NON-NEGATIVE ONLINE ESTIMATION FOR HAWKES PROCESS NETWORKS

Marc Piggott & Victor Solo

School of Electrical Engineering and Telecommunications
University of New South Wales
Sydney, AUSTRALIA
email: v.solo@unsw.edu.au

ABSTRACT

Networks of interacting Hawkes processes have emerged as useful models in neuroscience, geophysics, high frequency finance, and social network analysis. The Hawkes process is of fundamental importance, being a point process analog of an autoregression. Here we develop a fixed gain adaptive (aka online) distributed estimator for the parameters of a Hawkes process model. The stochastic intensity is modeled by a causal Laguerre basis expansion. The natural recursive structure of this basis is exploited to derive a new two time scale adaptive algorithm based on exponentially weighted least squares which preserves non-negativity constraints. Simulations illustrate the results.

Index Terms— Hawkes process, adaptive algorithm, NMF, distributed algorithm.

1. INTRODUCTION

Network Hawkes processes [15], [16], have found application in neural coding, high frequency finance [20], social network analysis [43], bioinformatics [2] and many other areas. Their utility is due to their mutually exciting structure: the occurrence of past points influences the probability of occurrence of future points via an impulse response function (which we call the Hawkes impulse response (HIR). In the language of point processes, the stochastic intensity is history dependent. We refer the reader to classical references [8], [19], [33].

There is a growing literature on estimation for Hawkes processes. Early approaches were based on maximum likelihood (MLE); see e.g. [27]. Expectation-Maximization (EM) methods followed, first in [7], while an alternative EM algorithm exploiting the cluster process structure [17] was given in [37] and subsequently extended in various directions [18], [32]. Recently, another EM-type algorithm incorporating group sparsity has been developed [40]. Bayesian approaches are given in [23], [31]. A popular nonparametric approach is the MMEL algorithm which solves an Euler-Lagrange equation for the exciting kernels [42]. Another line of attack is via approximation with time series models [11], [20]. Mean field approximations are used in [1], while explicit sparsity penalized algorithms are proposed for different aims in e.g. [14], [28], [30],[40], [42], [43].

Despite this, most existing algorithms are not suitable for online (aka adaptive) applications with streaming data and time varying parameters. The small literature on point process adaptive filtering, mostly from the neuroscience community, includes likelihood steepest descent, Kalman filter, recursive least squares (RLS) and sequential Monte-Carlo algorithms; see e.g. [10], [35], [38]. However, there are two main complications when it comes to the Hawkes process.

First, there are non negativity and stability constraints that must be obeyed. A simple approach to manage these issues is to use projections [13] but this can lead to slow convergence and even stalling.

In the traditional (non-point process) setting, constrained adaptive filtering without projections is a long studied, challenging problem [34]. Recently, an RLS dichotomous coordinate descent algorithm which incorporates box and norm constraints was proposed in [25], while [6] gives variants of non-negative least mean squares (LMS); see also [5] and references therein. Online non-negative matrix factorization (NMF) has also been studied [3], [4], [24], [39], [41].

Second, the memory in the Hawkes process results in two time scale behavior which must be addressed. The slow scale is the parameter adaptation; the fast scale is the underlying system state which encapsulates the history dependence. In [13], partly to minimize this problem, the authors make the extremely stringent assumptions that: (i) each node has the same HIR; (ii) the HIRs are first order.

We address all these issues by developing a novel two time scale exponentially weighted least squares (EWLS) algorithm which uses ideas from the online NMF literature [3]. We model the HIR in a Laguerre basis, whose recursive structure is exploited to complete our two time scale algorithm. The new algorithm allows the HIRs to vary from node to node

This work was partly supported by the Australian Research Council

and can easily handle higher order models; so the stringent assumptions of [13] are completely overcome.

2. HAWKES-LAGUERRE MODEL

Consider an d node network represented by a multivariate Hawkes process whose vector counting process $N(t) = (N_1(t), \ldots, N_d(t))^T$, where $N_k(t)$ is a scalar counting process corresponding to the number of events at node k in $(0, t]$ [15], [16]. The vector instantaneous firing rate is history dependent[1] and denoted by $\lambda(t) = (\lambda_1(t), \ldots, \lambda_d(t))^T$. This is also called the vector stochastic intensity, and here we operate under the standard assumptions, e.g. orderliness [8], [19] (which would be better named, no-simultaneity). The Hawkes-Laguerre model for the vector stochastic intensity is

$$\lambda_n(t) = c_n + \Sigma_{k=1}^d \int_0^t g_{nk}(t-u) dN_k(u)$$

being thus made up contributions from every other node. The Laguerre expansion for the HIR $g_{nk}(v)$ [36], [29], [30] is

$$\begin{aligned} g_{nk}(v) &= \Sigma_{l=1}^p \alpha_{n,k,l} \phi_l(v) \\ \phi_l(v) &= \beta e^{-\beta v} \frac{(\beta v)^{\ell-1}}{(\ell-1)!} \end{aligned}$$

where β is a user chosen inverse time constant [2]. Note the scaling which ensures $\int_0^\infty \phi_l(v) dv = 1$. Substituting all this in gives the following expression for the stochastic intensity.

$$\begin{aligned} \lambda_n(t) &= c_n + \sum_{k=1}^d \sum_{\ell=1}^p \alpha_{n,k,\ell} x_{k,\ell}(t) \\ x_{k,\ell}(t) &= \int_0^t \phi_\ell(t-u) \, dN_k(u) \end{aligned}$$

A critical new observation is that the recursive structure of the Laguerre basis allows one to derive natural adaptive algorithms for the Hawkes process; see appendix A.

3. FIXED GAIN NON-NEGATIVE EWLS

3.1. Least Squares Origin

In the Hawkes-Laguerre setting, the least squares formulation is studied in e.g. [29], [30]. Consider node k. Observations at node k at time $t_i = i\Delta$ are denoted by $y_{k,i} = \frac{1}{\Delta}(N_k(t_i) - N_k(t_{i-1}))$. Next, we define vectors

$$\begin{aligned} \theta_k^T &= [c_k, \alpha_{k,1}^T, \ldots, \alpha_{k,d}^T] \\ \alpha_{k,j}^T &= [\alpha_{k,j,1}, \ldots, \alpha_{k,j,p}] \\ w_i^T &= [1, x_1^T(t_{i-1}), \ldots, x_d^T(t_{i-1})] \\ x_k^T(t_{i-1}) &= [x_{k,1}(t_{i-1}), \ldots, x_{k,p}(t_{i-1})]. \end{aligned}$$

[1]The history is formally $\mathcal{H}^t = \sigma(\{N_s, 0 \le s \le t\})$, and we omit the notation to reduce clutter.

[2]For simplicity we have taken β and p to be independent of n and k, the algorithm we derive applies with trivial modifications in the general case.

Note that $\lambda_i(\theta_k) = w_i^T \theta_k$, and that $y_{k,i} - \lambda_i(\theta_k)$ is approximately a martingale difference "noise". We restrict the entries of θ to be nonnegative (written $\theta \ge 0$) to ensure a nonnegative stochastic intensity estimate, and no inhibitory behavior. We thus consider the nonnegative least squares problem for node k at (discrete) time t:

$$arg. \min_{\theta_k \ge 0} J_{k,t}(\theta_k) := \frac{1}{2} \sum_{i=1}^t \left(y_{k,i} - w_i^T \theta_k \right)^2.$$

By vectorizing, the multiplicative update NMF algorithm [22] can be used here. To write the update, we introduce

$$W^{(t)} = \begin{bmatrix} w_1^T \\ \vdots \\ w_t^T \end{bmatrix}.$$

Starting from the initial guess $\theta_{k,0}$, it updates its a^{th} component by

$$\begin{aligned} (\theta_{k,1})_a &= (\theta_{k,0})_a \frac{\sum_{i=1}^t \left(W^{(t)}\right)_{i,a} y_{k,i}}{\sum_{i=1}^t \left(W^{(t)}\right)_{i,a} \left(W^{(t)} \theta_{k,0}\right)_i} \\ &= (\theta_{k,0})_a \frac{\sum_{i=1}^t (w_i)_a y_{k,i}}{\sum_{i=1}^t (w_i)_a w_i^T \theta_{k,0}} \\ &= (\theta_{k,0})_a \frac{\sum_{i=1}^t (w_i)_a y_{k,i}}{\left(\left(\sum_{i=1}^t w_i w_i^T \right) \theta_{k,0} \right)_a} \end{aligned}$$

This is iterated to arrive at a minimizer of $J_{k,t}$. However, in the online setting we have a sequence of cost functions $J_{k,1}, J_{k,2}, \ldots$ etc. At each time, t, we could run the above NMF algorithm to convergence but this is not an online solution. Instead, at time t we take a single step of the NMF update corresponding to $J_{k,t}$.

$$(\theta_{k,t})_a = (\theta_{k,t-1})_a \frac{(S_{k,t})_a}{(P_t \theta_{t-1})_a} \quad (3.1)$$

$$S_{k,t} = S_{k,t-1} + w_t y_{k,t}, \quad S_{k,0} = 0 \quad (3.2)$$

$$P_t = P_{t-1} + w_t w_t^T, \quad P_0 = 0. \quad (3.3)$$

3.2. EWLS Completion

Note that (3.1)-(3.3) gives no control over the learning rate. We thus propose an EWLS version. Setting $0 < \rho < 1$, and $\epsilon > 0$ small, we arrive at

$$(\theta_{k,t})_a = \max\left((\theta_{k,t-1})_a \frac{(S_{k,t})_a}{(P_t \theta_{k,t-1})_a}, \epsilon \right) \quad (3.4)$$

$$P_t = \rho P_{t-1} + (1-\rho) w_t w_t^T, \quad P_0 = 0 \quad (3.5)$$

$$S_{k,t} = \rho S_{k,t-1} + (1-\rho) w_t y_{k,t}, \quad S_0 = 0 \quad (3.6)$$

w_t is updated via (5.1)-(5.2) from appendix A; from there it also follows that w_j is uniformly bounded when $\beta\Delta < 1$. The

use of max in (3.4) is crucial, otherwise a parameter which hits 0 would always remain 0, i.e. one would lose adaptivity. It also improves robustness to initial transients, since a component of θ may be inadvertently attracted toward 0.

3.3. Initialization

The initialization of (3.4) is subtle. Note that $(S_{k,t})_a = 0$ $(a = 2, \ldots, dp+1)$ until the second point is observed. So, (3.4) is not useful the second point occurs at node k. It is thus best to wait until every node has at least 2 points before employing (3.4), though we run (3.5)-(3.6) from time 0.

3.4. Equilibrium Points and Behavior

Rewriting the updates (3.5)-(3.6) as follows yields valuable insight into the algorithm,

$$\begin{aligned}(P_t\theta_{k,t-1})_a &= (1-\rho)\sum_{j=1}^{t}\rho^{t-j}(w_j)_a\,\lambda_j(\theta_{k,t-1})\\ (S_{k,t})_a &= (1-\rho)\sum_{j=1}^{t}\rho^{t-j}(w_j)_a\,y_{k,j}\end{aligned}$$

where $\lambda_j(\theta_{k,t-1}) = w_j^T\theta_{k,t-1}$. Suppose that the true underlying parameters are constant, say $\theta_k^* \geq \epsilon$. Note that if $\theta_{k,t-1} = \theta_k^*$ then since $\lambda_j(\theta_k^*) \approx y_{k,j} + o(\Delta)$ (definition of stochastic intensity), and since w_j is uniformly bounded, one would expect $\theta_{k,t} \approx \theta_{k,t-1} = \theta_k^*$ for infinitesimal Δ. We thus conjecture that the true solution is in some sense an equilibrium point of (3.4)-(3.6).

Also, note that when the memory parameters are small i.e. $(\theta_{k,t-1})_{2:dp+1} = \epsilon$, the update is approximately (for small ϵ)

$$(\theta_{k,t})_1 = \frac{(S_{k,t})_1}{1-\rho^t}$$

which is a scaled (non-negative) LMS algorithm for a Poisson process with gain $\mu = 1-\rho$; see appendix B.

4. SIMULATIONS

We simulated multivariate Hawkes processes using a thinning algorithm [26] which combines a modified[3] version of his algorithm 2 with his algorithm 1. The simulations are verified using the procedure in [12]. The parameters are shown in Figure 1. We chose $m = 10000$ (number of time constants of data), $\beta = 15$ (inverse time constant), $B = 2000$ (number of bins per time constant of data), $\ell = 100$ (number of correlation lengths per window), $L = \frac{1}{\beta(1-\alpha_{\max})} = 6.66$ (maximum correlation length considered, $\alpha_{\max} = 0.99$), $\gamma = \frac{1}{\ell L} = 1.510^{-4}$ (time constant associated with window) and $\epsilon = 0.0001$. These quantities imply, $T = \frac{m}{\beta} = 666.66$ (total simulation interval in seconds), $\Delta = \frac{1}{\beta B} = 3.3310^{-5}$ (bin size in seconds), $\rho = 1-\gamma\Delta$.

Figures 2 and 3 show the parameter estimates for node 1 and 4 respectively.

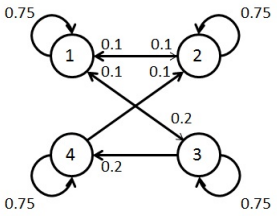

Fig. 1. d=4 node network Hawkes process: $p = 1$; all self HIR parameters = 0.75; all background rates =1; inter-nodal HIR parameters as shown.

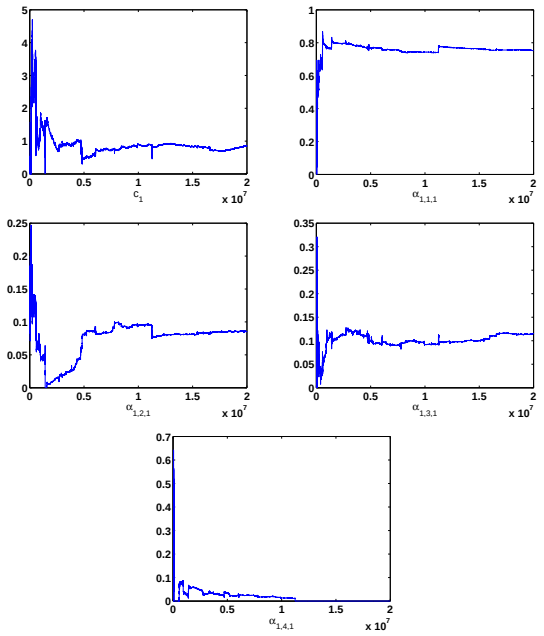

Fig. 2. Node 1 parameter estimates settling to steady state fluctuations near the true values.

5. CONCLUSION

We have proposed an adaptive algorithm for network Hawkes processes, where the HIRs are expanded in terms of Laguerre bases. The algorithm enforces non negativity constraints, and easily manages the two time scale structure of the underlying process by using the recursive nature of the Laguerre bases. Simulations showed the efficacy of the algorithm in a small network setting. Extensions enforcing the stability constraint and employing sparsity will be pursued elsewhere as will be the challenging task of algorithm stability analysis.

[3]Corrections are required in steps 2, 7 and 8 of algorithm 2 in [26].

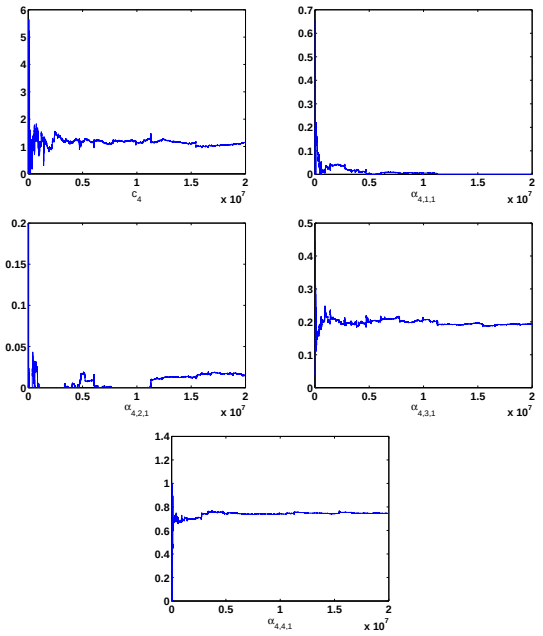

Fig. 3. Node 4 parameter estimates settling to steady state fluctuations near the true values.

A: Online Computation of $x_{k,\ell}$

One can proceed by formulating a state space model as in [36], but a more efficient route is via the following lemma.

Lemma 1. If $f(t) = \int_0^t g(t-u)\, dN_u$ then

$$\frac{df}{dt} = g(0)\frac{dN}{dt} + \int_0^t \frac{\partial g(t-u)}{\partial t} dN_u.$$

Proof. Follows from Leibniz's rule.

We now apply the lemma.

(1) $\underline{x_{k,1}(t)}$. Here $g(u) = \beta e^{-\beta u}$ thus

$$\begin{aligned} g(0) &= \beta \text{ and } \frac{\partial g}{\partial u} = -\beta g(u) \\ \frac{dx_{k,1}(t)}{dt} &= \beta \frac{dN_k}{dt} - \beta x_{k,1}(t). \end{aligned}$$

(2) $\underline{x_{k,2}(t)}$. Here $g(u) = \beta^2 u e^{-\beta u}$ and so

$$g(0) = 0, \frac{dg}{du} = \beta\left(\beta e^{-\beta u}\right) - \beta^2\left(\beta u e^{-\beta u}\right).$$

This yields

$$\int_0^t \frac{\partial}{\partial t} g(t-u)\, dN_k(u) = \beta \int_0^t \beta e^{-\beta(t-u)} dN_k(u)$$
$$- \beta^2 \int_0^t \beta(t-u) e^{-\beta(t-u)} dN_k(u).$$

Putting this together,

$$\frac{dx_{k,2}(t)}{dt} = \beta x_{k,1}(t) - \beta x_{k,2}(t).$$

(3) $x_{k,\ell}(t), \ell \geq 2$.

Following the same line of argument we find

$$\frac{dx_{k,\ell}(t)}{dt} = \beta x_{k,\ell-1}(t) - \beta x_{k,\ell}(t),$$

which is then discretized using backward Euler as

$$\begin{aligned} x_{k,1}(t_i) &\approx (1-\beta\Delta) x_{k,1}(t_{i-1}) \\ &+ \beta\left(N_k(t_i) - N_k(t_{i-1})\right) \qquad (5.1) \\ x_{k,\ell}(t_i) &\approx (1-\beta\Delta) x_{k,\ell}(t_{i-1}) \\ &+ \beta\Delta x_{k,\ell-1}(t_{i-1}), \qquad \ell \geq 2. \qquad (5.2) \end{aligned}$$

B: Steepest Descent - Poisson Case

In the single node Poisson case with rate λ^*, we define

$$J(\lambda) := \frac{1}{2}\mathbb{E}\left[y_t - \lambda\right]^2$$

and note by stationarity,

$$\mathbb{E}[Y_t] = \lambda_* = arg.\min_{\lambda \geq 0} J(\lambda).$$

To get an LMS-type steepest descent algorithm, let $0 < \rho < 1$, note $\frac{dJ}{d\lambda} = \lambda - \mathbb{E}[y_t]$ and replace $\mathbb{E}[y_t]$ with y_t:

$$\begin{aligned} \lambda_t &= \lambda_{t-1} - (1-\rho)\frac{\partial J}{\partial \lambda}(\lambda_{t-1}) \\ &\approx \lambda_{t-1} - (1-\rho)(\lambda_{t-1} - y_{t-1}) \\ &= \rho\lambda_{t-1} + (1-\rho) y_{t-1}. \end{aligned}$$

This is exactly the $(S_{k,t})_1$ update (3.6) if we set $\lambda_0 = 0$.

6. REFERENCES

[1] E. Bacry, S. Gaïffas, L. Mastromatteo, J-F. Muzy, "Mean-field inference of Hawkes point processes", J. Phys. A: Math. Theor. Vol. 49, 174006, 2016.

[2] P. Reynaud-Bouret, S. Schbath, "Adaptive estimation for Hawkes processes; application to genome analysis", Ann. Stat., Vol. 38, No. 5, pp. 2781–2822, 2010.

[3] S. S. Bucak, B. Gunsel, "Incremental subspace learning via non-negative matrix factorization", Patt. Rec., Vol. 42, pp. 788–797, 2009.

[4] B. Cao, D. Shen, J-T. Sun, X. Wang, Q. Yang, Z. Chen, "Detect and track latent factors with online nonnegative matrix factorization", Proc. 20th IJCAI, Hyderabad, pp. 2689–2694, 2007.

[5] J. Chen, "System identification under non-negativity constraints applications in adaptive filtering and hyper spectral image analysis", PhD, Univ. de Technologie de Troyes, 2013.

[6] J. Chen, C. Richard, J. C. M. Bermudez, P. Honeine, "Variants of non-negative least-mean-square algorithm and convergence analysis", IEEE Trans. Sig. Proc., Vol. 62, No. 15, pp. 3990–4005, 2014.

[7] E. S. Chornoboy, L. P. Schramm, A. F. Karr, "Maximum likelihood identification of neural point process systems", Adv. Appl. Probab., Vol. 34, pp. 267–280, 2002.

[8] D. J. Daley, D. Vere-Jones, "An Introduction to the theory of point processes, volume I: Elementary theory and methods", 2nd ed, Springer-Verlag, New York, 2003.

[9] J. Duchi, S. Shalev-Shwartz, Y. Singer, T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions", Proc. ICML, Helsinki, Finland, pp. 272–279, 2008.

[10] U. T. Eden, L. M. Frank, R. Barbieri, V. Solo, E. N. Brown, "Dynamic analysis of neural encoding by point process adaptive filtering", Neural Comput., Vol. 16, No. 5, pp. 971–998, 2004.

[11] M. Eichler, R. Dahlhaus, J. Dueck, "Graphical modeling for multivariate Hawkes processes with nonparametric link functions", Vol. 38, No. 2, pp. 225–242, 2017.

[12] F. Gerhard, R. Haslinger, G. Pipa, "Applying the multivariate time-rescaling theorem to neural population models", Neural Comput., Vol. 23, No. 6, pp. 1452–1483, 2011.

[13] E. C. Hall, R. M. Willett, "Tracking dynamic point processes on networks", IEEE Trans. Inf. Theory., Vol. 62, No. 7, pp. 4327–4346, 2016.

[14] N. R. Hansen, P. Reynaud-Bouret, V. Rivoirard, "Lasso and probabilistic inequalities for multivariate point processes", Bernoulli, Vol. 21, No. 1, pp. 83–143, 2015.

[15] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes", Biometrika, Vol. 58, No. 1, pp. 83–90, 1971.

[16] A. G. Hawkes, "Point spectra of some mutually exciting point processes", J. Roy. Stat. Soc., Series B, Vol. 33, No. 3, pp. 438–443, 1971.

[17] A. G. Hawkes, D. Oakes, "A cluster process representation of a self-exciting process", J. Appl. Prob., Vol. 11, No. 3, pp. 493–503, 1974.

[18] R. A. Hegemann, E. A. Lewis, A. L. Bertozzi, "An Estimate & Score Algorithm for simultaneous parameter estimation and reconstruction of incomplete data on social networks", Secur. Inform. Vol. 2, No. 1, 2013.

[19] A. F. Karr, *Point processes and their statistical inference*, 2nd Ed., Marcel Dekker, New York, 1991.

[20] M. Kirchner, "An estimation procedure for the Hawkes process", Quant. Fin., Vol. 17, No. 4, pp. 571–595, 2017.

[21] K. Lange, *MM Optimization Algorithms*, SIAM, 2016.

[22] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization", NIPS, pp. 556–562, 2000.

[23] S. W. Linderman, R. P. Adams, "Discovering latent network structure in point process data", Proc. ICML., Beijing, China, pp. II-1413–1421, 2014.

[24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, "Online learning for matrix factorization and sparse coding", J. Mach. Learn. Res., Vol. 11, pp. 19–60, 2010.

[25] V. H. Nascimento, Y. V. Zakharov, "RLS adaptive filter with inequality constraints", IEEE Sig. Proc. Lett., Vol. 23, No. 5, pp. 752–756, 2016.

[26] Y. Ogata, "On Lewis' Simulation Method for Point Processes", IEEE Trans. Inf. Theory, Vol. IT–27, No. 1, pp. 23–31, 1981.

[27] T. Ozaki, "Maximum likelihood estimation of Hawkes' self-exciting point process", Ann. Inst. Stat. Math., Vol. 31, No. 1, pp. 145–155, 1979.

[28] S. A. Pasha, V. Solo, "Hawkes-Laguerre reduced rank model for point processes,", IEEE ICASSP, Vancouver, BC, pp. 6098–6102, 2013.

[29] S. A. Pasha, V. Solo, "Topology identification of dynamic point process networks", Proc. IEEE ICASSP, Florence, pp. 375–378, 2014.

[30] S. A. Pasha, V. Solo, "Topology identification for sparse dynamic point process networks", Proc. IEEE CDC, Los Angeles, pp. 1786–1791, 2014.

[31] J. G. Rasmussen, "Bayesian inference for Hawkes processes", Methodol. Comput. Appl. Probab., Vol. 15, pp. 623–642, 2013.

[32] A. Simma, M. Jordan, "Modeling events in time using cascades of Poisson processes", Proc. 26th Conf. UAI., Catalina Island, pp. 546–555, 2010.

[33] D. L. Snyder, M. I. Miller, *Random point processes in time and space*, 2nd Ed., Springer-Verlag, New York, 1991.

[34] V. Solo, "Adaptive estimation of positive parameters", Proc. 32nd IEEE Conf. Decis. Control, Vol. 3, pp. 2613–2614, 1993.

[35] V. Solo, "Averaging analysis of a point process adaptive algorithm", J. Appl. Prob., Vol. 41, pp. 361–372, 2004.

[36] V. Solo, A. Pasha, "A test for independence between a point process and an analogue signal", J. Time Series Anal., Vol. 33, No. 5, pp. 824–840, 2012.

[37] A. Veen, F. P. Schoenberg, "Estimation of space-time branching process models in seismology using an EM-type algorithm", J. Amer. Stat. Assoc., Vol. 103, No. 482, pp. 614–624, 2008.

[38] Y. Wang, A. R. Paiva, J. C. Principe, J. C. Sanchez, "Sequential Monte Carlo point-process estimation of kinematics from neural spiking activity for brain-machine interfaces", Neural Comput., Vol. 21, No. 10, pp. 2894–2930, 2009.

[39] F. Wang, C. Tan, A. C. König, P. Li, "Efficient document clustering via online nonnegative matrix factorizations", Proc. 11th SIAM Intl. Conf. Data Mining, 2011.

[40] H. Xu, M. Farajtabar, H. Zha, "Learning granger causality for Hawkes processes", Proc. ICML, New York, Vol. 48, pp. 1660–1669, 2016.

[41] R. Zhao, V. Y. F. Tan, "Online nonnegative matrix factorization with outliers", Proc. IEEE. ICASSP, pp. 2662–2666, 2016.

[42] K. Zhou, H. Zha, L. Song, "Learning triggering kernels for multi-dimensional Hawkes processes", Proc. ICML, Atlanta, GA., Vol. 28, pp. III-1301–III-1309, 2013.

[43] K. Zhou, H. Zha, L. Song, "Learning social infectivity in sparse low-rank networks using multidimensional Hawkes processes", AISTATS, 2013.