TOPOLOGY IDENTIFICATION OF DYNAMIC POINT PROCESS NETWORKS

Syed Ahmed Pasha and Victor Solo

School of Electrical Engineering & Telecommunications
University of New South Wales
NSW 2052 Australia

ABSTRACT

Recently, there has been a growing interest in dynamic networks for understanding interactions and information flows. A fundamental problem is the identification of the links or the network topology. In comparison with its time series counterpart, the problem has received little attention in the point process literature. But with high-dimensional point process data becoming available in a number of application areas such as communication networks and neural coding, topology identification has become crucial for understanding the information flows. Here we discuss for the first time topology identification of a dynamic network of interacting Hawkes processes. Cortical recordings from cats are used to identify the interaction of neurons in the primary visual cortex.

***Index Terms*—** Point process, stochastic intensity, topology identification, sparse estimation, penalized least squares

1. INTRODUCTION

The recent interest in dynamic networks is widespread [1]. Applications include sociological networks [2], biological networks [3], economic networks [4], ecology [5], epidemiology [6], genetic networks [7] and more recently networks involving point processes [8].

But the point process network literature is in its infancy. Applications typically involve counting interactions such as phone calls, emails [9] and neural spike trains [10]. [8] develop methods for the analysis of point process networks with directed interactions. [10] develop a sparsity approach to fitting a network model with limited history dependence. [11] develop methods for the estimation of time-varying Poisson networks that have some limited history dependence.

In the time series literature, [12] used the group LASSO method [13] to recover a sparse network modeled by the autoregressive model. For high frequency data, [14] proposed using the causal Laguerre basis representation for the transfer function between two nodes to avoid very high-order autoregressive models.

This work was partly supported by the ARC Discovery grant DP130103081.

While the Laguerre parameterization of the point process stochastic intensity is not a new idea (see [15, 16], also [17] for Laguerre parameterization of the log-intensity), the topology identification problem under such a general representation has not been attempted.

Here we discuss for the first time topology identification of a dynamic network of interacting Hawkes processes [18, 19]. Hawkes processes are the natural generalisation of autoregressive processes to the point process domain and yield rich dynamics.

In the remainder of the paper we discuss the Hawkes-Laguerre point process model in section 2. We discuss topology identification for the Hawkes-Laguerre model in section 3. The estimation procedure is used to identify the neuronal network in the cat primary visual cortex (section 4). The paper concludes with some final comments in section 5.

Notation. Given $x \in \mathbb{R}^n$ and $P \in \mathbb{R}^{n\times n}$, by $\|x\|_P$ we mean $(x^T Px)^{\frac{1}{2}}$ and $(\alpha)_+ \equiv \max(0, \alpha)$.

2. HAWKES-LAGUERRE POINT PROCESS MODEL

We observe a d-dimensional multivariate point process N_t consisting of counting processes $N_{k,t}, k = 1, \cdots, d$. Here $N_{k,t} = \#$ events of the k-th process up to time and including t. Under No-Simultaneity [20] or orderliness [21] (i.e. in a small time interval with high probability only zero or one event of any type can occur) we can define the vector stochastic intensity using $dN_{k,t} = N_{k,t+\delta} - N_{k,t}$ as,

$$\mu_{k,t} = P(dN_{k,t} = 1|\mathcal{H}^t) = \mu_{k,t}\delta + o(\delta), k = 1, ..., d$$

where $\mathcal{H}^t$ is the history of the vector process $N_s, 0 < s \leq t$. The observations can then be modelled by

$$Y_{k,t} = \mu_{k,t} + e_{k,t},\ k = 1, ..., d$$

where $Y_{k,t} = \frac{1}{\delta} dN_{k,t}$ and $e_{k,t}$ is a martingale increment noise.

The Hawkes-Laguerre model for the stochastic intensity is given by

$$\mu_{k,t} = c_k + \Sigma_1^d \int_0^t h_{k,j}(u) dN_{j,t-u}$$

where c_k is the unknown background firing rate of the k-th component of the multivariate process. The transfer function $h_{k,j}(\cdot)$ is expanded in Laguerre polynomials which is a causal basis [22] and has the significant advantage that ensuring positivity of the intensity function is straightforward. It also has the great advantage that it can be fitted using linear methods. The Laguerre expansion is

$$h_{k,j}(u) = \Sigma_1^{p_{k,j}} \beta_{k,j,l}(u\beta_{k,j,0})^{l-1} e^{-\beta_{k,j,0}u} \beta_{k,j,0}$$

where $\frac{1}{\beta_{k,j,0}}$ is a user chosen time constant. This is similar to [15, 16] with a slight modification (namely the scaling term $\beta_{k,j,0}$ on the end) which improves scaling issues.

Continuing, the stochastic component of $\mu_{k,t}$ is then

$$\begin{aligned} & \int_0^t h_{k,j}(u) dN_{j,t-u} \\ = & \Sigma_1^{p_{k,j}} \beta_{k,j,l} \int_0^t (u\beta_{k,j,0})^{l-1} e^{-\beta_{k,j,0}u} \beta_{k,j,0} dN_{j,t-u} \\ = & \Sigma_1^{p_{k,j}} \beta_{k,j,l} \psi_{k,j,l;t} \end{aligned}$$

where crucially, $\psi_{k,j,l;t}$ can be precomputed and so can be assumed known. We thus have a model linear in the parameters

$$\mu_{k,t} = c_k + \Sigma_1^d \Sigma_1^{p_{k,j}} \beta_{k,j,l} \psi_{k,j,l;t} \qquad (1)$$

Note that the Laguerre expansion provides a valid causal basis expansion for any positive $\beta_{k,j,0}$ and so $\beta_{k,j,0}$ can be specified by the user and need not be estimated.

3. POINT PROCESS NETWORK IDENTIFICATION

We develop a least squares solution for group sparse estimation of the dynamic point process network.

We partition the interval $0 < t \leq T$ into tiny bins of width δ so that $dN_{k,t} = N_{k,t+\delta} - N_{k,t}$ is 0 or 1 with very high probability. Also, $T = K\delta$ and $t = n\delta$.

Then the point process data together with the Laguerre model (1) can be expressed as a standard linear regression model

$$y_k = X_k \beta_k + e_k$$

where $y_k = (Y_{k,n} - c_k, ..., Y_{k,0} - c_k)^T$. β_k can be partitioned into groups $\beta_k = (\beta_{k,1}^T, ..., \beta_{k,d}^T)^T$, then $\beta_{k,j}$ are the coefficients of the directed link from node j to node k with corresponding elements of X_k given by the j-th column of $X_k = (X_{k,1}, ..., X_{k,d})$ where

$$X_{k,j} = \begin{pmatrix} \psi_{k,j,1;n} & \cdots & \psi_{k,j,p_{d,d};n} \\ \vdots & \ddots & \vdots \\ \psi_{k,j,1;0} & \cdots & \psi_{k,j,p_{d,d};0} \end{pmatrix}$$

The network identification problem under the l_1 penalized least squares criterion for grouped variables is the following convex optimization problem

$$\min J(\beta_k), \; k = 1, ..., d$$

with

$$J = \frac{1}{2}\|y_k - X_k\beta_k\|^2 + \lambda \Sigma_1^d \|\beta_{k,j}\|_{P_{k,j}} \qquad (2)$$

where $\lambda > 0$ is the penalty parameter. With $\beta_{k,j} \in \mathbb{R}^{p_{k,j}}$, we will take $P_{k,j} = p_{k,j} I$ as in [13, 14].

The second term in (2) is not differentiable at $\beta_{k,j} = \mathbf{0}$ but the sub-gradient method [23] immediately gives the necessary conditions.

Taking the derivative with respect to $\beta_{k,j}$ of J and setting to zero gives

$$-X_{k,j}^T(y_k - X_k\beta_k) + \lambda\sqrt{p_{k,j}}\bar{\beta}_{k,j} = \mathbf{0}$$

where $\bar{\beta}_{k,j} = \frac{\beta_{k,j}}{\|\beta_{k,j}\|}$ for $\beta_{k,j} \neq \mathbf{0}$ and $\|\beta_{k,j}\| < 1$ for $\beta_{k,j} = \mathbf{0}$ which gives the second condition

$$\|X_{k,j}^T(y_k - X_k\beta_k)\| < \lambda\sqrt{p_{k,j}}$$

In the particular case of orthonormal $X_{k,j}$, the solution is [13]

$$\beta_{k,j} = \left(1 - \lambda\frac{\sqrt{p_{k,j}}}{\|Z_{k,j}\|}\right)_+ Z_{k,j} \qquad (3)$$

where $Z_{k,j} = X_{k,j}^T(y_k - X_k\beta_{k,-j})$ with $\beta_{k,-j} = (\beta_{k,1}^T, ..., \beta_{k,j-1}^T, \mathbf{0}, \beta_{k,j+1}^T, ..., \beta_{k,d}^T)^T$. Given the estimate of β_k,

$$c_k = \frac{1}{K+1}\Sigma_0^K(y_k - X_k\beta_k) \qquad (4)$$

Equations (3),(4) are solved iteratively as described below.

Note that the criterion (2) promotes sparsity at the group level so that some groups will have zero coefficients and some all non-zero coefficients. Sparsity within a group can be promoted by an l_1 optimization for each directed edge with group sizes of 1.

3.1. Computational Details

As already mentioned, X_k is known and needs to be computed just once. $X_{k,j}$ will not be orthonormal in general. While this is not necessary, the algorithm in such a case has better numerical properties. For this reason, orthonormalizing $X_{k,j}$ is preferred.

For $k = 1, ..., d$, the computations can be done concurrently. For a given k and an estimate of c_k, β_k, the algorithm proceeds in a cyclic descent manner where each iteration of the algorithm involves cycling through $j = 1, ..., d$ for computing $\beta_{k,j}$. This is followed by the c_k update step (4). The algorithm terminates on convergence of the estimates.

Note that in the $\beta_{k,j}$ update step (3), $Z_{k,j}$ is computed as

$$\begin{aligned} Z_{k,j} &= X_{k,j}^T e_k + \beta_{k,j} \\ e_k &= y_k - X_k\beta_k \end{aligned}$$

The estimation problem remains manageable in the high-dimensional setting since the procedure has a complexity $\mathcal{O}(\max_k \Sigma_1^d p_{k,j})$ and $p_{k,j}$ is typically not high.

4. DATA ANALYSIS

The neural data come from a study to understand cortical circuit functions and network dynamics in cats from simultaneous recordings of the neural activity [24].

Several multi-channel electrode arrays were used to record simultaneously from single units spanning the cortical layers in cats. Visually evoked responses in the primary visual cortex were recorded from just 10 cells. Here we analyze spontaneous activity recorded from neurons in the cat cortical area 17 [25]. These data have previously been analyzed in [24, 26, 27] as well as [28] to detect recurring patterns in the spike trains. But no previous work has dealt with the kind of network analysis we pursue here.

The data comprise neuronal recordings of spontaneous activity from 25 neurons (channels) recorded simultaneously for about 2.7 min. For analysis we considered recordings in the interval $(0, 10]$ s and only considered channels with 100 or more counts in the interval. This led to $d = 10$ channels with about 150 counts per channel. A raster plot of the neuronal firings in each channel is shown in Fig. 1.

The transfer functions $h_{k,j}(\cdot),\ k, j = 1, ..., d$ were expanded in $p_{k,j} = 5$ terms of the Laguerre polynomials with the time constant $\frac{1}{\beta_{k,j,0}} = 0.5$ s. The counting increments $dN_{k,t}$ were constructed using a discretization step $\delta = 0.8$ ms. The starting values of the parameters $(c_k, \beta_k),\ k = 1, ..., d$ in the optimization algorithm were obtained from the standard least squares solution. With the l_1 penalty parameter $\lambda = 50$ for $k = 1, ..., d$, the sparse network shown in Fig. 2 was identified. The self-exciting link at each node has been dropped to avoid clutter and bi-directional interactions between 2 neurons are indicated by the undirected edge.

The parameter estimates (c_k, β_k) for $k = 1, ..., d$ were found to settle after about 100 iterations. The magnitude of the coefficients $\beta_{k,j,l}$ for $l = 5$ was relatively small compared to the coefficients for $l < 5$ and for a given l, $\beta_{k,k}$ had the largest magnitude, indicating dominance of the self-exciting term at each node.

In comparison to the fully connected network with $d^2 = 100$ links and $d + d^2 p = 510$ parameters, the l_1 penalized least squares removed 58 links and yielded 220 parameters.

We find that neurons 4 and 19 do not influence the neural activity in other neurons. Neurons 14 and 25 influence activity in neurons $2, 13$ and $3, 4$ respectively. Similarly, neuron 21 influences activity in all neurons except $3, 4, 15$ and so on.

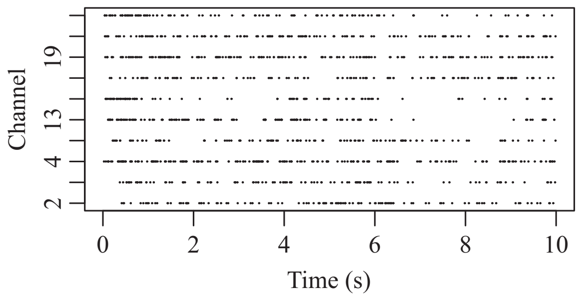

Fig. 1. Raster Plot of Neural Firings in the Cat Cortical Area 17.

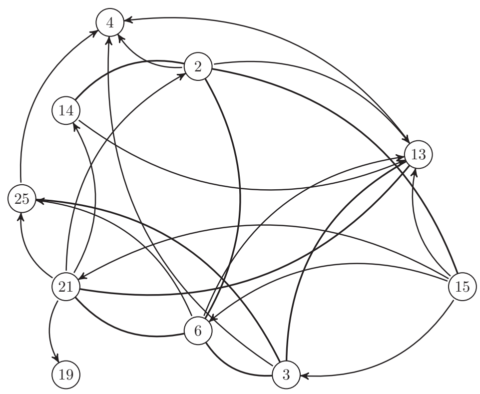

Fig. 2. Neuronal Network Topology in the Cat Cortical Area 17.

5. CONCLUSIONS

In this paper we have addressed the topology identification problem for the dynamic point process model. The modelling is based on the Hawkes process which has dependence on its past. The Laguerre representation ensures positivity of the intensity function and can be fitting using linear methods. We use the group sparse penalized least squares approach for topology identification and test the algorithm on some neural recordings from cats.

In this work, the number of Laguerre polynomial terms in the transfer functions $h_{k,j}(\cdot),\ k, j = 1, ..., d$ were fixed to $p = 5$ and the l_1 penalty parameter λ in (2) was chosen by trial and error. Future work will consider model selection based on the Bayesian Information Criterion to determine $p_{k,j}, \lambda$. Future work will also consider promoting sparsity within a group using more general penalty functions in the optimization [29] as well as a likelihood in place of the least squares term.

We will also be investigating use of more formal statistical graphical models [30].

Acknowledgement. Neural data were recorded by Tim Blanche in the laboratory of Nicholas Swindale, University of British Columbia, and downloaded from the NSF-funded CRCNS Data Sharing website.

6. REFERENCES

[1] ED Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*, Springer, New York, 2009.

[2] DJ Watts and SH Strogatz, "Collective dynamics of 'small-world' networks," *Nature London*, vol. 393, pp. 440–442, 1998.

[3] A. Wagner and D. Fell, "The small world inside large metabolic networks," *Proc. Royal Society London Series B*, vol. 268, pp. 1803–1810, 2001.

[4] M. Naylor, L. Rose, and B. Moyle, "Topology of foreign exchange markets using hierarchical structure methods," *Physica A*, vol. 382, pp. 199–208, 2007.

[5] D. Urban and T. Keitt, "Landscape connectivity: A graph-theoretic perspective," *Ecology*, vol. 82, no. 5, pp. 1205–1218, 2001.

[6] C Moore and MEJ Newman, "Epidemics and percolation in small world networks," *Physical Review E*, vol. 61, no. 5, pp. 5678–5682, 2000.

[7] TS Gardner, S Shimer, and JJ Collins, "Inferring microbial genetic networks," *ASM News*, vol. 70, no. 3, pp. 121–126, 2004.

[8] PO Perry and PJ Wolfe, "Point process modeling for directed interaction networks," *Jl. Royal Stat. Soc.: Series B*, vol. 75, pp. 821–849, 2013.

[9] JR Tyler, DM Wilkinson, and BA Huberman, "E-mail as spectroscopy: Automated discovery of community structure within organizations," *Inform. Soc.*, vol. 21, pp. 143–153, 2005.

[10] RC Kelly, RE Kass, MA Smith, and TS Lee, "Accounting for network effects in neuronal responses using l_1 regularized point process models," *Adv. Neural Inf. Process. Syst.*, vol. 23(2), pp. 1099–1107, 2010.

[11] R. Shyamsundar, T. Graepel, and R. Herbrich, "Poisson-networks: A model for structured point processes," in *AI AND STATISTICS 10*, 2005.

[12] A. Bolstad, BD Van Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Sig. Proc.*, vol. 59, pp. 2628–2641, 2011.

[13] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[14] AJ Seneviratne and V. Solo, "Topology identification of a sparse dynamic network," in *Proc. IEEE CDC*, 2012, pp. 1518–1523.

[15] Y. Ogata and H. Akaike, "On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes," *J. Roy. Statist. Soc. Series B (Methodological)*, pp. 102–107, 1982.

[16] V. Solo and S.A. Pasha, "Point process principal components analysis via geometric optimization," *Neural Computation*, vol. 25(1), pp. 101–122, 2013.

[17] V. Solo and A. Pasha, "Testing for independence between a point process and an analog signal," in *Proc. IEEE ICASSP*, 2010, pp. 3762–3765.

[18] A.G. Hawkes, "Point spectra of some mutually exciting point processes," *J. Roy. Statist. Soc.*, vol. 33(3), 1971.

[19] A.G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58(1), pp. 83–90, 1971.

[20] V. Solo, "High dimensional point process system identification: PCA and dynamic index models," in *Proc. IEEE CDC*, 2006, pp. 829–833.

[21] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes, Volume I (2nd. ed.)*, Springer-Verlag, New York, 2003.

[22] B. Wahlberg, "System identification using laguerre models," *IEEE Trans. Automatic Control*, vol. 36, no. 5, pp. 551–562, 1991.

[23] S. Alliney and S.A. Ruzinsky, "An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation," *IEEE Trans. Sig. Proc.*, vol. 42, no. 3, pp. 618–627, 1994.

[24] K Koepsell, TJ Blanche, N Swindale, and BA Olshausen, "Modeling the influence of local network activity on neuron spiking responses in primary visual cortex," *Computational and Systems Neuroscience*, 2008.

[25] T.J. Blanche, M.A. Spacek, J.F. Hetke, and N.V. Swindale, "Polytrodes: high density silicon electrode arrays for large scale multiunit recording," *J. Neurophys.*, vol. 93(5), pp. 2987–3000, 2005.

[26] S.A. Pasha and V. Solo, "Hawkes-Laguerre reduced rank model for point processes," in *Proc. IEEE ICASSP*, 2013, pp. 6098–6102.

[27] S.A. Pasha and V. Solo, "Hawkes-Laguerre dynamic index models for point processes," in *Proc. IEEE CDC, in-press*, 2013.

[28] M.D. Humphries, "Spike-train communities: Finding groups of similar spike trains," *J. Neurosci.*, vol. 31(6), pp. 2321–2336, 2011.

[29] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *J. Comp. Graph. Stat.*, vol. 22, no. 2, pp. 231–245, 2013.

[30] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, J. Wiley, 1990.