2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Modelling Inhibitory Effects with a Nonlinear Hawkes Model

Syed Ahmed Pasha and Victor Solo

Abstract— **Point processes have a wide range of current applications in areas such as high frequency finance, neural coding, and streaming data. The Hawkes model is widely used due to its flexibility and relatively manageable model fitting. However it is not capable of modelling inhibition though this occurs widely in practice. Here we introduce a new nonlinear Hawkes model and develop model fitting, and tuning parameter selection. We illustrate it with a simulation and application to data where it reveals a strong inhibitory effect.**

I. INTRODUCTION

In a growing number of application areas the data take the form of point processes. For example, spiking activity in the brain [1], high-frequency financial market movements [2], genomics [3], streaming data [4] and email records [5]. All these data share two important properties: they exhibit history dependence and are characterized by bursts of activity localized in time.

The Hawkes process [6], a point process conditional intensity model that accommodates these properties, has been applied successfully in numerous areas. In seismology, the Hawkes process has been used to study aftershocks following an earthquake [7]. In [8], a limited memory Hawkes process was used to model functional relationships between neurons. Interactions of motifs along a genome have been modeled with a bivariate Hawkes process [3]. Patterns of criminal behaviour in burglary and gang violence in Los Angeles have been studied employing the Hawkes process [9]. In [10], the authors demonstrated that the clustering properties of the trades-through in a limit order book can be well modelled with a Hawkes process. In [11], the Hawkes process has been used to model viral dynamics in social media systems and tested on some You-tube data. Similarly a Hawkes process has been employed to infer the underlying network of social interactions [12], [5].

In all of the aforementioned works, the Hawkes models are constructed to model self-excitatory behaviour. But there is growing evidence from a number of application areas to indicate the presence of both self-excitatory as well as inhibitory effects. For example, inhibitory effects between neurons are crucial for regulating neuronal activity [13], [14]. Despite these observations, modeling of such behavior has received little attention in the point process literature.

In [15], the conditional intensity is the composite function of the so-called scaled *softplus* function and the exponential Hawkes model. The positivity constraint on the model parameters is relaxed but this approach requires estimating an additional scale parameter for each scalar process. In [16], the conditional intensity is the composition of the sigmoid function and the unconstrained Hawkes exponential model. A maximum likelihood approach where the conditional intensity can take negative values has been discussed but the optimization maximizes an approximate log-likelihood ratio [17]. In [18], a maximum likelihood approach based on the notion of an *underlying intensity* (allowed to be negative) and *restart time* (instant when the intensity becomes non-negative) has been discussed but the exact maximum likelihood computation is limited to the exponential model only. A log-linear parameterization of the conditional intensity function (CIF) naturally ensures positivity with negative parameter values but is known to be highly sensitive to model fitting. In [19], this is achieved by pruning the number of parameters by assuming all scalar processes have the same baseline intensity and using a quasi-Newton method. The nonparametric approaches [20], [21] allow some inhibitory effect but do not ensure positivity of the CIF.

An important consequence of allowing inhibitory effects is that the cluster process interpretation [22] of the Hawkes process is not valid [23].

In this paper, we discuss a model for the conditional intensity that accommodates modeling of both self-excitatory as well as inhibitory effects. More specifically, we make the following contributions:

(i) We develop a maximum likelihood procedure for model fitting and discuss an information theoretic approach to model selection.
(ii) For optimization by gradient ascent, we employ an adaptive step size selection method which is computationally more efficient than backtracking line search. This delivers a significant speed up.
(iii) The maximum likelihood procedure is demonstrated via a simulation example and tested on some neural data from the cat primal visual cortex. The data shows a strong inhibitory component.

The remainder of the paper is structured as follows. In section II, we review the Hawkes-Laguerre conditional intensity model and propose an approach to make it flexible enough to accommodate inhibitory interaction. In section III a maximum likelihood procedure for model fitting and computational details are discussed. In section IV, we present a simulation example and some data analysis. The paper concludes with some final comments in section V.

This work was supported partly by the Australian Research Council.

S.A. Pasha is with the Department of Electrical & Computer Engineering, Air University, Islamabad, Pakistan s.pasha@mail.au.edu.pk

V. Solo is with the School of Electrical Engineering & Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia v.solo@unsw.edu.au

II. CONDITIONAL INTENSITY MODELS

Suppose N_t is a scalar counting process, i.e., $N_t = \#$ events of a process up to and including time t. Under no simultaneity of events, the stochastic (conditional) intensity is the probability of an event in the next small time interval given the history up to the present time. Let λ_t denote the conditional intensity, then it is given by,

$$\mathbb{P}(\mathrm{d}N_t = 1|\mathcal{H}^t) = \lambda_t\delta + o(\delta)$$

where $\mathrm{d}N_t = N_{t+\delta} - N_t$ is a counting increment, $\mathcal{H}^t$ is the history of the counting process N_t and $\lim_{\delta\to 0}\frac{o(\delta)}{\delta} = 0$.

A. Linear Hawkes Model

The Hawkes process [6], models the conditional intensity as a filtered version of the counting increments (jumps) in the counting process with a linear causal filter; thus, incorporating history dependence. Formally, the Hawkes process is given by [6],

$$\lambda_t = c + \int_0^t g(t-u)\mathrm{d}N_u, \tag{1}$$

where c is a deterministic component of background rate which corresponds to events generated by a Poisson process. The (convolution) integral is a stochastic component where $g(u)$ is a causal impulse response. We call it the Hawkes impulse response (HIR) [24].

For the marginal rate, $\lambda_e = \mathbb{E}(\lambda_t)$ to be meaningful, we must have $c > 0$ and the Hawkes' stability condition $H = \int_0^\infty g(u)du < 1$ must be satisfied [6]. We call H the memory parameter [24].

For modelling, it is necessary to parameterize $g(u)$. We have used Laguerre polynomials in our previous work, as we do now due to their enormous flexibility in modeling the dynamics. For a comparative study which shows the superiority of the Laguerre approach to the others see [24].

The Laguerre parameterization of $g(u)$ is

$$\begin{aligned} g(u) &= \Sigma_{m=1}^q \alpha_m\phi_m(u) \quad (2)\\ \phi_m(u) &= e^{-\beta u}\frac{(\beta u)^{m-1}}{(m-1)!}\beta \end{aligned}$$

with $\int_0^\infty \phi_m(u)\mathrm{d}u = 1$. β is the reciprocal of a user chosen time constant and $0 < q < \infty$ is the number of Laguerre polynomial terms for $g(u)$. Using the Laguerre parameterization, $H = \int_0^\infty g(u)\mathrm{d}u = \Sigma_{m=1}^q \alpha_m$. So, under the requirement $\alpha_m \geq 0$ the stability condition becomes simply $0 < H < 1$.

Substituting (2) in (1), we have a linear (in the parameter) Hawkes-Laguerre (LHL) model

$$\lambda_t = \xi_t^T\theta, \tag{3}$$

where $\xi_t = [1, x_1(t), x_2(t), ..., x_q(t)]^T$ with $x_m(t) = \int_0^t \phi_m(t-u)\mathrm{d}N_u$, $1 \leq m \leq q$, and $\theta = [c, \alpha_1, ..., \alpha_q]^T$.

A simple approach to preserve positivity of the conditional intensity is to impose positivity constraints on the impulse response parameters α_m, $1 \leq m \leq q$. However, this limits the ability of the Hawkes process to model inhibitory effects in addition to self-excitatory behavior. Explicit conditions that allow some of the parameters to assume negative values can be derived that guarantee positivity of the conditional intensity but these become too complicated to ensure except for a very small model order q.

B. Nonlinear Hawkes-Laguerre Model

Here, we take a different approach. We let,

$$\lambda_t = |\xi_t^T\theta|, \tag{4}$$

This circumvents both issues, i.e., avoids positivity constraints on α_m, $1 \leq m \leq q$ to accommodate inhibitory effects in addition to self-excitatory behavior and avoids conditions on the parameters that need to be satisfied to preserve positivity of the conditional intensity. We call the resultant model the nonlinear Hawkes-Laguerre (NHL) model.

A potential drawback of this approach is the loss of the linear structure which is appealing in (3). Notwithstanding, in the sequel, we develop a maximum likelihood procedure and discuss computational strategies that deliver an efficient algorithm.

In previous work, we have investigated the alternative 'log-linear' model

$$\ln\lambda_t = \xi_t^T\theta$$

with unconstrained parameters for modeling inhibitory behavior. However we found ensuring stability of the model and convergence of a maximum likelihood procedure to be very challenging. The proposed model (4) does not suffer from a convergence problem. We discuss stability below.

For future reference, we note the derivatives,

$$\begin{aligned} \frac{\mathrm{d}\lambda_t}{\mathrm{d}\theta} &= \frac{\xi_t^T\theta}{|\xi_t^T\theta|}\xi_t, \quad \lambda_t \neq 0\\ &= \mathrm{sgn}(\xi_t^T\theta)\xi_t \quad (5)\\ \frac{\mathrm{d}^2\lambda_t}{\mathrm{d}\theta\mathrm{d}\theta^T} &= 2\delta(\xi_t^T\theta)\xi_t\xi_t^T \quad (6) \end{aligned}$$

where $\delta(.)$ is the Dirac delta function.

Since the Hessian (6) involves the delta function, this precludes the development of a Newton method. In the next section, we pursue a gradient-based approach for maximum likelihood estimation.

Although [25] have provided checkable stability conditions for nonlinear Hawkes models we do not pursue stability here. This is because, although our model falls within the class of models they treat, checking their stability condition turns out to be a non-trivial problem.

III. MAXIMUM LIKELIHOOD ESTIMATION

The point process log-likelihood ratio is given by [26],

$$\mathcal{L} = \int_0^T \ln\lambda_t\mathrm{d}N_t - \int_0^T \lambda_t\mathrm{d}t$$

The gradient with respect to the parameter vector θ is

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta} = \int_0^T \frac{1}{\lambda_t}\frac{\mathrm{d}\lambda_t}{\mathrm{d}\theta}\mathrm{d}N_t - \int_0^T \frac{\mathrm{d}\lambda_t}{\mathrm{d}\theta}\mathrm{d}t$$

Substituting for $\frac{\mathrm{d}\lambda_t}{\mathrm{d}\theta}$ from (5),

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta} = \int_0^T \frac{\mathrm{sgn}(\xi_t^T\theta)}{\lambda_t}\xi_t \mathrm{d}N_t - \int_0^T \mathrm{sgn}(\xi_t^T\theta)\xi_t \mathrm{d}t \quad (7)$$

We develop a gradient ascent approach to maximum likelihood estimation.

A. Gradient Ascent

The gradient ascent step is given by,

$$\theta^{(i+1)} = \theta^{(i)} + \gamma^{(i)} \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta}\Big|_{\theta=\theta^{(i)}}, \quad \theta^{(0)} = \theta_o. \quad (8)$$

$\gamma^{(i)}$ is a step size determined using a line search method. Backtracking line search (or Armijo's rule) [27], [28], is an efficient alternative to exact line search. It is an iterative procedure that involves function evaluation to test a stopping condition. If the function evaluation is not cheap, backtracking line search can be slow. In such a case, the Barzilai-Borwein line search [29], [30], can be considered which is not an iterative procedure. We used it in our scalar work, but a lot of trial and error was required to get it to work in the vector case.

B. Barzilai-Borwein Line Search

The adaptive step size is computed in such a manner that the gradient update rule (8) approximates the Newton method without explicitly computing the Hessian. With minimum overhead, the Barzilai-Borwein line search can often significantly improve the performance of the gradient method. Since the Barzilai-Borwein line search is a two-point method, requiring the previous iterate and gradient, we apply backtracking line search to determine the initial step size only.

Using starting values, θ_o and γ_o, the initial step size $\gamma^{(0)}$ is determined using backtracking line search by setting $\gamma^{(0)} \leftarrow b\gamma^{(0)}$ until

$$\mathcal{L}(\theta_o + \gamma^{(0)}\nabla\mathcal{L}(\theta_o)) > \mathcal{L}(\theta_o) + a\gamma^{(0)}||\nabla(\mathcal{L}(\theta_o))||^2 \quad (9)$$

is satisfied where $\nabla\mathcal{L}(\theta) := \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta}$. $b \in (0,1)$ controls the shrinkage of the step size and $a \in (0,1)$ controls the relaxation of the gradient to satisfy (9). The gradient ascent step (8) delivers $\theta^{(1)}$ and the gradient $\nabla\mathcal{L}(\theta^{(1)})$ can be computed. The step size $\gamma^{(i)}$, $i \geq 1$ is computed using the Barzilai-Borwein line search as [29], [30],

$$\gamma^{(i)} = \frac{(\theta^{(i)} - \theta^{(i-1)})^T(\nabla\mathcal{L}(\theta^{(i)}) - \nabla\mathcal{L}(\theta^{(i-1)}))}{||\nabla\mathcal{L}(\theta^{(i)}) - \nabla\mathcal{L}(\theta^{(i-1)})||^2} \quad (10)$$

In [29], the authors have shown that the gradient method with step size given by (10) is R-superlinearly convergent for the quadratic problem. For more general problems there is no convergence guarantee and $\gamma^{(i)}$ given by (10) may be too small or too large. Therefore, we need to ensure that $\gamma^{(i)}$ satisfies

$$0 < \underline{\gamma} \leq \gamma^{(i)} \leq \bar{\gamma}, \quad \text{for all } i. \quad (11)$$

The update step (8) is repeated until $||\nabla\mathcal{L}(\theta^{(i)})|| < \epsilon$ for some small positive value ϵ.
The algorithm is summarized below.

Algorithm 1 Gradient Ascent

Require: $T_j,\ j = 1,...,N_T,\ \theta_o,\ \gamma_o, \underline{\gamma}, \bar{\gamma} > 0,\ a, b \in (0,1)$ and $\varepsilon \ll 1$.

Precompute $x_m(t)$ for $m = 1,2,...,q$; set $\gamma^{(0)} = \gamma_o$ and evaluate $\nabla\mathscr{L}(\theta_o)$.
while $\mathscr{L}(\theta_o + \gamma^{(0)}\nabla\mathscr{L}(\theta_o)) < \mathscr{L}(\theta_o) + a\gamma^{(0)}||\nabla(\mathcal{L}(\theta_o))||^2$ **do**
  $\gamma^{(0)} \leftarrow b\gamma^{(0)}$
end while
Set $\theta^{(1)} \leftarrow \theta_o + \gamma^{(0)}\nabla\mathscr{L}(\theta_o)$ and evaluate $\nabla\mathscr{L}(\theta^{(1)})$.
while $\|\nabla\mathscr{L}(\theta^{(i)})\| > \varepsilon$ **do**
  Compute $\gamma^{(i)}$ via (10) and ensure (11).
  Set $\theta^{(i+1)} \leftarrow \theta^{(i)} + \gamma^{(i)}\nabla\mathscr{L}(\theta^{(i)})$ and evaluate $\nabla\mathscr{L}(\theta^{(i+1)})$.
end while

C. Computational Details

In the gradient computation (7), the first term

$$\int_0^T \frac{\mathrm{sgn}(\xi_t^T\theta)}{\lambda_t}\xi_t \mathrm{d}N_t = \sum_{j=1}^{N_T} \frac{\mathrm{sgn}(\xi_{T_j}^T\theta)}{\lambda_{T_j}}\xi_{T_j}$$

can be computed exactly with ξ_{T_j}, $1 \leq j \leq N_T$ precomputed. The second term in (7) can be approximated using a small discretization step δ,

$$\int_0^T \mathrm{sgn}(\xi_t^T\theta)\xi_t \mathrm{d}t \approx \delta \sum_{k=1}^{n} \mathrm{sgn}(\xi_{t_k}^T\theta)\xi_{t_k}$$

where $t_k := k\delta$ for $1 \leq k \leq n$ with $T = n\delta$. ξ_{t_k}, $1 \leq k \leq n$ can be precomputed.

IV. SIMULATION & DATA ANALYSIS

In this section, we apply the new algorithm to simulated data and real data. And compare it to an EM fit based on a linear Hawkes-Laguerre model.

The aim of the simulation is to test whether the new gradient ascent algorithm is competitive/reliable on a problem for which we already have a reliable algorithm. For fitting linear Hawkes-Lauerre models based on a variety of ways of modelling the HIR, the EM algorithm is widely used and robust [24]. However because the EM is multiplicative it cannot easily fit negative parameters. Further it is not clear how to develop the EM algorithm for a nonlinear Hawkes-Laguerre model. For these reasons we have compared the new algorithm to EM on a linear Hawkes-Laguerre model. Note that when all parameters are positive the nonlinear Hawkes-Laguerre model reduces to a linear Hawkes-Laguerre model. As we shall see the new algorithm is very competitive.

With that in mind we were confident to apply the new algorithm to some neuronal data from the cat primary visual cortex.

A. Simulation

We consider a scalar Hawkes-Laguerre model (3) as follows

(i) Model order is $q = 3$.
(ii) HIR parameter values are $\alpha = (0.2, 0.3, 0.4)^T$.
(iii) A time constant of 1 s is taken ($\Rightarrow \beta = 1$).
(iv) Background rate $c = 0.1$.

The Hawkes stability condition, $\alpha^T 1 = 0.9 < 1$ is satisfied.

The data were simulated using Ogata's thinning procedure [31] to generate about 3,500 counts. The incremental counts of the time binned process using 100 time bins are shown in Fig. 1(a) which shows the self-exciting property of the model.

For system identification, the model order q was determined using the Bayesian Information Criterion (BIC) [32],

$$BIC = -2\mathcal{L} + (1+q)\log N_T$$

by fitting the conditional intensity model (4) for $1 \leq q \leq 6$. Fig. 1(b) shows the relative BIC values after subtracting the BIC value for $q = 0$, which corresponds to the Poisson process. The BIC minimizer coincides with the true model order $q = 3$.

The starting values in θ_o were set to 10^{-2} and $\gamma_o = 1$. To choose the initial step size using backtracking line search, we used typical values for the pair $(a, b) = (0.5, 0.9)$. Using the Barzilai-Borwein line search in subsequent iterations, we found that (10) frequently yielded a step size value that was either too small or negative and setting the lower bound $\underline{\gamma} = 10^{-3}$ yielded an acceptable convergence rate. The step size was never found to be too large and the upper bound $\bar{\gamma} = 1$ was never attained.

For $q = 3$, the log-likelihood ratio iterates are shown in Fig. 1(c) which increase monotonically and flatten after 10 iterations. The gradient ascent (GA) iterates are shown in Fig. 1(d). α_2 and α_3 take about 2000 iterations to converge for the stopping criterion, $||\nabla\mathcal{L}(\theta^{(i)})|| < 10^{-6}$. In Table I, we show the parameter estimates and the relative error at convergence. We find that $\alpha^T 1 = 0.924 < 1$.

The true and estimated Hawkes impulse response are shown in Fig. 1(e) where the contribution of the higher-order terms is clearly visible. The estimated impulse response shows slight deviation from the true response. The conditional intensity constructed from the parameter estimates is shown in Fig. 1(e).

The performance of the gradient ascent algorithm was compared to the EM algorithm [24] which uses a multiplicative update rule to ensure positivity of the impulse response parameters. The BIC values, log-likelihood iterates, and parameter estimates using the EM algorithm coincide with the corresponding values obtained using the gradient ascent algorithm and therefore are omitted. The computational effort (per iteration) of the gradient ascent algorithm is higher but takes fewer iterations to converge.

B. Neuronal Activity Data

The data comprise neuronal responses recorded in the cat primary visual cortex. Spontaneous activity was recorded

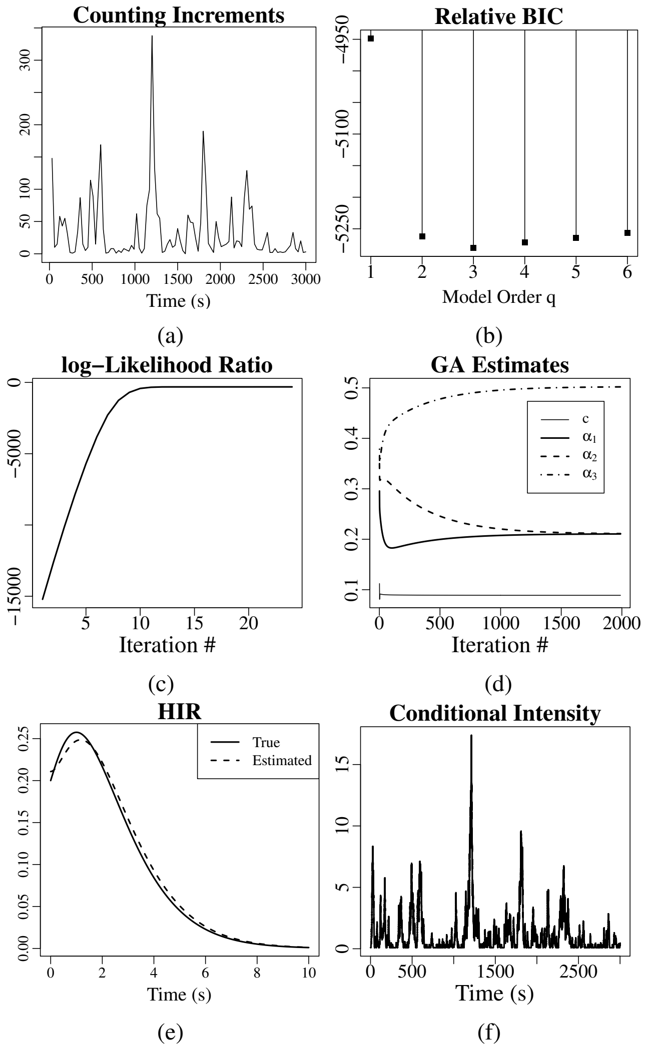

Fig. 1: Simulation: (a) Incremental counts in 100 bins, (b) Relative BIC values, (c) log-likelihood ratio iterates, (d) Gradient ascent iterates, (e) Hawkes impulse response, and (f) Conditional intensity estimate.

for about 162 s which generated 489 counts in channel 1. The incremental counts in 100 time bins are shown in Fig. 2(a). Model order selection was done using the BIC values assembled for $1 \leq q \leq 6$. Fig. 2(b) shows the relative BIC values for $1 \leq q \leq 3$ after subtracting the BIC value for $q = 0$. The minimum BIC value was attained for $q = 3$ (the last three values: 1475.49, 1120.85 and 30752.95 are not shown).

We call the fitted nonlinear Hawkes-Laguerre model the GA-NHL model. Below we compare it with a linear Hawkes-Laguerre model fitted using the EM algorithm which we call the EM-LHL fit.

For $q = 3$, Fig. 2(c) shows the log-likelihood iterates which increase monotonically and flatten after 10 iterations. The GA-NHL iterates are shown in Fig. 2(d). α_2 and α_3 take about 1000 iterations to converge. Note that α_2 starts at a positive value and converges to a negative value that is significantly different from zero. This suggests the presence

TABLE I: Simulation: Gradient Ascent Estimates and Relative Error.

	$c = 0.1$	$\alpha_1 = 0.2$	$\alpha_2 = 0.3$	$\alpha_3 = 0.4$
Estimates	0.089	0.211	0.210	0.503
Relative Error	0.111	0.054	0.298	0.256

TABLE II: Neuronal Activity: GA-NHL and EM-LHL Estimates.

	c	α_1	α_2	α_3
GA-NHL	0.240	0.890	-1.053	1.085
EM-LHL	0.279	0.531	0.000	0.378

of some inhibitory effect in addition to self-excitation which cannot be found using the EM-LHL fit. Table II shows the parameter estimates at convergence. The HIR is shown in Fig. 2(e) and drops rapidly due to the negative coefficient of α_2. The conditional intensity constructed from the GA-NHL parameter estimates is shown in Fig. 2(f).

We tested the EM algorithm on the neural data, by fitting a linear Hawkes model, resulting in the EM-LHL fit. BIC values were assembled for $1 \leq q \leq 6$. Fig. 3(a) shows the relative BIC values for $2 \leq q \leq 4$. The minimum BIC value using the EM-LHL was also attained for $q = 3$. However, the minimum (relative) BIC $= -93.9$ obtained using GA-NHL is lower than the minimum BIC $= -87.47$ obtained using EM-LHL demonstrating the superiority of the proposed nonlinear model. This then provides evidence that the inhibitory effect found by the GA-NHL is real. Note that the BIC value for $q = 0$ obtained using both algorithms is the same which justifies comparison of the relative BIC values.

The EM-LHL iterates are shown in Fig. 3(b). α_2 takes about 500 iterations to converge to zero. Table II shows the parameter estimates at convergence satisfying Hawkes stability condition, $\alpha^T 1 = 0.918 < 1$. The HIR is shown in Fig. 3(c) where the first-order (exponential decay) behaviour dominates with some higher-order effect. The conditional intensity constructed from the EM-LHL parameter estimates is shown in Fig. 3(d) which is roughly similar to GA-NHL Fig. 2(f) but has a lower marginal rate $\lambda_e = \mathbb{E}(\lambda_t)$. So the GA-NHL modelling of inhibition reveals a higher overall level of activity.

V. CONCLUSIONS AND FUTURE WORK

A. Conclusions

In this paper, we have presented a point process nonlinear Hawkes conditional intensity model (where the Hawkes impulse response is modelled compactly with Laguerre polynomials) that accommodates inhibitory behavior in addition to self-excitation.

We developed a gradient ascent approach to maximum likelihood estimation. Step size selection was done using the Barzilai-Borwein line search to improve computational efficiency. This provided a reliable algorithm.

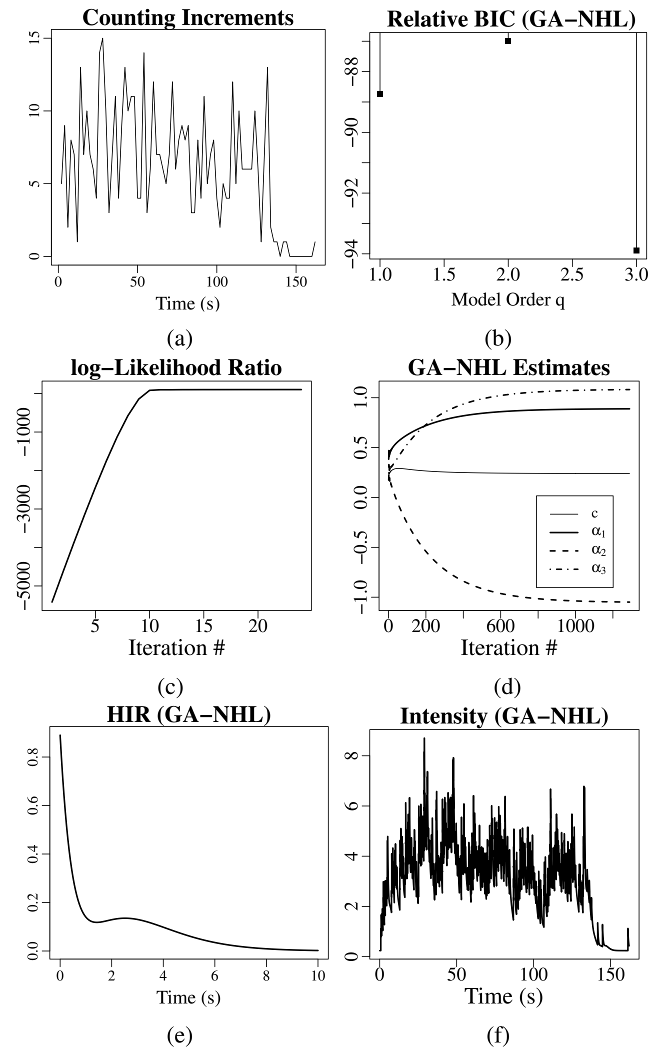

Fig. 2: Neuronal Activity: (a) Incremental counts in 100 bins, (b) Relative BIC values using GA-NHL, (c) log-likelihood ratio iterates, (d) GA-NHL iterates, (e) HIR using GA-NHL, (f) Conditional intensity estimate using GA-NHL.

In simulations, the algorithm (which we called gradient ascent-nonlinear Hawkes-Laguerre (GA-NHL)) was compared to a linear Hawkes-Laguerre model fitted with the EM algorithm (which we called EM-LHL) in order to assess its reliability and competitiveness. Note that when all parameters are positive the NHL reduces to the LHL. The GA-NHL was found to give near identical results.

Finally the GA-NHL algorithm was tested on some neural data recordings in the cat primary visual cortex . It found both self-excitatory as well as inhibitory effects which the EM-LH algorithm cannot uncover. It also exhibited a superior fit according to a BIC criterion.

B. Future Work

Future work will consider estimation of the β parameter by jointly selecting β and the impulse response model order q. Estimation of β in the maximum likelihood procedure will also be investigated. For more general optimization problems

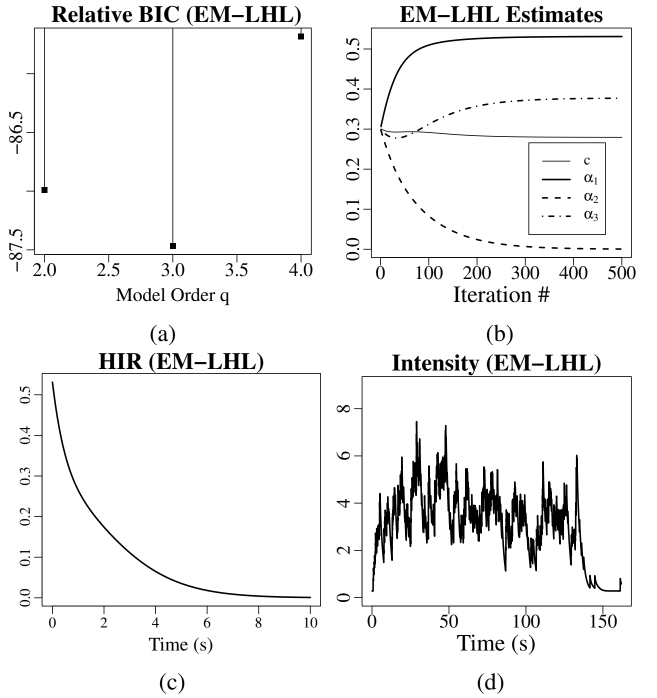

Fig. 3: Neuronal Activity: (a) Relative BIC values using EM-LHL algorithm, (b) EM-LHL algorithm iterates, (c) HIR using EM-LHL, and (d) Conditional intensity using EM-LHL.

than the quadratic problem, the Barzilai-Borwein line search is paired with non-monotone line search to improve convergence rate. This will be investigated in future work. We will also extend the method to multivariate nonlinear Hawkes modelling.

VI. Acknowledgments

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.

References

[1] F. Rieke, D. Warland, R. d. R. Van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. MIT press, 1999.

[2] N. Hautsch, *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*. Springer Science & Business Media, 2004.

[3] G. Gusto and S. Schbath, "FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, 2005.

[4] F. Chen and P. Hall, "Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling," *Journal of Applied Probability*, vol. 50, no. 4, pp. 1006–1024, 2013.

[5] J. R. Zipkin, F. P. Schoenberg, K. Coronges, and A. L. Bertozzi, "Point-process models of social network interactions: Parameter estimation and missing data recovery," *European Journal of Applied Mathematics*, vol. 27, no. 3, pp. 502–529, 2016.

[6] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971.

[7] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, vol. 83, no. 401, pp. 9–27, 1988.

[8] E. Chornoboy, L. Schramm, and A. Karr, "Maximum likelihood identification of neural point process systems," *Biological cybernetics*, vol. 59, no. 4-5, pp. 265–275, 1988.

[9] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.

[10] I. M. Toke and F. Pomponio, "Modelling trades-through in a limit order book using Hawkes processes," *Economics*, vol. 6, no. 1, 2012.

[11] D. MacKinlay, "Estimating self-excitation effects for social media using the hawkes process," *Departement Management, Technologie und Oekonomie, Eidgenoessische Technische Hochschule, Zurich*, 2015.

[12] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes," in *Artificial Intelligence and Statistics*, 2013, pp. 641–649.

[13] T. C. Smith and C. E. Jahr, "Self-inhibition of olfactory bulb neurons," *Nature Neuroscience*, vol. 5, no. 8, pp. 760–766, 2002.

[14] A. Maffei, S. B. Nelson, and G. G. Turrigiano, "Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation," *Nature neuroscience*, vol. 7, no. 12, pp. 1353–1359, 2004.

[15] H. Mei and J. M. Eisner, "The neural Hawkes process: A neurally self-modulating multivariate point process," *Advances in neural information processing systems*, vol. 30, 2017.

[16] N. Malem-Shinitski, C. Ojeda, and M. Opper, "Variational Bayesian inference for nonlinear Hawkes process with Gaussian process self-effects," *Entropy*, vol. 24, no. 3, p. 356, 2022.

[17] R. Lemonnier and N. Vayatis, "Nonparametric Markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014*. Springer, 2014, pp. 161–176.

[18] A. Bonnet, M. M. Herrera, and M. Sangnier, "Maximum likelihood estimation for Hawkes processes with self-excitation or inhibition," *Statistics & Probability Letters*, vol. 179, p. 109214, 2021.

[19] L. Carstensen, A. Sandelin, O. Winther, and N. R. Hansen, "Multivariate Hawkes process models of the occurrence of regulatory elements," *BMC bioinformatics*, vol. 11, pp. 1–19, 2010.

[20] P. Reynaud-Bouret and S. Schbath, "Adaptive estimation for Hawkes processes; application to genome analysis," *The Annals of Statistics*, vol. 38, no. 5, p. 2781, 2010.

[21] E. Bacry and J.-F. Muzy, "First-and second-order statistics characterization of Hawkes processes and non-parametric estimation," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2184–2202, 2016.

[22] A. G. Hawkes and D. Oakes, "A cluster process representation of a self-exciting process," *Journal of Applied Probability*, vol. 11, no. 3, pp. 493–503, 1974.

[23] M. Costa, C. Graham, L. Marsalle, and V. C. Tran, "Renewal in Hawkes processes with self-excitation and inhibition," *Advances in Applied Probability*, vol. 52, no. 3, pp. 879–915, 2020.

[24] B. I. Godoy, V. Solo, and S. A. Pasha, "Truncated Hawkes point process modeling: System theory and system identification," *Automatica*, vol. 113, p. 108733, 2020.

[25] P. Bremaud and L. Massoulle, "Stability of nonlinear hawkes processes," *The Annals of Probability*, vol. 24, p. 1563–1588, 1996.

[26] I. Rubin, "Regular point processes and their detection," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 547–557, 1972.

[27] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[28] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2016.

[29] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.

[30] R. Fletcher, "On the Barzilai-Borwein method," in *Optimization and Control with Applications*. Springer, 2005, pp. 235–256.

[31] Y. Ogata, "On Lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, 1981.

[32] H. Linhart and W. Zucchini, *Model Selection*. John Wiley & Sons, 1986.