A Multivariate Hawkes Process with Gaps in Observations

Triet M. Le*

August 9, 2016

Abstract

Given a collection of entities (or nodes) in a network and our intermittent observations of activities from each entity, an important problem is to learn the hidden edges depicting directional relationships among these entities. Here, we study causal relationships (excitations) that are realized by a multivariate Hawkes process. The multivariate Hawkes process (MHP) and its variations (spatial-temporal point processes) have been used to study contagion in earthquakes, crimes, neural spiking activities, the stock and foreign exchange markets, etc. In this paper, we consider the case with intermittent observations (and hence gaps.) We propose a variational problem for detecting sparsely hidden relationships with a multivariate Hawkes process that takes into account the gaps from each entity (MHPG). We bypass the problem of dealing with a *large* amount of missing events by introducing a *small* number of unknown boundary conditions. In the case where our observations are sparse (e.g. from 10% to 30%), we show through numerical simulations that robust recovery with MHPG is still possible even if the lengths of the observed intervals are small but they are chosen accordingly. In these cases, the proposed MHPG outperforms the classical MHP in parameter estimations. *The numerical results also show that the knowledge of gaps is very crucial in discovering the underlying patterns and hidden relationships.*

1 Introduction

Point processes have demonstrated to be promising tools for extracting dynamic patterns and discovering hidden relationships in event data. Variations of (spatial-temporal, univariate, multivariate) point processes have been applied to event data from many different fields in science. For instance, self-exciting point processes have been used in seismology to model contagion of earthquakes [17, 18, 19, 26] , and in anthropology to study the spread of crimes and violence acts [16, 14, 22, 21]. The multivariate Hawkes process (a parametric version of the self-exciting point process) has been applied to financial data to study contagion and influential entities in financial networks [1, 2, 4, 9, 8] , and also to social media networks [23, 27, 11, 15, 10]. The multivariate

*NGA Research, National Geospatial-Intelligence Agency, 7500 GEOINT Dr., Springfield, VA 22150, Email: Triet.M.Le@nga.mil

Hawkes process with inhibition has also been used in neuroscience to make inference of functional connectivity from neural spiking activities [13, 20], among others. The common assumption in these work is that all events over a long-enough time interval of interest are observed. However, for reasons associated to the weather or the environment, etc., events are only observed intermittently. Thus the challenges are: 1) how to recover robustly the underlying parameters in the presence of gaps; 2) how to distribute the gaps for optimal recovery given the available resources.

In technical terms, let N be the number of entities (or nodes) within a network. For each entity m ranging from 1 to N, let $E_m = \{t_{m,i}\}$ be the set of observed events that are contained in some interval of interest, say $(0, T]$. Here each $t_{m,i}$ in $(0, T]$ (assuming $t_{m,i-1} < t_{m,i}$) represents a time-stamp when an event from entity m occurs. Consider first the case where all events in $(0, T]$ from all entities are observed. We recall the following definitions of point processes.

Definition 1 (Poisson Process [5])**.** For each entity m, denote the finite collection of disjoint intervals in $(0, T]$ by $\{(c_{m,k}, d_{m,k}]\}_{k=1}^{K_m}$. Let $N(c_{m,k}, d_{m,k}]$ be the number of observed events from entity m that are contained in $(c_{m,k}, d_{m,k}]$. We say the collection of observed events in E_m follows a **homogeneous** Poisson process with some constant intensity $\lambda_m \geq 0$ if the following probability holds

$$P(N(c_{m,k}, d_{m,k}] = n) = \frac{[\lambda_m(d_{m,k} - c_{m,k})]^n e^{-\lambda_m(d_{m,k} - c_{m,k})}}{n!}. \tag{1}$$

Suppose now instead of a constant λ_m, we have a positive (integrable) function $\lambda_m : (0, T] \to \mathbb{R}^+$. Then we say the set of events E_m follows an **inhomogeneous** Poisson process with intensity function $\lambda_m(t)$ if

$$P(N(c_{m,k}, d_{m,k}] = n) = \frac{[\Lambda_{m,k}]^n e^{-\Lambda_{m,k}}}{n!}, \tag{2}$$

where

$$\Lambda_{m,k} = \int_{c_{m,k}}^{d_{m,k}} \lambda_m(t)\ dt.$$

The interpretations of (1) (or (2)) are as follows:

1. The number of events in $(c_{m,k}, d_{m,k}]$ follows a Poisson distribution with mean and variance $\lambda_m(d_{m,k} - c_{m,k})$ (or $\int_{c_{m,k}}^{d_{m,k}} \lambda_m(t)\ dt$).
2. The number of events in disjoint intervals or from different entities are independent random variables. In other words, an occurrence of an event from entity n has no influence on future events from entity m or on entity n itself.

The multivariate Hawkes process introduces *directional dependencies among entities and events* into the definition of the (conditional) intensity function $\lambda_m(t)$. The word *'conditional'* is used because $\lambda_m(t)$ is conditioned on prior events.

Definition 2 (Multivariate Hawkes Process [12])**.** We say the collection of observed events in $E_m, m = 1, \cdots, N$, follows a multivariate Hawkes process if for all t in $(0, T]$, the conditional intensity function (CIF) $\lambda_m(t)$ for entity m is given by

$$\lambda_m(t) = u_m + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j} \in E_n, t_{n,j} < t} b_m e^{-b_m(t - t_{n,j})}. \tag{3}$$

Here, the background $u_m \geq 0$ is a homogeneous Poisson process, and it is included here to promote independent random events. Since $\int_0^\infty b_m e^{-b_m t}\, dt = 1$ for $b_m > 0$, the matrix $a = (a_{m,n})_{N \times N}$ with the entry $a_{m,n} \geq 0$ depicts how an event $t_{n,j}$ from entity n will trigger or excite future events from entity m. A multivariate Hawkes process is stationary if and only if the largest eigenvalue of the matrix a in absolute value is strictly bounded above by 1 [5]. $1/b_m$ (the width of the exponential function) is the timescale providing the likelihood when the next event from entity m occurs.

To incorporate inhibition into the multivariate Hawkes process, one allows $a_{m,n}$ to be negative [20]. However, in this paper we focus on the case where $a_{m,n}$ is non-negative.

It is possible to consider an inhomogeneous Poisson process for the background u, and to have a different time scale or mode of excitation for each pair of entities. For simplicity, we focus on the single mode of excitation case given by (3). Also, each event $t_{m,i}$ may have a different mark or jump size $M_{m,i} \geq 0$. Here we consider all events to be the same, namely $M_{m,i} = 1$.

Remark 1. *$\lambda_m(t)$ in (3) satisfies the following mean-reverting dynamics*

$$d\lambda_m(t) = b_m(u_m - \lambda_m(t))dt + \sum_{n=1}^{N} a_{m,n} b_m dN_n(t), \text{ for } t \in (0, T], \tag{4}$$

with the boundary condition $\lambda_m(0) = u_m$. In general, the solution to (4) has the form

$$\lambda_m(t) = u_m + (\lambda_m(0) - u_m)e^{-b_m t} + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j} \in E_n, t_{n,j} < t} b_m e^{-b_m(t - t_{n,j})}.$$

Figure 1 shows simulations of two univariate point processes ($N = 1$) in the interval $(0, T]$, with $T = 10$. Figure 1-(a) shows a homogeneous Poisson process with constant $\lambda = 1$, and figure 1-(b) shows a (univariate) Hawkes process with $u = 1, a = 0.5$ and $b = 2$. Recall the CIF of a univariate Hawkes process (equation (3) with $N = 1$) is defined as

$$\lambda(t) = u + a \sum_{0 < t_i < t} b e^{-b(t - t_i)}. \tag{5}$$

Note the dynamics of $\lambda(t)$ in figure 1-(b) as events (in blue spikes) evolve. Based on the definition of a homogeneous Poisson process, we expect that events are uniformly distributed (as evident in figure 1-(a).) The same phenomenon doesn't hold for a Hawkes process whenever $ab > 0$. This is evident in figure 1-(b) as events are more clustered as a result of self-excitation and hence there are more bustiness in the intensity function.

Figure 2 shows a simulation of a multivariate Hawkes process ($N = 2$) with $u = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}, a = \begin{pmatrix} 0.25 & 0.75 \\ 0 & 0.25 \end{pmatrix}, b = \begin{pmatrix} 10 \\ 1 \end{pmatrix}$. $a_{1,2} = 0.75$ implies that events from entity 2 is very contagious toward entity 1. On the other hand, events from entity 1 has no influence on entity 2 since $a_{2,1} = 0$. These effects can be seen in the evolution of the CIFs (in blue). An event from entity 2 creates a jump in the CIF of entity 1 which causes a series of events to follow, but not vice versa.

MHP: Let N be the number of entities. In a complete observation setting, for each entity m we are given the set of observed events $E_m = \{t_{m,i}\}$ and the task is to

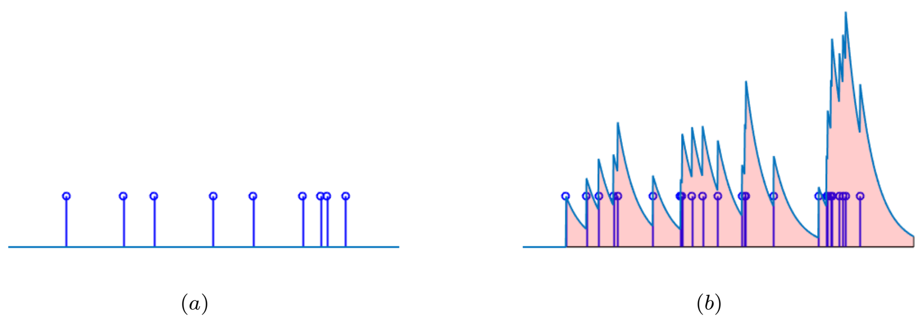


Figure 1: Simulations of two univariate point processes in the interval $(0, T]$, $T = 10$: (a): Homogeneous Poisson process with constant intensity $\lambda = 1$, (b): A univariate Hawkes process depicting the dynamic of $\lambda(t)$ given in (5) as events (blue spikes) evolve. Here $u = 1, a = 0.5$ and $b = 2$.

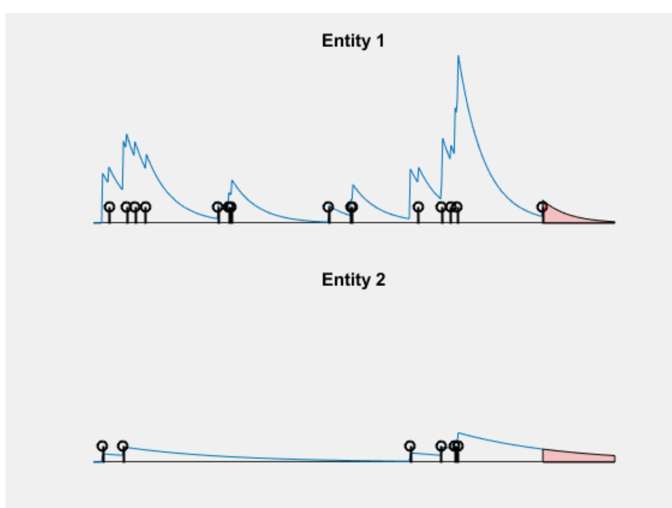


Figure 2: A simulation of a multivariate Hawkes process ($N = 2$): In this figure, entity 2 is very influential toward entity 1 (since $a_{1,2} = 0.75$) and entity 1 has no influence on entity 2 (since $a_{2,1} = 0$.) Both entities have the same amount of self-excitation ($a_{1,1} = a_{2,2} = 0.25$.)

learn the parameters $u = (u_m)_{N\times 1}, a = (a_{m,n})_{N\times N}$ and $b = (b_m)_{N\times 1}$ that completely describes the conditional intensity function $\lambda_m(t)$ in (3). The common approach is to minimize the $(-)$log-likelihood function [5]:

$$\min_{u,a,b}\left\{L(u,a,b) = \sum_{m=1}^{N}\left[\int_0^T \lambda_m(t)\ dt - \sum_{t_{m,i}\in E_m}\log(\lambda_m(t_{m,i}))\right] + G(a)\right\}, \tag{6}$$

with the constraint $u_m \geq 0$ and $b_m \geq 0$. The second term $G(a)$ is the prior or regularization on the matrix a. For instance, to impose sparsity on interactions, one can use the Lasso constraint [24] $G(a) = \mu\sum_{m,n}|a_{m,n}|$ for some $\mu > 0$.

Remark 2 (Spatial-Temporal Interpretation)**.** *Suppose each entity is a grid cell* (m,n) *on a two dimensional rectangle, and the directional interaction that cell* (m',n') *has toward* (m,n) *is determined by* $a_{m-m',n-n'}$*. Then we can rewrite the CIF for the Hawkes process in* (3) *as follows*

$$\lambda_{m,n}(t) = u_{m,n} + \sum_{m',n'} a_{m-m',n-n'} \sum_{t_{m',n',j}<t} b_{m,n}e^{-b_{m,n}(t-t_{m,n,j})}, \tag{7}$$

where $t_{m,n,i}$ *is the event that falls inside grid* (m,n)*. Let* $p = m-m'$ *and* $q = n-n'$*. Here, the stability constraint becomes* $0 \leq \sum_{p,q} a_{p,q} < 1$*. Having* $a_{p,q} = 0$ *whenever* $d(p,q) := \sqrt{p^2+q^2} > R$ *for some* $R > 0$ *enforces each grid cell to only interact with other grid cells that are at most a distance* R *from itself.*

For reasons associated to the weather or the environment, one has intermittent observations (and hence gaps in observations.) Figure 3 shows an example of intermittent observations for a network of two entities. The shaded intervals represent the observational gaps. We observe events in blue and do not observe events in red from each entity. For this simulation, we use the same parameters as in figure 2.

For each entity m, let $\{(c_{m,k}, d_{m,k}]\}_{k=1}^{K_m}$ be the collection of disjoint observed intervals (e.g. unshaded intervals in figure 3) that are contained in $(0,T]$, and let $O_m = \{t_{m,i}\}$ be the corresponding partially observed events (e.g. events in blue from figure 3). Motivated by remark 1, for each t that belongs to one of the observed intervals $(c_{m,k}, d_{m,k}]$ we approximate the conditional intensity function for entity m as

$$\begin{aligned}\bar{\lambda}_m(t) &= u_m + (\bar{\lambda}_m(c_{m,k}) - u_m)e^{-b_m(t-c_{m,k})}\\ &\quad + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in O_n, c_{m,k}\leq t_{n,j}<t} b_m e^{-b_m(t-t_{n,j})},\end{aligned}$$

where $\bar{\lambda}_{m,k} := \bar{\lambda}_m(c_{m,k})$ are extra unknowns. By an abuse of notation, denote the vector $\bar{\lambda} := \{\lambda_{m,k}\}$. To learn the underlying parameters u, a and b, we propose to solve the following minimization problem (MHPG)

$$\begin{aligned}J(u,a,b,\bar{\lambda}) = \sum_{m=1}^{N}\sum_{k=1}^{K_m}&\left[\int_{c_{m,k}}^{d_{m,k}} \bar{\lambda}_m(t)\ dt - \sum_{t_{m,i}\in(c_{m,k},d_{m,k}]} \log\left(\bar{\lambda}_m(t_{m,i})\right)\right]\\ &+ G(a) + P(\bar{\lambda}),\end{aligned}$$

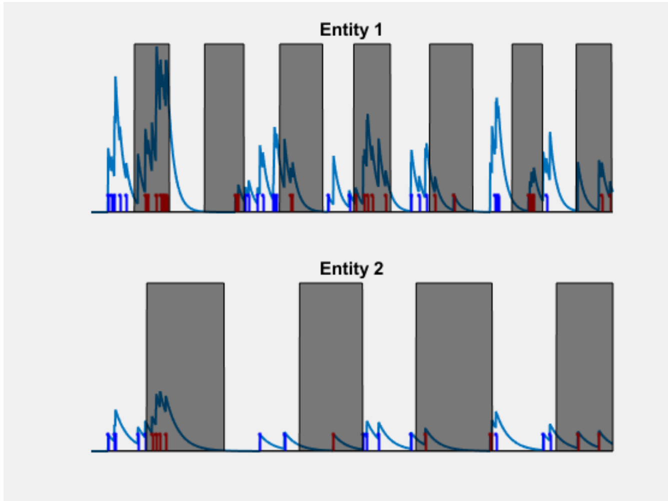


Figure 3: A simulation (using the same parameters as in figure 2) showing the intermittent observations for each entity. Blue spikes are observed events and red spikes are unobserved events.

where $G(a)$ and $P(\bar{\lambda})$ are appropriate constraints or regularizations on the matrix a and the vector $\bar{\lambda}$.

Using simulated data for a network of two entities with the fraction of observation $p = 0.3$, figure 4 shows that the proposed method MHPG is able to capture the underlying statistics better than MHP. See section 3 for further details and comparisons.

The paper is organized as follows. In section 2 we go over a formulation to incorporate the gaps into the modeling of the conditional intensity function $\lambda_m(t)$ and propose a variational model with appropriate regularizations and constraints to learn the underlying parameters u, a and b. In section 3, we provide a numerical study using simulated data to show that the proposed method robustly recovers the underlying parameters in the presence of large amount of missing events ($\geq 70\%$), and outperforms the classical MHP. Appendix A gives a detailed description of the numerical implementation of the proposed model.

2 Mutivariate Hawkes Process with Gaps (MHPG)

Let E_m be the complete set of events (including observed and unobserved) in $(0, T]$ generated by entity m. Suppose now that we do not have complete observations, that is let $\{(c_{m,k}, d_{m,k}]\}_{k=1}^{K_m}$ be the collection of disjoint observed intervals for entity m that are contained in $(0, T]$, and let O_m be the set of the corresponding partially observed events.

Let $t \in (c_{m,k}, d_{m,k}]$ and recall from (3), we have (assuming $\lambda_m(0) = u_m$)

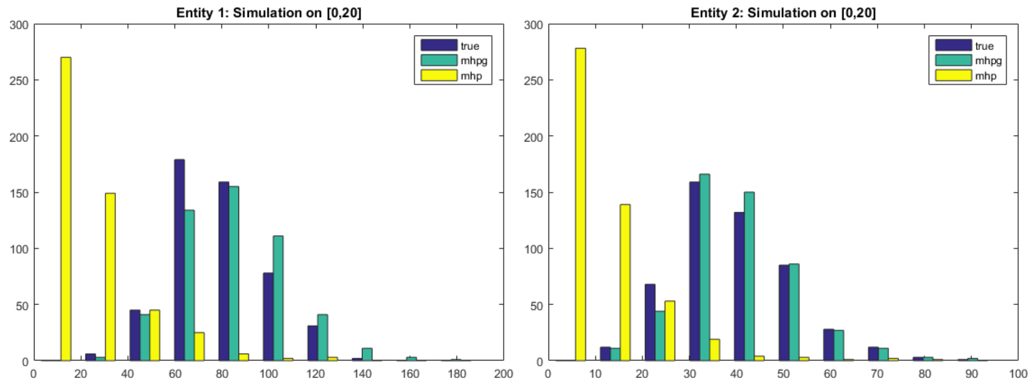


Figure 4: Comparisons of the event count histograms of MHP and MHPG with the ground truth for an interval of length 20 times the timescale of entity 1. See example 1 for further details.

$$\lambda_m(t) = u_m + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in E_n, t_{n,j}<t\le d_{m,k}} b_m e^{-b_m(t-t_{n,j})} \tag{8}$$

$$= u_m + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in E_n, t_{n,j}<c_{m,k}} b_m e^{-b_m(t-t_{n,j})} \tag{9}$$

$$+ \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in E_n, c_{m,k}\le t_{n,j}<t\le d_{m,k}} b_m e^{-b_m(t-t_{n,j})}. \tag{10}$$

Since by (3),

$$\lambda_m(c_{m,k}) = u_m + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in E_n, t_{n,j}<c_{m,k}} b_m e^{-b_m(c_{m,k}-t_{n,j})}.$$

Substituting $\lambda_m(c_{m,k})$ into (10), we get

$$\begin{aligned}\lambda_m(t) &= u_m + (\lambda_m(c_{m,k}) - u_m)e^{-b_m(t-c_{m,k})} \\ &+ \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in E_n, c_{m,k}\le t_{n,j}<t\le d_{m,k}} b_m e^{-b_m(t-t_{n,j})}.\end{aligned} \tag{11}$$

The boundary value $\lambda_{m,k} := \lambda_m(c_{m,k})$ is an extra unknown since it may depend on the unobserved events that are contained in the gap $(d_{m,k-1}, c_{m,k}]$. The third term on the right-hand side of the last equation is summing over (observed and unobserved) events from entity n that are contained in $(c_{m,k}, d_{m,k}]$. Clearly, if m and n have the same observed intervals, then we have

$$\{t_{n,j} \in E_n : c_{m,k} \le t_{n,j} \le d_{m,k}\} = \{t_{n,j} \in O_n : c_{m,k} \le t_{n,j} \le d_{m,k}\}. \tag{12}$$

In general, we consider the following approximation of the CIF for entity m:

$$\bar{\lambda}_m(t) = u_m + (\lambda_m(c_{m,k}) - u_m)e^{-b_m(t-c_{m,k})} + \sum_{n=1}^{N} a_{m,n} \sum_{t_{n,j}\in O_n, c_{m,k}\le t_{n,j}<t\le d_{m,k}} b_m e^{-b_m(t-t_{n,j})}. \tag{13}$$

Note that the representation of $\bar{\lambda}_m(t)$ in (13) is exactly equal to $\lambda_m(t)$ in (11) whenever (12) holds. For the general case where the observed intervals for m and n have some overlapping, then (13) provides an approximation to $\lambda_m(t)$. In section 3, we show that by taking the intersection of the observed intervals and remove events that do not belong to this intersection, one can achieve better reconstruction.

MHPG: Denote $\bar{\lambda}_{m,k} := \lambda_m(c_{m,k})$ and by an abuse of notation let $\bar{\lambda} = \{\bar{\lambda}_{m,k}\}$. We propose to learn the parameters u, a, b and $\bar{\lambda}$ by minimizing the following functional:

$$J(u,a,b,\bar{\lambda}) = \sum_{m=1}^{N}\sum_{k=1}^{K_m}\left[\int_{c_{m,k}}^{d_{m,k}} \bar{\lambda}_m(t)\ dt - \sum_{t_{m,i}\in(c_{m,k},d_{m,k}]} \log\left(\bar{\lambda}_m(t_{m,i})\right)\right] + G(a) + P(\bar{\lambda}), \tag{14}$$

where $G(a)$ and $P(\bar{\lambda})$ are priors (or regularizations) on a and $\bar{\lambda}$ respectively. If all entities have the same observed intervals, then using techniques from [5], one can show that J from (14) is the $(-)$log-likelihood function. From the graph/network point of view, the LASSO constraint [24] on the matrix a,

$$G(a) = \mu \sum_{m,n=1}^{N} |a_{m,n}|, \tag{15}$$

enforces sparsity on a. In other words, each entity only interacts with a few other entities within the network. Theorem 4.1 from [6] provides theoretical results for the marginal distributions of $\{\lambda_m(t)\}_{t>0}$. The (-)log of these distributions can be used to define $P(\bar{\lambda})$. Given the complicated forms of these distributions, we consider instead the following constraint on $\bar{\lambda}_{m,k}$:

$$u_m \le \bar{\lambda}_{m,k} \le Cu_m, \tag{16}$$

for some $C \ge 1$. This can be viewed as having $\bar{\lambda}_{m,k}$ following a uniform distribution on $[u_m, Cu_m]$.

Remark 3. *Theorem 4.1 from [6] provides theoretical results for the marginal distributions of $\{\lambda_m(t)\}_{t>0}$. The (-)log of these distributions can be used to define $P(\bar{\lambda})$. Given the complicated forms of these distributions, we consider instead the uniform distribution and hence equation* (16)*.*

Although equation (11) can be derived directly from equation (4) for $t \in (c_{m,k}, d_{m,k}]$, the technique described in equation (8)-(10) can be applied to a much more general case. Denote $g(t) = be^{-bt}$ for $b, t > 0$. Transforming equation (8) to equation (13) is possible because of the fact that e^{-bt} satisfies the semi-group property. The same technique can also be used for $g(t) = Q(t)s(t)$, where $Q(t)$ is a polynomial and $s(t)$ is any function satisfying the semi-group property (namely $s(t_1 + t_2) = s(t_1)s(t_2)$.)

By approximating a power-law function with a sum of exponentials, this technique can also be applied there. Indeed, let $g(t) = c_b t e^{-bt}$, where c_b is chosen such that $\int_0^\infty g(t)\ dt = 1$. For simplicity consider the univariate self-exciting point process with $\lambda(t), t \in (0, T]$, given by

$$\lambda(t) = u + a \sum_{0<t_i<t} g(t - t_i) = u + a \sum_{0<t_i<t} c_b (t - t_i) e^{-b(t-t_i)}. \tag{17}$$

Take $t \in (c_k, d_k]$ and assuming $t_i \neq c_k$, (17) becomes

$$\lambda(t) = u + a \sum_{0<t_i<c_k} c_b (t - t_i) e^{-b(t-t_i)} + a \sum_{c_k<t_i<t} c_b c_b (t - t_i) e^{-b(t-t_i)}. \tag{18}$$

Let $A = a \sum_{0<t_i<c_k} c_b (t - t_i) e^{-b(t-t_i)}$, then

$$\begin{aligned} A &= a \sum_{0<t_i<c_k} c_b (t - c_k + c_k - t_i) e^{-b(t-c_k+c_k-t_i)} \\ &= a \sum_{0<t_i<c_k} c_b (c_k - t_i) e^{-b(c_k-t_i)} \left[e^{-b(t-c_k)} \right] \\ &+ a \sum_{0<t_i<c_k} c_b e^{-b(c_k-t_i)} \left[(t - c_k) e^{-b(t-c_k)} \right]. \end{aligned}$$

Apply the last equation to (18), we get

$$\begin{aligned} \lambda(t) = u &+ (\lambda(c_k) - u) e^{-b(t-c_k)} + (\widetilde{\lambda}(c_k) - u) \left[(t - c_k) e^{-b(t-c_k)} \right] \\ &+ a \sum_{c_k<t_i<t} c_b c_b (t - t_i) e^{-b(t-t_i)}, \text{ for } t \in (c_k, d_k], \end{aligned} \tag{19}$$

where $\widetilde{\lambda}(c_k) = u + a \sum_{0<t_i<c_k} c_b e^{-b(c_k-t_i)}$ has the form of a univariate Hawkes process. In this case the extra unknowns are $\{\lambda(c_k)\}$ and $\{\widetilde{\lambda}(c_k)\}$. In general, for $g(t) = Q(t)e^{-bt}$ with $Q(t)$ being a polynomial of degree M, there will be $M+1$ extra unknown boundary values to solve.

3 Numerical Results

In this section we compare the performance of MHP and MHPG given in (6) and (14) respectively using (15) for the regularization on the matrix a, and (16) for the constraint of $\bar{\lambda}_{m,k}$. We use the following algorithm to generate observed intervals $(c_k, d_k]$ for each entity.

Algorithm 1 (Algorithm for generating observed intervals)**.** Given a fraction of observation $0 < p < 1$, and $0 < \tau_1 < \tau_2$ representing the lower and upper bound for the lengths of the observed intervals.

1. Set $c_1 = 0$ and $d_1 = e_1$ where e_1 is a uniform random number in (τ_1, τ_2).
2. Suppose c_{k-1} and d_{k-1} are computed.
3. Set $c_k = d_{k-1} + n_k$, where n_k is a uniform random number in $(\frac{\tau_1}{2p}, \frac{\tau_2}{2p})$.

4. Set $d_k = c_k + e_k$, where e_k is a uniform random number in (τ_1, τ_2).
5. Proceed until either $c_k \geq T$ or $d_k \geq T$.

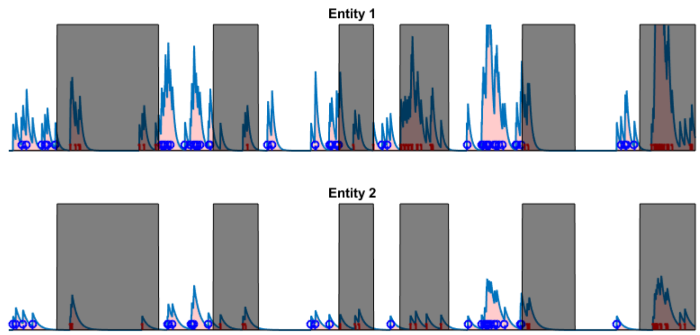
In all of the following examples we consider a simple network of two entities with the true parameters

$$u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ a = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{pmatrix}, \ b = \begin{pmatrix} 10 \\ 5 \end{pmatrix}. \tag{20}$$

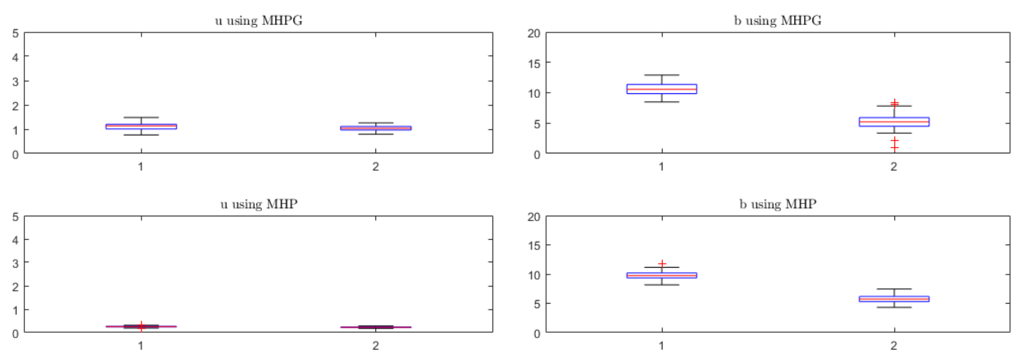
In this case, the eigenvalues of a are its diagonal entries. To obtain the set of observed events for each entity, we generate the events using the algorithm from [7]. For each entity m we then select only the events that are contained in the observed intervals $(c_{m,k}, d_{m,k}]$ for entity m. The timescales for entities 1 and 2 are $\frac{1}{10}$ and $\frac{1}{5}$ respectively. Here, we pick $T = 500$, which means that we are picking a period of 5000 times the timescale of entity 1. If we view $1/b_1$ as one minute, then $T = 500$ is approximately half a week.

In the numerical results below, examples 1-2 show the importance of knowing precisely the observed intervals for each entity. Example 2 shows that the uncertainty in parameter estimation increases as we observe less overall (e.g the fraction of observation is $p = 0.1$.) The best scenario for MHPG in (14) is when all the entities in consideration have the same observed intervals. Example 3 shows that effect when this assumption is violated. A solution to deal with this violation is taking the intersection of the observed intervals and remove events that do not belong to this intersection (which we call biased events). Example 4 shows that by removing biased events, one obtains a slightly better reconstruction of the parameters. In example 5, we use the same fraction of observations as in example 1, but the lengths of the observed intervals are decreased. As a result, we see a slight increase in the uncertainty of the reconstructed parameters.

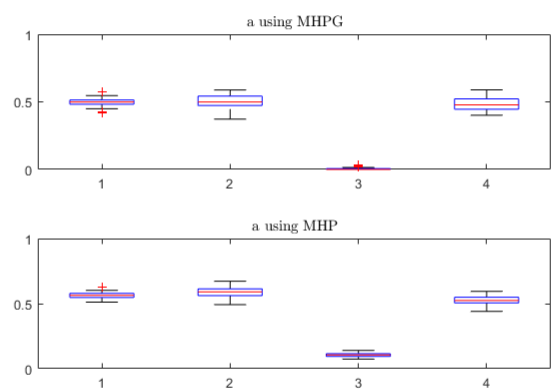
Example 1. In this example, we consider the case where all entities have the same collection of observed intervals $\{(c_k, d_k]\}$ generated using algorithm 1 with $p = 0.3$, $\tau_1 = 1/b_1$ and $\tau_2 = 30/b_1$. For instance, if $1/b_1$ represents one minute, then the lengths of the observed intervals would range from 1 to 30 minutes. For each simulation, there are 1100 total events on average. Figure 5-(i) shows a simulation of a Hawkes process using the parameters in (20). Here observed events are in blue and red events in shaded intervals are unobserved. Both entities have the same distribution of gaps. Boxplots from figure 5-(ii)-(iv) show the reconstruction performance of MHP in (6) and MHPG in (14) using algorithm 2 for 100 simulations. We observe that both MHP and MHPG can recover the timescale variable b fairly well. However, MHPG provides beter reconstruction for the background u and the relationship matrix a. The reconstructed parameters can then be used to forecast the number of events that each entity may anticipate within a certain time window. Using the *median values* of the 100 reconstructions for u, a and b for each model, figures 6-(i)-(iv) compare the histograms of event counts for two intervals of lengths 5 and 20. In both cases, the histograms of event counts using the reconstructed parameters from MHPG closely match with the histograms of event counts using the true parameters. Since the estimations of the background parameter u from MHP are significantly smaller than the true background, the forecasts produce less events.



(i)



(ii) (iii)



(iv)

Figure 5: (i): A simulation of a Hawkes process using the true parameters in (20); shaded intervals represent gaps in observations (here $p = 0.3$). Boxplots in (ii),(iii) and (iv) show estimation of u, b and a respectively using MHP in (6) and MHPG in (14). See (20) for the ground truths.

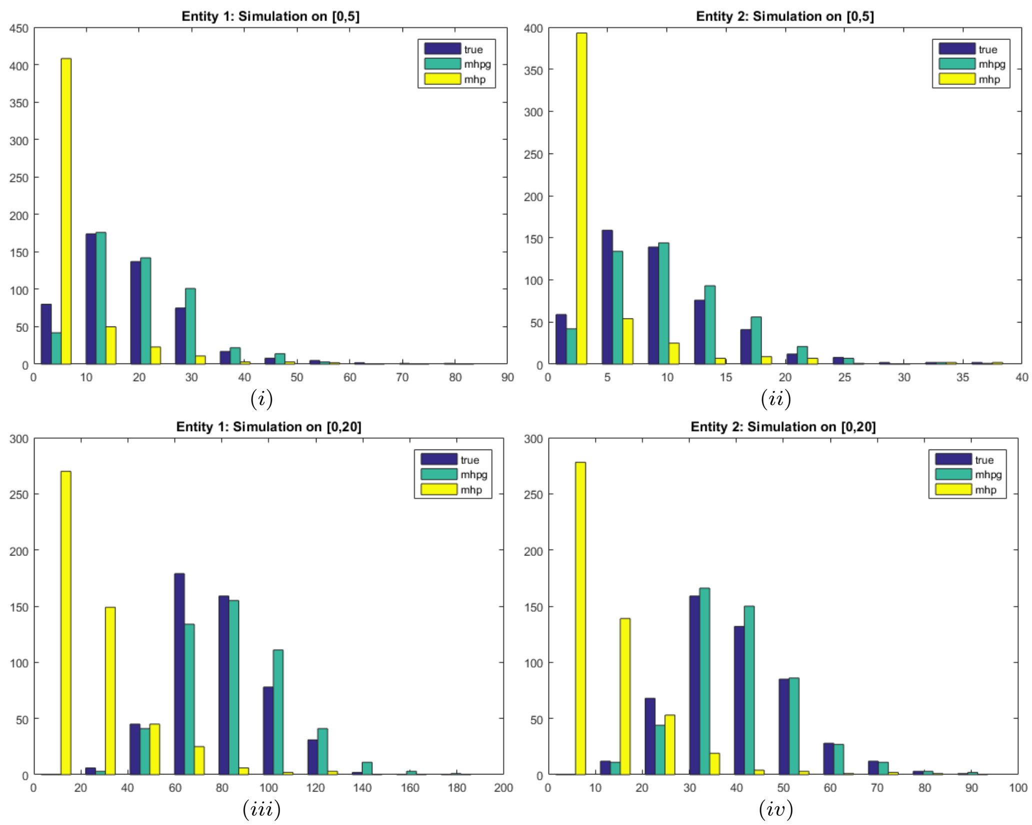


Figure 6: Comparisons of event-count histograms of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG (in green) fits the ground truth (in blue) much better then MHP (in yellow).

Example 2. Clearly, when we observe less our uncertainty of parameter estimation increases. In this example, the setup is the same as in example 1, but now $p = 0.1$. As a result, there are 300 total events on average for each simulation. The boxplots in figures 7-(ii)-(iv) show a greater variation in the estimated parameters. However, the median values of the 100 reconstructions are still close to the ground truths. This is evident from the results in figure 8.

Example 3. The set up is the same as in example 1 but now for each entity m we generate a separate collection of observed intervals $\{(c_{m,k}, d_{m,k}]\}$ using algorithm 1 with $p = 0.3$, $\tau_{m,1} = 1/b_m$ and $\tau_{m,2} = 30/b_m$. For each simulation, there are 1100 total events on average. 9-(i) shows a simulation of a Hawkes process using the parameters in (20). Here observed events are in blue and red events in shaded intervals are unobserved. Both entities have different distribution of gaps. The boxplots in figure 9-(iv) shows that in this case MHP increases the self-excitation components $a_{m,m}$ and decreases the mutual-excitation components $a_{m,n}, m \neq n$. A similar phenomenon happens for MHPG but with less deviations form the ground truths. Note that since entity 2 is not influenced by entity 1, the gaps from entity 1 do not effect the computation of $a_{2,2}$. As seen in figures 9-(ii)-(iv), the reconstructions of u_2, b_2 and $a_{2,2}$ using MHPG via algorithm 2 are as accurate as in example 1. However, the reconstructions for $u_1, b_1, a_{1,1}$ and $a_{2,1}$ are less accurate. These inaccuracies are shown in figures 10-(i) and (iii), where the histograms of event counts for entity 1 using MHPG deviates slightly (but still less than MHP) from the histograms of event counts using the ground truth parameters. Again, here we use the medians of the 100 reconstructed parameters to simulate the events.

Example 4. In this example, we first generate the observed intervals as in example 3 with $p = 0.3$. We then take the intersection of the observed intervals of the two entities, and consider only events that belong to this intersection. After taking the intersection, the fraction of observation becomes $p = 0.15$ on average which is larger than the fraction of observation in example 2. In figure 11: (i) shows the histogram (very non-uniform) of the resulting lengths of the observed intervals, and (ii) shows a snapshot of the resulting intervals and the observed events (in blue). Boxplots in (iii)-(v) show the reconstruction of u, a and b using MHP and MHPG via algorithm 2. The performance for MHPG is better than results from example 3. Figures 12 compares the histograms of event counts of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG provides a better histogram fit than MHP. However, the performance of MHPG is not as good as in example 2 even though the fraction of observation is larger. The presence of too many small observed intervals doesn't improve the parameter estimation.

Example 5. In this example, we reduce the lengths of the observed intervals so that $\tau_{m,1} = 1/b_1$ and $\tau_{m,2} = 20/b_1$ with $p = 0.3$. As seen in figures 13 and 14, MHPG still provides a good estimate of the truth parameters. In comparison with example 1, here we get a slightly larger variation (uncertainty) in the estimated parameters as depicted by the boxplots in figure 13. Using the medians of the 100 reconstructed parameters, figure 14 shows that MHPG still provides a better histogram fit than MHP.

In conclusion, we present here a simple technique for modeling the CIF of a multivariate Hawkes process that incorporates observational gaps in (13). The proposed

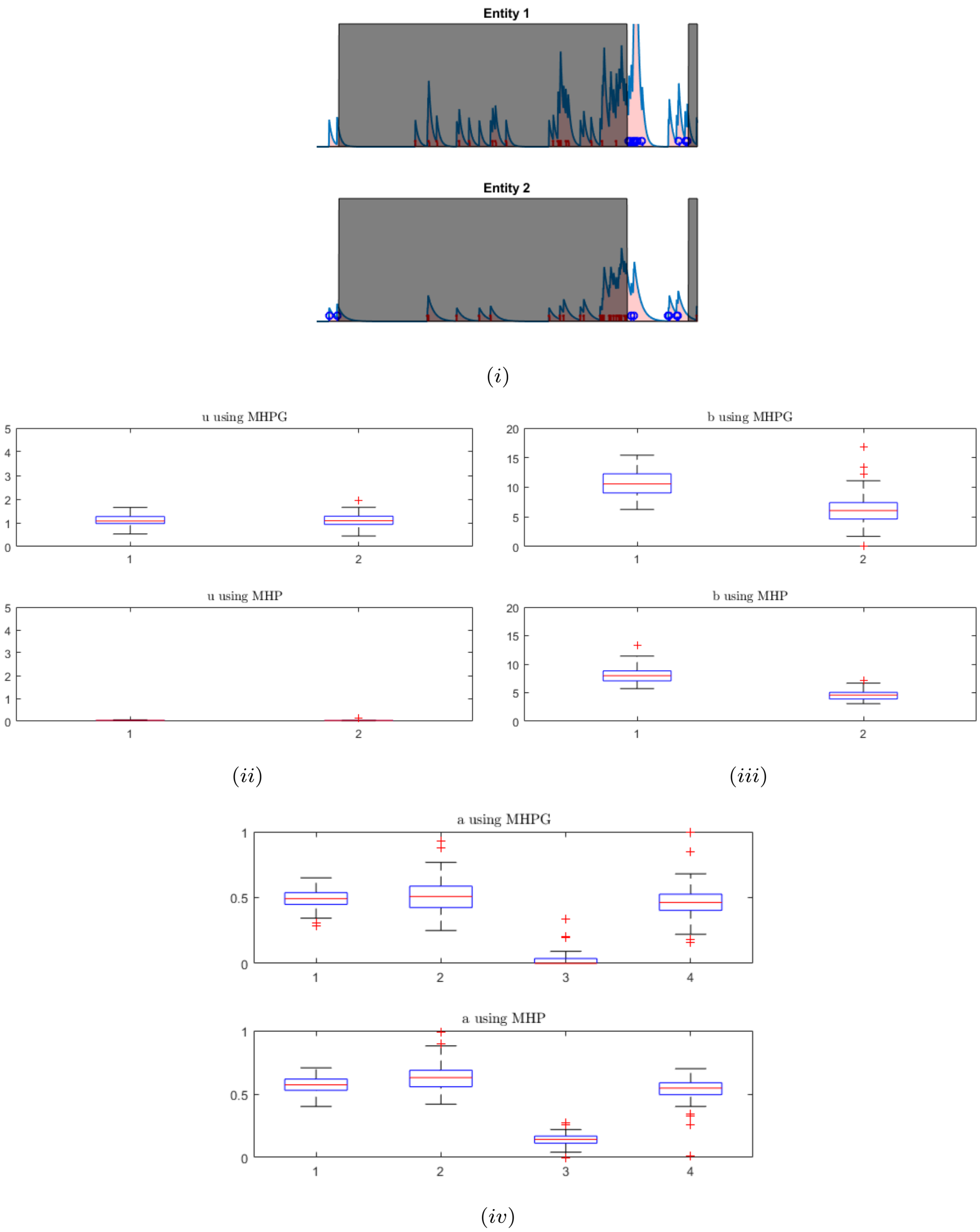


Figure 7: (i): A simulation of a Hawkes process using the true parameters in (20); shaded intervals represent gaps in observations (here $p = 0.1$.) Boxplots in (ii),(iii) and (iv) show estimation of u, b and a respectively using MHP in (6) and MHPG in (14). See (20) for the ground truths.

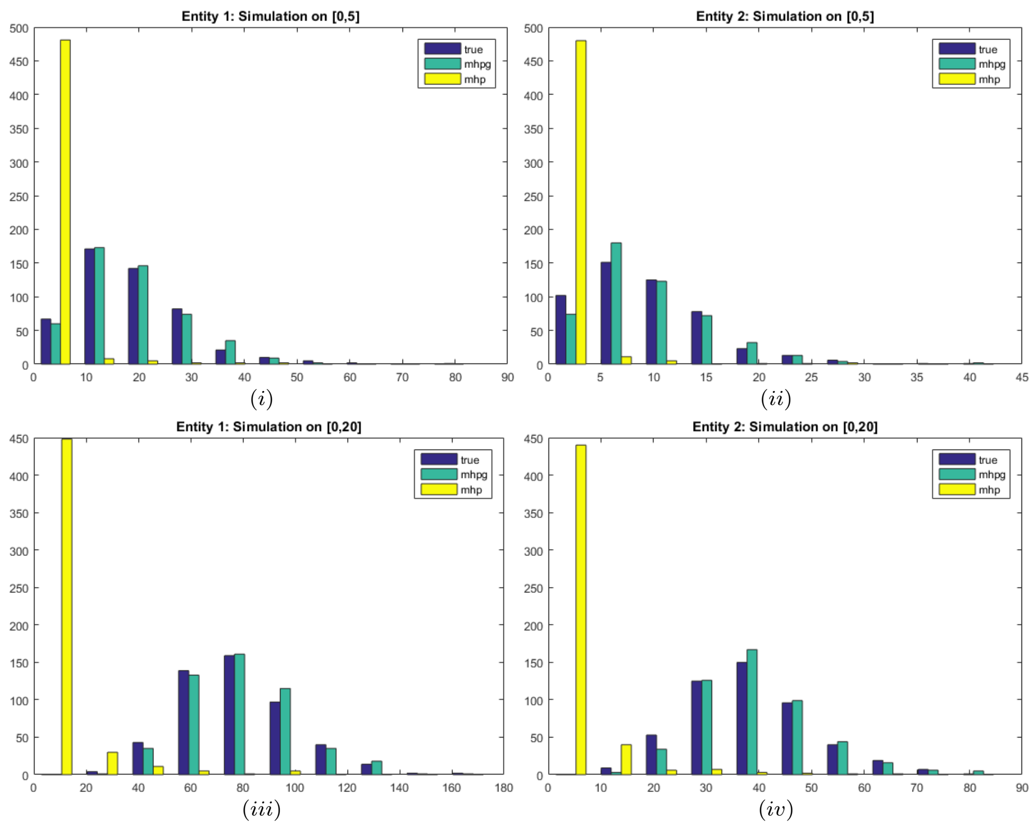
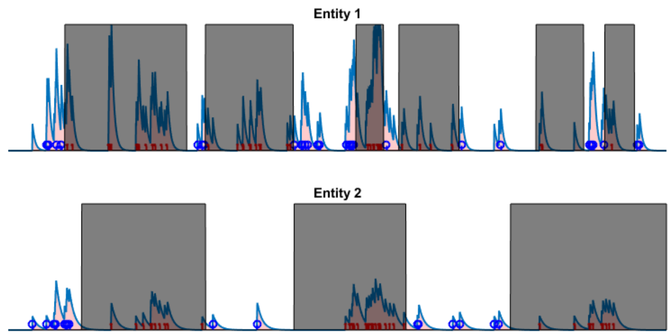
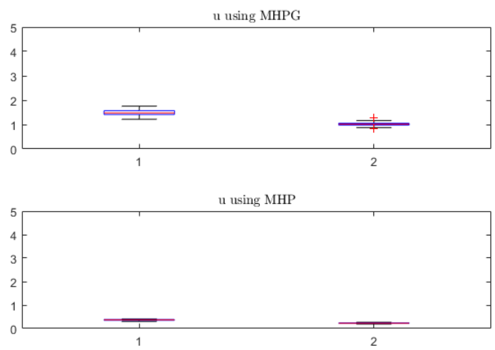


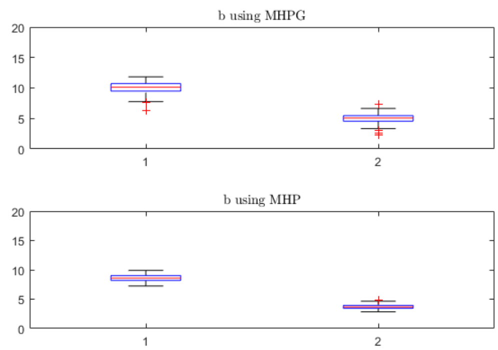
Figure 8: Comparisons of event-count histograms of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG (in green) fits the ground truth (in blue) much better then MHP (in yellow).



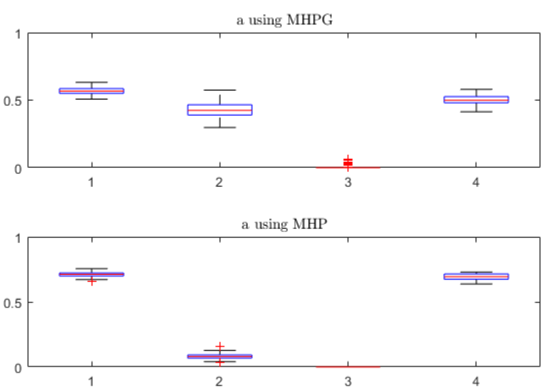
(i)



(ii)



(iii)



(iv)

Figure 9: (i): A simulation of a Hawkes process using the true parameters in (20); shaded intervals represent gaps in observations (here $p = 0.3$). Boxplots in (ii),(iii) and (iv) show estimation of u, b and a respectively using MHP in (6) and MHPG in (14). See (20) for the ground truths.

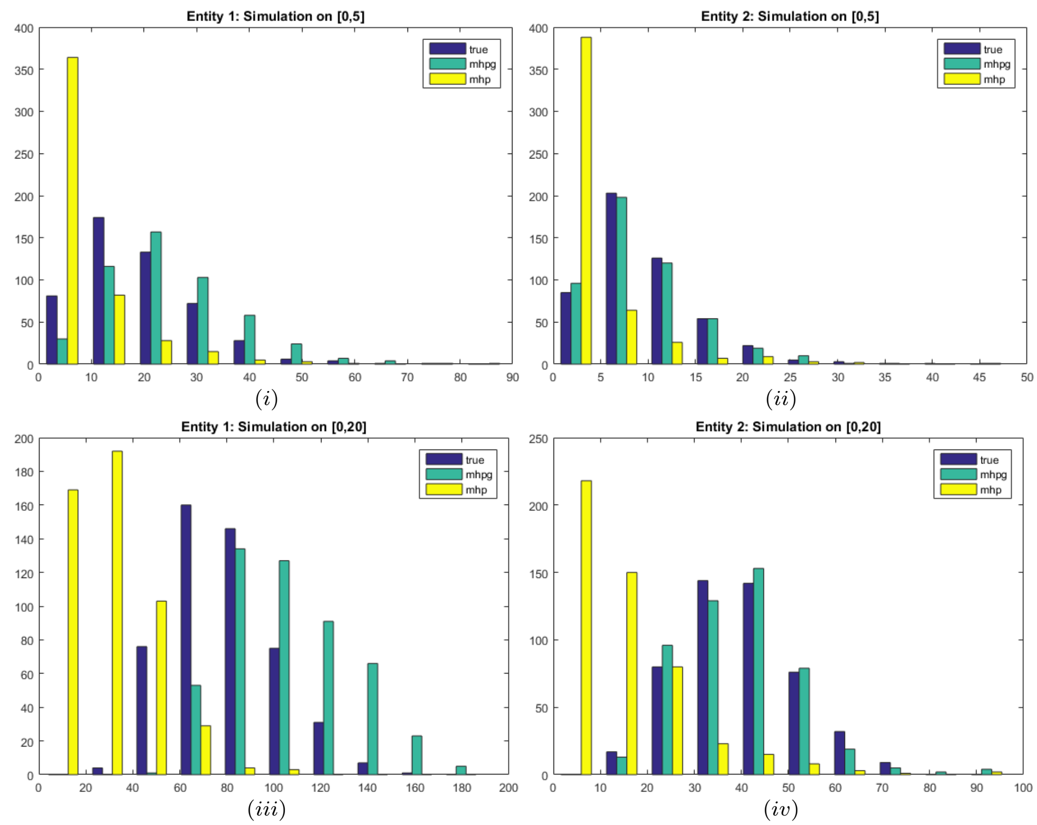


Figure 10: Comparisons of event-count histograms of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG (in green) fits the ground truth (in blue) much better then MHP (in yellow).

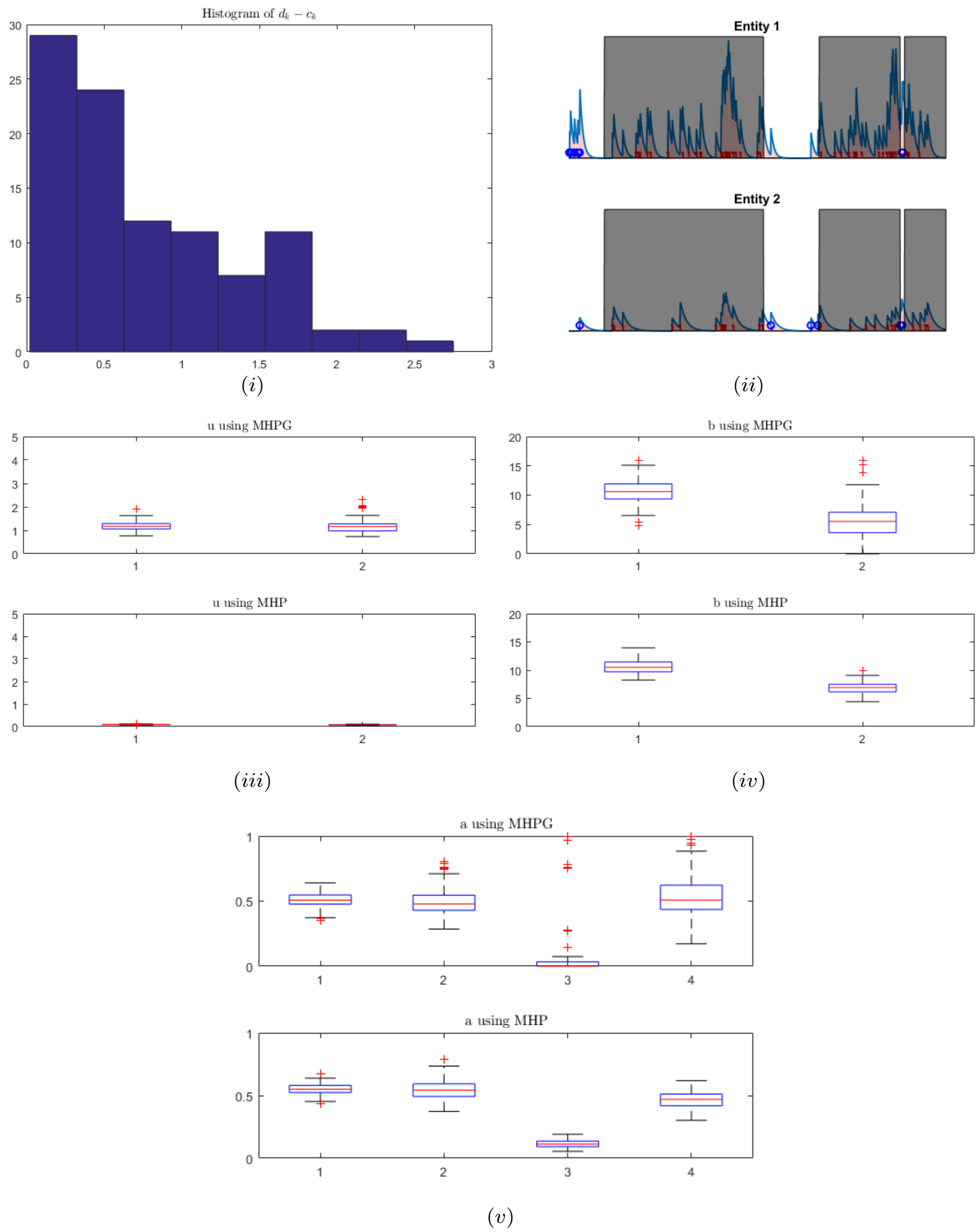


Figure 11: A simulation of a Hawkes process using the true parameters in (20). (i) shows the histogram of the lengths of the observed intervals. (ii) shows a snapshot of the observed intervals. Boxplots in (iii)-(v) show the reconstructed parameters using MHP and MHPG. See (20) for the ground truths.

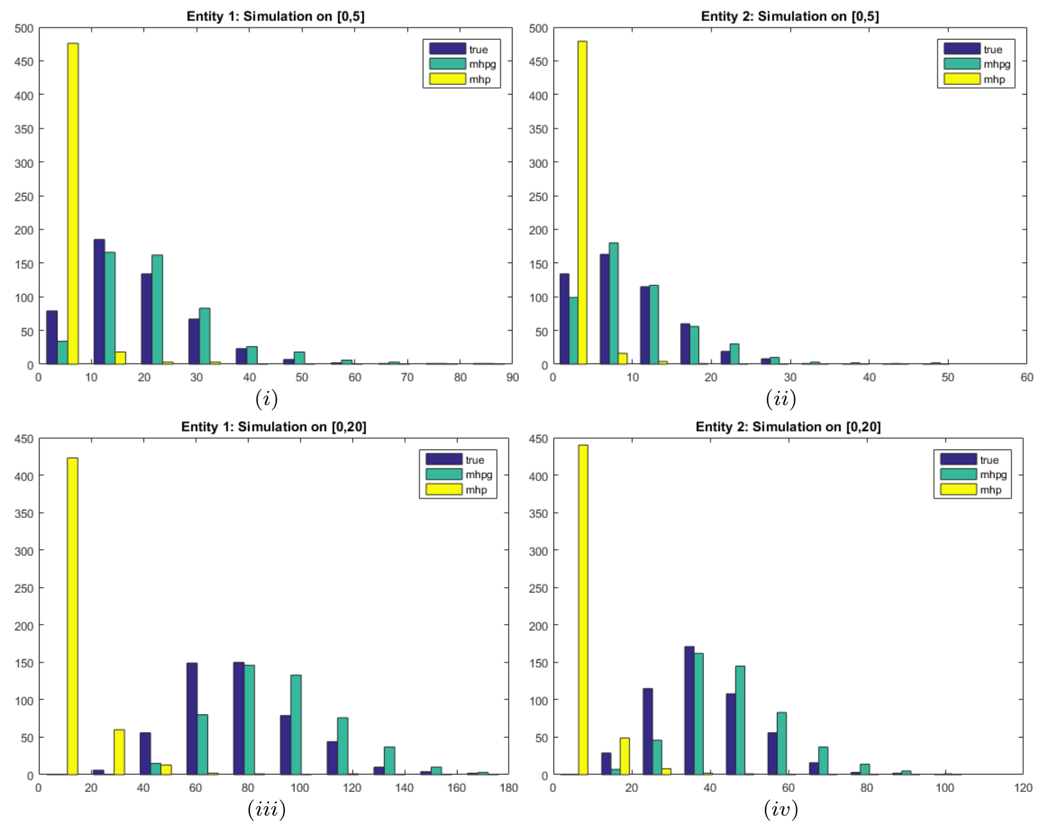
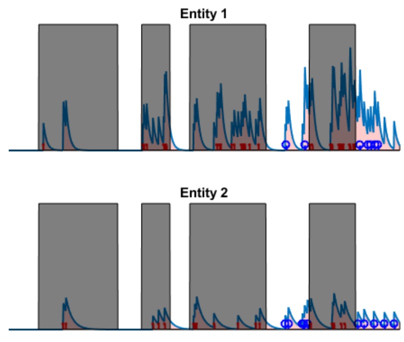
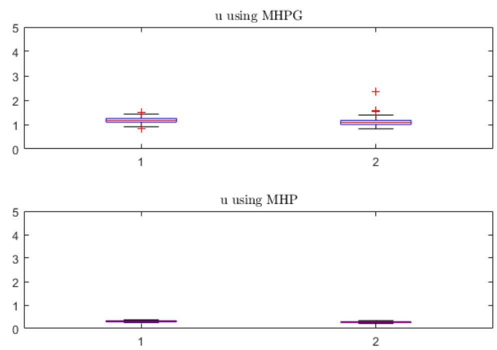


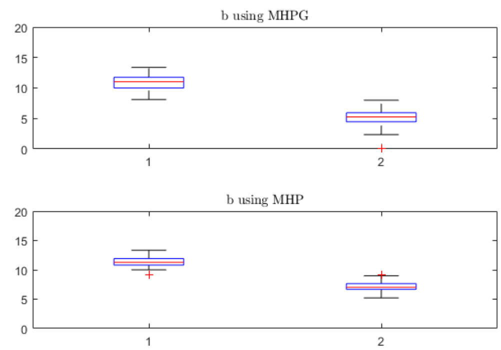
Figure 12: Comparisons of event-count histograms of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG (in green) fits the ground truth (in blue) much better then MHP (in yellow).



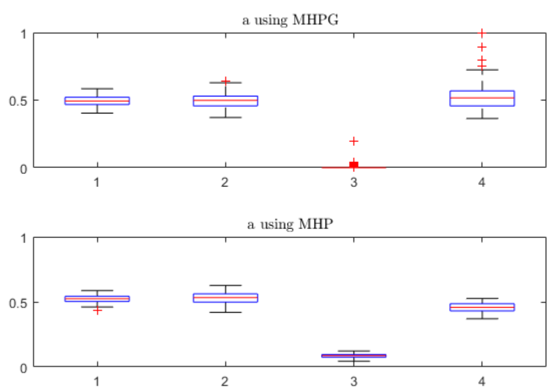
(i)



(ii)



(iii)



(iv)

Figure 13: (i): A simulation of a Hawkes process using the true parameters in (20); shaded intervals represent gaps in observations (here $p = 0.3$.) Boxplots in (ii),(iii) and (iv) show estimation of u, b and a respectively using MHP in (6) and MHPG in (14). See (20) for the ground truths.

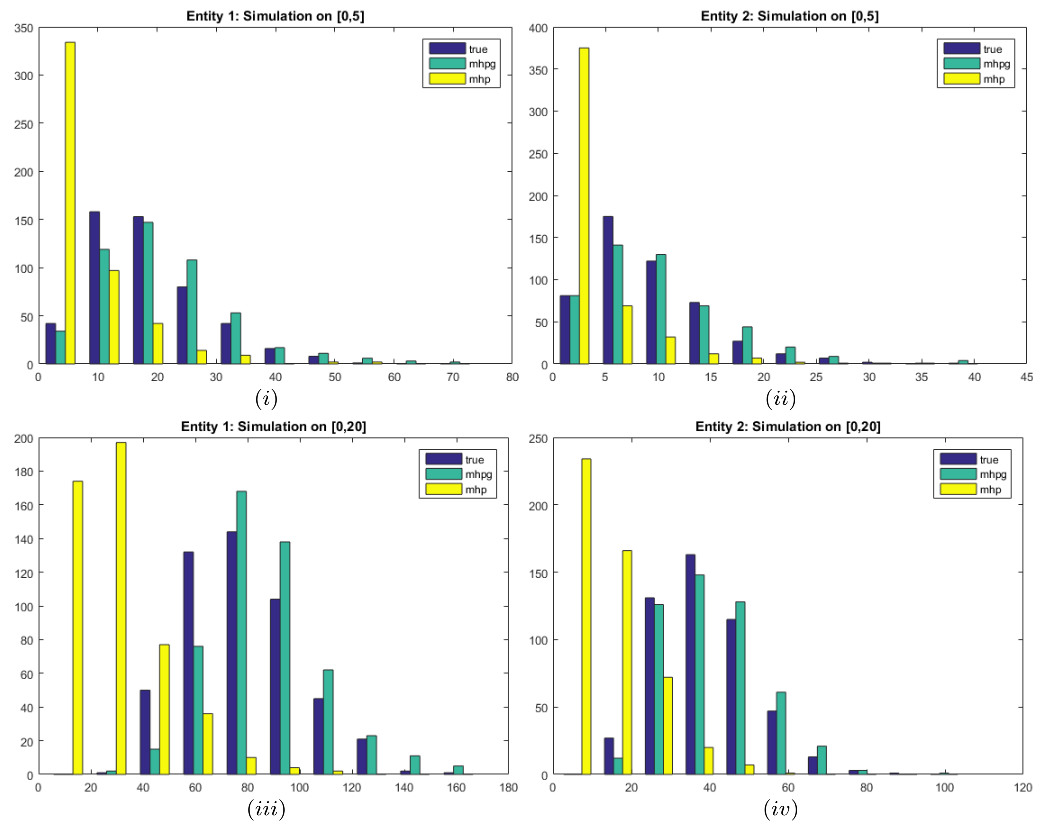


Figure 14: Comparisons of event-count histograms of MHP and MHPG with the ground truths for two intervals of length 5 and 20 times the timescale $1/b_1$. MHPG (in green) fits the ground truth (in blue) much better then MHP (in yellow).

minimizing energy (14) simplifies the problem of having a large amount of missing events by introducing a much smaller number of unknown boundary values, e.g. $\lambda_{m,k}$. In our numerical study, a constraint such as (16) is sufficient for a stable reconstruction of the underlying parameters u, a and b. We also show that MHPG provides a better reconstruction by only considering events that belong to the intersection of the observed intervals from each entity. The experiments show that the uncertainty of parameter estimations depends both on the fraction and persistence of observations. Theoretically, it is still an open problem to determine the quantitative relationships between the underlying parameters of the Hawkes process and the distribution of intermittent observed intervals. From the graph/network point of view, the constraint such as (15) on the matrix a enforces sparsity on the interactions among the entities. In other words, weak causal links are removed from the network. For analyzing spatial-temporal event data, modeling the CIF as in (7) with some smoothing on a can be used. To incorporate gaps, one can then proceed as in (13).

A Numerical Implementation

Denote $\bar{\lambda}_{m,k,i} = \bar{\lambda}_m(t_{m,i})$ for some $t_{m,i} \in (c_{m,k}, d_{m,k}]$. Also, let

$$\bar{\Lambda}_{m,k} = \int_{c_{m,k}}^{d_{m,k}} \bar{\lambda}_m(t)\ dt.$$

We have

$$\begin{aligned}\bar{\Lambda}_{m,k} &= u_m(d_{m,k} - c_{m,k}) + \frac{\bar{\lambda}_{m,k} - u_m}{b_m}\left(1 - e^{-b_m(d_{m,k} - c_{m,k})}\right) \\ &+ \sum_{n=1}^{N} a_{m,n} B_{m,n,k}, \text{ and} \\ \bar{\lambda}_{m,k,i} &= u_m + (\bar{\lambda}_{m,k} - u_{m,k}) e^{-b_m(t_{m,i} - c_{m,k})} \\ &+ \sum_{n=1}^{N} a_{m,n} A_{m,n,k,i},\end{aligned}$$

where

$$\begin{aligned}A_{m,n,k,i} &= \sum_{c_{m,k} \le t_{n,j} < t_{m,i} \le d_{m,k}} b_m e^{-b_m(t_{m,i} - t_{n,j})} \\ &= A_{m,n,k,i-1} e^{-b_m(t_{m,i} - t_{m,i-1})} \\ &+ \sum_{c_{m,k} \le t_{m,i-1} \le t_{n,j} < t_{m,i}} b_m e^{-b_m(t_{m,i} - t_{n,j})},\end{aligned} \tag{21}$$

which can be computed recursively and

$$B_{m,n,k} = \sum_{c_{m,k} \le t_{n,j} \le d_{m,k}} \left(1 - e^{-b_m(d_{m,k} - t_{n,j})}\right).$$

The minimizing energy we are interested in is:

$$J(u,a,b,\bar{\lambda}) = \sum_{m=1}^{N}\sum_{k=1}^{K_m}\left[\bar{\Lambda}_{m,k} - \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]}\log(\bar{\lambda}_{m,k,i})\right] + \mu\sum_{m,n=1}^{N}|a_{m,n}| = L(u,a,b,\bar{\lambda}) + G(a), \tag{22}$$

with the constraint that $u_m \leq \bar{\lambda}_{m,k} \geq Cu_m$.

The functional J in (22) is not convex, in particular, with respect to b. There are numerous successful numerical schemes that have been proposed to compute a minimizer for non-convex functionals. See for instance the numerical scheme PALM [3], or Block Prox-Linear Method [25], among others. The method we use here follows PALM but instead of using gradient descend which is very slow in practice we use the fixed point method.

Computing u_m: We have

$$\frac{\partial J}{\partial u_m} = \sum_{k=1}^{K_m}\left[\frac{\partial\bar{\Lambda}_{m,k}}{\partial u_m} - \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]}\frac{1}{\bar{\lambda}_{m,k,i}}\frac{\partial\bar{\lambda}_{m,k,i}}{\partial u_m}\right],$$

where

$$\begin{aligned}\frac{\partial\bar{\Lambda}_{m,k}}{\partial u_m} &= \frac{1}{b_m}\left[b_m(d_{m,k}-c_{m,k}) - \left(1-e^{-b_m(d_{m,k}-c_{m,k})}\right)\right]\\ &= \frac{1}{b_m}\left[e^{-b_m(d_{m,k}-c_{m,k})} - (1-b_m(d_{m,k}-c_{m,k})\right],\end{aligned}$$

and

$$\frac{\partial\bar{\lambda}_{m,k,i}}{\partial u_m} = \left(1-e^{-b_m(t_{m,i}-c_{m,k})}\right).$$

Setting $\frac{\partial J}{\partial u_m} = 0$, we see that a minimizer u_m must satisfies

$$u_m = \left[\sum_{k=1}^{K_m}\frac{u_m}{\bar{\lambda}_{m,k,i}}\frac{\partial\bar{\lambda}_{m,k,i}}{\partial u_m}\right] / \left[\sum_{k=1}^{K_m}\frac{\partial\bar{\Lambda}_{m,k}}{\partial u_m}\right]. \tag{23}$$

Computing $a_{m,n}$: We have

$$\begin{aligned}\frac{\partial L}{\partial a_{m,n}} &= \sum_{k=1}^{K_m}\frac{\partial\bar{\lambda}_{m,k}}{\partial a_{m,n}} - \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]}\frac{1}{\bar{\lambda}_{m,k,i}}\frac{\partial\bar{\lambda}_{m,k,i}}{\partial a_{m,n}}\\ &= \sum_{k=1}^{K_m}B_{m,n,k} - \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]}\frac{A_{m,n,k,i}}{\bar{\lambda}_{m,k,i}}\\ &= \sum_{k=1}^{K_m}B_{m,n,k} - \frac{1}{a_{m,n}}\sum_{t_{m,i}\in[c_{m,k},d_{m,k}]}\frac{a_{m,n}A_{m,n,k,i}}{\bar{\lambda}_{m,k,i}}.\end{aligned}$$

Setting

$$\bar{a}_{m,n} = \left[\sum_{t_{m,i} \in [c_{m,k}, d_{m,k}]} \frac{a_{m,n} A_{m,n,k,i}}{\bar{\lambda}_{m,k,i}} \right] / \left[\sum_{k=1}^{K_m} B_{m,n,k} \right]$$

We then solve

$$a_{m,n} = shrink_{\mu} \left(\bar{a}_{m,n} \right). \tag{24}$$

Computing b_m: We have

$$\frac{\partial J}{\partial b_m} = \sum_{k=1}^{K_m} \left[\frac{\partial \bar{\lambda}_k}{\partial b_m} - \sum_{t_{m,i} \in [c_{m,k}, d_{m,k}]} \frac{1}{\bar{\lambda}_{m,k,i}} \frac{\partial \bar{\lambda}_{m,k,i}}{\partial b_m} \right],$$

where

$$\begin{aligned} \frac{\partial \bar{\lambda}_k}{\partial b_m} &= (\bar{\lambda}_{m,k} - u_m) \Big[-\frac{1}{b_m^2} (1 - e^{-b_m (d_{m,k} - c_{m,k})}) \\ &+ \frac{1}{b_m} (d_{m,k} - c_{m,k}) e^{-b_m (d_{m,k} - c_{m,k})} \Big] \\ &+ \sum_{n=1}^{N} a_{m,n} \frac{\partial B_{m,n,k}}{\partial b_m}. \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \bar{\lambda}_{m,k,i}}{\partial b_m} &= -(\bar{\lambda}_{m,k} - u_m)(t_{m,i} - c_{m,k}) e^{-b_m (t_{m,i} - c_{m,k})} \\ &+ \sum_{n=1}^{N} a_{m,n} \frac{\partial A_{m,n,k,i}}{\partial b_m} \end{aligned}$$

where

$$\begin{aligned} \frac{\partial A_{m,n,k,i}}{\partial b_m} &= \sum_{c_{m,k} \le t_{n,j} < t_{m,i} \le d_{m,k}} e^{-b_m (t_{m,i} - t_{n,j})} \\ &- b_m \sum_{c_{m,k} \le t_{n,j} < t_{m,i} \le d_{m,k}} (t_{m,i} - t_{n,j}) e^{-b_m (t_{m,i} - t_{n,j})} \\ &= \frac{\partial A^{(1)}_{m,n,k,i}}{\partial b_m} - b_m \frac{\partial A^{(2)}_{m,n,k,i}}{\partial b_m}. \end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial J}{\partial b_m} &= \sum_{k=1}^{K_m} \frac{\partial \bar{\lambda}_{m,k}}{\partial b_m} \\
&- \sum_{k=1}^{K_m} \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]} \frac{1}{\bar{\lambda}_{m,k,i}} \cdot \\
&\quad \left(-(\bar{\lambda}_{m,k} - u_m)(t_{m,i} - c_{m,k}) e^{-b_m(t_{m,i}-c_{m,k})} + \sum_{n=1}^{N} a_{m,n} \frac{\partial A^{(1)}_{m,n,k,i}}{\partial b_m} \right) \\
&- b_m \sum_{k=1}^{K_m} \left[\sum_{t_{m,i}\in[c_{m,k},d_{m,k}]} \frac{\left(\sum_{n=1}^{N} a_{m,n} \frac{\partial A^{(2)}_{m,n,k,i}}{\partial b_m} \right)}{\bar{\lambda}_{m,k,i}} \right] \\
&= A_1 - A_2 - b_m A_3.,
\end{aligned}$$

Setting $\frac{\partial J}{\partial b_m} = 0$ implies that a minimizer b_m must satisfies,

$$b_m = \frac{A_1 - A_2}{A_3}. \tag{25}$$

Computing $\bar{\lambda}_{m,k}$:

$$\frac{\partial J}{\partial \bar{\lambda}_{m,k}} = \frac{1}{b_m}(1 - e^{-b_m(d_{m,k}-c_{m,k})}) - \sum_{t_{m,i}\in[c_{m,k},d_{m,k}]} \frac{1}{\bar{\lambda}_{m,k,i}} e^{-b_m(t_{m,i}-c_{m,k})}.$$

Setting $\frac{\partial J}{\partial \bar{\lambda}_{m,k}} = 0$ implies that a minimizer $\bar{\lambda}_{m,k}$ must satisfies

$$\bar{\lambda}_{m,k} = b_m \left[\sum_{t_{m,i}\in[c_{m,k},d_{m,k}]} \frac{\bar{\lambda}_{m,k} e^{-b_m(t_{m,i}-c_{m,k})}}{\bar{\lambda}_{m,k,i}} \right] / \left[1 - e^{-b_m(d_{m,k}-c_{m,k})} \right], \tag{26}$$

with the constraint $u_m \leq \bar{\lambda}_{m,k} \geq C u_m$, for some $C \geq 1$.

Algorithm 2 (Algorithm for parameter estimation)**.** Given $E_m = \{t_{m,i}\} \subset \cup_{k=1}^{K_m} (c_{m,k}, d_{m,k}], m = 1, \cdots, N$, some $\mu > 0$ and dt = small.

1. Initial guess: $u_m^0 = 1, a_{m,n}^0 = 0.5/N, b_m = 1000, \bar{\bar{\lambda}}^0 = \{\bar{\lambda}_{m,k}^0 = u_m^0\}$.
2. Suppose u^ℓ, a^ℓ, b^ℓ and $\bar{\bar{\lambda}}^\ell$ are known.
3. While not convergent
   (a) Compute $u^{\ell+1}$ using u^ℓ, a^ℓ, b^ℓ and $\bar{\lambda}^\ell$ via (23).
   (b) Compute $a^{\ell+1}$ using $u^{\ell+1}, a^\ell, b^\ell$ and $\bar{\lambda}^\ell$ via (24).
   (c) Compute $b^{\ell+1}$ using $u^{\ell+1}, a^{\ell+1}, b^\ell$ and $\bar{\lambda}^\ell$ via (25).
   (d) Compute $\bar{\lambda}^{\ell+1}$ using $u^{\ell+1}, a^{\ell+1}, b^{\ell+1}$ and $\bar{\lambda}^\ell$ via (26) using the constraint (16) with $C = 20$.
4. End while.

References

[1] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

[2] Shahriar Azizpour and Kay Giesecke. Self-exciting corporate defaults: Contagion vs. frailty. Technical report, Stanford University working paper series, 2008.

[3] Jerome Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[4] Clive G. Bowsher. Modeling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.

[5] Daryl J Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes, volume I: Elementary Theory and Methods of Probability and its Applications*. Springer, New York,, 2003.

[6] Angelos Dassios and Hongbiao Zhao. A dynamic contagion process. *Advances in applied probability*, pages 814–846, 2011.

[7] Angelos Dassios and Hongbiao Zhao. Exact simulation of hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 18(62), 2013.

[8] Paul Embrechts and Matthias Kirchner. Hawkes graphs. *arXiv preprint arXiv:1601.01879*, 2016.

[9] Paul Embrechts, Thomas Liniger, Lu Lin, et al. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.

[10] Jalal Etesami, Negar Kiyavash, Kun Zhang, and Kushagra Singhal. Learning network of multivariate hawkes processes: A time series approach. *arXiv preprint arXiv:1603.04319*, 2016.

[11] Eric C Hall and Rebecca M Willett. Tracking dynamic point processes on networks. *arXiv preprint arXiv:1409.0031*, 2014.

[12] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[13] Sanggyun Kim, David Putrino, Soumya Ghosh, and Emery N Brown. A granger causality measure for point process models of ensemble neural spiking activity. *PLoS computational biology*, 7(3):e1001110, 2011.

[14] Erik Lewis, George Mohler, P Jeffrey Brantingham, and Andrea L Bertozzi. Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25(3):244–264, 2012.

[15] Naoki Masuda, Taro Takaguchi, Nobuo Sato, and Kazuo Yano. Self-exciting point process modeling of conversation event sequences. In *Temporal Networks*, pages 245–264. Springer, 2013.

[16] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

[17] Y. Ogata and K. Shimazaki. Transition from aftershock to normal activity: The 1965 rat islands earthquake aftershock sequence. *Bulletin of the Seismological Society of America*, 74(5):1757–1765, 1984.

[18] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[19] Yosihiko Ogata. Seismicity analysis through point-process modeling: A review. *Pure and applied geophysics*, 155(2-4):471–507, 1999.

[20] Patricia Reynaud-Bouret, Vincent Rivoirard, and Christine Tuleau-Malot. Inference of functional connectivity in neurosciences via hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*.

[21] M.B. Short, G.O. Mohler, P.J. Brantingham, and G.E. Tita. Gang rivalry dynamics via coupled point process networks. *Discrete & Continuous Dynamical Systems-Series B*, 19(5), 2014.

[22] Aiden Sidebottom. Repeat burglary victimization in malawi and the influence of housing type and area-level affluence. *Security journal*, 25(3):265–281, 2-12.

[23] Alexey Stomakhin, Martin B Short, and Andrea L Bertozzi. Reconstruction of missing data in social networks based on temporal patterns of interactions. *Inverse Problems*, 27(11):115013, 2011.

[24] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. of Royal Stat. Society, Series B (Methodological)*, pages 267–288, 1996.

[25] Yangyang Xu and Wotao Yin. A globally convergent algorithm for nonconvex optimization based on block coordinate update. *arXiv preprint arXiv:1410.1386*, 2014.

[26] J. zhuang, O. Yosihiko, and Vere-Jones D. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Society*, 97(458):369–380, 2002.

[27] Joseph R Zipkin, Frederic P Schoenberg, Kathlyn Coronges, and Andrea L Bertozzi. Point-process models of social network interactions: parameter estimation and missing data recovery. *UCLA CAM Report*, (14-65), August, 2014.