \quad (2.14)$$

(see Appendix A). Also we assume that the initial state probability vector $\pi = (\pi_m)$ is defined from $\pi_i = P(z_1 = u_i)$.

3 Parameter Estimation

Overview

Now we address the problem of how to estimate the parameters of our model given real data measurements. Our emphasis lies in defining a procedure that obtains the estimates rather than focusing on the detailed development and analysis of the technique employed. We will, therefore, rely heavily on the literature and focus only on the novel aspects.

We summarize the estimation procedure as follows:

Two Step Estimation:

(i) Classical HMM identification [7] of the transition probability matrix and associated relevant parameters, (*e.g.*, output levels q_1, q_2 and noise variance σ_w^2).

(ii) Novel optimal parameter fit according to the partially coupled model using gradient descent adaptation.

As motivation for the above, we note that the first step is independent of the "preferred" Markov model parameterization (2.6), *i.e.*, it imposes no bias towards the structure (2.10) we seek. Incorporating the first step also has the advantage of using existing theory (and software) with no modification. The second step is a well defined procedure which may be employed on more general systems than those that we consider, *i.e.*, it is not particularly restricted to the parameterization that we have developed. This feature enables the procedure to be easily modified in the event that, for example, the model were to be extended or changed.

Hidden Markov Modelling

We begin with a brief decription of the standard HMM formulation which we utilize. The complete HMM parameter set is usually denoted

$$\lambda \triangleq \Big\{\mathbf{A}, \mathbf{b}, \pi\Big\}.$$

The matrix $\mathbf{A}$ is the transition probability matrix with entries. (To accommodate the standard matrix indexing conventions both the i and j indices can be incremented.)

$$a_{ij} \triangleq P\Big(z_{k+1} = u_j \Big|\ z_k = u_i\Big),\ \forall k,$$

where $i, j \in \{0, 1, 2, \cdots, L\}$, which gives complete information regarding ths statistical properties the transition between the states. The vector $\mathbf{b}$ was defined in (2.13) and depends on the parameters $\{q_1, q_2, \sigma_w\}$. This vector relates actual observations to the hidden state. The third element of λ, π, defines the initial probability distribution across the states.

Some important things to note regarding standard HMM theory are:

(i) extensive work has been done developing computationally tractable recursions for evaluating maximum likelihood parameter estimates;

(ii) these recursions crucially rely on estimating parameters corresponding to the "naive" parameterization, for example, the parameters a_{ij} are directly estimated;

(iii) the estimation objective is to find the model estimate $\widehat{\lambda} = \Big\{\widehat{\mathbf{A}}, \widehat{\mathbf{b}}, \widehat{\pi}\Big\}$ that maximizes the probability of the complete observation record Y_K, *i.e.*,

$$\widehat{\lambda} \triangleq \Big\{\widehat{\mathbf{A}}, \widehat{\mathbf{b}}, \widehat{\pi}\Big\} = \arg\max_{\widetilde{\lambda}} P\Big(Y_K \Big|\ \widetilde{\lambda}\Big). \quad (3.1)$$

In terms of the parameters that we seek the parameters q_1, q_2 and σ_w^2 can be directly estimated from the above standard HMM framework (also the number of pores, L, can be estimated using similar ideas), whereas we need new methods to deal with the problem of estimating the parameters in Θ which enter in a nonlinear fashion. This estimation procedure is treated next.

Parameter Fitting

In the above HMM parameter estimation problem we assume we have available optimal estimates of the transition probability matrix coefficients a_{ij}. Since our model developed in section 2 impinges primarily on the transition probability matrix then estimating the parameters in the set Θ (2.7) concern only the a_{ij} parameters. Let the maximum likelihood matrix estimate generated from the data by the HMM processing have components given by $\widehat{a}_{ij}$ (3.1) and let the partially coupled Markov chain model have transition probability matrix denoted $a_{ij}(\Theta)$. We seek to estimate Θ.

We pose the following parameter fitting problem:

$$\begin{aligned}\Theta^\star &\triangleq \arg\min_{\Theta} \frac{1}{2}\Big\|\mathbf{A}(\Theta) - \widehat{\mathbf{A}}\Big\|_F^2 \\ &= \arg\min_{\Theta} \underbrace{\frac{1}{2}\sum_{i=0}^{L}\sum_{j=0}^{L}\Big(a_{ij}(\Theta) - \widehat{a}_{ij}\Big)^2}_{\mathcal{J}(\Theta)} \end{aligned} \quad (3.2)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix (which is just a euclidean norm of the matrix interpreted as a stacked vector).

The explicit solution of the above problem is unavailable since the parameters Θ enter in a nonlinear fashion. However, a simple adaptive procedure has been developed which solves the problem iteratively. The technique is a gradient descent (search) strategy. Explicitly, consider the difference equation

$$\widetilde{\Theta}(k+1) = \widetilde{\Theta}(k) - \mu\ \frac{\partial \mathcal{J}}{\partial \Theta}\bigg|_{\Theta=\widetilde{\Theta}(k)} \quad (3.3)$$

where

$$\frac{\partial \mathcal{J}}{\partial \Theta} \equiv \nabla_\Theta \mathcal{J} = \left[\ \frac{\partial \mathcal{J}}{\partial \zeta}\ \ \frac{\partial \mathcal{J}}{\partial \rho}\ \ \frac{\partial \mathcal{J}}{\partial \xi}\ \ \frac{\partial \mathcal{J}}{\partial \eta}\ \ \frac{\partial \mathcal{J}}{\partial \delta}\ \ \frac{\partial \mathcal{J}}{\partial \kappa}\ \right]'.$$

Using (3.2) the partial derivatives in turn are written

$$\frac{\partial \mathcal{J}}{\partial \zeta} = \sum_{i=0}^{L}\sum_{j=0}^{L}\Big(a_{ij}(\Theta) - \widehat{a}_{ij}\Big)\frac{\partial a_{ij}(\Theta)}{\partial \zeta}$$

and similarly with respect to the remaining parameters in Θ. Finally in (3.3), μ represents a small positive stepsize parameter.

At this point we highlight the need for efficient means to compute both the components $a_{ij}(\Theta)$ and the partial derivatives, *e.g.*, $\partial a_{ij}(\Theta)/\partial\zeta$, above. This important numerical aspect dealing with the development of recursive formulae is treated in Appendix B.

When running the update equation (3.3) we initialize with

$$\widetilde{\Theta}(0) \triangleq [\ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ \ 0.5\ ]'$$

since the parameters to be estimated are known to be probabilities, *i.e.*, take values between 0 and 1. The stepsize is typically taken as $\mu = 0.01$ or smaller (to ensure numerical stability). Also in the algorithm we project the parameters of $\widehat{\Theta}$ to the range $[0, 1]$ to ensure that they remain probabilities.

4 Real Biological Data

Figure 1 shows a segment of a real data record of the channel current from an excised membrane patch of a cultured hippocampal neuron. The pores, when open, permitted the flow of sodium ions across the membrane resulting in the upward deflections in the current trace. The record shows clearly 3 uniformly distributed levels which implies, by the previous theory, that the number of pores is $L = 2$. Hidden Markov Model identification techniques for the full 200 000 sample measurement yielded the transition probability matrix:

$$\begin{pmatrix} 0.9359 & 0.0622 & 0.0019 \\ 0.0179 & 0.9421 & 0.0399 \\ 0.0002 & 0.0306 & 0.9692 \end{pmatrix}.$$

The parameter identification stage yielded the estimates: $\zeta = 0.967$, $\rho = 0.984$, $\kappa = 0.011$, $\xi = 0.921$, $\eta = 0.991$ and $\delta = 0.730$. This measurement indicates that the 2 pores involved were acting essentially independently.

5 Conclusions

We have developed a Markov chain model which incorporates partial coupling amongst a set of binary Markov chains. (Such binary chains can be given a simple physical interpretation as the opening and closing, in a probabilistic manner, of a pore in a biological membrane.) We developed a parametric model to describe these dynamics, and then moved on to show how the parameters could be estimated by combining standard HMM identification techniques and gradient descent adaptation in a novel way.

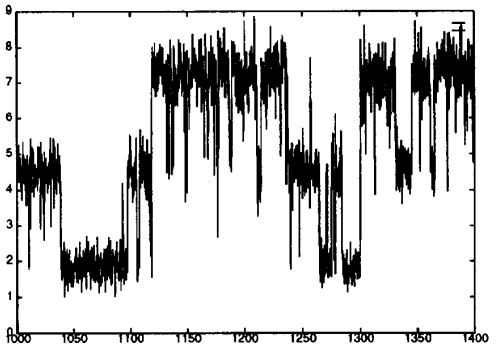

Figure 1: Neuron Membrane Ion Current vs Time

The model and identification techniques were developed with a view to a specific application in processing laboratory measurements of the current in an ion channel. We applied the methods to real biological data. The success of the identification is seen as the first stage in developing a physical model of the mechanism of such biological channels.

References

[1] Krouse, M.E., Schneider, G.T., and Gage, P.W., "A large anion-selective channel has seven conductance levels", *Nature* 319, pp.58-60, 1986.

[2] Premkumar, L.S., Gage, P.W., and Chung, S.H., "Coupled potassium channels induced by arachidonic acid in cultered neurons", *Proc. R. Soc. Lond. B* B 242, pp.17-22, 1990.

[3] Hymel, L., Strriessnig, J., Glossmann, H, and Schindler, H., "Purified skeletal muscle 1,4-dihydropyridine receptor forms phosphorylation dependent oligomeric calcium channels in planar bilayers", *Proc. Natl. Acad. Sci. U.S.A.* 85, pp.4290-4294, 1988.

[4] Colquhoun, D., and Sigworth, F.J., "Fitting and statistical analysis of single channel records.", In: B. Sakmann and E. Neher (Eds.), Single Channel Recording, Plenum, New York, pp.191-263, 1983.

[5] Chung, S. H., Krishnamurthy, V., and Moore J.B., "Adaptive processing techniques based on Hidden Markov Models for characterizing very small channel currents buried in noise and deterministic interferences", *Phil. Trans. R. Soc. Lond.* B 334, pp.357-384, 1991.

[6] Colquhoun, D., and Hawkes, A.G., "On the stochastic properties of single ion channels", *Phil. Trans. R. Soc. Lond.* B 334, pp.205-235, 1981.

[7] Rabiner, L. R., "A tutorial on Hidden Markov Models and selected applications in speech recognition", *Proc. IEEE*, vol.77, pp.257-285, 1989.

A Markov Property

Suppose the process is known to be in (an arbitrary) aggregated state i (*i.e.*, i pores are closed; exactly which pores is not important) at time k. Then the Markov question is equivalent to showing the probability of transiting to (an arbitrary) aggregated state j at time $k+1$ is unaffected by any additional knowledge of the process before time k. Since all pores are identical (a symmetry property) having i pores closed is equivalent to being in the particular state

$$s_k^{(1)} = q_2,\, s_k^{(2)} = q_2, \cdots,\, s_k^{(i)} = q_2,$$

$$s_k^{(i+1)} = q_1,\, s_k^{(i+2)} = q_1, \cdots,\, s_k^{(L)} = q_1.$$

This explicit state has state index $2^i - 1$.

Given the transition is made to an aggregated state with j closed pores we see that there are $\binom{L}{i}$ possible transitions from $S_k = 2^i - 1$ (small abuse of notation) which we will index by members from the set $\mathcal{R}_j \subset \{0, 1, 2, \cdots, 2^L - 1\}$. Naturally this set has cardinality $|\mathcal{R}_j| = \binom{L}{j}$. By considering the (nonequal) sum of the probabilities of these transitions we see, utilizing (2.9), that

$$P(z_{k+1} = u_j \mid z_k = u_i) = \sum_{r \in \mathcal{R}_j} P\Big(S_{k+1} = r \;\Big|\; S_k = 2^i - 1\Big).$$

The utility of this expression is expressing the transition properties of the aggregated states in terms of the (nonaggregated) states. Since by assumption the $\{S_k\}$ process is Markovian we can conclude the same for the aggregated process.

B Parameterized Matrices

Transition Probability Matrix

The objective is to compute the components of (2.10). Only the uncoupled portion presents a difficulty so we will focus on this. To simplify notation we will represent the components of $\mathbf{A}^{(I)}$ for L pores by $a_{mn}^{[L]}$.

The boundary conditions, using this notation, are

$$a_{00}^{[1]} = \zeta, \quad a_{01}^{[1]} = 1 - \zeta, \quad a_{10}^{[1]} = 1 - \rho, \quad a_{11}^{[1]} = \rho.$$

When considering the transition probabilities of a collection of L independent pores we can regard this as two groups; one containing $L-1$ pores and the other an isolated independent pore. With this observation in mind we can easily develop two sets of recursions. The first set presumes that the isolated pore is initially open meaning the index m is restricted to the range $0 \le m \le L-1$, whence

$$a_{mn}^{[L]} = \begin{cases} \zeta\, a_{mn}^{[L-1]} + (1-\zeta)\, a_{m(n-1)}^{[L-1]}, & 1 \le n \le L-1 \\ \zeta\, a_{m0}^{[L-1]}, & n = 0 \\ (1-\zeta)\, a_{m(L-1)}^{[L-1]}, & n = L. \end{cases} \tag{B.1}$$

The second set of recursions presumes the isolated pore is initially closed meaning we can find equations for $m = L$. This set is entirely analogous to those in (B.1) and will not be given.

As an example, for $L = 2$ we have

$$\begin{aligned} a_{11}^{[2]} &= \zeta\, a_{11}^{[1]} + (1-\zeta)\, a_{10}^{[1]} \\ &= \zeta\,\rho + (1-\zeta)(1-\rho) \\ a_{10}^{[2]} &= \zeta\, a_{10}^{[1]} \\ &= \zeta\,(1-\rho) \end{aligned}$$

which tallies with (2.11). Continuing with, say, $L = 3$ we then compute, for example,

$$\begin{aligned} a_{11}^{[3]} &= \zeta\, a_{11}^{[2]} + (1-\zeta)\, a_{10}^{[2]} \\ &= \zeta\,\Big(\zeta\,\rho + (1-\zeta)(1-\rho)\Big) + (1-\zeta)\,\zeta\,(1-\rho) \\ &= \zeta\,(3\,\zeta\,\rho - 2\,\rho + 2 - 2\,\zeta) \end{aligned}$$

Partial Derivatives

In the gradient descent adaptation we need to compute partial derivatives with respect to the parameters in Θ. With (2.10) written in the form

$$\mathbf{A} = (1-\kappa)\mathbf{A}^{(I)} + \kappa\mathbf{A}^{(C)}$$

we obtain

$$\frac{\partial \mathbf{A}}{\partial \zeta} = (1-\kappa)\frac{\partial \mathbf{A}^{(I)}}{\partial \zeta}, \quad \frac{\partial \mathbf{A}}{\partial \rho} = (1-\kappa)\frac{\partial \mathbf{A}^{(I)}}{\partial \rho}$$

$$\frac{\partial \mathbf{A}}{\partial \kappa} = \mathbf{A}^{(C)} - \mathbf{A}^{(I)}, \quad \frac{\partial \mathbf{A}}{\partial \delta} = \kappa\frac{\partial \mathbf{A}^{(C)}}{\partial \delta},$$

$$\frac{\partial \mathbf{A}}{\partial \xi} = \kappa\frac{\partial \mathbf{A}^{(C)}}{\partial \xi} \quad \text{and} \quad \frac{\partial \mathbf{A}}{\partial \eta} = \kappa\frac{\partial \mathbf{A}^{(C)}}{\partial \eta}.$$

We will focus only on the uncoupled contribution since the terms involving δ, ξ and η lead to easily determined sparse matrices. By differentiating the recursions like (B.1) it is straightforward to develop recursions for quantities like

$$\frac{\partial a_{ij}^{[L]}}{\partial \zeta}, \quad \text{and} \quad \frac{\partial a_{ij}^{[L]}}{\partial \rho}$$

in terms of lower order terms of the same form and lower order transition probabilities. The addition boundary conditions that need to apply in this case are readily determined to be

$$\frac{\partial a_{00}^{[1]}}{\partial \zeta} = 1, \quad \frac{\partial a_{01}^{[1]}}{\partial \zeta} = -1, \quad \frac{\partial a_{10}^{[1]}}{\partial \zeta} = 0, \quad \frac{\partial a_{11}^{[1]}}{\partial \zeta} = 0,$$

$$\frac{\partial a_{00}^{[1]}}{\partial \rho} = 0, \quad \frac{\partial a_{01}^{[1]}}{\partial \rho} = 0, \quad \frac{\partial a_{10}^{[1]}}{\partial \rho} = -1, \quad \frac{\partial a_{11}^{[1]}}{\partial \rho} = 1.$$

□