A regularized point process generalized linear model for assessing the connectivity in the cat motor cortex

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
A Regularized Point Process Generalized Linear Model for Assessing the
Functional Connectivity in the Cat Motor Cortex

Zhe Chen, David F. Putrino, Demba E. Ba, Soumya Ghosh, Riccardo Barbieri, and Emery N. Brown

Abstract—Identification of multiple simultaneously recorded neural
spike train recordings is an important task in understanding neuronal
dependency, functional connectivity, and temporal causality in neural
systems. An assessment of the functional connectivity in a group
of ensemble cells was performed using a regularized point process
generalized linear model (GLM) that incorporates temporal smoothness
or contiguity of the solution. An efficient convex optimization algorithm
was then developed for the regularized solution. The point process
model was applied to an ensemble of neurons recorded from the cat
motor cortex during a skilled reaching task. The implications of this
analysis to the coding of skilled movement in primary motor cortex is
discussed.

I. INTRODUCTION

Identifying a neuronal system via multivariate neural spike trains
recorded from ensemble neurons has many valuable implications
for understanding the system from a statistical perspective, and
has been used for establishing statistical associations or causality
between neurons, or finding spatiotemporal correlations, or studying
the functional connectivity [3], [8], [9], [10], [12], [14]. A statistical
treatment of multiple spike trains is to use the theory of stochastic
multivariate point processes. Statistical inference for point process
observations often starts with a certain class of statistical model,
followed by parameter estimation by either maximum likelihood or
Bayesian inference procedure [1], [2], [8], [14], [16].

The point process generalized linear model (GLM) [10], [16] has
been widely used for characterizing functional (spiking) dependence
among ensemble neurons. Recent new developments allow for the
incorporation of the Bayesian inference [14] or modeling common
dependencies, functional connectivity, and temporal causality in neural
systems. An assessment of the functional connectivity in a group
of ensemble cells was performed using a regularized point process
generalized linear model (GLM) that incorporates temporal smoothness
or contiguity of the solution. An efficient convex optimization algorithm
was then developed for the regularized solution. The point process
model was applied to an ensemble of neurons recorded from the cat
motor cortex during a skilled reaching task. The implications of this
analysis to the coding of skilled movement in primary motor cortex is
discussed.

II. A POINT PROCESS MODEL FOR MULTIPLE SPIKE TRAINS

Let \( c = 1, \ldots, C \) denote the index of a multivariate (C-
dimensional) point process. For the \( c \)th point process, let \( N_c(t) \)
denote the counting process up to time \( t \), and let \( dN_c(t) \) denote
the indicator variable, which equals to 1 if there is an event (spike)
at time \( t \) and 0 otherwise. Therefore, the multiple neural spike train
data are completely characterized by a multivariate point process
\( N_{1:C}(0: T) \).

This work was supported by NIH Grants DP1-OD003646, R01-
DA015644 and R01- HL084502.

The authors are with the Neuroscience Statistics Research Laboratory,
Massachusetts General Hospital, Harvard Medical School, Boston, MA
02114, USA. S. Ghosh is with the School of Medicine and Pharmacology,
Perth, University of Western Australia, Australia. (Email: zhechen@mit.edu)

In modeling the neural spike train point process, the conditional
intensity function (CIF) is used to characterize the instantaneous
firing probability of a discrete event (i.e., spike) [1]:

\[
\lambda_c(t|\theta, H_{0:t}) = \lim_{\Delta \to 0} \frac{\Pr\{N_c(t + \Delta) - N_c(t) = 1|H_{0:t}\}}{\Delta},
\]

(1)

where \( H_{0:t} \) denotes all of ensemble neuronal firing history and
any other information up to time \( t \). Where \( \Delta \) is sufficiently small,
the product CIF \( \lambda_c(t|\theta, H_{0:t}) \Delta \) tells approximately the probability
of observing a spike within the interval \( [t, t + \Delta] \):

\[
\Pr\{N_c(t + \Delta) - N_c(t) = 1|H_{0:t}\} \approx \lambda_c(t|H_{0:t})\Delta.
\]

(2)

Here, we restrict ourselves to the cases where \( \sum_{c=1}^{C} dN_c(t) \leq 1 \)
at any time \( t \), i.e., no joint firing is allowed in the continuous-time
setting (in the case of discrete-time setting, no joint firing is allowed
at the finest temporal scale under consideration). Let \( \theta \) denote
the ensemble unknown parameters in the parametric form of function
\( \lambda_c^{C}_{\theta_{c}} \). Specifically, we express the CIF in the following log-
linear form [10]:

\[
\log \lambda_c(t) = \alpha x_c(t) = \sum_{j=0}^{d} \alpha_j x_{j}^c(t) + \sum_{i=1}^{C} \sum_{k=1}^{K} \alpha_{i,k} x_{i,k}^c(t),
\]

where \( \dim(\alpha) = d + 1 \) (where \( d = C \times K \)) denotes total number
of parameters in the augmented parameter vector \( \alpha = \{\alpha_{i,k}\} \),
and \( x_c(t) = \{x_{i,k}^c(t)\} \), where \( x_{i,k}^c(t) = 1 \) and \( x_{i,k}^c(t) \) denotes
the spike count information from cell \( i \) at the \( k \)th time-lag history windows.

Let \( \theta = \{\alpha_c\} \), where \( \dim(\theta) = C(1 + d) \). By assuming
that the spike trains are mutually conditionally independent, the
continuous-time log-likelihood of observed data is given by:

\[
L(\theta) = \sum_{c=1}^{C} \left\{ \int_0^T -\log(\lambda_c(t|\theta))dt \right\} + \int_0^T \log(\lambda_c(t|\theta))dN_c(t).
\]

By discretization of (3), we also obtain the discrete-time log-
likelihood function, in which the integration will be replaced by a
finite sum. From (3) it is clear that \( -L(\theta) \) is convex with respect
to \( (w.e.l.) \) to each \( \lambda_c \). In addition, the index \( c \) is uncoupled from
each other in the network log-likelihood function, which implies that
we can optimize the function separately for each spike train
\( N_c(0 : T) \) once \( \lambda_c \) is specified. For simplicity, from now on we
will drop off the index \( c \) at \( \lambda_c \) and \( \alpha_c \) when no confusion occurs.

III. REGULARIZATION AND OPTIMIZATION

When the size of parameter space is large, it is often desirable to impose
certain prior knowledge (such as spatial sparsity) or physically plausible constraint (such as temporal smoothness) on the parameters [15], [6]. This can be done by the so-called “regularization” to improve the generalization ability of the model (on unseen data) while fitting finite training data. Regularization can be interpreted as imposing a prior on the parameter space in terms of Bayesian inference, and the log-likelihood will be interpreted as the

\footnote{Note that here we use the simplified notation: \( \lambda_c(t) = \lambda_c(t|\alpha, H_{0:t}) \).}
log posterior density of the parameters [14]. Regularization seeks to maximize a penalized log-likelihood function, which consists of a log-likelihood function plus a penalty function weighted by a regularization parameter. Specifically, we propose to maximize the following penalized log-likelihood function using $\ell_2$ regularization:

$$L(\alpha) = \int_0^T -\lambda(t|\alpha)dt + \int_0^T \log(\lambda(t|\alpha))dN(t) - \rho\alpha^T Q\alpha$$

where $\rho > 0$ denotes a regularization parameter, and $Q$ denotes a user-defined positive semi-definite matrix (to be defined below). When $Q = I$ (identity matrix), then the standard “ridge regression” is recovered. Maximization of the penalized log-likelihood (4) can be solved by the expectation-maximization (EM) algorithm [5], below we propose an alternative yet more efficient solution.

A standard way to optimize the penalized log-likelihood is through the Newton method. Specifically, let $H(\alpha)$ and $g(\alpha)$ denote the Hessian matrix and gradient vector of the parameter vector $\alpha$, respectively, computed from the $L(\alpha)$ given by (4); the iterative Newton update equation (at the $n$th step) is given by

$$\alpha_{n+1} = \alpha_n - H^{-1}(\alpha_n)g(\alpha_n)$$

where $y = [dN(1), \ldots, dN(T)], X = [x(1), \ldots, x(T)], W(\alpha) = \text{diag}(w_1, \ldots, w_T)$ is a $T \times T$ diagonal weighting matrix ($w_t = \lambda(t|\alpha)$). Since $L(\alpha)$ is concave w.r.t. $\alpha$, the maximum likelihood estimation reduces to a convex optimization problem. Equation (5) can also be formulated as iteratively solving a linear quadratic system: $X^T W(\alpha_n) X + \rho Q \alpha_{n+1} = X^T W(\alpha_n) b$, where $b = X \alpha_n + W^{-1}(\alpha_n)(y - \hat{y}(\alpha_n))$. For such a convex optimization problem, efficient iterative algorithms such as reweighted least squares (RWLS) [11] or conjugate gradient [7] can be applied.

Next, for the purpose of regularization we want to impose a physiologically-inspired constraint on the parameters in $\alpha$. The motivation is that spiking history dependence often has a local smoothness between the neighboring temporal windows. When the motivation is that spiking history dependence often has a local smoothness, let us define a new quadratic penalty function:

$$\sum_{c,k} (\alpha_{c,k} - \overline{\alpha}_{c,k})^2 = \sum_{c,k} \gamma (\alpha_{c,k} - \overline{\alpha}_{c,k-1})^2,$$

which penalizes the local variance of $\{\alpha_{c,k}\}$. Let $\overline{\alpha}_c$ denote the short-term average vector for the corresponding parameter $\alpha_c = [\alpha_{c,1}, \ldots, \alpha_{c,k}]$; then we further have

$$\|\alpha_c - \overline{\alpha}_c\|^2 = \|\alpha_c - S\overline{\alpha}_c\|^2,$$

where a smoothing matrix $S$ is introduced to represent $\overline{\alpha}_c$ in terms of $\alpha_c$. Note that the exponentially moving average $\overline{\alpha}_c$ can be viewed as a convolution product between the sequences $\{\alpha_{c,k}\}$ and a template. Suppose the template vector has an exponential-decay property with length $\ell=4$, such that template $\equiv [(1 - \gamma), \gamma(1 - \gamma), \gamma^2(1 - \gamma), \gamma^3(1 - \gamma)]$. Note that the convolution smoothing operation can also be conveniently expressed as a matrix product operation: $\overline{\alpha}_c = S\alpha_c$, where $S$ is a Toeplitz matrix with the right-shifted template appearing at each row given as follows:

$$S = \begin{bmatrix}
    0 & 0 & 0 & \cdots & 0 \\
    \gamma & 0 & 0 & \cdots & 0 \\
    0 & \gamma & 0 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \ddots & \vdots \\
    \gamma^3(1 - \gamma) & \gamma^2(1 - \gamma) & \gamma(1 - \gamma) & 1 - \gamma & 0 \\
\end{bmatrix}\ .$$

Finally, we obtain the regularization matrix $Q = P^T P$, where in light of (7) P has a block-Toeplitz structure:

$$P = \begin{bmatrix}
    1 - S & 0 & 0 & \cdots & 0 \\
    0 & 1 - S & 0 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \ddots & \vdots \\
    0 & 0 & \cdots & 0 & 1 - S \\
\end{bmatrix}. $$

Our smoothing operator can be seen as an extension of the contingent smoothness operator (as in [15]), where the term $(\alpha_{c,k} - \overline{\alpha}_{c,k})^2$ in (8) is replaced by $(\alpha_{c,k} - \alpha_{c,k-1})^2$ (i.e., the local mean $\overline{\alpha}_{c,k}$ is replaced by its intermediate neighbor $\overline{\alpha}_{c,k-1}$ without moving averaging). Nevertheless, our “smoothness” operator is more general and also accommodates [6] as a special case. Like ours, the regularization matrix $Q$ in [6] also has a block-Toeplitz structure. Since the regularized minimization problem is convex w.r.t. $\alpha$, so the final regularized solution remains globally optimal.

Upon convergence of the algorithm (using a criterion that the log-likelihood change in two subsequent updates is less than $10^{-4}$), the goodness-of-fit of the point process model is evaluated based on the Time-Rescaling Theorem and Kolmogorov-Smirnov (KS) test [1]. The autocorrelation function of the rescaled time series is also computed to test the independence. It should be pointed out that our optimization procedure is insensitive to the initial parameter vector $\alpha$, and it typically converges within 10 iterations. In our experiment, we empirically set $\gamma=0.5$, and the regularization parameter $\rho$ can be selected by cross-validation on the KS statistic or the deviance.

IV. DATA AND RESULTS

The real spike train data used in this study were recorded from 13 neurons in forelimb (7/13) and hindlimb (6/13) areas of the primary motor cortex (M1) of awake behaving cats in an extracellular recording preparation [13]. During recording periods, the animals performed a skilled reaching task. The data analyzed in this study were taken from one recording session in one cat where the same task was repeated for 42 independent trials. Each complete trial recording lasted 3 seconds, which contains “baseline”, “pre-motion”, “reaching”, “withdraw”, and “feed” [13]. Thirteen single units were isolated, including 8 regular-spiking (RS, pyramidal) cells and 5 fast-spiking (FS, interneuron) cells (# 3, 4, 5, 8, 10), which were classified based on baseline firing rate and spike wavelength [4]. Based on whether the cat was able to accurately grasp the food pellet in one, or required more than one attempt, the
42 trials were divided into 21 successful and 21 unsuccessful trials, respectively.

A total of $13 \times 9$ spiking history windows were empirically chosen (each cell contains 9 preceding history of spike counts), and $d_{trim}(\alpha)$ is 118 for each neuron. The empirically selected 9 firing history windows (up to preceding 40 ms) consist of the spike counts in the past 1-3, 4-6, 7-9, 10-12, 13-15, 16-20, 21-25, 26-30, 31-40 ms, respectively. Because neuronal spiking activity is conditionally independent (Eq. 3), individual neurons were fit with separate point process GLMs, followed by a KS test. For each neuron, the regularization coefficient $\rho$ was selected by cross-validation. As an example, we show a few snapshots of dynamic change in functional connectivity between 13 motor neurons during reaching movement (Fig. 2). Figure 3 also shows the estimation results of one cell (#2) during baseline and reaching periods for both successful trials. Both KS plots fall within 95% confidence bounds, indicating that the point process model provides a good description of the spike trains. In this example, it also appears that during baseline, the spiking dependence coefficients between many cell pairs are close to 0 (e.g., $3 \rightarrow 2, 7 \rightarrow 2, 11 \rightarrow 2, 13 \rightarrow 2$). The same cell pairs show more significant nonzero connectivity during reaching movement, suggesting that the neuronal interactions between these cell pairs are task-related. In order to characterize the time-varying functional connectivity among ensemble neurons, the excitatory (E) and inhibitory (I) connectivity ratio was computed within specific temporal windows. The ratio was defined as the total number of significant nonzero (positive or negative) coefficients against the total number of pairwise connections. The result on performance-related reaching movement is shown in Fig. 4. The average (E+I) connectivity ratio is about 0.35 over time. A close examination of Fig. 4 reveals that the number of excitatory and inhibitory connections are quite balanced (in both successful and unsuccessful trials), indicating the state of neuronal network is balanced by both excitation and inhibition. However, during successful trials, the numbers of excitatory and inhibitory connections appear more synchronized (in the same mode of increase or decrease in connectivity ratio) than in unsuccessful trials. Cell pairs were further classified according to subtypes (RS-FS, RS-RS, FS-FS) in order to investigate the incidence of functional interactions in these three groups. It was found that in both baseline and reach conditions, FS-FS pairs were most likely to display significant temporal spiking dependency, followed by RS-FS and RS-RS pairs. This observation was consistent with the findings discussed in [4]. In addition, when the total number of significant temporal spiking dependent events was compared between the baseline and reaching conditions, more interactions occurred during reaching in all three cell-pair groups, as well as for both successful and unsuccessful trials. This is in agreement with descriptions of task-related correlated neural activity [13]. In comparing skill-related successful vs. unsuccessful trials for each cell (data not shown), it was common to observe opposite excitatory vs. inhibitory effects among some cell pairs (Fig. 5). This phenomenon was seen in both RS and FS cells. This provides “statistical” evidence that the strength, timing and pseudo-pyramidal effect of functional interactions between task-related cell pairs may play a role in coding for performance-related skill.

V. Discussion

We have developed a regularized point process GLM for characterizing functional connectivity of ensemble motor neurons during a reaching movement task. The introduction of “temporal smoothness” regularization into the model is important in that first, it significantly reduces the variance of the parameter estimate (due to space limit, we cannot show the non-regularized estimation results), thereby improving the generalization ability for the unseen spike train data in cross-validation (see e.g., Fig. 6); and second, it imposes a physiologically plausible constraint (as a prior) on the solution, making the interpretations of our results more meaningful.

In our preliminary analysis, we observed a dynamic temporal spiking dependency (beyond the standard 2nd-order cross-correlogram analysis) within M1 neuronal ensembles (Figs. 2-4). It was also found that significant (nonzero) statistical dependence between neuron pairs were seen more common during task performance, but not during the baseline period (Fig. 3). Interestingly, while the outcome of task performance did not appear to affect the incidence of functional interactions, the differences observed in neuronal interactions during successful and successful trials suggest that temporal coding in ensemble neurons may influence task performance.

The above findings provide a promising direction in interpreting the physiology and temporal coding among recorded motor neurons. An in-depth statistical analysis for more spike train data is currently under investigation. With further analysis, we hope that the statistical analysis within the regularized point process GLM framework
Fig. 3. An example of estimated GLM coefficients \( \{ \alpha_{c,k} \} \) for 9 history-dependent components from all 13 cells during baseline (top) and reaching movement (bottom). A→B assumes cell A triggers the target cell B firing with a unidirectional spiking dependence. The KS plot and the deviance convergence curve are also shown at the last two panels in each case. Shaded areas represent 95% confidence bounds.

Fig. 4. The time-varying significant nonzero connectivity ratio for both excitatory (E) and inhibitory (I) connections during reaching movement within successful and unsuccessful trials.

Fig. 5. Two representative examples of estimated cell pair interactions between unsuccessful and successful trials during reaching movement.

Fig. 6. The training and cross-validated KS statistics vs. the regularization coefficient \( \rho \) (cell #2, among 21 successful trials during reaching movement). In this case, the suboptimal \( \rho \) is chosen to be 30 (for \( \gamma = 0.5 \)).

may shed some light on discovering the functional connectivity among ensemble neurons. More specifically, this may lead to developing computational tools which allow to make inferences regarding task performance based on the nature of neuronal interactions.

REFERENCES


