Extended Kalman Filtering of Point Process Observation

Yousef Salimpour*, *Member*, *IEEE*, Hamid Soltanian-Zadeh, *Senior Member*, *IEEE*

***Abstract*— A temporal point process is a stochastic time series of binary events that occurs in continuous time. In computational neuroscience, the point process is used to model neuronal spiking activity; however, estimating the model parameters from spike train is a challenging problem. The state space point process filtering theory is a new technique for the estimation of the states and parameters. In order to use the stochastic filtering theory for the states of neuronal system with the Gaussian assumption, we apply the extended Kalman filter. In this regard, the extended Kalman filtering equations are derived for the point process observation. We illustrate the new filtering algorithm by estimating the effect of visual stimulus on the spiking activity of object selective neurons from the inferior temporal cortex of macaque monkey. Based on the goodness-of-fit assessment, the Kalman particle filter provides more accurate state estimate than the conventional methods.**

***Index Terms* — Stochastic filtering, Kalman filtering, Point process, Generalized linear model, Spike train, Peristimulus time histogram, Inferior temporal cortex.**

I. Introduction

ESTABLISHING and quantifying the correlation of neural spiking activity with an external stimulus is the focus of much investigation. Neurons generate series of spikes in response to the stimulus. A spike train is a stochastic process composed of a sequence of binary events that occur in continuous time. The point process theory is used as a stochastic framework to model the nonlinear property of neural spike train; however, due to dynamical behavior of neural systems many challenges are still remain[1], [2].

Point process framework is commonly used to model neuronal spiking activity. This framework allows dynamic modeling which is an important tool in computational neuroscience for studying neural dynamics [2]. Neural receptive field plasticity, neural coding analyses, neural spike train decoding, neural prostheses, analyses of learning, and control algorithms design for brain machine interfaces are examples of the neural dynamics [3], [4].

Since point process filtering method assumes that the posterior density of the state given the discrete observation is Gaussian and the recursive estimation is derived with this assumption [4], therefore effective state estimation restricted to Gaussian random variables. One approach for applying state space estimation algorithm for the nonlinear and dynamic system is use of extended Kalman filter [5].

In this paper, in order to generalize the point process filtering approach for the states of nonlinear system, we combine the extended Kalman filtering theory with point process modeling. Based on the point process observation, the extended Kalman filtering recursive equations are derived and applied for the estimation of the model parameters which were considered as states of the system. We illustrate the properties of the hybrid filtering process by filtering the spiking activity of neurons from inferior temporal cortex of the macaque monkey while the animal performs the passive fixation task.

II. Materials and Methods

A. State-Space Model with Point Process Observations

A stochastic neural point process can be completely characterized by its conditional intensity function which is a strictly positive function that gives a history-dependent generalization of the rate function of a Poisson process. We use the conditional intensity function to characterize the spike train as a point process [6]. We assume that on an interval $(0,T)$, J spikes are fired by the single neuron at times $t_1, t_2 \ldots\ldots, t_J$ for $t \in (0,T)$ the conditional intensity function is defined as:

$$\lambda(t|\theta(t), N_{0:t}) = \lim_{\Delta t \to 0}\left(\frac{Pr(N(t+\Delta t) - N(t) = 1|\theta(t), N_{0:t})}{\Delta t}\right) \quad (1)$$

where $\lambda(t|\theta(t), N_{0:t})$ is a conditional probability, The $N(t)$ is the number of spikes fired by the neuron in $(0,t), N_{0:t}$ includes the neuron's spiking history up to time t, and $\theta(t)$ is interested parameter to be estimated [6]. Because the conditional intensity function completely defines the point process, to model the neural spike train in terms of a point process, it suffices to define its conditional intensity function. The parametric models are used to express the conditional intensity as a function of covariates of interest [5], [7]. The conditional intensity function enables us to write the counting process $N(t)$ as an observation equation for the state space approach,

$$N(t) = \int_0^t \lambda(\tau|\theta(\tau), N_{0:\tau})d\tau + \eta(t) \quad (2)$$

Manuscript received June 22, 2010. This work is supported by the Computational Neuroscience group in the School of Cognitive Sciences, Institute for Studies in Fundamental Sciences , Tehran, Iran.

*Yousef Salimpour is a PhD candidate in Neuroscience and Neuroengineering in School of Cognitive Sciences, Institute for Studies in Fundamental Sciences (IPM) and Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, Tehran, Iran (e-mail: salimpour@ipm.ir).

Hamid Soltanian-Zadeh is with the Department of Radiology, Henry Ford Health System, Detroit, MI 48202 USA (e-mail: hamids@rad.hfh.edu) and School of Cognitive Sciences, Institute for Studies in Fundamental Sciences (IPM) , Tehran, Iran (e-mail: zadeh-hs@ipm.ir). And also with the Control and Intelligent Processing, Center of Excellence (CIPCE) Department of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran (e-mail: hszadeh@ut.ac.ir)

where $\eta(t)$ is a zero mean error process that is the residual between a point process and its expectation [6].

In order to represent the point process model, we discretize the time interval $(0, T)$ by dividing it into K intervals of width $\Delta = TK^{-1}$, such that there is at most one spike per interval. For $k = 1, \ldots, K$, let ΔN_k be the indicator of a spike in the interval $((k-1)\Delta, k\Delta)$, which is one if there is a spike and zero otherwise. We let $N_{1:k} = [\Delta N_1, \ldots, \Delta N_k]$ denote the spiking activity and $\theta_{1:k} = [\theta_1, \ldots, \theta_k]$ be the values of θ in $(0, k\Delta)$. It follows from the theory of point processes, that by taking the discrete approximation of the joint probability density of the spike train on the interval $((k-1)\Delta, k\Delta)$, the probability mass function of the observation equation for our state-space model is defined as [6]:

$$P(\Delta N_k | \theta_k, N_{1:k-1}) = exp(\Delta N_k \, log(\lambda_k \Delta) - \lambda_k \Delta) \quad (3)$$

To apply the extended Kalman filter, we construct a discrete time version of the observation Equation (2) for a fine partition of the observation interval, linearize its expected value as a function of the state process by using the linear terms of a Taylor expansion about the one-step prediction mean, and add Gaussian white noise errors. The resulting approximate observation equation is:

$$\Delta N_k = \lambda_k \Delta + (\theta_k - \theta_{k|k-1}) \Delta \, \frac{\partial \lambda}{\partial \theta} \Big|_{\theta_{k|k-1}} + \upsilon_k \quad (4)$$

The Gaussian error term υ_k should be selected so as to have similar statistical properties of the observation distribution. The variance of the discrete time approximation to the point process model is $\lambda_k \Delta$, which is unknown. Since Δ is sufficiently fine the $\upsilon_k \sim N(0, \lambda_k \Delta)$ might be a good choice. The state equation is the Gaussian linear stochastic dynamical system as follow, where ω_k is a zero-mean Gaussian noise with covariance matrix Q_k.

$$\theta_k = \theta_{k-1} + \omega_k \quad (5)$$

B. *Extended Kalman Filtering of Point Process*

We model the conditional intensity function in terms of the state process as

$$log \, \lambda^s \, (k\Delta | \theta_k, N_{1:k-1}) = \sum_{r=1}^{R} \theta_r \, g_r(k\Delta) \quad (6)$$

which is a kind of generalized linear models where the $g_r(\mathrm{k}\Delta)$ are a set of functions that model the stimulus specific effect on spiking activity. The $g_r(\mathrm{k}\Delta)$ is 1 in the interval $((k-1)\Delta, k\Delta)$ and 0 elsewhere. Under this parametric model, the spiking activity on different trials is independent and effect of history is included in state process. Consequently, the time varying rate function based on the averaging of the trials in specific time bins is a special case of our model.

We apply the Kalman filtering method for parameter estimation. It follows from the theory of point processes, that by taking the discrete approximation of the joint probability density of the spike train on the specific interval $((k-1\Delta, k\Delta)$, the probability mass function of the observation equation for our state-space model is defined as:

$$P(\Delta N_k | \theta_k, N_{1:k-1}) = exp(\Delta N_k \, log(\lambda_k \Delta) - \lambda_k \Delta) \quad (7)$$

A standard approach for formulating state-space estimation algorithms uses the Bayes' rule and Chapman–Kolmogorov equations. For the model defined in Equation (4) and (5).The a *priori* pdf is obtained from Equation (8) and the a *posteriori* pdf is obtained from Equations (9).

$$P(\theta_k | N_{1:k-1}) = \int P(\theta_k | \theta_{k-1}) P(\theta_{k-1} | N_{1:k-1}) d\theta_{k-1} \quad (8)$$

$$P(\theta_k | N_{1:k}) = \frac{P(\Delta N_k | \theta_k, N_{1:k-1}) P(\theta_k \, | N_{1:k-1})}{\int P(\Delta N_k | \theta_k) P(\theta_k | N_{1:k-1}) d\theta_k} \quad (9)$$

Equations (8) and (9) are a recursive system for computing the posterior density $P(\theta_k | N_{1:k})$. The first term in the numerator of Equation (9) is the probability mass function of the observation process in Equation (3), the second term is the one-step prediction density defined in Equation (8) and the denominator is a normalizing constant that ensures that the posterior probability density integrates to one. The challenge of this problem is to evaluate Equation (8) and (9) for the observation and system models in Equations (4) and (5).

Let $\theta_{k|k-1}$ and $W_{k|k-1}$ define the mean vector and covariance matrix of the Gaussian approximation in Equation (8), and $\theta_{k|k}$ and $W_{k|k}$ be the mean vector and covariance matrix of the Gaussian approximation in Equation (9). The state transition model in Equation (5) is sufficient to compute the one step prediction probability densities,

$$\theta_{k|k-1} = \theta_{k-1|k-1} \quad (10)$$

We write the posterior probability in the interval $((k-1)\Delta, k\Delta)$ by applying a Gaussian approximation.

$$P(\theta_k | N_{1:k}) \approx exp(\Delta N_k \, log(\lambda_k \Delta) - \lambda_k \Delta) \; \times \; exp\left(-\frac{1}{2}(\theta_k - \theta_{k|k-1})^T (W_{k|k-1})^{-1} (\theta_k - \theta_{k|k-1})\right)$$
$$\approx exp\left(-\frac{1}{2}(\theta_k - \theta_{k|k})^T (W_{k|k})^{-1} (\theta_k - \theta_{k|k})\right) \quad (11)$$

The maximum a posterior estimate of the state is defined by the $\frac{\partial log P(\theta_k | N_{1:k})}{\partial \theta_k}\Big|_{\theta_k = \theta_k{}^{MAP}} = 0$ and this relation should be approximately true for any value of $\theta_k{}^{MAP}$. We can therefore choose any specific point to evaluate this expression. Evaluating at $\theta_k{}^{MAT} = \theta_{k|k-1}$ and rearranging the Equation gives,

$$\theta_{k|k} = \theta_{k|k-1} + W_{k|k} \frac{\partial log \lambda_k}{\partial \theta_k}\Big|_{\theta_{k|k-1}} (\Delta N_\mathrm{k} - \lambda_k \Delta) \quad (12)$$

Since $W_{k|k-1} = Cov[\theta_k - \theta_{k|k-1}^{MAP}]$, we have

$$W_{k|k-1} = W_{k-1|k-1} + Q_k \quad (13)$$

and also we know that, $W_{k|k} = Cov[\theta_k - \theta_{k|k}^{MAP}]$. Based on Equation (12) we can derive the recursive equation for $W_{k|k}$ respectively. In kalman filtering framework the updated a posterior covariance is

$$W_{k|k-1} = Cov[\theta_k - \theta_{k|k-1} - W_{k|k}\frac{\partial log\lambda_k}{\partial \theta_k}\Big|_{\theta_{k|k-1}}(\Delta N_k - \lambda_k\Delta)] \quad (14)$$

Which can be estimated based on Equation (14) the final relation as follows:

$$W_{k|k} = \left[1 - \frac{\left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)^T(\lambda_k\Delta)\left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)W_{k|k-1}}{\left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)^T(\lambda_k\Delta)\left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)W_{k|k-1} + 1}\right]W_{k|k-1} \quad (15)$$

Rearranging the above equation, it reduces to

$$(W_{k|k})^{-1} = (W_{k|k-1})^{-1} + \left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)^T(\lambda_k\Delta)\left(\frac{\partial log\lambda_k}{\partial \theta_k}\right)\Big|_{\theta_{k|k-1}} \quad (16)$$

Thus far, the Kalman filter is completely derived for point process observation.

The histogram representation of recorded data across the repeated trials is standard analysis in neuronal data. Under the state space model with kalman filtering estimation algorithm, we can compute the probability density of a histogram constructed with desired bin width. For a spike train in time interval $(0, T)$ given two times $0 \le t_1 \le t_2 \le T$ The smoothed histogram based on conditional intensity function definition is

$$\Lambda(t_2 - t_1) = \int_{t_1}^{t_2}\lambda(\tau)d\tau \quad (17)$$

and hence, the smoothed rate function estimation is

$$\widehat{\Lambda}(t_2 - t_1) = \int_{t_1}^{t_2}\hat{\lambda}(\tau)d\tau \approx \sum_{t_1 \le k\Delta \le t_2}\lambda(k\Delta|\theta_k, N_{1:k-1})\Delta. \quad (18)$$

The confidence intervals for the smoothed estimate of the rate function can be efficiently computed by Monte Carlo methods.

C. *Goodness-of-fit Tests*

We use the time-rescaling theorem to construct a goodness-of-fit test for a neural spike data model. Given a point process with conditional intensity function $\lambda\,(t|\theta(t), N_{0:t})$ and occurrence times $t_1, t_2, \ldots, t_J$ where $t_j \in (0, T)$, if we define $z_j = exp(-\int_{t_{j-1}}^{t_j}\lambda(\tau)d\tau\,)$, then these z_j are independent, exponential random variables with rate parameter one [10]. A common approach to measuring agreement between the model and the data is to construct a Kolmogorov-Smirnov (KS) plot. The KS plot is a plot of the empirical cumulative distribution function (CDF) of the rescaled times against an exponential CDF. If the conditional intensity model accurately describes the observed spiking data, then the empirical and model CDFs should roughly coincide, and the KS plot should follow a 45° line. If the conditional intensity model fails to account for some aspect of the spiking behaviour, then that lack of fit will be reflected in the KS plot as a significant deviation from the 45° line. Confidence bounds for the degree of agreement between a model and the data may be constructed using the distribution of the Kolmogorov–Smirnov statistic [8].

III. RESULTS

In order to illustrate some of the properties of the likelihood space, the neural data of spiking activity from the inferior temporal cortex neurons of a macaque monkey is used. Each stimulus is presented for 250-ms and followed by 250-ms inter-stimulus blank interval. A 100-ms interval before stimulus presentation is recorded for the purpose of baseline activity study. Category selective neurons are entered in this study and the face selectivity is the most important feature for the neuron selection. A sample of stimuli presented to the monkey while performing a passive fixation task, the raster plot of the spiking activity of a face neuron, and projection of the spike train onto the likelihood space are shown in Figure 1.

The conventional method for computing the time varying rate function is the procedure called peristimulus time histogram (PSTH). In order to find the PSTH, we align the spike sequence with the onset of stimulus that repeated several times, then, divide the observation period into small bins size and then, count the number of spikes from all sequences that fall in the bin, and finally, we draw a bar-graph histogram with the bar-height of count normalized to the time bin and number of repetition in units of spikes per second. The result of PSTH calculation is shown in Figure 1.

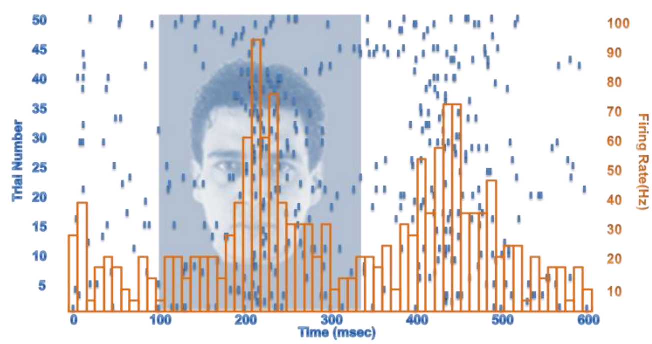
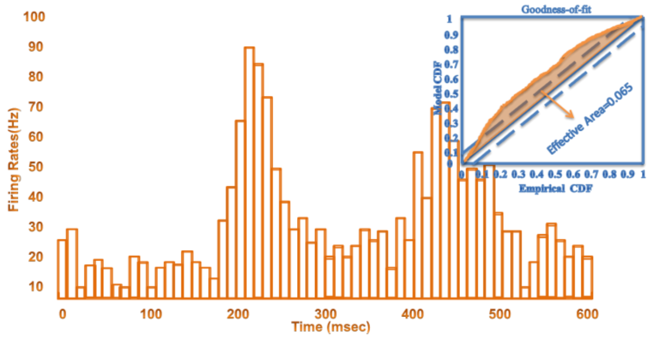
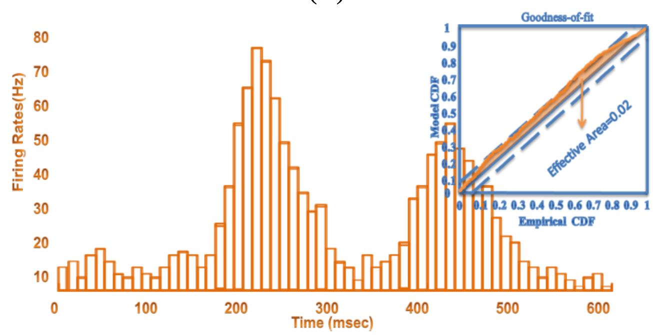


Figure 1. Sample responses of a neuron from inferior temporal cortex of a macaque monkey while the animal is doing the passive fixation task. The raster plot and the peristimulus time histogram as an estimation of time varying rate function are shown for a human face.

We use the same spike trains to estimate time varying rate function based on peristimulus time histogram with assessment of model accuracy assessment. The estimated functions with goodness- of -fit measure are shown in Figure 2.A. We generalize the filtering approach by combining the extended Kalman filter with point process modeling. We apply the new hybrid filtering approach on the same neural data. The results of time varying rate function with no assumption on the model parameters with goodness- of -fit measure are shown in Figure 2.B. We calculate the effective area between KS plot and the 45° line in order to have quantitative criteria for comparing the goodness- of -fit as illustrated in Figure 2.



(A)



(B)

Figure 2. (A) Estimation of time varying rate function based on peristimulus histogram approach with goodness- of -fit measure. (B) The results of time varying rate function estimation by applying the extended Kalman filtering approach with goodness- of -fit measure. The effective area between KS plot and the 45° line is calculate for having a quantitative criterion for goodness- of -fit comparison.

IV. Discussion and Conclusion

The development of kalman methods to construct filter algorithms for broad classes of observation and state models is an active area of signal processing research. To extend these methods to the analysis of stochastic dynamical systems observed through point processes, we use the combination of Kalman filtering and point process modeling algorithms. We derive the iterative Kalman filtering Equations for the point process observation and use them for updating the states at the measurement time.

This might be an efficient way for filtering of the real data from the biological systems such as neurons, typically involving elements of non-Gaussianity, high dimensionality and nonlinearity. To illustrate the properties of the derived filtering algorithm, we try to estimate the states of the neuron as a dynamic nonlinear system. The spike train is our observation from the neural system and state transition is the first order random walk model. We define the effect of input stimulus on spiking activity as states of system and try to optimal estimate.

Our finding shows that the extended Kalman filtering method may be used to construct filtering algorithm for state estimation using point process models of neural systems. Furthermore, it shows that using the combination of Kalman filter with particle filter can give more accurate estimation algorithm. The importance of these finding is the estimation of the neural response to the different covariate such as input stimulus which is essential in neuroscience based studies. This method could be extended to the population of neurons by estimating the conditional intensity model for the populations of neurons that might be useful for investigating the neural mechanism of stimulus encoding in nervous system in the population level.

While this study establishes the feasibility of constructing likelihood space for the neuronal populations as a linear stochastic dynamical system with point process observation models, an important extensions for the current framework is a possibility to extend the current algorithm to the nonlinear state space model for computing smoothed state estimate [9], [10].

Acknowledgment

We thank Dr. Hossein Esteky, the head of the School of Cognitive Sciences and the director of the Vision Lab, for providing experimental data and scientific support. This work was supported in part by the Computational Neuroscience Group in School of Cognitive Sciences, the Institute for Studies in Fundamental Sciences, Tehran, Iran.

References

[1] Brown EN, Kass RE, Mitra PP. "Multiple neural spike train data analysis: state-of-the-art and future challenges." Nature Neurosci , 456-461, 2004.

[2] Czanner G, Eden UT, Wirth S, Yanike M, Suzuki WA, Brown EN. Analysis of between-trial and within-trial neural spiking dynamics. Journal of Neurophysiology, 2008, 99: 2672-2693.

[3] Barbieri R, Wilson MA, Frank LM, and Brown EN. "An analysis of hippocampal spatiotemporal representations using a Bayesian algorithm for neural spike train decoding."*IEEE Trans Neural SystRehabilEng*13: 131-136, 2005.

[4] Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T., "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking."*IEEE Transactions on Signal Processing*50 (2): 174–188, 2002.

[5] Ergun A, Barbieri R, Eden UT, Wilson MA, and Brown EN." Construction of Point Process Adaptive Filter Algorithms for Neural Systems Using Sequential Monte Carlo Methods,"*IEEE Trans Biomed Eng* 54: 419-428, 2007.

[6] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Process*, 2nd ed. New-York: Springer-Verlag, 2003.

[7] Truccolo W, Eden UT, Fellow MR, Donoghue JP, Brown EN. "A point process framework for relating neural spiking activity for spiking history, neural ensemble and extrinsic covariate effects,"*J Neurophys 93: 1074-1089*, 2005.

[8] Brown EN, Barbieri R, Ventura V, Kass RE, and Frank LM.,"The time-rescaling theorem and its application to neural spike train data analysis,"*Neural Comput*14: 325-346, 2002.

[9] A. C. Smith and E. N. Brown, “Estimating a state-space model from point process observations,” *Neural Comp.*, vol. 15, pp. 965–991, 2003.

[10] S. J. Godsill, A. Doucet, and M. West, “Monte Carlo smoothing for nonlinear time series,” *JASA*, vol. 99, no. 465, pp. 156–168, 2004.