CONF-9605108--1

BNL-62775

# ON THE PROCESSING OF PIECEWISE–CONSTANT SIGNALS BY HIERARCHICAL MODELS WITH APPLICATION TO SINGLE ION CHANNEL CURRENTS

*Petar M. Djurić,* Jong-Kae Fwu,* Slobodan Jovanović,** and Kelvin Lynn***

* Department of Electrical Engineering
State University of New York
Stony Brook, NY 11794-2350
djuric@sbee.sunysb.edu
fwu@sbee.sunysb.edu

** Physics Department
Brookhaven National Laboratory
Upton, NY 11973
jovanovic@bnlarm.bnl.gov
kgl@aps.org



## ABSTRACT

A new approach for processing of piecewise-constant signals is proposed. It is based on modeling the observed data as a sum of a random signal and noise. The random signal has a Gibbs distribution, and the noise is Gaussian. A MAP criterion is derived for joint estimation of the number of signal levels and reconstruction of the signal. The criterion comprises of three terms, one corresponding to the likelihood of the data and two to penalties. One penalty term penalizes for unnecessary transitions, and the other, for unnecessary levels. The method has been tested on synthesized data and applied to single ion channel recordings.

## 1. INTRODUCTION

Piecewise constant signals arise frequently in many areas of science and engineering. They are characterized by several constant levels and are commonly corrupted by unknown noise. In many applications, the number of levels and their associated values are not known. The signals themselves also change levels randomly. The main processing task is that of detecting the number of levels and reconstructing the noiseless signal.

There are a variety of methods for addressing this problem. In this paper, we propose a procedure based on hierarchical models. The observed data are represented as a sum of two random processes, one corresponding to the signal and the other to the noise. The unobservable (noiseless) signal is modeled by a Gibbs distribution and the noise by a Gaussian distribution. We apply a maximum a posteriori (MAP) criterion to obtain the optimal number of levels and the estimate of the associated signal. The resulting criterion is a penalized likelihood function with terms that can be easily interpreted. The method is implemented iteratively in stages by using a tree-structure (TS) initialization scheme [5] and the Iterated Conditional Modes (ICM) method [2]. In the first stage, it is assumed that the signal has only one level, in the second two levels, in the third three, and so on. After the completion of each stage, the best sequence is selected for that stage and compared to the best sequence from the previous stages. If based on the derived criterion the sequence from the latest stage is better, it is kept, otherwise it is removed. The procedure ends with the processing of the hypothesis associated with the maximum number of levels.

We are interested to apply our method to patch clamping recordings. One of the main objectives of patch clamping is the study of ion permeation mechanisms in biological membranes. Patch clamping allows for the isolation of small patches of membranes and involves measurement of ion channel currents. The ion channels are large proteins embedded in the membranes of all living cells. These macromolecules form pores across the cell membrane, and in certain conformations, they allow the flow of ions into or out of the cell, thereby controlling the electrical function of the cell. The measured currents are noisy piecewise constant signals which reflect the gating kinetics of the individual channels. Previous attempts to process such signals include the half amplitude analysis [8], mean-variance histograms [8], the stategram [7], various forms of the Hinkley detector [3], and Bayesian methods based on Markov chain prior distributions [4]. For an overview of statistical analysis methods in the study of ion channels, see [1].

In this paper, we provide the derivation of our procedure, interpret the results, and show the necessary steps for its implementation. We also test the method on synthesized signals and apply it to single ion channel currents.

This work was supported by the National Science Foundation under Award No. MIP-9506743 and, in part, by U.S. DOE under contract no. DE-AC02-76CH00016.

## 2. PROBLEM STATEMENT AND TERMINOLOGY

Let $\mathbf{X}_k$ be a signal which represents a random sequence of $N$ samples, with $k$ being the number of possible signal values (levels), and let $\mathbf{x}_k$ be a realization of $\mathbf{X}_k$. The set of various levels is denoted by $\mathcal{S}_k = \{s_1, s_2, \cdots s_k\}$, and the collection of all possible signal sequencies $\mathbf{x}_k$ by the set $\mathcal{X}_k$, where $\mathcal{X}_k = \{\mathbf{x}_k = [x_k[1]\ x_k[2] \cdots x_k[N]]^T : x_k[i] \in \mathcal{S}_k,\ 1 \le i \le N\}$. The signal remains constant with a value equal to $s_i \in \mathcal{S}_k$, $i = 1, 2, \cdots k$, for some time, which is referred to as the dwell time, and then moves to a different level. It then remains there for another dwell time, then changes levels again, etc. The signal $\mathbf{x}_k$ is unobservable. Instead, we observe the noise corrupted version $\mathbf{y}$, which is obtained as a sum of $\mathbf{x}_k$ and a white noise vector $\mathbf{w}$, i.e.,

$$\mathbf{y} = \mathbf{x}_k + \mathbf{w}. \tag{1}$$

The number of signal levels $k$, the signal levels $\mathcal{S}_k$, and the noise parameters are unknown. Also unknown are the dwell times of the signal. Based on the model (1) and the assumptions

- the signal $\mathbf{X}_k$ has a finite number of well defined levels $k$, where $k \le K$, for a known $K$,
- the signal is described by a Gibbs distribution,
- the noise samples are independent and zero mean Gaussian with variances that depend on the signal levels, and
- the introduced random processes are time reversible,

the objective is to determine from $\mathbf{y}$ the number of levels $k$, $k \in \{1, 2, \cdots, K\}$, and estimate the signal $\mathbf{x}_k$, which includes the signal levels $s_i$, $i = 1, 2, \cdots, k$, and the associated dwell times.

One of the key assumptions is the Gibbs distribution of the random signal $\mathbf{X}_k$. If $\mathbf{x}_k$ is one outcome from the set $\mathcal{X}_k$, its probability is given by [2], [6],

$$p(\mathbf{x}_k) = \frac{e^{U(\mathbf{x}_k)}}{Z_k} \tag{2}$$

where $Z_k$ is a normalizing constant, and $U(\mathbf{x}_k)$ is an energy function. With the assumption (2), we adopt a neighborhood system $\mathcal{N}$, which is a collection of subsets of $\{1, 2, \cdots, N\}$, that is $\mathcal{N} = \{\mathcal{N}_i : i \in \{1, 2, \cdots, N\}\}$, where $\mathcal{N}_i$ denotes the neighbors of $i$ and satisfies the conditions (a) $i \notin \mathcal{N}_i$ and (b) $i \in \mathcal{N}_j$ iff $j \in \mathcal{N}_i$. Thus, $\mathbf{X}_k$ is a one-dimensional Markov random field with respect to $\mathcal{N}$.

Finally, we introduce the sequence of $k$ different labels $\mathbf{l}_k$, where $l_k[i] \in \{1, 2, \cdots, k\}$. This sequence is associated with the signal sequence $\mathbf{x}_k$ via

$$\mathbf{x}_k = \Phi(\mathbf{l}_k, \mathcal{S}_k) \tag{3}$$

where $\Phi(\cdot)$ is a function that uniquely maps the label sequence $\mathbf{l}_k$ to the signal $\mathbf{x}_k$ according to $x_k[i] = s_{l[i]}$, $i \in \{1, 2, \cdots, N\}$. For example, if $\mathbf{l}_k = [1\ 1\ 2\ 1\ 3\ \cdots]^T$, $\mathbf{x}_k = [s_1\ s_1\ s_2\ s_1\ s_3\ \cdots]^T$. In quantifying the probability $p(\mathbf{x}_k)$, we choose energy functions that allow us to write

$$p(\mathbf{x}_k) = p(\mathbf{l}_k). \tag{4}$$

## 3. ESTIMATION CRITERION

We want to apply the maximum a posteriori probability (MAP) criterion for estimating $k$. The marginalized MAP estimate is given by

$$\hat{k} = \arg\max_k p(k|\mathbf{y}) \tag{5}$$

where $p(k|\mathbf{y})$ is the posterior probability mass function of $k$ given the observed data $\mathbf{y}$. This criterion will be very difficult to implement because its solution is based on

$$\hat{k} = \arg\max_k \left\{ \sum_{\mathbf{l}_k} \int_{\Theta_k} f(\mathbf{y}|\mathbf{l}_k, \theta_k) f(\theta_k) p(\mathbf{l}_k) p(k) d\theta_k \right\} \tag{6}$$

where $f(\mathbf{y}|\mathbf{l}_k, \theta_k)$ is the density of the data given the label sequence $\mathbf{l}_k$ and the signal and noise parameters $\theta_k$ with prior density $f(\theta_k)$, $\Theta_k$ is the parameter space of $\theta_k$, and $p(k)$ is the prior probability of $k$ levels. Note that the summation in (6) has an extremely large number of terms.

Another MAP estimator jointly provides the MAP estimates of $k$ and $\mathbf{x}_k$, and it is expressed by

$$\hat{\mathbf{x}}_{\hat{k}} = \arg\max_{\mathbf{x}_k, k} p(\mathbf{x}_k|\mathbf{y}). \tag{7}$$

This criterion can readily be evaluated by using one of several existing iterative techniques.

Finally, we introduce a third criterion, which we find more appropriate than (7) for reasons to be explained below. The form of the criterion is

$$\hat{\mathbf{l}}_{\hat{k}} = \arg\max_{\mathbf{l}_k, k} p(\mathbf{l}_k|\mathbf{y}) \tag{8}$$

or

$$\hat{\mathbf{l}}_{\hat{k}} = \arg\max_{\mathbf{l}_k, k} \left\{ \int_{\Theta_k} f(\mathbf{y}|\mathbf{l}_k, \theta_k) f(\theta_k) p(\mathbf{l}_k) p(k) d\theta_k \right\}. \tag{9}$$

Clearly, the criterion (9) selects the joint MAP estimates of $k$ and the label sequence $\mathbf{l}_k$. Its main difference from (7) is that it integrates out the unknown signal and noise parameters. Of course, once $k$ and $\mathbf{l}_k$ are estimated, it is trivial to determine the signal levels $s_i$ and obtain the estimate of $\mathbf{x}_k$.

The evaluation of (9) is easier than that of (6) but is still computationally intractable. One difficulty is the evaluation of the normalizing constants $Z_k$ needed due to $p(\mathbf{l}_k)$. To alleviate this problem, we substitute $p(\mathbf{l}_k)$ by the pseudolikelihood [2]

$$q(\mathbf{l}_k) = \prod_i p(l_k[i]|\partial l_k[i], \boldsymbol{\phi}) \tag{10}$$

where $\boldsymbol{\phi}$ denotes the parameters of the Gibbs distribution, and $\partial l_k[i]$ is the set of neighboring samples of $l_k[i]$.

Another problem in evaluating (9) is the computation of the integral $\int_{\boldsymbol{\Theta}_k} f(\mathbf{y}|\mathbf{l}_k, \boldsymbol{\theta}_k) f(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k$. However, by Taylor expending $f(\mathbf{y}|\mathbf{l}_k, \boldsymbol{\theta}_k)$ around the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_k$, and using asymptotic assumptions, we can approximate it by

$$\int_{\boldsymbol{\Theta}_k} f(\mathbf{y}|\mathbf{l}_k, \boldsymbol{\theta}_k) f(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k \dot{\propto} \frac{f(\mathbf{y}|\hat{\mathbf{l}}_k, \hat{\boldsymbol{\theta}}_k)}{\prod_{j=1}^{k} n_j} \tag{11}$$

where $\hat{\mathbf{l}}_k$ is the estimated label sequence, and $n_j$ are the total number of samples whose label is $l_k[\cdot] = j$.

With these approximations and the assumptions that the additive noise is zero mean Gaussian with level dependent variance and $p(k) = 1/K$, the MAP criterion results in

$$\hat{\mathbf{l}}_{\hat{k}} = \arg\min_{\mathbf{l}_k, k} \left\{ \sum_{j=1}^{k} \frac{n_j}{2} \ln \hat{\sigma}_j^2 - \ln q(\mathbf{l}_k) + \sum_{j=1}^{k} \ln n_j \right\} \tag{12}$$

where $\hat{\sigma}_j^2$ is the estimated variance of the samples labeled as $l_k[\cdot] = j$. The interpretation of the three terms in (12) is straightforward. The first term is the likelihood which decreases with increasing $k$. The second and third terms are penalties for introducing signal transitions from one level to another and additional levels, respectively, and they grow with the number of transitions and $k$. Note that penalization for the parameters of the Gibbs distribution is not necessary because we assume the same parameters are present for all the models used in (12).

## 4. IMPLEMENTATION OF THE PROPOSED CRITERION

The criterion (12) can be implemented by applying the iterated conditional modes (ICM) algorithm [1]. For good performance, the ICM requires relatively good initial conditions. To obtain them we propose to use the tree-structure (TS) method introduced in [5], which is a completely data driven scheme. The overall (TS-ICM) procedure consists of the following steps. First, assume that $k = 1$, estimate the only level $\hat{s}_1$, and evaluate the criterion function. Set $k = 2$, and as initial conditions use $\hat{s}_1^{(0)} = \hat{s}_1 - \epsilon$ and $\hat{s}_2^{(0)} = \hat{s}_1 + \epsilon$, where $\epsilon$ is some small number, and $\hat{s}_1$ is the result for $k = 1$. Apply the ICM method, estimate $\hat{s}_1$ and $\hat{s}_2$, and evaluate the criterion. Set $k = 3$ and use as initial conditions $\hat{s}_1^{(0)} = \hat{s}_1 - \epsilon$, $\hat{s}_2^{(0)} = \hat{s}_1 + \epsilon$, and $\hat{s}_3^{(0)} = \hat{s}_2$, with $\hat{s}_1$ and $\hat{s}_2$ being the results for $k = 2$, and continue with the ICM and the evaluation of the criterion. As another possible set of initial conditions use $\hat{s}_1^{(0)} = \hat{s}_1$, $\hat{s}_2^{(0)} = \hat{s}_2 - \epsilon$, and $\hat{s}_3^{(0)} = \hat{s}_2 + \epsilon$, and again apply the ICM and evaluate the criterion. As a final sequence of three levels, $\hat{\mathbf{l}}_3$, choose the one that has the smaller criterion value. Next set $k = 4$, and continue along the same lines until the testing of $k = K$ levels is completed. The solution is the label sequence $\hat{\mathbf{l}}_{\hat{k}}$ that yields the smallest criterion value. Once $\hat{\mathbf{l}}_{\hat{k}}$ is determined, the signal $\hat{\mathbf{x}}_{\hat{k}}$ is easily estimated.

## 5. SIMULATION RESULTS

Our method was tested on synthesized data and applied to real patch clamp recordings. The synthesized data records had 2000 samples, three different levels, and 6 level changes. The signal-to-noise ratio (SNR) was defined by $\min|\Delta/\sigma|$, where $\Delta$ is the difference between the levels of two adjacent segments and $\sigma$ the standard deviation of the noise, which was the same for every segment. The SNR was varied between 1 and 5 in steps of 1. For each SNR, there were 100 trials. The probability $p(l_k[i]|\partial l_k[i], \boldsymbol{\phi})$ in (10) was defined by

$$p(l_k[i] = m|\partial l_k[i], \boldsymbol{\phi}) = \frac{e^{\beta u_m[i]}}{\sum_{j=1}^{k} e^{\beta u_j[i]}} \tag{13}$$

where $\boldsymbol{\phi} = \beta$, and $u_m[i]$ denotes the number of neighbors of $i$ having the label $m$. The number of neighbors in the experiment was equal to four, that is, two neighbors on each side. The results are shown in Table 1. For SNR's of 2 and higher, the method had very good performance. Figure 1 at the top shows a typical realization with SNR=2 and at the bottom, the histogram of the detected level changes in 100 trials. The peaks of the histogram are at the correct locations of the signal transitions.

Figure 2 at the top displays a real patch clamp recording, which is quite complex and has several conductance levels. We applied our procedure to these data to determine the number of levels and estimate

$\hat{x}_k$. The maximum number of hypothesized levels was 15. The procedure found 10 levels and estimated the noise-free signal shown at the bottom of Figure 2.

| SNR/k | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 97 | 3 | 0 | 0 |
| 2 | 0 | 0 | 98 | 1 | 1 |
| 3 | 0 | 0 | 98 | 2 | 0 |
| 4 | 0 | 0 | 100 | 0 | 0 |
| 5 | 0 | 0 | 99 | 1 | 0 |

Table 1: The entries represent the number of times k levels were estimated in 100 trials.

## 6. REFERENCES


[1] F. G. Ball and J. A. Rice, "Stochastic models for ion channels: Introduction and bibliography," Mathematical Bioscience, vol. 112, pp. 189-206, 1992.

[2] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259-302, 1986.

[3] S. Draber and R. Schultze, "Detection of jumps in single-channel data containing subconductance levels," *Biophysical Journal*, vol. 67, pp. 1404-1413, 1994.

[4] D. R. Fredkin and J. A Rice, "Bayesian restoration of single channel patch-clamp recordings," Biometrics, vol. 48, pp. 427-448, 1992.

[5] J.-K. Fwu and P. M. Djurić, "Unsupervised vector image segmentation by a tree structure-ICM algorithm," submitted to the *IEEE Transactions on Medical Imaging*.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian reconstruction of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721- 741, 1984.

[7] S. Jovanovic, "A system for on-line analysis of single ion channel currents," *Master Thesis*, State University of New York, 1994.

[8] J. B. Patlak, "Measuring kinetics of complex single ion channel data using mean-variance histograms," *Biophysical Journal*, vol. 65, pp. 29-42, 1993.


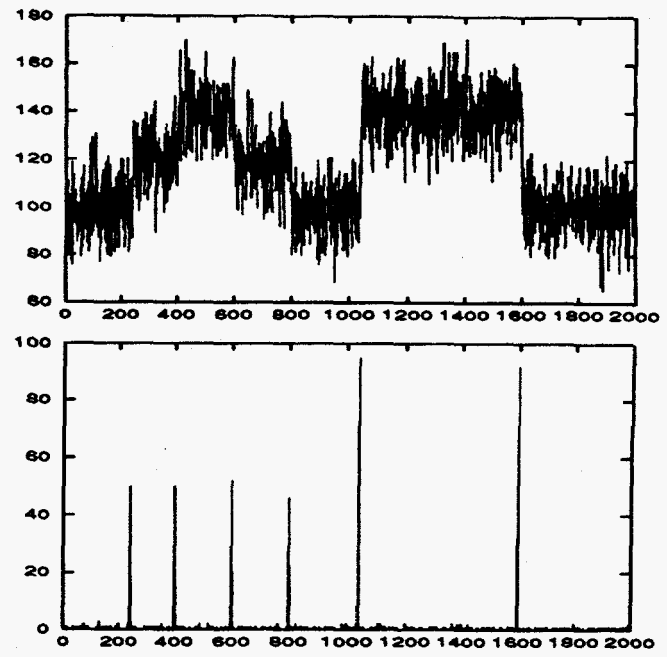


Figure 1: Top: A realization with 3 levels and 6 transitions for SNR=2. Bottom: Histogram of the estimated transitions in 100 trials.

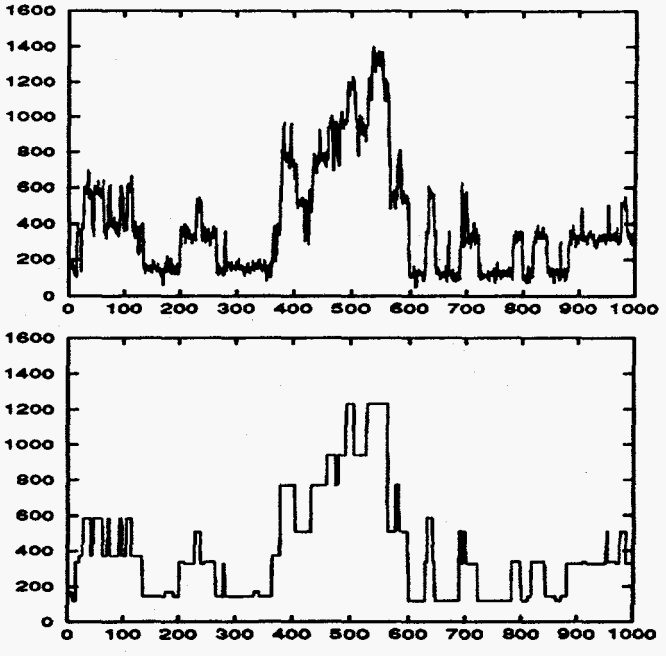


Figure 2: Top: Real data with unknown number of levels. Bottom: Estimated signal with 10 levels.

## DISCLAIMER