The Malliavin-Stein method for the multivariate compound Hawkes process

Mahmoud Khabou

Malliavin-Stein method for the multivariate compound Hawkes process

Mahmoud Khabou*

September 15, 2021

Abstract

In this paper, we provide upper bounds on the d_2 distance between a large class of functionals of a multivariate compound Hawkes process and a given Gaussian vector. This is proven using Malliavin's calculus defined on an underlying Poisson embedding. The upper bound is then used to infer the speed of convergence of Central Limit Theorems for the multivariate compound Hawkes process with exponential kernels as the observation time T goes to infinity.

1 Introduction

Hawkes processes have been used to model events that exhibit self-exciting properties. Initially introduced in 1971 [7] to model seismic activities, Hawkes processes became popular in other fields like credit risk [6] or micro-structure in finance [2]. In recent years, the need to model systems with interacting components has been growing, that is how the multivariate Hawkes process emerged. In addition to its self-exciting properties, the multivariate Hawkes process can model mutually-exciting phenomena and it has been used in a wide array of fields such that neuro-science [11], [12], social networks [1], cyber-security [4] and financial econometrics [5].

In this paper, we consider the multivariate compound process (or total loss process) $(\boldsymbol{L}_t)_{t\geq 0}$ defined component-wise at a time t as

$$\begin{cases} L_t^1 = \sum_{k=1}^{H_t^1} Y_k^1, \quad (Y_k^1)_{k=1,\cdots} \text{ are } i.i.d, \\ \vdots \\ L_t^d = \sum_{k=1}^{H_t^d} Y_k^d, \quad (Y_k^d)_{k=1,\cdots} \text{ are } i.i.d, \end{cases} \tag{1}$$

where $(Y_k^j)_{k\in\mathbb{N}^*,j=1,\cdots,d}$ are $i.i.d$ random variables and $\boldsymbol{H}_t = \left(H_t^1,\cdots,H_t^d\right)$ is a family of d point processes each having an intensity $(\lambda_t^1,\cdots,\lambda_t^d)$ such that

$$\lambda_t^i \mathrm{d}t = \mathbb{E}\left[H_{t+\mathrm{d}t}^i - H_t^i | \mathcal{F}_{t^-}\right], \quad i = 1,\cdots,d.$$

The process $\boldsymbol{L}$ is called a multivariate compound Hawkes process if its intensity $\boldsymbol{\lambda}$ follows the dynamics

$$\begin{cases} \lambda_t^1 = \mu^1 + \int_{[0,t)} \sum_{k=1}^d \Phi_{1k}(t-s)\mathrm{d}L_s^k, \\ \vdots \\ \lambda_t^d = \mu^d + \int_{[0,t)} \sum_{k=1}^d \Phi_{dk}(t-s)\mathrm{d}L_s^k, \end{cases} \tag{2}$$

*INSA de Toulouse, IMT UMR CNRS 5219, Université de Toulouse, 135 avenue de Rangueil 31077 Toulouse Cedex 4 France. Email: mahmoud.khabou@insa-toulouse.fr

where $\boldsymbol{\mu} = (\mu^1, \cdots, \mu^d) \in \mathbb{R}_+^d$ plays the role of a baseline intensity and $\Phi = (\Phi_{ik})_{i,k=1,\cdots,d}$ is a family of non-negative integrable kernels on $\mathbb{R}_+$.
This dynamics can also be expressed under the following matrix form

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{[0,t)} \Phi(t-s)\mathrm{d}\boldsymbol{L}_s.$$

In some applications, the question about the Hawkes process' longtime behaviour naturally arises. In the special case $Y_k^j \equiv 1$, and under the condition that the spectral radius of the matrix $S := \|\Phi\|_1$ is strictly less than one ($\rho(S) < 1$), Bacry *et al.* [3] proved the following martingale Central Limit Theorem *(CLT)* for $(\boldsymbol{L}_t)_{t\geq 0} = (\boldsymbol{H}_t)_{t\geq 0}$

$$\left(\frac{\boldsymbol{H}_{Tv} - \int_0^{Tv} \boldsymbol{\lambda}_t \mathrm{d}t}{\sqrt{T}}\right)_{v\in[0,1]} \underset{T\to+\infty}{\Longrightarrow} \left(\Sigma^{1/2} W_v\right)_{v\in[0,1]}$$

where the convergence takes place in law for the Skorokhod topology and where Σ is a diagonal matrix that depends only on $\boldsymbol{\mu}$ and S. From this martingale result, they derived the alternative CLT

$$\left(\frac{\boldsymbol{H}_T - \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t}{\sqrt{T}}\right)_{v\in[0,1]} \underset{T\to+\infty}{\Longrightarrow} \left((I_d - S)^{-1}\Sigma^{1/2} W_v\right)_{v\in[0,1]}.$$

Until very recently, the quantification of the speed of this convergence has not been thoroughly studied. In the univariate 1-marginal case ($v = 1$), Hillairet *et al.* [8] have found a Berry-Esséen type bound on the Wassestein distance between the normalized one-dimensional Hawkes process and its Gaussian limit

$$d_W\left(\frac{H_T - \int_0^T \mathbb{E}[\lambda_t]\mathrm{d}t}{\sqrt{T}}, \mathcal{N}(0,\tilde{\sigma}^2)\right) = O\left(\frac{1}{\sqrt{T}}\right), \quad \tilde{\sigma}^2 = \frac{\mu}{(1-\|\phi\|_1)^3}$$

in case the kernel is an exponential $\phi(u) = \alpha e^{-\beta u}$ (with $0 < \alpha < \beta$) or an Erlang function $\phi(u) = \alpha u e^{-\beta u}$ (with $0 < \alpha < \beta^2$). This result has been derived thanks to an approach introduced by Nourdin and Peccati [9] which combines Malliavin's calculus with Stein's method.

In this article we prove a generalization of the quantification result to the multivariate case using a version of Malliavin's calculus that is adapted to higher dimensions. Following the lines of [13], we prove an upper bound on the distance between a vector of divergences with respect to the multivariate Poisson process and a given centered multivariate Gaussian in the d_2 metric. Unlike the Wasserstein metric which remains relevant in the multivariate normal space [9], the d_2 metric is more suitable for the multi-dimensional Poisson space.

After defining compound multivariate Hawkes process as the result of the thinning of Poisson measures in section 2.3, we introduce the elements of multivariate Malliavin's calculus on the Hawkes process in sections 2.4 and 2.5. Finally, in section 3.1, we prove a general bound on a class of multivariate Hawkes functionals. As a first application in section 3.2 we give a bound on the Wasserstein distance between the multivariate and multimarginal normalized martingale and its Gaussian limits. Then we use those results to show that if the kernels take the exponential form

$$\Phi_{ij}(u) = \alpha_{ij} e^{-\beta_i u},$$

and under the following assumptions

Assumption 1 (Stability 1)**.** *The spectral radius of the matrix* $B^{-1}A\,\mathrm{diag}(m^1, \cdots, m^d)$ *satisfies*

$$\rho\left(B^{-1}A\,\mathrm{diag}(m^1, \cdots, m^d)\right) < 1,$$

with $A = (\alpha_{ij})_{ij}$, $B^{-1} = \mathrm{diag}(\frac{1}{\beta_1}, \cdots, \frac{1}{\beta_d})$ *and* $(m^1, \cdots, m^d) = (\mathbb{E}[Y^1], \cdots, \mathbb{E}[Y^d])$.

Assumption 2 (Stability 2)**.** *Set* $V = B - A\,\mathrm{diag}(m^1, \cdots, m^d)$*. The eigenvalues of* V *are positive.*

Assumption 3 (Third moment)**.** *The measures* $\nu^1, \cdots, \nu^d$ *of* $Y_1^1, \cdots, Y_1^d$ *have finite third moments.*

the follwoing result holds:

Theorem 1.1. *Let $\boldsymbol{L}_T$ be a multivariate compound Hawkes process.*
Set $\boldsymbol{Y'}_T = \frac{\boldsymbol{L}_T - \operatorname{diag}(m^1,\cdots,m^d)\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}T}{\sqrt{T}}$.
Set $\tilde{C} = \left(J\sqrt{C}\right)\left(J\sqrt{C}\right)^{\top} = JCJ^{\top}$, *where* $J = \left(I_d - \operatorname{diag}(m^1,\cdots,m^d)B^{-1}A\right)^{-1}$ *and* $C = \operatorname{diag}(\sigma_1^2,\cdots,\sigma_d^2)$ *where for any* $j = 1,\cdots,d$

$$\sigma_j^2 = \int x^2\nu^j(\mathrm{d}x)\left[\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}\right]^j.$$

Let $\tilde{\boldsymbol{G}} \sim \mathcal{N}\left(0,\tilde{C}\right)$. *Then, under the Assumptions 1, 2, 3, there exists a constant K that does not depend on T such that*

$$d_2(\boldsymbol{Y'}_T, \tilde{\boldsymbol{G}}) \leq \frac{K}{\sqrt{T}},$$

for any $T > 0$.

Proof. Cf section 3.3. □

Remark 1. *Assumption 1 is a special case of the assumption*

$$\rho\left(\|\Phi\|_1\operatorname{diag}(m^1,\cdots,m^d)\right) < 1,$$

since in the case of an exponential kernel one has

$$\|\Phi\|_1 = \left(\|\Phi_{ij}\|_1\right)_{ij} = \left(\|\alpha_{ij}e^{\beta_i\cdot}\|_1\right)_{ij} = \left(\frac{\alpha_{ij}}{\beta_i}\right)_{ij} = B^{-1}A.$$

In the case of a uni-variate compound Hawkes process ($d = 1$), this is equivalent to assuming

$$\|\Phi\|_1 m = \frac{\alpha}{\beta}m < 1.$$

Remark 2. *If the memory parameters $(\beta_i)_{i=1,\cdots,d}$ are the same for every particle (i.e $\Phi_{ij}(u) = \alpha_{ij}e^{-\beta u}$), Assumptions 1 and 2 become equivalent.*

For instance, we take the bivariate compound Hawkes process $\boldsymbol{L}_T = \left(L_T^1, L_T^2\right)$ where

$$\begin{cases} L_t^1 = \sum_{k=1}^{H_t^1} Y_k^1, \\ \\ L_t^2 = \sum_{k=1}^{H_t^2} Y_k^2, \end{cases}$$

with *i.i.d* claims $(Y_k^i)_{i=1,2,k\in\mathbb{N}}$ with common exponential distribution $\mathcal{E}(1)$. The intensities are assumed to follow the dynamics

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{[0,t)} Ae^{-\beta(t-s)}\mathrm{d}\boldsymbol{L}_s.$$

For stability, it is enough to choose the parameters such that $\rho(A) < \beta$. For instance, by setting

$$A = \begin{pmatrix} \frac{1}{2} & 2 \\ 2 & \frac{1}{2} \end{pmatrix},$$

any $\beta > \frac{5}{2}$ satisfies Assumptions 1 and 2. Assumption 3 is satisfied because the exponential distribution has moments of every order.
Figure 1 illustrates the convergence of the compensated loss (*cf.* Theorem 1.1) to its centered Gaussian limit $\mathcal{N}(0,\tilde{C})$ (also defined in Theorem 1.1)

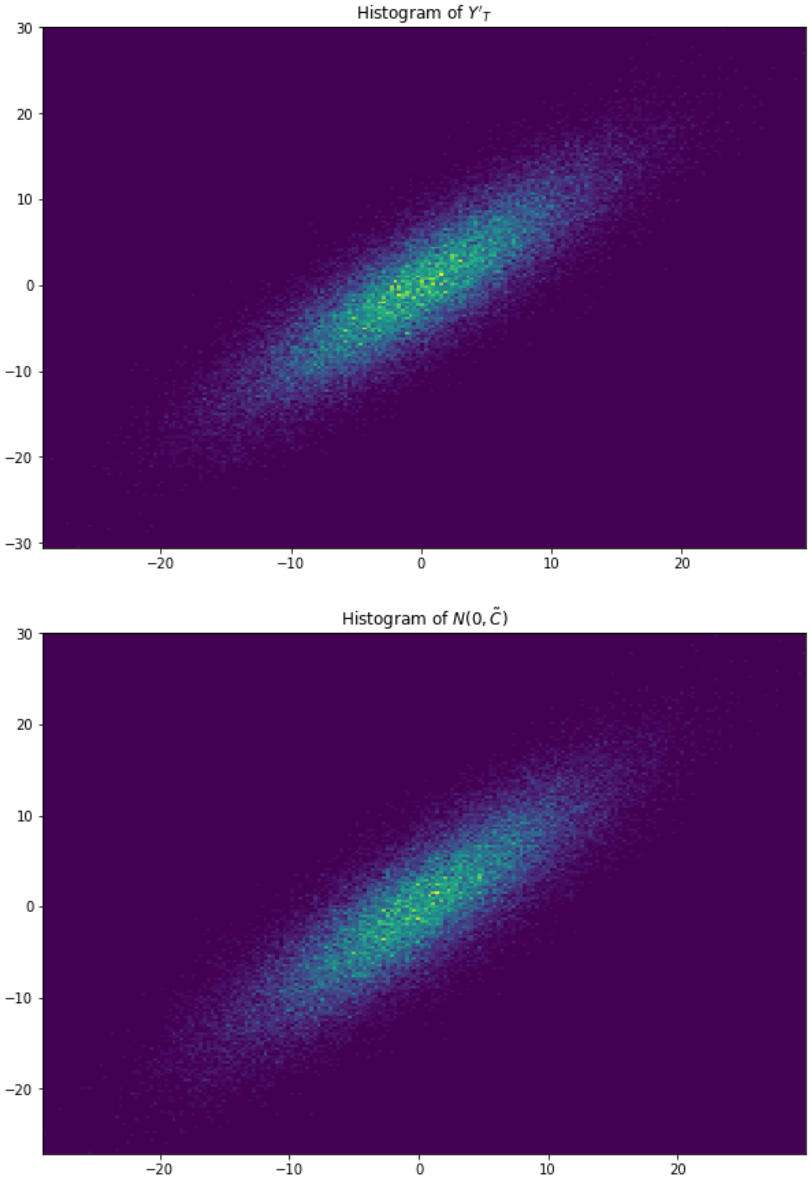

Figure 1: $2-D$ histograms of $\boldsymbol{Y'}_T$ and $\mathcal{N}(0, \tilde{C})$ for $\beta = 4$, $\boldsymbol{\mu} = (2,3)^\top$ and $T = 1000$ for $n = 40000$ Monte Carlo samples.

On the other hand, showing the speed of convergence in the d_2 metric is a bit more complicated, mainly for two reasons:

1. We do not know for which function the upper bound is reached for the sake of simulation.
2. Since the quantity $\mathbb{E}[f(\boldsymbol{Y'}_T)]$ cannot be computed directly, is has to be approximated with a Monte Carlo estimation $\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{Y'}^k_T)$. This means that for large times, the term in $\frac{1}{\sqrt{T}}$ can be eclipsed by the slow decay of the Monte Carlo estimator which is in $\frac{1}{\sqrt{n}}$.

We can still illustrate the behaviour of $\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{Y'}^k_T) - \mathbb{E}[f(\boldsymbol{G})]$ for $\boldsymbol{G} \sim \mathcal{N}(0, \tilde{C})$, where f is a "well behaved" function for which $\mathbb{E}[f(\boldsymbol{G})]$ is know explicitly. For instance, the following figure shows the evolution in time for $f(\boldsymbol{x}) = e^{-\frac{1}{4}\|\boldsymbol{x}\|^2}$

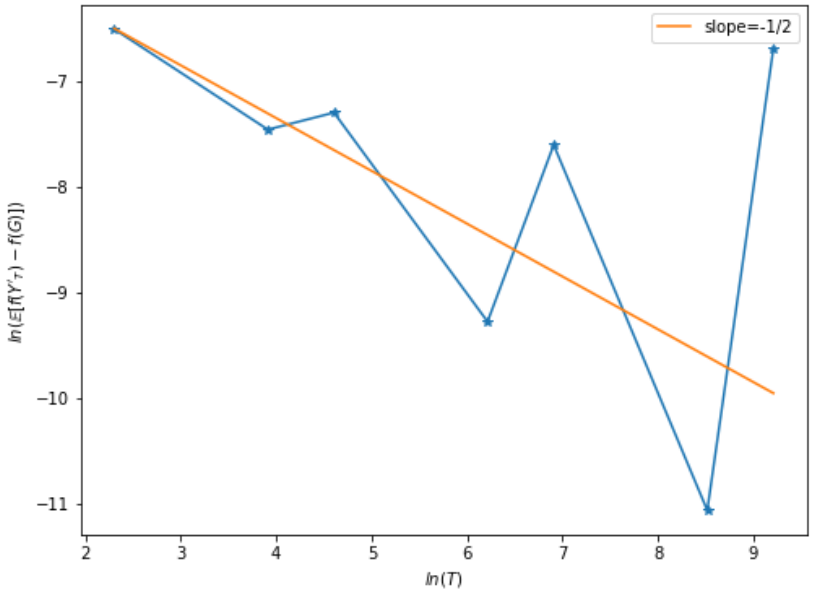

Figure 2: Monte Carlo estimator $\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{Y'}_T^k) - \mathbb{E}[f(\boldsymbol{G})]$ (in blue) for $\beta = 6$, $T = [10, 50, 100, 500, 1000, 5000, 10000]$ and $n = 150000$. Notice how for large T, the Monte Carlo estimator's error becomes stronger.

2 Notations and preliminaries

In this section we generalize the mathematical framework introduced in Section 2 of [8]. We then proceed to define the multivarite Hawkes loss as the result of a thinning procedure from a Poisson measure. We finally recall some elements of Stein's method in

2.1 General notations

1. Let d be an integer. For any two vectors $\boldsymbol{u} = \left(u^i\right)_{i=1,\cdots,d}$ and $\boldsymbol{v} = \left(v^i\right)_{i=1,\cdots,d}$ in $\mathbb{R}^d$, the product $\boldsymbol{uv}$ is defined as the vector of $\mathbb{R}^d$ such that $\boldsymbol{uv} = \left(u^i v^i\right)_{i=1,\cdots,d}$

2. Let $\theta \in \mathbb{R}$ and $\mathbb{1}_{\theta \leq u^i}$ be the indicator of the set $\{\theta \leq u^i\}$. We define $\mathbf{1}_{\theta \leq \boldsymbol{u}}$ as the element of $\mathbb{R}^d$ such that $\mathbf{1}_{\theta \leq \boldsymbol{u}} = \left(\mathbb{1}_{\theta \leq u^i}\right)_{i=1,\cdots,d}$.

3. Similarly, if $\boldsymbol{\nu}$ is a vector of d measures and $\boldsymbol{f}$ is a vector of d functions, we define $\int \boldsymbol{f}(x)\boldsymbol{\nu}(\mathrm{d}x) = \left(\int f^i(x)\nu^i(\mathrm{d}x)\right)_{i=1,\cdots,d}$.

4. The inner product $\langle \cdot, \cdot \rangle$ corresponds to the Euclidian inner product and $\| \cdot \|$ is its norm.

5. The operator norm of a matrix $A \in \mathcal{M}_d(\mathbb{R})$ is

$$\|A\|_{op} := \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

6. For every function $g : \mathbb{R}^d \to \mathbb{R}$, let

$$\|g\|_{Lip} := \sup_{\boldsymbol{x} \neq \boldsymbol{y}} \frac{|g(\boldsymbol{x}) - g(\boldsymbol{y})|}{\|\boldsymbol{x} - \boldsymbol{y}\|}.$$

7. If $g \in \mathcal{C}^1(\mathbb{R}^d)$ (continuously differentiable) then we write

$$M_2(g) := \sup_{\boldsymbol{x} \neq \boldsymbol{y}} \frac{\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|}{\|\boldsymbol{x} - \boldsymbol{y}\|}.$$

8. Similarly, if $g \in \mathcal{C}^2(\mathbb{R}^d)$ (twice continuousy differentiable) then

$$M_3(g) := \sup_{\boldsymbol{x} \neq \boldsymbol{y}} \frac{\| \operatorname{Hess} f(\boldsymbol{x}) - \operatorname{Hess} f(\boldsymbol{y}) \|_{op}}{\|\boldsymbol{x} - \boldsymbol{y}\|}.$$

2.2 Elements of stochastic analysis on the multivariate Poisson space

Let d be a positive integer. Let $\nu^1, \cdots, \nu^d$ be a family of integrable probability measures on $\mathbb{R}_+$ such that $\nu^i(\{0\}) = 0$ for every $i = 1, \cdots, d$, and define $m^i = \int_{\mathbb{R}_+} x\nu^i(\mathrm{d}x)$.
In this section, every component of the multivariate compound Hawkes process is obtained through the thinning of 3−component Poisson measure.
Let the space of configurations Ω^d, where

$$\Omega := \left\{ \omega^j = \sum_{i=1}^{n} \delta_{(t_i, \theta_i, x_i)},\ 0 = t_0 < t_1 < \cdots < t_n,\ (\theta_i, x_i) \in \mathbb{R}_+ \times \mathbb{R},\ n \in \mathbb{N} \cup \{+\infty\} \right\}.$$

Let $\mathcal{F}$ be the σ-field associated to the vague topology on Ω^d, and $\mathbb{P}$ the Poisson measure under which the family

$$\boldsymbol{N} = \left(N^j\right)_{j=1,\cdots,d}$$

where

$$N^j\left([0,t] \times [0,b] \times (-\infty, y]\right)(\omega^j) := \omega^j\left([0,t] \times [0,b] \times (-\infty, y]\right), \quad (t,b,y) \in \mathbb{R}_+^3, \quad j = 1, \cdots, d$$

is a family of independent homogeneous Poisson processes with intensity measures $\mathrm{d}t \otimes \mathrm{d}\theta \otimes \mathrm{d}\nu^j$, that is,

$$\begin{aligned}\mathbb{P}\big[N^1([0,t_1] \times [0,b_1] \times (-\infty, y_1]) = n_1, \cdots, N^d([0,t_d] \times [0,b_d] \times (-\infty, y_d]) = n_d\big] \\ = \prod_{j=1}^{d} \frac{\left(t_j b_j \nu_j\big((-\infty, y_j]\big)\right)^{n_j} \exp\big(-t_j b_j \nu_j\big((-\infty, y_j]\big)\big)}{n_j!}.\end{aligned}$$

We set $\mathbb{F}^{\boldsymbol{N}} = \left(\mathcal{F}_t^{\boldsymbol{N}}\right)_{t \geq 0}$ to be the natural filtration of $\left(N^j\right)_{j=1,\cdots,d}$. The expectation with respect to $\mathbb{P}$ is denoted by $\mathbb{E}[\cdot]$. For $t \geq 0$, we denote by $\mathbb{E}_t[\cdot]$ the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t^{\boldsymbol{N}}]$.

Remark 3. *It is possible to define all the Poisson measures on the same probability space Ω^d by taking*

$$\tilde{N}^j\left([0,t] \times [0,b] \times (-\infty, y]\right)(\omega) = N^j\left([0,t] \times [0,b] \times (-\infty, y]\right)(\omega^j), \quad (t,b,y) \in \mathbb{R}_+^3,$$

for any $\omega = \left(\omega^1, \cdots, \omega^d\right) \in \Omega$.
From now on we confound $(\tilde{N}^j)_{j=1,\cdots,d}$ with $(N^j)_{j=1,\cdots,d}$.

We now generalize the operators defined in [8] for the 1−dimensional setting. We start with the component-wise shift operator.

Definition 2.1 (Shift operator)**.** *Let $j \in [\![1, d]\!]$. We define for (t, θ, x) in $\mathbb{R}_+^3$ the measurable maps*

$$\begin{array}{rccl} \varepsilon^{j+}_{(t,\theta,x)}: & \Omega & \to \Omega \\ & (\omega^1, \cdots, \omega^j, \cdots, \omega^d) & \mapsto \left(\omega^1, \cdots, \varepsilon^{j+}_{(t,\theta,x)}(\omega^j), \cdots, \omega^d\right), \end{array}$$

where for any A in $\mathcal{B}(\mathbb{R}_+^2 \times \mathbb{R})$

$$\left(\varepsilon^{j+}_{(t,\theta,x)}(\omega^j)\right)(A) := \omega^j\left(A \setminus (t,\theta,x)\right) + \mathbb{1}_A(t,\theta,x),$$

with

$$\mathbb{1}_A(t,\theta,x) := \begin{cases} 1, & \text{if } (t,\theta,x) \in A, \\ 0, & \text{else.} \end{cases}$$

Definition 2.2 (Malliavin's derivative). *For F in $\mathcal{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, we define $D^j F$ the Malliavin's derivative of F as*

$$D^j_{(t,\theta,x)} F := F \circ \varepsilon^{j+}_{(t,\theta,x)} - F, \quad (t,\theta,x) \in \mathbb{R}^3_+.$$

If $\boldsymbol{F} = (F^1, \cdots, F^n)$ for some $n \geq 2$ where $F^i \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P}) \quad \forall i = 1, \cdots, n,$

$$D^j_{(t,\theta,x)} \boldsymbol{F} := (F^i \circ \varepsilon^{j+}_{(t,\theta,x)} - F^i)_{i=1,\cdots,n}, \quad (t,\theta,x) \in \mathbb{R}^3_+$$

Definition 2.3. *Let $\mathcal{I}$ be the sub-sigma field of $\mathcal{B}(\mathbb{R}^3_+) \otimes \mathcal{F}^N$ of stochastic processes $Z := (Z_{(t,\theta,x)})_{(t,\theta,x)\in\mathbb{R}^3_+}$ in $\mathcal{L}^1(\Omega \times \mathbb{R}^3_+, \mathbb{P} \otimes \mathrm{d}t \otimes \mathrm{d}\theta \otimes \nu)$ such that*

$$D^j_{(t,\theta,x)} Z_{(t,\theta,x)} = 0, \quad \text{for a.a. } (t,\theta,x) \in \mathbb{R}^3_+, \quad \text{for all } j = 1, \cdots, d.$$

Definition 2.4 (Divergence operator). *We set $\mathcal{S}$ the set of stochastic processes $Z := (Z_{(t,\theta,x)})_{(t,\theta,x)\in\mathbb{R}^3_+}$ in $\mathcal{I}$ such that for every $j \in [\![1, d]\!]$:*

$$\mathbb{E}\left[\int_{\mathbb{R}^3_+} \left|Z_{(t,\theta,x)}\right|^2 \mathrm{d}t\mathrm{d}\theta\nu^j(\mathrm{d}x)\right] + \mathbb{E}\left[\left(\int_{\mathbb{R}^3_+} Z_{(t,\theta,x)} N^j(\mathrm{d}t, \mathrm{d}\theta, \mathrm{d}x)\right)^2\right] < +\infty,$$

where $\int_{\mathbb{R}^3_+} Z_{(t,\theta,x)} N^j(\mathrm{d}t, \mathrm{d}\theta, \mathrm{d}x)$ is understood in the sense of the Stieltjes integral.

For Z in $\mathcal{S}$, we set the divergence operator with respect to N^j as

$$\delta^j(Z) := \int_{\mathbb{R}^3_+} Z_{(t,\theta,x)} N^j(\mathrm{d}t, \mathrm{d}\theta, \mathrm{d}x) - \int_{\mathbb{R}^3_+} Z_{(t,\theta,x)} \mathrm{d}t\mathrm{d}\theta\nu^j(\mathrm{d}x). \tag{3}$$

For a vector $\boldsymbol{Z} \in \mathcal{S}^d$, the divergence operator with respect to $\boldsymbol{N}$ is defined as

$$\delta^{\boldsymbol{N}}(\boldsymbol{Z}) = \sum_{i=1}^d \delta^i(Z^i),$$

and if $Z = (\boldsymbol{Z}^{\cdot 1}, \cdots, \boldsymbol{Z}^{\cdot n}) \in \mathcal{S}^{d\times n}$, is a matrix, its divergence is defined as

$$\begin{aligned}\delta^{\boldsymbol{N}}(Z) &= \left(\delta^{\boldsymbol{N}}(\boldsymbol{Z}^{\cdot 1}), \cdots, \delta^{\boldsymbol{N}}(\boldsymbol{Z}^{\cdot n})\right), \\ &= \sum_{i=1}^d \left(\delta^i(Z^{i1}), \cdots, \delta^i(Z^{in})\right).\end{aligned}$$

2.3 Definition of the multivariate compound Hawkes process

In this subsection we give a definition of the multivariate compound Hawkes process in the Markov framework. Consider the events times $\tau^j_1, \tau^j_2, \cdots$ associated with the $j-$th component and define the counting process

$$H^j_t = \sum_{i\geq 1} \mathbb{1}_{\tau^j_i \leq t}.$$

Now assume that each event τ^j_i corresponds to a random "loss" $Y^j_i \sim \nu^j$ such that the variables $(Y^j_i)_{i\geq 1}$ are independent and identically distributed *(i.i.d)*.
The compound process L^j_t of the total loss attributed to the $j-$th component is defined as

$$L^j_t = \sum_{i\geq 1} \mathbb{1}_{\tau^j_i \leq t} Y^j_i.$$

To $(H^j)_{j=1,\cdots,d}$ and $(L^j)_{j=1,\cdots,d}$ we associate a predictable intensity vector $(\lambda^j)_{j=1,\cdots,d}$ such that

$$\mathbb{P}\left[H^j_{t+\mathrm{d}t} - H^j_t = 1 | \mathcal{F}_{t-}\right] = \lambda^j_t \mathrm{d}t$$

which tells us how likely it is for H^j to jump between t and $t+\mathrm{d}t$, right before t.
The process $\boldsymbol{L}_t = (L_t^1, \cdots, L_t^d)_{t\geq 0}$ is called a multivariate compound Hawkes process if its intensity vector $\boldsymbol{\lambda}_t$ follows the dynamics

$$\begin{aligned}\lambda_t^j &= \mu^j + \int_{[0,t)} \sum_{k=1}^d \Phi_{jk}(t-s)\mathrm{d}L_s^k, \\ &= \mu^j + \sum_{k=1}^d \sum_{\tau_i^k < t} \Phi_{jk}(t-\tau_i^k)Y_i^k,\end{aligned}$$

where $(\mu^1, \cdots, \mu^d) \in \mathbb{R}_+^d$ and $\left(\Phi_{i,j}\right)_{i,j=1,\cdots,d}$ are non-negative integrable functions.

In this article, we restrict ourselves to the case where the kernels Φ are a family of exponential functions

$$\Phi_{jk}(u) = \alpha_{jk} e^{-\beta_j u}$$

where $(\beta_j)_{j=1,\cdots,d} \in (\mathbb{R}_+)^d$ and $A = (\alpha_{ij})_{i,j=1,\cdots,d} \in \mathcal{M}_d(\mathbb{R}_+)$. The intensity can be expressed under matrix form

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \int_{[0,t)} e^{-B(t-s)} A \mathrm{d}\boldsymbol{L}_s, \tag{4}$$

with $B = \mathrm{diag}(\beta_1, \cdots, \beta_d)$. Furthermore, we recall Assumptions 1, 2, 3.

We now introduce an equivalent definition of the multivariate compound Hawkes process, presented as the result of a stochastic differential equation *(SDE)* with respect to a family of random Poisson measures.

Theorem 2.1. *Let $\boldsymbol{N} = \left(N^1, \cdots, N^d\right)$ be a family of independent Poisson measures as presented in (Section 2.2). Let $\boldsymbol{\mu} \in \mathbb{R}_+^d$ and $A \in \mathcal{M}_d(\mathbb{R}_+)$ and $(\beta_1, \cdots, \beta_d) \in \mathbb{R}_+^d$ such that the Assumption 1 is verified. The SDE below admits a unique solution $(\boldsymbol{L}, \boldsymbol{H}, \boldsymbol{\lambda})$ with $\boldsymbol{H}$ and $\boldsymbol{L}$ (resp. $\boldsymbol{\lambda}$) $\mathbb{F}^{\boldsymbol{N}}$-adapted (resp. $\mathbb{F}^{\boldsymbol{N}}$-predictable)*

$$\begin{cases} L_t^j = \displaystyle\int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} x \mathbb{1}_{\{\theta \leq \lambda_s^j\}} N^j(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}x), \quad t \geq 0, \quad j = 1, \cdots, d, \\ H_t^j = \displaystyle\int_{(0,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbb{1}_{\{\theta \leq \lambda_s^j\}} N^j(\mathrm{d}s, \mathrm{d}\theta, \mathrm{d}x), \quad t \geq 0, \quad j = 1, \cdots, d, \\ \lambda_t^j = \mu^j + \int_{(0,t)} \sum_{k=1}^d \alpha_{jk} e^{-\beta_j(t-u)} \mathrm{d}L_u^k, \quad t \geq 0 \quad j = 1, \cdots, d. \end{cases} \tag{5}$$

We set $\mathbb{F}^{\boldsymbol{H}} := (\mathcal{F}_t^{\boldsymbol{H}})_{t\geq 0}$ (respectively $\mathbb{F}^{\boldsymbol{L}} := (\mathcal{F}_t^{\boldsymbol{L}})_{t\geq 0}$) the natural filtration of $\boldsymbol{H}$ (respectively of $\boldsymbol{L}$) and $\mathcal{F}_\infty^{\boldsymbol{H}} := \lim_{t\to+\infty} \mathcal{F}_t^{\boldsymbol{H}}$ (respectively $\mathcal{F}_\infty^{\boldsymbol{L}} := \lim_{t\to+\infty} \mathcal{F}_t^{\boldsymbol{L}}$). Obviously $\mathcal{F}_t^{\boldsymbol{H}} \subset \mathcal{F}_t^{\boldsymbol{L}} \subset \mathcal{F}_t^{\boldsymbol{N}}$ as $\boldsymbol{H}$ is completely determined by the jump times of $\boldsymbol{H}$ which are exactly those of $\boldsymbol{L}$.

Proof. This is merely a $d-$dimensional version of Theorem 2.12 in [8]. □

Definition 2.5. *Let $\boldsymbol{L}$ and $\boldsymbol{\lambda}$ be the processes defined in Theorem 2.1. The multivariate Hawkes martingale $\boldsymbol{M}$ is the process defined as*

$$\boldsymbol{M}_T = \boldsymbol{L}_T - \mathrm{diag}\left(m^1, \cdots, m^d\right) \int_0^T \boldsymbol{\lambda}_t \mathrm{d}t$$

for any $T \geq 0$.

2.4 Malliavin's analysis of the multivariate compound Hawkes process

This subsection is a generalization of Section 2.3 of [8] where we examine the impact of the $i-$th Malliavin's derivative on the $j-$th component of $\boldsymbol{L}$, $\boldsymbol{H}$ and $\boldsymbol{\lambda}$.

Definition and Proposition 2.1 (Lemma 2.15 in [8]). *Let t and v in $\mathbb{R}_+$ and (θ, θ_0, x) in $\mathbb{R}_+^3$. For every $j \in [\![1, d]\!]$ it holds that :*

$$\begin{aligned}&\mathbb{1}_{\{\theta \le \lambda_t^j\}}(\boldsymbol{L}_v \circ \varepsilon^{j+}_{(t,\theta,x)}, \boldsymbol{\lambda}_v \circ \varepsilon^{j+}_{(t,\theta,x)})_{v \ge 0}\\ &= \mathbb{1}_{\{\theta_0 \le \lambda_t^j\}}(\boldsymbol{L}_v \circ \varepsilon^{j+}_{(t,\theta_0,x)}, \boldsymbol{\lambda}_v \circ \varepsilon^{j+}_{(t,\theta_0,x)})_{v \ge 0}.\end{aligned}$$

which entails that for every -eventually vectorial- $\mathcal{F}_\infty^{\boldsymbol{L}}$-measurable random variable $\boldsymbol{F}$

$$\mathbb{1}_{\{\theta \le \lambda_t^j\}} D^j_{(t,\theta,x)} \boldsymbol{F} = \mathbb{1}_{\{\theta_0 \le \lambda_t^j\}} D^j_{(t,\theta_0,x)} \boldsymbol{F}, \quad \mathbb{P} - a.s.$$

where the derivative is applied to each component of $\boldsymbol{F}$. The last equation allows us to define

$$D^j_{(t,\lambda_t^j,x)} \boldsymbol{F} := \mathbb{1}_{\{\theta \le \lambda_t^j\}} D^j_{(t,\theta,x)} \boldsymbol{F}, \quad \forall (\theta, x) \in \mathbb{R}_+^2$$

as well as its vector version

$$\begin{aligned}D_{(t,\boldsymbol{\lambda}_t,x)} \boldsymbol{F} &= \left(D^j_{(t,\lambda_t^j,x)} \boldsymbol{F}\right)_{j=1,\cdots,d},\\ &:= \left(\mathbb{1}_{\{\theta \le \lambda_t^j\}} D^j_{(t,\theta,x)} \boldsymbol{F}\right)_{j=1,\cdots,d}, \quad \forall (\theta, x) \in \mathbb{R}_+^2.\end{aligned}$$

Proof. For each $i \in [\![1, d]\!]$ and $v \ge t$, we have conditionally on $\theta_0 \le \lambda_t^j$

$$L^i_v \circ \varepsilon^{+j}_{(t,\theta_0,x)} = L^i_{t-} + x \mathbb{1}_{\{\theta_0 \le \lambda_t^j\}} \mathbb{1}_{i=j} + \int_{(t,v] \times \mathbb{R}} \int_{\mathbb{R}_+} y \mathbb{1}_{\{\theta \le \lambda^i_u \circ \varepsilon^{+j}_{(t,\theta_0,x)}\}} N^i(\mathrm{d}\theta, \mathrm{d}u, \mathrm{d}y),$$

and $L^i_v \circ \varepsilon^{+j}_{(t,\theta_0,x)} = L^i_v$ if $t > v$. Similarly for $\boldsymbol{\lambda}$, we have

$$\begin{aligned}\lambda^i_v \circ \varepsilon^{+j}_{(t,\theta_0,x)} &= \left(\mu_i + \int_{(0,t)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} \mathrm{d}L^k_u + \int_{[t,v)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} \mathrm{d}L^k_u\right) \circ \varepsilon^{+j}_{(t,\theta_0,x)}\\ &= \mu_i + \int_{(0,t)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} \mathrm{d}L^k_u + x\alpha_{ij} e^{-\beta(v-t)} + \int_{(t,v)} \sum_{k=1}^d e^{-\beta(v-u)} \alpha_{ik} \mathrm{d}(L^k_u \circ \varepsilon^{+j}_{(t,\theta_0,x)}).\end{aligned}$$

In other words, $(\boldsymbol{L} \circ \varepsilon^{+j}_{(t,\theta_0,x)}, \boldsymbol{\lambda} \circ \varepsilon^{+j}_{(t,\theta_0,x)})$ solves the same (path-wise and in the SDE sense) equation for any θ_0 such that $\theta_0 \le \lambda_t$.
For the second equality, *cf.* the proof of Proposition 2.16 in [8]. □

Remark 4. *It is also possible to include the process $\boldsymbol{H}$ in these results. From now on, $\boldsymbol{H}$ will be omitted and we will focus exclusively on $\boldsymbol{L}$ and $\boldsymbol{\lambda}$.*

Now we give the Malliavin's derivative of the multivariate compound Hawkes process as well as its intensity. To do so we start with introducing some notations. The vector $\boldsymbol{e}_i$ is the element of $\mathbb{R}^d$ that has 1 in the $i-$th component and zero elsewhere.

Proposition 2.1. *Let $t \ge 0$ and $x \in \mathbb{R}$. For every $j \in [\![1, d]\!]$, we have*

$$(D^j_{(t,\lambda_t^j,x)} \boldsymbol{L}_s, D^j_{(t,\lambda_t^j,x)} \boldsymbol{\lambda}_s) = \begin{cases} (x\boldsymbol{e}_j + \hat{\boldsymbol{L}}^{j,t,x}_s, \hat{\boldsymbol{\lambda}}^{j,t,x}_s), & s \ge t,\\ (0,0), & s < t \end{cases}$$

where the equality is understood path-wise and in the SDE sense and where $(\hat{\boldsymbol{L}}^{j,t,x}_s, \hat{\boldsymbol{\lambda}}^{j,t,x}_s)_{s \ge t} = \left((\hat{L}^{i,j,t,x}_s, \hat{\lambda}^{i,j,t,x}_s)_{s \ge t}\right)_{i=1,\cdots,d}$ is the unique solution to the SDE

$$\begin{cases} \hat{L}^{i,j,t,x}_s = \displaystyle\int_{(t,s] \times \mathbb{R}_+^2} y \mathbb{1}_{\{\lambda^i_u \le \theta \le \lambda^i_u + \hat{\lambda}^{i,j,t,x}_u\}} N^i(\mathrm{d}u, \mathrm{d}\theta, \mathrm{d}y), \quad s \ge t,\\ \hat{\lambda}^{i,j,t,x}_s = x\alpha_{ij} e^{-\beta_i(s-t)} + \displaystyle\sum_{k=1}^d \int_{(t,s)} \alpha_{ik} e^{-\beta_i(s-u)} \mathrm{d}\hat{L}^{k,j,t}_u, \quad s > t,\ \hat{\lambda}^{i,j,t,x}_t = x\alpha_{ij}, \end{cases} \tag{6}$$

or under the matrix form

$$\begin{cases} \hat{\boldsymbol{L}}_s^{j,t,x} = \displaystyle\int_{(t,s]\times\mathbb{R}_+^2} y\mathbf{1}_{\{\boldsymbol{\lambda}_u \le \theta \le \boldsymbol{\lambda}_u + \hat{\boldsymbol{\lambda}}_u^{j,t,x}\}} \boldsymbol{N}(\mathrm{d}u, \mathrm{d}\theta, \mathrm{d}y), \quad s \ge t, \\ \hat{\boldsymbol{\lambda}}_s^{j,t,x} = xe^{-B(s-t)}\boldsymbol{A}_{\cdot j} + \displaystyle\int_{(t,s)} e^{-B(s-u)} A \mathrm{d}\hat{\boldsymbol{L}}_u^{j,t,x}, \quad s > t,\ \hat{\boldsymbol{\lambda}}_t^{j,t,x} = x\boldsymbol{A}_{\cdot j}, \end{cases} \tag{7}$$

where the first integral is understood as a vector of integrals with respect to each Poisson measure and the indicator of a vector is understood as a vector of indicators.
The process $(\hat{\boldsymbol{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x})_{s\in[t,+\infty)}$ *is a generalized multivariate compound Hawkes process, with an intensity vector that is not bounded away from* 0*.*

Proof. It is sufficient to apply the same procedure of Proposition 2.19 's proof in [8] to each component of the vectors. Note that in this proposition, the subscript contains x, because in our case the size of the jump has an impact on the $\hat{\boldsymbol{\lambda}}$'s behaviour, since it is an integral with respect to $\mathrm{d}\hat{\boldsymbol{L}}$ and not merely $\mathrm{d}\hat{\boldsymbol{H}}$. □

Remark 5. *Note that the process* $\left(\hat{\boldsymbol{L}}_s^{j,t,x}, \hat{\boldsymbol{\lambda}}_s^{j,t,x}\right)_{s\ge t}$ *defined above is equal in distribution to the (generalized) Hawkes process* $\left(\tilde{\boldsymbol{L}}_s^{j,t,x}, \tilde{\boldsymbol{\lambda}}_s^{j,t,x}\right)_{s\ge t}$ *defined as a solution to the SDE*

$$\begin{cases} \tilde{\boldsymbol{L}}_s^{j,t,x} = \displaystyle\int_{(t,s]\times\mathbb{R}_+^2} y\mathbf{1}_{\{\mathbf{0} \le \theta \le \tilde{\boldsymbol{\lambda}}_u^{j,t,x}\}} \tilde{\boldsymbol{N}}(\mathrm{d}u, \mathrm{d}\theta, \mathrm{d}y), \quad s \ge t, \\ \tilde{\boldsymbol{\lambda}}_s^{j,t,x} = xe^{-B(s-t)}\boldsymbol{A}_{\cdot j} + \displaystyle\int_{(t,s)} e^{-B(s-u)} A \mathrm{d}\tilde{\boldsymbol{L}}_u^{j,t,x}, \quad s > t,\ \tilde{\boldsymbol{\lambda}}_t^{j,t,x} = x\boldsymbol{A}_{\cdot j}, \end{cases} \tag{8}$$

where $\tilde{\boldsymbol{N}}$ *is a family of Poisson measures independent from* $\boldsymbol{N}$ *but have the same distributions.*

We conclude this section by stating the multivariate integration by parts (IBP) formula for the compound Hawkes martingale.

2.5 Multivariate Stein's method

Stein's method is based on an alternative characterization of the Gaussian distribution. The combination of this characterization with elements of Malliavin's calculus (known as the Nourdin-Peccati approach) provides us with a way to estimate the distance between a random variable and a Gaussian.

Lemma 2.2. *Let* C *be a* $d\times d$ *real symmetric positive definite matrix and* $\boldsymbol{Y}$ *an* $\mathbb{R}^d$ *random variable. Then* $\boldsymbol{Y} \sim \mathcal{N}(0, C)$ *if and only if for every twice differentiable function* $f : \mathbb{R}^d \to \mathbb{R}$ *such that* $\mathbb{E}\big[|\langle C, \mathrm{Hess}\, f(\boldsymbol{Y})\rangle_{H.S}| + |\langle \boldsymbol{Y}, \nabla f(\boldsymbol{Y})\rangle|\big] < +\infty$*:*

$$\mathbb{E}\big[\langle C, \mathrm{Hess}\, f(\boldsymbol{Y})\rangle_{H.S} - \langle \boldsymbol{Y}, \nabla f(\boldsymbol{Y})\rangle\big] = 0$$

where $\langle A, B\rangle_{H.S} = tr(AB^T) = \sum A_{ij}B_{ij}$ *is the Hilbert-Schmidt matrix inner product and* Hess *is the Hessian operator.*

Proof. Cf. [8] for example. □

This shows that if for a sufficiently large class of functions f one has

$$\mathbb{E}\big[\langle C, \mathrm{Hess}\, f(\boldsymbol{Y})\rangle_{H.S} - \langle \boldsymbol{Y}, \nabla f(\boldsymbol{Y})\rangle\big] \simeq 0$$

in some sense, then the random variable $\boldsymbol{Y}$'s distribution is fairly "close" to $\mathcal{N}(0, C)$.
We now introduce the function classes as well as the metric that we will use to quantify how close the given distance of a random variable to a Gaussian distribution.

Definition 2.6. *1. The distance* d_2 *between two integrable random variables* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *is given by*

$$d_2(\boldsymbol{X}, \boldsymbol{Y}) := \sup_{f\in\mathcal{H}} |\mathbb{E}[f(\boldsymbol{X})] - \mathbb{E}[f(\boldsymbol{Y})]|,$$

where

$$\mathcal{H} := \{g \in \mathcal{C}^2(\mathbb{R}^d),\ \textit{such that}\ \|g\|_{Lip} \le 1\ \textit{and}\ M_2(g) \le 1\}.$$

2. For a fixed matrix $C \in \mathcal{S}_d^{++}(\mathbb{R})$ we define $\mathcal{F}_C$ to be the functional space

$$\mathcal{F}_C := \{g \in \mathcal{C}^2(\mathbb{R}^d),\ \textit{such that } M_2(g) \leq \|C^{-1}\|_{op}\|C\|_{op}^{1/2}\ ,\ M_3(g) \leq \frac{\sqrt{2\pi}}{4}\|C^{-1}\|_{op}^{3/2}\|C\|_{op}\}.$$

Combining the definitions with the multivariate Stein's equation, it is possible to prove (following the lines of Lemma 2.17 in [13]) the following estimate of the d_2 distance between a centered variable $\boldsymbol{F}$ and $\boldsymbol{G} \sim \mathcal{N}_d(0, C)$:

$$d_2(\boldsymbol{F}, \boldsymbol{G}) \leq \sup_{f \in \mathcal{F}_C} \left|\mathbb{E}\big[\langle C, \operatorname{Hess} f(\boldsymbol{F})\rangle_{H.S} - \langle \boldsymbol{F}, \nabla f(\boldsymbol{F})\rangle\big]\right|.$$

Remark 6. *Unlike for the univariate Poisson space case (cf. [8]) or the multivariate variables on a Gaussian space [10], the Wasserstein distance d_W is not well adapted to our computations. Nevertheless, if $d_2(\boldsymbol{F}_T, \boldsymbol{G}) \longrightarrow 0$ as T goes to infinity implies that $\boldsymbol{F}_T$ converges to the Gaussian $\boldsymbol{G}$ in distribution (cf. Remark 2.16 in [13]) .*

3 Main results

3.1 General bound

In this section, we give a bound on the d_2 distance between any random variable that can be expressed as a divergence with respect to the Poisson measure.

Lemma 3.1. *Let n be an integer and let $\boldsymbol{F} = (F^1, \cdots, F^n)$ be an $\mathcal{F}_\infty^{\boldsymbol{L}}$ such that $\mathbb{E}[\|\boldsymbol{F}\|^2] \leq +\infty$. For all $\phi \in \mathcal{C}^3(\mathbb{R}^n)$ with bounded derivatives, for any $k = 1, \cdots, n$, there exists a random $\bar{\boldsymbol{F}}$ such that*

$$D^k_{(t,\theta,x)}\partial_i\phi(\boldsymbol{F}) = \langle \nabla\partial_i\phi(\boldsymbol{F}), D^k_{(t,\theta,x)}\boldsymbol{F}\rangle + \frac{1}{2}\langle D^k_{(t,\theta,x)}\boldsymbol{F}, \operatorname{Hess}\partial_i\phi(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle,\ (t,\theta,x) \in \mathbb{R}^3_+$$

where

$$\left|\langle \boldsymbol{y}, \operatorname{Hess}\partial_k\phi(\bar{\boldsymbol{F}})\boldsymbol{y}\rangle\right| \leq \|\boldsymbol{y}\|^2 M_3(\phi),\ \boldsymbol{y} \in \mathbb{R}^d$$

.

Proof. The equality is merely an application of the multivariate Taylor-Young Theorem, combined with the fact that

$$\begin{aligned} D^k_{(t,\theta,x)}\partial_i\phi(\boldsymbol{F}) &= \partial_i\phi(\boldsymbol{F} \circ \varepsilon^{k+}_{(t,\theta,x)}) - \partial_i\phi(\boldsymbol{F}), \\ &= \partial_i\phi(\boldsymbol{F} + D^k_{(t,\theta,x)}\boldsymbol{F}) - \partial_i\phi(\boldsymbol{F}). \end{aligned}$$

When it comes to the upper bound on the rest, we have using Cauchy-Schwarz

$$\begin{aligned} \left|\langle \boldsymbol{y}, \operatorname{Hess}\partial_i\phi(\bar{\boldsymbol{F}})\boldsymbol{y}\rangle\right| &\leq \|\boldsymbol{y}\|\|\operatorname{Hess}\partial_i\phi(\bar{\boldsymbol{F}})\boldsymbol{y}\|, \\ &\leq \|\boldsymbol{y}\|^2\|\operatorname{Hess}\partial_i\phi(\bar{\boldsymbol{F}})\|_{op}. \end{aligned}$$

Using Schwarz's Theorem we have that

$$\begin{aligned} \operatorname{Hess}\partial_i\phi(\bar{\boldsymbol{F}}) &= \partial_i \operatorname{Hess}\phi(\bar{\boldsymbol{F}}), \\ &= \lim_{h\to 0} \frac{\operatorname{Hess}\phi(\bar{\boldsymbol{F}} + h\boldsymbol{e}_i) - \operatorname{Hess}\phi(\bar{\boldsymbol{F}})}{h}, \end{aligned}$$

and the result follows using the fact that the norm $\|\cdot\|_{op}$ is continuous. □

This lemma will be useful in proving the following result.

Theorem 3.2. *Let $n \in \mathbb{N}^*$. Let $Z = \left(Z^{ki}_{(t,\theta,x)}\right)_{(t,\theta,x)\in\mathbb{R}^3_+}$ be a stochastic process in $\mathcal{S}^{d\times n}$. Set $\boldsymbol{F} = \delta^{\boldsymbol{N}}(Z)$ the divergence of Z. Then, letting $\boldsymbol{G} \sim \mathcal{N}(0, C)$ (for $C \in \mathcal{S}_n^{++}(\mathbb{R})$) we have*

$$\begin{aligned} d_2(\boldsymbol{F}, \boldsymbol{G}) \leq & \|C^{-1}\|_{op}\|C\|_{op}^{1/2}\sum_{i,j=1}^n \mathbb{E}\left[\left|C_{ij} - \sum_{k=1}^d \int_{\mathbb{R}^3_+} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j \mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right|\right] \\ & + \frac{\sqrt{2\pi}}{8}\|C^{-1}\|_{op}^{3/2}\|C\|_{op}\sum_{i,j=1}^n\sum_{k=1}^d \mathbb{E}\left[\int_{\mathbb{R}^3_+}\left|Z^{ki}_{(t,\theta,x)}\right|\left\|D^k_{(t,\theta,x)}\boldsymbol{F}\right\|^2 \mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right]. \end{aligned}$$

If in particular $Z^{ki}_{(t,\theta,x)} = \mathbb{1}_{\theta\leq\lambda^k_t}U^{ki}_{(t,x)}$*, the upper bound takes the form*

$$d_2(\boldsymbol{F},\boldsymbol{G}) \leq \|C^{-1}\|_{op}\|C\|^{1/2}_{op}\sum_{i,j=1}^{n}\mathbb{E}\left[\left|C_{ij} - \sum_{k=1}^{d}\int_{\mathbb{R}^2_+}\lambda^k_t U^{ki}_{(t,x)}D^k_{(t,\lambda^k_t,x)}F^j\mathrm{d}t\nu^k(\mathrm{d}x)\right|\right]$$
$$+\frac{\sqrt{2\pi}}{8}\|C^{-1}\|^{3/2}_{op}\|C\|_{op}\sum_{i,j=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}^2_+}\lambda^k_t\left|U^{ki}_{(t,x)}\right|\left\|D^k_{(t,\lambda^k_t,x)}\boldsymbol{F}\right\|^2\mathrm{d}t\nu^k(\mathrm{d}x)\right].$$

Proof. We start by recalling the bound on the d_2 distance between $\boldsymbol{F}$ and $\boldsymbol{G}$

$$d_2(\boldsymbol{F},\boldsymbol{G}) \leq \sup_{f\in\mathcal{F}_C}\left|\mathbb{E}\big[\langle C,\mathrm{Hess}\, f(\boldsymbol{F}_T)\rangle_{H.S} - \langle\boldsymbol{F},\nabla f(\boldsymbol{F})\rangle\big]\right|.$$

In [10], the following technique is used. For any $f\in\mathcal{F}_C$ and $\eta>0$ set $f_\eta(x) = \mathbb{E}[f(x+\sqrt{\eta}N)]$ where N is a centered Gaussian of unit variance. It is easy to see that

1. $f_\eta\in\mathcal{C}^\infty(\mathbb{R}^n)$.
2. $\|f-f_\eta\|_\infty\longrightarrow 0$ when η goes to zero.
3. $M_2(f_\eta)\leq M_2(f)$ and $M_3(f_\eta)\leq M_3(f)$ using Young's convolution inequality.

Thus we can assume that $f\in\mathcal{C}^\infty(\mathbb{R}^n)$ and the computations yield

$$\begin{aligned}\mathbb{E}\big[\langle C,\mathrm{Hess}\, f(\boldsymbol{F})\rangle_{H.S} - \langle\boldsymbol{F},\nabla f(\boldsymbol{F})\rangle\big] &= \mathbb{E}\big[\sum_{i,j=1}^{n}C_{ij}\partial^2_{ij}f(\boldsymbol{F}) - \sum_{i=1}^{n}F^i\partial_i f(\boldsymbol{F})\big],\\ &= \sum_{i,j=1}^{n}C_{ij}\mathbb{E}\big[\partial^2_{ij}f(\boldsymbol{F})\big] - \sum_{i=1}^{n}\mathbb{E}\big[F^i\partial_i f(\boldsymbol{F})\big].\end{aligned}$$

By definition, $\boldsymbol{F}$ is defined as the divergence of the matrix Z, thus for each $i\in[\![1,n]\!]$

$$F^i = \sum_{k=1}^{d}\delta^k(Z^{ki}),$$

which entails that

$$\mathbb{E}\big[F^i\partial_i f(\boldsymbol{F})\big] = \sum_{k=1}^{d}\mathbb{E}\big[\delta^k(Z^{ki})\partial_i f(\boldsymbol{F})\big].$$

Set $\mathbb{E}_{\neq k}[\cdot] = \mathbb{E}[\cdot|N^1,\cdots,N^{k-1},N^{k+1},\cdots,N^d]$, which stands for the expected value knowing all the counting measures except for the $k-$th one. This notation is introduced in order use the integration by parts formula in [14] which is available for univariate processes. Hence

$$\begin{aligned}\mathbb{E}\big[F^i\partial_i f(\boldsymbol{F})\big] &= \sum_{k=1}^{d}\mathbb{E}\big[\mathbb{E}_{\neq k}[\delta^k(Z^{ki})\partial_i f(\boldsymbol{F})]\big],\\ &= \sum_{k=1}^{d}\mathbb{E}\big[\mathbb{E}_{\neq k}\left[\int\int\int_{\mathbb{R}^3_+}Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}\partial_i f(\boldsymbol{F})\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right]\big],\\ &= \sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}^3_+}Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}\partial_i f(\boldsymbol{F})\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right].\end{aligned}$$

Using Lemma 3.1, we have that

$$\begin{aligned}D^k_{(t,\theta,x)}\partial_i f(\boldsymbol{F}) &= \langle\nabla\partial_i f(\boldsymbol{F}), D^k_{(t,\theta,x)}\boldsymbol{F}\rangle + \frac{1}{2}\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle,\\ &= \sum_{j=1}^{n}\partial^2_{ij}f(\boldsymbol{F})D^k_{(t,\theta,x)}F^j + \frac{1}{2}\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle,\end{aligned}$$

for some random $\bar{\boldsymbol{F}}$. Hence

$$\begin{aligned}\mathbb{E}\big[F^i\partial_i f(\boldsymbol{F})\big] =& \sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}\sum_{j=1}^{n}\partial^2_{ij}f(\boldsymbol{F})D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right]\\ &+\sum_{k=1}^{d}\frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right],\\ =&\sum_{j=1}^{n}\mathbb{E}\left[\partial^2_{ij}f(\boldsymbol{F})\sum_{k=1}^{d}\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right]\\ &+\sum_{k=1}^{d}\frac{1}{2}\mathbb{E}\left[\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right].\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}\big[\langle C,\mathrm{Hess}\,f(\boldsymbol{F})\rangle_{H.S}-\langle\boldsymbol{F},\nabla f(\boldsymbol{F})\rangle\big] =& \sum_{i,j=1}^{n}\mathbb{E}\left[\partial^2_{ij}f(\boldsymbol{F})\left(C_{ij}-\sum_{k=1}^{d}\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right)\right]\\ &-\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right].\end{aligned}$$

By taking the absolute value, using the triangular inequality and the fact that for any $f\in\mathcal{F}_C$ and any $(i,j)\in[\![1,d]\!]^2$

$$\left\|\partial^2_{ij}f\right\|_\infty \le M_2(f)\le\|C^{-1}\|_{op}\|C\|^{1/2}_{op}$$

we obtain the bound

$$\begin{aligned}|\mathbb{E}\big[\langle C,\mathrm{Hess}\,f(\boldsymbol{F})\rangle_{H.S}-\langle\boldsymbol{F},\nabla f(\boldsymbol{F})\rangle\big]| \le&\|C^{-1}\|_{op}\|C\|^{1/2}_{op}\sum_{i,j=1}^{n}\mathbb{E}\left[\left|C_{ij}-\sum_{k=1}^{d}\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right|\right]\\ &+\frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^3}|Z^{ki}_{(t,\theta,x)}|\left|\langle D^k_{(t,\theta,x)}\boldsymbol{F},\mathrm{Hess}\,\partial_i f(\bar{\boldsymbol{F}})D^k_{(t,\theta,x)}\boldsymbol{F}\rangle\right|\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right],\\ \le&\|C^{-1}\|_{op}\|C\|^{1/2}_{op}\sum_{i=1}^{n}\mathbb{E}\left[\left|C_{ij}-\sum_{k=1}^{d}\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right|\right]\\ &+\frac{\sqrt{2\pi}}{8}\|C^{-1}\|^{3/2}_{op}\|C\|_{op}\sum_{i=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^3}|Z^{ki}_{(t,\theta,x)}|\left\|D^k_{(t,\theta,x)}\boldsymbol{F}\right\|^2\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right],\end{aligned}$$

and finally

$$\begin{aligned}d_2(\boldsymbol{F},\boldsymbol{G}) \le&\|C^{-1}\|_{op}\|C\|^{1/2}_{op}\sum_{i,j=1}^{n}\mathbb{E}\left[\left|C_{ij}-\sum_{k=1}^{d}\int_{\mathbb{R}_+^3} Z^{ki}_{(t,\theta,x)}D^k_{(t,\theta,x)}F^j\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right|\right]\\ &+\frac{\sqrt{2\pi}}{8}\|C^{-1}\|^{3/2}_{op}\|C\|_{op}\sum_{i=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^3}|Z^{ki}_{(t,\theta,x)}|\left\|D^k_{(t,\theta,x)}\boldsymbol{F}\right\|^2\mathrm{d}t\mathrm{d}\theta\nu^k(\mathrm{d}x)\right].\end{aligned}$$

For the second point, assume that $\forall k\in[\![1,d]\!]$ and $\forall i\in[\![1,n]\!]$

$$Z^{ki}_{(t,\theta,x)} = \mathbb{1}_{\theta\le\lambda^k_t}U^{ki}_{(t,x)}.$$

Keeping in mind that according to Definition 2.1, we have

$$\mathbb{1}_{\theta\le\lambda^k_t}D^k_{(t,\theta,x)}F^j = D^k_{(t,\lambda^k_t,x)}F^j$$

whenever θ is less than λ_t^k, thus

$$d_2(\boldsymbol{F},\boldsymbol{G}) \leq \|C^{-1}\|_{op}\|C\|_{op}^{1/2}\sum_{i,j=1}^{n}\mathbb{E}\left[\left|C_{ij}-\sum_{k=1}^{d}\int_{\mathbb{R}_+^2}\lambda_t^k U_{(t,x)}^{ki}D_{(t,\lambda_t^k,x)}^k F^j \mathrm{d}t\nu^k(\mathrm{d}x)\right|\right]$$
$$+\frac{\sqrt{2\pi}}{8}\|C^{-1}\|_{op}^{3/2}\|C\|_{op}\sum_{i=1}^{n}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^2}\lambda_t^k|U_{(t,x)}^{ki}|\left\|D_{(t,\lambda_t^k,x)}^k\boldsymbol{F}\right\|^2\mathrm{d}t\nu^k(\mathrm{d}x)\right].$$

□

3.2 Bounds on the CLTs

The following theorem is our first main result where we give the bound on the d_2 distance between a vector formed by the normalized Hawkes martingale

$$\boldsymbol{F}_T=\frac{\boldsymbol{M}_T}{\sqrt{T}}=\frac{\boldsymbol{L}_T-\mathrm{diag}(m^1,\cdots,m^d)\int_0^T\boldsymbol{\lambda}_t\mathrm{d}t}{\sqrt{T}}$$

and its Gaussian limit as T goes to infinity.

Theorem 3.3. *Fix p and d in $\mathbb{N}^*$. Let $(\boldsymbol{L}_t)_{t\geq 0}$ be a compound multivariate Hawkes process whose intensity $\boldsymbol{\lambda}$ follows the dynamics 4 and let $0<v_1<\cdots<v_p\leq 1$ be p distinct positive numbers. Set*

$$\boldsymbol{\Gamma}_T=\left(F_{v_1T}^1,\cdots,F_{v_pT}^1,\cdots,F_{v_1T}^d,\cdots,F_{v_pT}^d\right)\in\mathbb{R}^{p\cdot d},$$

and

$$\hat{C}=\mathrm{diag}(C^1,\cdots,C^d)\in\mathcal{S}_{p\cdot d}(\mathbb{R})$$

the block diagonal matrix such that $\forall n\in[\![1,d]\!]$

$$C_{ij}^n=C_{ji}^n=\int x^2\nu^n(\mathrm{d}x)\sqrt{\frac{v_i}{v_j}}\left[\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}\right]^n,\quad\forall 1\leq i\leq j\leq p.$$

Let $\boldsymbol{G}\sim\mathcal{N}(0,\hat{C})$. Then there is a constant $K>0$ independent from T such that

$$d_2(\boldsymbol{\Gamma}_T,\boldsymbol{G})\leq\frac{K}{\sqrt{T}}.$$

Proof. Throughout this proof, K is a positive constant that does not depend on T and that is susceptible to change from one line to the other. We also set $v_0=v_p$.
For each $i\in[\![1,p\cdot d]\!]$ and $k\in[\![1,d]\!]$, we define the matrix process

$$Z_{(t,\theta,x)}^{k,i}=\mathbb{1}_{\theta\leq\lambda_t^k}\mathbb{1}_{k=i\div p+1}\mathbb{1}_{t\leq v_{i\%p}T}\frac{x}{\sqrt{v_{i\%p}T}},$$

where $i\%p$ (respectively $i\div p$) is the remainder (respectively the quotient) after dividing i by p. It is possible to write the set of integers from 1 to $p\cdot d$ as a partition of d disjoint intervals $S_1\cup\cdots\cup S_d$ each of cardinal p. Thus for any $n=1,\cdots,d$, $n=i\div p+1$ if and only if $i\in S_n$.
We assume $i\in S_n$ and compute the $i-$th component of the divergence of Z

$$\begin{aligned}(\delta(Z))^i&=\sum_{k=1}^{d}\delta^k(Z^{ik}),\\&=\delta^n(Z^{ni}),\\&=\int_{\mathbb{R}_+^3}Z_{(t,\theta,x)}^{ni}\left(N^n(\mathrm{d}t,\mathrm{d}\theta,\mathrm{d}x)-\mathrm{d}t\mathrm{d}\theta\nu^n(\mathrm{d}x)\right),\\&=\frac{1}{\sqrt{v_{i\%p}T}}\int_0^{v_{i\%p}T}\int_{\mathbb{R}_+^2}x\mathbb{1}_{\theta\leq\lambda_t^n}\left(N^n(\mathrm{d}t,\mathrm{d}\theta,\mathrm{d}x)-\mathrm{d}t\mathrm{d}\theta\nu^n(\mathrm{d}x)\right),\\&=\Gamma_T^i.\end{aligned}$$

Using the second equality of Theorem 3.2, the d_2 distance is bounded by

$$
\begin{aligned}
d_2(\mathbf{\Gamma}_T, \boldsymbol{G}) \leq & \|\hat{C}^{-1}\|_{op}\|\hat{C}\|_{op}^{1/2} \sum_{i,j=1}^{pd} \mathbb{E}\left[\left|\hat{C}_{ij} - \sum_{k=1}^{d}\int_{\mathbb{R}_+^2} \lambda_t^k \mathbb{1}_{t\leq v_{i\%p}T}\mathbb{1}_{k=i\div p+1}\frac{x}{\sqrt{v_{i\%p}T}} D^k_{(t,\lambda_t^k,x)}\Gamma_T^j \mathrm{d}t\nu^k(\mathrm{d}x)\right|\right] \\
& + \frac{\sqrt{2\pi}}{8}\|\hat{C}^{-1}\|_{op}^{3/2}\|\hat{C}\|_{op} \sum_{i=1}^{pd}\sum_{k=1}^{d}\mathbb{E}\left[\int_{\mathbb{R}_+^2} \lambda_t^k \mathbb{1}_{t\leq v_{i\%p}T}\mathbb{1}_{k=i\div p+1}\frac{x}{\sqrt{v_{i\%p}T}} \left\|D^k_{(t,\lambda_t^k,x)}\mathbf{\Gamma}_T\right\|^2 \mathrm{d}t\nu^k(\mathrm{d}x)\right], \\
\leq & \|\hat{C}^{-1}\|_{op}\|\hat{C}\|_{op}^{1/2} \sum_{n_1,n_2=1}^{d}\sum_{i\in S_{n_1}}\sum_{j\in S_{n_2}} \mathbb{E}\left[\left|\hat{C}_{ij} - \int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right] \\
& + \frac{\sqrt{2\pi}}{8}\|\hat{C}^{-1}\|_{op}^{3/2}\|\hat{C}\|_{op} \sum_{n_1=1}^{d}\sum_{i\in S_{n_1}}\mathbb{E}\left[\int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} \left\|D^{n_1}_{(t,\lambda_t^{n_1},x)}\mathbf{\Gamma}_T\right\|^2 \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right], \\
\leq & \|\hat{C}^{-1}\|_{op}\|\hat{C}\|_{op}^{1/2} \sum_{n_1,n_2=1}^{d}\sum_{i\in S_{n_1}}\sum_{j\in S_{n_2}} \mathbb{E}\left[\left|\hat{C}_{ij} - \int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right] \\
& + \frac{\sqrt{2\pi}}{8}\|\hat{C}^{-1}\|_{op}^{3/2}\|\hat{C}\|_{op} \sum_{n_1,n_2=1}^{d}\sum_{i\in S_{n_1}}\sum_{j\in S_{n_2}}\mathbb{E}\left[\int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} \left|D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j\right|^2 \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right], \\
\leq & K \sum_{n_1,n_2=1}^{d}\sum_{i\in S_{n_1}}\sum_{j\in S_{n_2}}\left(A_1^{i,j} + A_2^{i,j}\right),
\end{aligned}
$$

where

$$
A_1^{i,j} = \mathbb{E}\left[\left|\hat{C}_{ij} - \int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right],
$$

and

$$
A_2^{i,j} = \mathbb{E}\left[\int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}} \left|D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j\right|^2 \mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right].
$$

In both terms we have dependence on n_1 and n_2 which are respectively functions of i and j. For $i, j = 1, \cdots, d$. we will treat each term separately.

Term $A_1^{i,j}$

First we start by computing the Malliavin's derivative in the $i-$th direction of Γ_T^j. By linearity of the derivative operator

$$
\begin{aligned}
D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j &= D^{n_1}_{(t,\lambda_t^{n_1},x)}F^{n_2}_{v_{j\%p}T}, \\
&= \frac{1}{\sqrt{v_{j\%p}T}} D^{n_1}_{(t,\lambda_t^{n_1},x)}\left(L^{n_2}_{v_{j\%p}T} - m^{j\%p}\int_0^{v_{j\%p}T}\lambda_s^{n_2}\mathrm{d}s\right),
\end{aligned}
$$

which yields using Proposition 2.1

$$
\begin{aligned}
D^{n_1}_{(t,\lambda_t^{n_1},x)}\Gamma_T^j &= \frac{\mathbb{1}_{t\leq v_{j\%p}T}}{\sqrt{v_{j\%p}T}}\left(x\mathbb{1}_{n_1=n_2} + \hat{L}^{n_2,n_1,t,x}_{v_{j\%p}T} - m^{j\%p}\int_t^{v_{j\%p}T}\hat{\lambda}_s^{n_2,n_1,t,x}\mathrm{d}s\right), \\
&= \frac{\mathbb{1}_{t\leq v_{j\%p}T}}{\sqrt{v_{j\%p}T}}\left(x\mathbb{1}_{n_1=n_2} + \hat{M}^{n_2,n_1,t,x}_{v_{j\%p}T}\right).
\end{aligned}
$$

Thus

$$\begin{aligned}
A_1^{i,j} =&\mathbb{E}\left[\left|\hat{C}_{ij} - \int_{\mathbb{R}_+}\int_0^{v_{i\%p}T} \lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}}\frac{\mathbb{1}_{t\leq v_{j\%p}T}}{\sqrt{v_{j\%p}T}}\left(x\mathbb{1}_{n_1=n_2} + \hat{M}^{n_2,n_1,t,x}_{v_{j\%p}T}\right)\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right],\\
=&\mathbb{E}\left[\left|\hat{C}_{ij} - \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_{\mathbb{R}_+}\int_0^{(v_{i\%p}\wedge v_{j\%p})T} x\lambda_t^{n_1}\left(x\mathbb{1}_{n_1=n_2} + \hat{M}^{n_2,n_1,t,x}_{v_{j\%p}T}\right)\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right],\\
\leq&\mathbb{E}\left[\left|\hat{C}_{ij} - \frac{\mathbb{1}_{n_1=n_2}}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_{\mathbb{R}_+}\int_0^{(v_{i\%p}\wedge v_{j\%p})T} x^2\lambda_t^{n_1}\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right]\\
&+\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\left|\int_{\mathbb{R}_+}\int_0^{(v_{i\%p}\wedge v_{j\%p})T} x\lambda_t^{n_1}\hat{M}^{n_2,n_1,t,x}_{v_{j\%p}T}\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right],\\
\leq&\mathbb{E}\left[\left|\hat{C}_{ij} - \frac{\mathbb{1}_{n_1=n_2}}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_{\mathbb{R}_+}\int_0^{(v_{i\%p}\wedge v_{j\%p})T} x^2\mathbb{E}[\lambda_t^{n_1}]\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right]\\
&+\frac{\mathbb{1}_{n_1=n_2}}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_{\mathbb{R}_+}x^2\nu^{n_1}(\mathrm{d}x)\mathbb{E}\left[\left|\int_0^{(v_{i\%p}\wedge v_{j\%p})T}\lambda_t^{n_1} - \mathbb{E}[\lambda_t^{n_1}]\mathrm{d}t\right|\right]\\
&+\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\left|\int_{\mathbb{R}_+}\int_0^{(v_{i\%p}\wedge v_{j\%p})T} x\lambda_t^{n_1}\hat{M}^{n_2,n_1,t,x}_{v_{j\%p}T}\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right|\right],\\
\leq&A_{1,1}^{i,j} + \mathbb{1}_{n_1=n_2}A_{1,2}^{i,j} + A_{1,3}^{i,j}.
\end{aligned}$$

We start with the term $A_{1,1}^{i,j}$. If i and j are not in the same interval S_{n_1} (*i.e.* $n_1 = n_2$), $\hat{C}_{ij} = 0$, thus $A_{1,1}^{i,j} = 0$. Note that due to symmetry arguments i and j are exchangeable, which allows us to assume that i and j are in the same interval and that $i \leq j$ which is equivalent to $i\%p \leq j\%p$. In this case, we have

$$\hat{C}_{ij} = C^{n_1}_{i\%pj\%p} = \int_{\mathbb{R}_+} x^2\nu^{n_1}(\mathrm{d}x)\sqrt{\frac{v_{i\%p}}{v_{j\%p}}}\left[\left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}\right]^{n_1}.$$

According to Lemma 4.2 it is possible to put the intensity's expectation under the form

$$\mathbb{E}[\boldsymbol{\lambda}_t] = \left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu} + Qe^{-tV}\boldsymbol{\mu},$$

where Q and V are matrices such that all the eigen-values of V are positive. Thus

$$\begin{aligned}
\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_0^{v_{i\%p}T}\mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t &= \sqrt{\frac{v_{i\%p}}{v_{j\%p}}}\left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu} + \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}Q\int_0^{v_{i\%p}T}e^{-Vt}\mathrm{d}t\boldsymbol{\mu},\\
&= \sqrt{\frac{v_{i\%p}}{v_{j\%p}}}\left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu} + \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}QV^{-1}\left(I_d - e^{-VTv_{i\%p}}\right)\boldsymbol{\mu},
\end{aligned}$$

which means that

$$\begin{aligned}
A_{1,1}^{i,j} =&\mathbb{E}\left[\left|\hat{C}_{ij} - \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_{\mathbb{R}_+}x^2\nu^{n_1}(\mathrm{d}x)\int_0^{v_{i\%p}T}\mathbb{E}[\lambda_t^{n_1}]\mathrm{d}t\right|\right],\\
\leq&\mathbb{E}\left[\left|\hat{C}_{ij} - \int_{\mathbb{R}_+}x^2\nu^{n_1}(\mathrm{d}x)\left[\left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}\right]^{n_1}\right|\right]\\
&+\mathbb{E}\left[\left|\int_{\mathbb{R}_+}x^2\nu^{n_1}(\mathrm{d}x)\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\left[QV^{-1}\left(I_d - e^{-VTv_{i\%p}}\right)\boldsymbol{\mu}\right]^{n_1}\right|\right],\\
=&\int_{\mathbb{R}_+}x^2\nu^{n_1}(\mathrm{d}x)\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\left|\left[QV^{-1}\left(I_d - e^{-VTv_{i\%p}}\right)\boldsymbol{\mu}\right]^{n_1}\right|\right],\\
=&O\left(\frac{1}{T}\right)\quad\text{because } e^{-VTv_{i\%p}}\longrightarrow 0.
\end{aligned}$$

For the term $A_{1,2}^{i,j}$ we use the second equality of Lemma 4.3 and the fact that for a given vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|$ where $\|.\|$ is the euclidean norm.

$$\begin{aligned}
\mathbb{E}\left[\left\|\int_0^{v_{i\%_p}T} \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t\right\|\right] &= \mathbb{E}\left[\left\|\int_0^{v_{i\%_p}T} e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)} A\boldsymbol{M}_s \mathrm{d}s\right\|\right], \\
&\leq \mathbb{E}\left[\int_0^{v_{i\%_p}T} \left\|e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)} A\boldsymbol{M}_s\right\| \mathrm{d}s\right], \\
&\leq \int_0^{v_{i\%_p}T} \left\|e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)} A\right\|_{op} \mathbb{E}\left[\|\boldsymbol{M}_s\|\right] \mathrm{d}s, \\
&\leq \int_0^{v_{i\%_p}T} \left\|e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)} A\right\|_{op} \mathbb{E}\left[\|\boldsymbol{M}_s\|^2\right]^{1/2} \mathrm{d}s,
\end{aligned}$$

where the last inequality comes from Cauchy-Schwarz.
Keeping in mind Ito's isometry we have

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{M}_s\|^2\right] &= \mathbb{E}\left[\sum_{i=1}^d \left|M_s^i\right|^2\right], \\
&= \sum_{i=1}^d \mathbb{E}\left[\left|M_s^i\right|^2\right], \\
&= \sum_{i=1}^d \mathbb{E}\left[\left[M^i\right]_s\right], \quad \text{where } \left[M^i\right]_s \text{ is the quadratic variation} \\
&= \sum_{i=1}^d \int_{\mathbb{R}_+} x^2 \nu^i(\mathrm{d}x) \int_0^s \mathbb{E}\left[\lambda_u^i\right] \mathrm{d}u, \\
&\leq K(s+1),
\end{aligned}$$

which implies that $\mathbb{E}\left[\|\boldsymbol{M}_s\|^2\right]^{1/2} \leq K(\sqrt{s}+1)$.
Using the fact that the operator norm is sub-multiplicative and Lemma 4.1 we have that

$$\begin{aligned}
\left\|e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)} A\right\|_{op} &\leq \left\|e^{-\left(B-A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(v_{i\%_p}T-s)}\right\|_{op} \|A\|_{op}, \\
&\leq K\left(1+(v_{i\%_p}T-s)^{d-1}\right) e^{-\rho_d(v_{i\%_p}T-s)}.
\end{aligned}$$

Combining these inequalities yields

$$\begin{aligned}
A_{1,2}^{i,j} &= \frac{1}{\sqrt{v_{i\%_p}v_{j\%_p}}T}\mathbb{E}\left[\left|\int_0^{v_{i\%_p}T} \lambda_t^i - \mathbb{E}[\lambda_t^i]\mathrm{d}t\right|\right], \\
&\leq \frac{1}{\sqrt{v_{i\%_p}v_{j\%_p}}T}\mathbb{E}\left[\left\|\int_0^{v_{i\%_p}T} \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t\right\|\right], \\
&\leq \frac{K}{\sqrt{v_{i\%_p}v_{j\%_p}}T}\int_0^{v_{i\%_p}T} e^{-\rho_d(v_{i\%_p}T-s)}\left(1+(v_{i\%_p}T-s)^{d-1}\right)(\sqrt{s}+1)\mathrm{d}s, \\
&\leq O\left(\frac{1}{\sqrt{T}}\right), \ \textit{cf.} \text{ the bound on } A_{1,2} \text{ in [8].}
\end{aligned}$$

For the term $A_{1,3}^{i,j}$ we start by noticing that the integral with respect to $\nu^i(\mathrm{d}x)$ is equivalent to taking the

expectation of a random variable X of law ν^i. Hence

$$\begin{aligned}
A_{1,3}^{i,j} &= \mathbb{E}\left[\left|\frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\int_0^{v_{i\%p}T}\int_{\mathbb{R}_+} x\lambda_t^{n_1}\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\nu^i(\mathrm{d}x)\mathrm{d}t\right|\right], \\
&= \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\left|\mathbb{E}_X\left[\int_0^{v_{i\%p}T} X\lambda_t^{n_1}\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\mathrm{d}t\right]\right|\right], \\
&\leq \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\mathbb{E}_X\left[\int_0^{v_{i\%p}T} X\lambda_t^{n_1}\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\mathrm{d}t\right]^2\right]^{1/2},
\end{aligned}$$

and by Jensen's inequality

$$\begin{aligned}
A_{1,3}^{i,j} &\leq \frac{1}{\sqrt{v_{i\%p}v_{j\%p}}T}\mathbb{E}\left[\mathbb{E}_X\left[\left(\int_0^{v_{i\%p}T} X\lambda_t^{n_1}\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,X}\mathrm{d}t\right)^2\right]\right]^{1/2}, \\
&\leq \frac{1}{\sqrt{v_{i\%p}v_{j\%p}T}}\left(\frac{1}{T}I_T\right)^{1/2},
\end{aligned}$$

where

$$I_T := \mathbb{E}\left[\mathbb{E}_X\left[\left(\int_0^{v_{i\%p}T} X\lambda_t^{n_1}\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,X}\mathrm{d}t\right)^2\right]\right]. \tag{9}$$

Using lemma 4.5 we have that

$$I_T \leq KT,$$

and therefore

$$A_{1,3}^{i,j} \leq O\left(\frac{1}{\sqrt{T}}\right).$$

This shows that

$$A_1^{i,j} = O\left(\frac{1}{\sqrt{T}}\right). \tag{10}$$

Term $A_2^{i,j}$

Keeping in mind that

$$A_2^{i,j} = \mathbb{E}\left[\int_{\mathbb{R}_+}\int_0^{v_{i\%p}T}\lambda_t^{n_1}\frac{x}{\sqrt{v_{i\%p}T}}\left|D_{(t,\lambda_t^{n_1},x)}^{n_1}\Gamma_T^j\right|^2\mathrm{d}t\nu^{n_1}(\mathrm{d}x)\right],$$

and that

$$D_{(t,\lambda_t^{n_1},x)}^{n_1}\Gamma_T^j = \frac{\mathbb{1}_{t\leq v_{j\%p}T}}{\sqrt{v_{j\%p}T}}\left(x\mathbb{1}_{n_1=n_2} + \hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right),$$

we have

$$\begin{aligned}
A_2^{i,j} &= \frac{1}{\sqrt{v_{i\%p}}v_{j\%p}T^{3/2}}\mathbb{E}\left[\int_0^{v_{i\%p}T\wedge v_{j\%p}T}\int_{\mathbb{R}_+} x\lambda_t^{n_1}\left(x\mathbb{1}_{n_1=n_2} + \hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right)^2\nu^{n_1}(\mathrm{d}x)\mathrm{d}t\right], \\
&\leq \frac{2}{\sqrt{v_{i\%p}}v_{j\%p}T^{3/2}}\mathbb{E}\left[\int_0^{v_{i\%p}T}\int_{\mathbb{R}_+} x^3\lambda_t^{n_1}\mathbb{1}_{n_1=n_2} + x\lambda_t^{n_1}\left|\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right|^2\nu^{n_1}(\mathrm{d}x)\mathrm{d}t\right], \\
&= \frac{2K\cdot\mathbb{1}_{n_1=n_2}}{T^{3/2}}\int_{\mathbb{R}_+} x^3\nu^{n_1}(\mathrm{d}x)\int_0^{v_{i\%p}T}\mathbb{E}[\lambda_t^{n_1}]\mathrm{d}t + \frac{2K}{T^{3/2}}\int_0^{v_{i\%p}T}\int_{\mathbb{R}_+} x\mathbb{E}\left[\lambda_t^{n_1}\left|\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right|^2\right]\nu^{n_1}(\mathrm{d}x)\mathrm{d}t, \\
&:= A_{2,1}^i + A_{2,2}^{i,j}.
\end{aligned}$$

According to Lemma 4.2 and to the fact that $\left(B - A\,\mathrm{diag}(m^1, \cdots, m^d)\right)$ has only positive eigenvalues we have that $\int_0^{v_{i\%p}T} \mathbb{E}[\lambda_t^i]\mathrm{d}t \leq KT$ and hence

$$A_{2,1}^i = O\left(\frac{1}{\sqrt{T}}\right).$$

For the final term it holds that

$$\begin{aligned}
A_{2,2}^{i,j} &= \frac{2K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x\mathbb{E}\left[\lambda_t^{n_1} \left|\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right|^2\right] \nu^{n_1}(\mathrm{d}x)\mathrm{d}t, \\
&= \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x\mathbb{E}\left[\lambda_t^{n_1}\mathbb{E}_t\left[\left|\hat{M}_{v_{j\%p}T}^{n_2,n_1,t,x}\right|^2\right]\right] \nu^{n_1}(\mathrm{d}x)\mathrm{d}t, \\
&= \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x\mathbb{E}\left[\lambda_t^{n_1}\mathbb{E}_t\left[\int_t^{v_{j\%p}T} \hat{\lambda}_s^{n_2,n_1,t,x} ds\right]\right] \nu^{n_1}(\mathrm{d}x)\mathrm{d}t.
\end{aligned}$$

Solving the SDE 14 with initial condition $\mathbb{E}_t\left[\tilde{\boldsymbol{\lambda}}_t^{n_1,t,x}\right] = x\boldsymbol{A}_{\cdot i}$ we have that

$$\mathbb{E}_t\left[\tilde{\lambda}_s^{n_2,n_1,t,x}\right] = x\left(e^{-\left(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\right)(s-t)}\boldsymbol{A}_{\cdot n_1}\right)^{n_2},$$

which is integrable. It follows that

$$\begin{aligned}
A_{2,2}^{i,j} &\leq \frac{K}{T^{3/2}} \int_0^{v_{i\%p}T} \int_{\mathbb{R}_+} x^2\mathbb{E}\left[\lambda_t^{n_1}\right] \nu^{n_1}(\mathrm{d}x)\mathrm{d}t, \\
&\leq \frac{K}{T^{3/2}} \int_{\mathbb{R}_+} x^2\nu^{n_1}(\mathrm{d}x) K v_{i\%p}T, \\
&\leq O\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

And finally

$$A_2^{i,j} = O\left(\frac{1}{\sqrt{T}}\right). \tag{11}$$

Combining 10 and 11 we conclude that

$$d_2(\boldsymbol{\Gamma}_T, \boldsymbol{G}) = O\left(\frac{1}{\sqrt{T}}\right).$$

□

Corollary 3.3.1. *Fix $d \in \mathbb{N}^*$. Let $(\boldsymbol{L}_t)_{t\geq 0}$ be a compound multivariate Hawkes process whose intensity $\boldsymbol{\lambda}$ follows the dynamics 4. Assume Assumptions 1, 2 and 3 are in force.*
Set $C = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_d^2)$ where for any $j = 1, \cdots, d$

$$\sigma_j^2 = \int x^2\nu^j(\mathrm{d}x)\left[\left(B - A\,\mathrm{diag}(m^1, \cdots, m^d)\right)^{-1} B\boldsymbol{\mu}\right]^j$$

and let $\boldsymbol{G} \sim \mathcal{N}(0, C)$.
Then there exists a constant $K > 0$ independent from T such that

$$d_2(\boldsymbol{F}_T, \boldsymbol{G}) \leq \frac{K}{\sqrt{T}}.$$

Proof. This is merely an application of Theorem 3.3 for $p = 1$ and $v_1 = 1$. □

Remark 7. *The specific bounds $\|g\|_{Lip} \leq 1$ and $M_2(g) \leq 1$ could have been relaxed to $\|g\|_{Lip} \leq K_1$ and $M_2(g) \leq K_2$ where K_1 and K_2 are two arbitrary positive constants.*

As a final result in this section, we consider the slightly modified process

$$\boldsymbol{Y}_T = \frac{\boldsymbol{L}_T - \mathrm{diag}(m^1, \cdots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_s] \mathrm{d}s}{\sqrt{T}}$$

and we study its behaviour as T goes to infinity.

Theorem 3.4. *Set* $\tilde{C} = \left(J\sqrt{C}\right)\ {}^t\left(J\sqrt{C}\right) = JC\ {}^tJ$, *where* $J = \left(I_d - \mathrm{diag}(m^1, \cdots, m^d)B^{-1}A\right)^{-1}$ *and* C *is defined in Corollary 3.3.1.*
Let $\tilde{\boldsymbol{G}} \sim \mathcal{N}\left(0, \tilde{C}\right)$. *There exists a constant* K *that does not depend on* T *such that*

$$d_2(\boldsymbol{Y}_T, \tilde{\boldsymbol{G}}) \leq \frac{K}{\sqrt{T}},$$

for any $T > 0$.

Proof. Thanks to Lemma 4.6 we have that

$$J^{-1}\boldsymbol{Y}_T = \boldsymbol{F}_T + J^{-1}\boldsymbol{R}_T.$$

Let $f \in \mathcal{H}$ and set $f_J : \boldsymbol{x} \mapsto \frac{f(J\boldsymbol{x})}{\|J\|_{op}}$, this function is clearly in $\mathcal{C}^2(\mathbb{R}^d)$ and $\|f_J\|_{Lip} \leq 1$ and $M_2(f_J) \leq \|J\|_{op}$. Let $\tilde{\boldsymbol{G}} \sim \mathcal{N}(0, \tilde{C})$. We have

$$\begin{aligned}\left|\mathbb{E}\left[f(\boldsymbol{Y}_T)\right] - \mathbb{E}[f(\tilde{\boldsymbol{G}})]\right| &= \|J\|_{op} \left|\mathbb{E}\left[f_J(J^{-1}\boldsymbol{Y}_T)\right] - \mathbb{E}\big[f_J(J^{-1}\tilde{\boldsymbol{G}})\big]\right|, \\ &= \|J\|_{op} \left|\mathbb{E}\left[f_J(\boldsymbol{F}_T + J^{-1}\boldsymbol{R}_T)\right] - \mathbb{E}\left[f_J(\boldsymbol{G})\right]\right|,\end{aligned}$$

where $\boldsymbol{G} = J^{-1}\tilde{\boldsymbol{G}} \sim \mathcal{N}(0, C)$.
Using a Taylor expansion, there exists a random $\boldsymbol{Y}^*$ such that

$$f_J(\boldsymbol{F}_T + J^{-1}\boldsymbol{R}_T) = f_J(\boldsymbol{F}_T) + \left\langle \nabla f_J(\boldsymbol{Y}^*), J^{-1}\boldsymbol{R}_T \right\rangle,$$

hence

$$\begin{aligned}\left|\mathbb{E}\left[f(\boldsymbol{Y}_T)\right] - \mathbb{E}[f(\tilde{\boldsymbol{G}})]\right| &= \|J\|_{op} \left|\mathbb{E}\left[f_J(\boldsymbol{F}_T)\right] - \mathbb{E}\left[f_J(\boldsymbol{G})\right] + \mathbb{E}\left[\left\langle \nabla f_J(\boldsymbol{Y}^*), J^{-1}\boldsymbol{R}_T \right\rangle\right]\right|, \\ &\leq \|J\|_{op} \left|\mathbb{E}\left[f_J(\boldsymbol{F}_T)\right] - \mathbb{E}\left[f_J(\boldsymbol{G})\right]\right| + \|J\|_{op} \left|\mathbb{E}\left[\left\langle \nabla f_J(\boldsymbol{Y}^*), J^{-1}\boldsymbol{R}_T \right\rangle\right]\right|, \\ &\leq \|J\|_{op} \sup_{g \in \mathcal{H}_J} \left|\mathbb{E}\left[g(\boldsymbol{F}_T)\right] - \mathbb{E}\left[g(\boldsymbol{G})\right]\right| + \|J\|_{op} \left|\mathbb{E}\left[\left\langle \nabla f_J(\boldsymbol{Y}^*), J^{-1}\boldsymbol{R}_T \right\rangle\right]\right|,\end{aligned}$$

where $\mathcal{H}_J = \{f \in \mathcal{C}^2(\mathbb{R}^d)$ such that $\|f\|_{Lip} \leq 1$ and $M_2(f) \leq \|J\|_{op}\}$. Using Cauchy-Schwarz's inequality (twice) the second term is bounded as follows

$$\begin{aligned}\left|\mathbb{E}\left[\left\langle \nabla f_J(\boldsymbol{Y}^*), J^{-1}\boldsymbol{R}_T \right\rangle\right]\right| &\leq \left|\mathbb{E}\left[\|\nabla f_J(\boldsymbol{Y}^*)\| \left\|J^{-1}\boldsymbol{R}_T\right\|\right]\right|, \\ &\leq \left|\mathbb{E}\left[\|\nabla f_J(\boldsymbol{Y}^*)\|^2\right]^{1/2} \mathbb{E}\left[\left\|J^{-1}\boldsymbol{R}_T\right\|^2\right]^{1/2}\right|, \\ &\leq \left|\mathbb{E}\left[\|f_J\|_{Lip}^2\right]^{1/2} \mathbb{E}\left[\left\|J^{-1}\right\|_{op}^2 \|\boldsymbol{R}_T\|^2\right]^{1/2}\right|, \\ &= \left\|J^{-1}\right\|_{op} \mathbb{E}\left[\|\boldsymbol{R}_T\|^2\right]^{1/2}.\end{aligned}$$

Thanks to Corollary 3.3.1 and Remark 7 there exists a positive constant K (independent from f and T and that can change from one line to another) such that

$$\left|\mathbb{E}\left[f(\boldsymbol{Y}_T)\right] - \mathbb{E}[f(\tilde{\boldsymbol{G}})]\right| \leq K\left(\frac{1}{\sqrt{T}} + \mathbb{E}\left[\|\boldsymbol{R}_T\|^2\right]^{1/2}\right).$$

And using Lemma 4.4 we have that

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{R}_T\|^2\right] &= \mathbb{E}\left[\left\|\operatorname{diag}(m^1,\cdots,m^d)\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^{-1}\frac{\mathbb{E}[\boldsymbol{\lambda}_T]-\boldsymbol{\lambda}_T}{\sqrt{T}}\right\|^2\right],\\
&\leq \left\|\operatorname{diag}(m^1,\cdots,m^d)\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^{-1}\right\|_{op}^2 \frac{1}{T}\mathbb{E}\left[\|\mathbb{E}[\boldsymbol{\lambda}_T]-\boldsymbol{\lambda}_T\|^2\right],\\
&\leq \frac{K}{T}\sum_{i=1}^{d}\mathbb{E}\left[\left(\lambda_T^i - \mathbb{E}[\lambda_T^i]\right)^2\right], \quad \text{and by virtue of Lemma 4.4,}\\
&\leq \frac{K}{T}.
\end{aligned}$$

And finally

$$\left|\mathbb{E}\left[f(\boldsymbol{Y}_T)\right] - \mathbb{E}[f(\tilde{\boldsymbol{G}})]\right| = O\left(\frac{1}{\sqrt{T}}\right),$$

hence the result. □

3.3 Proof of Theorem 1.1

We start by recalling that

$$\boldsymbol{Y}_T = \frac{\boldsymbol{L}_T - \operatorname{diag}(m^1,\cdots,m^d)\int_0^T \mathbb{E}[\boldsymbol{\lambda}_s]\mathrm{d}s}{\sqrt{T}},$$

and that

$$\begin{aligned}
\boldsymbol{Y}'_T &= \frac{\boldsymbol{L}_T - \operatorname{diag}(m^1,\cdots,m^d)\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^{-1}B\boldsymbol{\mu}T}{\sqrt{T}},\\
&= \frac{\boldsymbol{L}_T - \operatorname{diag}(m^1,\cdots,m^d)V^{-1}B\boldsymbol{\mu}T}{\sqrt{T}},
\end{aligned}$$

where $V = B - A\operatorname{diag}(m^1,\cdots,m^d)$.
Using Lemma 4.2, we have that

$$\mathbb{E}[\boldsymbol{\lambda}_t] = V^{-1}B\boldsymbol{\mu} + e^{-Vt}\left(I_d - V^{-1}B\right)\boldsymbol{\mu},$$

hence

$$\begin{aligned}
\int_0^T \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t &= V^{-1}B\boldsymbol{\mu}T + \int_0^T e^{-Vt}\mathrm{d}t\left(I_d - V^{-1}B\right)\boldsymbol{\mu},\\
&= V^{-1}B\boldsymbol{\mu}T + \left(I_d - e^{-VT}\right)V^{-1}\left(I_d - V^{-1}B\right)\boldsymbol{\mu}.
\end{aligned}$$

We deduce that

$$\boldsymbol{Y'}_T = \boldsymbol{Y}_T - \boldsymbol{R'}_T,$$

with $\boldsymbol{R'}_T = \frac{\operatorname{diag}(m^1,\cdots,m^d)\left(I_d - e^{-VT}\right)\left(V^{-1} - V^{-2}B\right)\boldsymbol{\mu}}{\sqrt{T}}$.
Let $\tilde{\boldsymbol{G}} \sim \mathcal{N}\left(0,\tilde{C}\right)$ as defined in Theorem 1.1 and let $f \in \mathcal{H}$. Using a Taylor expansion, we have for some $\boldsymbol{X}$

$$f(\boldsymbol{Y'}_T) - f(\tilde{\boldsymbol{G}}) = f(\boldsymbol{Y}_T) - f(\tilde{\boldsymbol{G}}) - \langle\nabla f(\boldsymbol{X}), \boldsymbol{R'}_T\rangle.$$

Since $\sup_{\boldsymbol{x}\in\mathbb{R}^d}\|\nabla f(\boldsymbol{x})\|_2 \leq 1$ and $\boldsymbol{Y'}_T = O\left(\frac{1}{\sqrt{T}}\right)$, we deduce that

$$d_2(\boldsymbol{Y'}_T, \tilde{\boldsymbol{G}}) \leq d_2(\boldsymbol{Y}_T, \tilde{\boldsymbol{G}}) + O\left(\frac{1}{\sqrt{T}}\right),$$

which yields using Theorem 3.4

$$d_2(\boldsymbol{Y'}_T, \tilde{\boldsymbol{G}}) = O\left(\frac{1}{\sqrt{T}}\right).$$

4 Lemmata

Lemma 4.1. *Set $V = B - A\operatorname{diag}(m^1, \cdots, m^d)$. Assume that Assumption 2 is in force. Then there are positive constants K and ρ_d such that*

$$\left\|e^{-tV}\right\|_{op} \leq K(1 + t^{d-1})e^{-\rho_d t}$$

for any $t \geq 0$.

Proof. First, we call V's eigenvalues $\rho_1 \geq \cdots \geq \rho_d$ and we recall that they are positive. Using Jordan-Chevalley's decomposition we can write

$$V = P\operatorname{diag}(\rho_1, \cdots, \rho_d)P^{-1} + V_{nil}$$

where V_{nil} is a nilpotent matrix that commutes with $P\operatorname{diag}(\rho_1, \cdots, \rho_d)P^{-1}$. Let $t \geq 0$, taking the exponential yields

$$\begin{aligned} e^{-tV} &= e^{-tP\operatorname{diag}(\rho_1,\cdots,\rho_d)P^{-1} - tV_{nil}}, \\ &= e^{-tP\operatorname{diag}(\rho_1,\cdots,\rho_d)P^{-1}} e^{-tV_{nil}}, \\ &= P\operatorname{diag}(e^{-\rho_1 t}, \cdots, e^{-\rho_d t})P^{-1} \sum_{j=1}^{d-1} t^j \frac{(-V_{nil})^j}{j!}. \end{aligned}$$

Since the operator norm is sub-multiplicative and using the triangular inequality

$$\begin{aligned} \left\|e^{-tV}\right\|_{op} &\leq \|P\|_{op}\|P^{-1}\|_{op} \left\|\operatorname{diag}(e^{-\rho_1 t}, \cdots, e^{-\rho_d t})\right\|_{op} \sum_{j=1}^{d-1} t^j \frac{\|V_{nil}\|_{op}^j}{j!}, \\ &\leq K \left\|\operatorname{diag}(e^{-\rho_1 t}, \cdots, e^{-\rho_d t})\right\|_{op} \left(1 + t^{d-1}\right), \\ &\leq K \left\|\operatorname{diag}(e^{-\rho_1 t}, \cdots, e^{-\rho_d t})\right\|_{\infty} \left(1 + t^{d-1}\right), \end{aligned}$$

where the last inequality comes from the fact that all norms are equivalent in finite dimension. □

Lemma 4.2. *If $\boldsymbol{\lambda}$ follows the dynamics 4 then for each t*

$$\mathbb{E}[\boldsymbol{\lambda}_t] = \left(B - A\operatorname{diag}(m^1, \cdots, m^d)\right)^{-1} B\boldsymbol{\mu} + e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)t} \left(I_d - \left(B - A\operatorname{diag}(m^1, \cdots, m^d)\right)^{-1} B\right) \boldsymbol{\mu}.$$

Proof. First we prove that if the kernels are exponential, then $\boldsymbol{\lambda}$ is a Markov process that follows an SDE. By multiplying 4 by e^{Bt} we get

$$\begin{aligned} e^{Bt}\boldsymbol{\lambda}_t = e^{Bt}\boldsymbol{\mu} + \int_{[0,t)} e^{Bs} A\mathrm{d}\boldsymbol{L}_s, \quad \text{and by differentiating,} \\ e^{Bt}\left(\mathrm{d}\boldsymbol{\lambda}_t + B\boldsymbol{\lambda}_t \mathrm{d}t\right) = e^{Bt} B\boldsymbol{\mu}\mathrm{d}t + e^{Bt} A\mathrm{d}\boldsymbol{L}_t. \end{aligned}$$

Hence the SDE

$$\mathrm{d}\boldsymbol{\lambda}_t = B\left(\boldsymbol{\mu} - \boldsymbol{\lambda}_t\right)\mathrm{d}t + A\mathrm{d}\boldsymbol{L}_t. \tag{12}$$

By taking the expected value of 12 and keepind in mind that $\mathrm{d}\mathbb{E}[\boldsymbol{L}_s] = \operatorname{diag}(m^1, \cdots, m^d)\mathbb{E}[\boldsymbol{\lambda}_s]\mathrm{d}s$, the intensity's expectation is the solution of the multivariate ODE

$$\mathrm{d}\mathbb{E}[\boldsymbol{\lambda}_t] = B\boldsymbol{\mu}\mathrm{d}t - \left(B - A\operatorname{diag}(m^1, \cdots, m^d)\right)\mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t, \tag{13}$$

with initial condition $\mathbb{E}[\boldsymbol{\lambda}_0] = \mu$. Assumption 2 guarantees that all the eigenvalues of $\left(B - A\operatorname{diag}(m^1, \cdots, m^d)\right)$ are positive, thus the matrix is invertible and we get the result. □

Lemma 4.3. *The difference between the intensity and its expected value is*

$$\boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t] = \int_{[0,t)} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(t-s)} A\mathrm{d}\boldsymbol{M}_s,$$

where

$$\boldsymbol{M}_t = \boldsymbol{L}_t - \int_0^t \operatorname{diag}(m^1, \cdots, m^d)\boldsymbol{\lambda}_s \mathrm{d}s.$$

Its integral with respect to time is

$$\int_0^T \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t = \int_0^T e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(T-s)} A\boldsymbol{M}_s ds.$$

Proof. By taking the difference between 12 and 13 we can verify that $\boldsymbol{\lambda} - \mathbb{E}[\boldsymbol{\lambda}]$ is a solution of the SDE

$$\mathrm{d}\left(\boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\right) = -\left(B - A\operatorname{diag}(m^1, \cdots, m^d)\right)\left(\boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\right) + A\mathrm{d}\boldsymbol{M}_t$$

with the initial condition $\boldsymbol{\lambda}_0 - \mathbb{E}[\boldsymbol{\lambda}_0] = \mathbf{0}$. Solving the SDE (using variation of parameters) yields

$$\boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t] = \int_{[0,t)} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(t-s)} A\mathrm{d}\boldsymbol{M}_s.$$

For the second equality, we start by taking the integral with respect to time until the instant T

$$\begin{aligned}
\int_0^T \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t &= \int_0^T \int_{[0,t)} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(t-s)} A\mathrm{d}\boldsymbol{M}_s \mathrm{d}t, \\
&= \int_0^T \int_0^T \mathbb{1}_{s<t} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(t-s)} A\mathrm{d}\boldsymbol{M}_s \mathrm{d}t, \quad \text{and using Fubini's identity} \\
&= \int_0^T \int_s^T e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(t-s)} \mathrm{d}t A\mathrm{d}\boldsymbol{M}_s, \\
&= \int_0^T \int_0^{T-s} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)u} \mathrm{d}u A\mathrm{d}\boldsymbol{M}_s, \quad \text{using a change of variables} \\
&= \int_0^T \Phi(T-s) A\mathrm{d}\boldsymbol{M}_s,
\end{aligned}$$

where Φ is the anti-derivative of $u \to e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)u}$ that vanishes at zero. In absence of common jumps, the integration by parts formula is

$$0 = \Phi(0)A\boldsymbol{M}_T - \Phi(T)A\boldsymbol{M}_0 = \int_0^T \mathrm{d}\left(\Phi(T-s)\right) A\boldsymbol{M}_s + \Phi(T-s)A\mathrm{d}\boldsymbol{M}_s,$$

hence

$$\int_0^T \boldsymbol{\lambda}_t - \mathbb{E}[\boldsymbol{\lambda}_t]\mathrm{d}t = \int_0^T e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(T-s)} A\boldsymbol{M}_s \mathrm{d}s.$$

□

Lemma 4.4. *For any fixed n, m in $[\![1, d]\!]$ we have for any $T \geq 0$*

$$\mathbb{E}\left[\left(\lambda_T^n - \mathbb{E}[\lambda_T^n]\right)\left(\lambda_T^m - \mathbb{E}[\lambda_T^m]\right)\right] \leq K$$

where K is a positive constant independent from T.

Proof. Using the first equality of Lemma 4.3 we have

$$\left(\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]\right)\left(\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]\right)^\top = \int_{[0,T)} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(T-s)} A\mathrm{d}\boldsymbol{M}_s \int_{[0,T)} \mathrm{d}\boldsymbol{M}_s^\top A^\top e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^\top (T-s)},$$

which yields using the multivariate Ito isometry

$$\mathbb{E}\left[\left(\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]\right)\left(\boldsymbol{\lambda}_T - \mathbb{E}[\boldsymbol{\lambda}_T]\right)^\top\right] = \int_{[0,T)} e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)(T-s)} A\mathrm{d}\mathbb{E}\left[[\boldsymbol{M}]_s\right] A^\top e^{-\left(B - A\operatorname{diag}(m^1,\cdots,m^d)\right)^\top (T-s)},$$

where $d\left[\boldsymbol{M}\right]$ is the quadratic variation matrix infinitesimal growth

$$\begin{aligned}\mathrm{d}\big(\left[\boldsymbol{M}\right]_s\big)_{ij} &= \mathrm{d}\left[M^i, M^j\right]_s, \\ &= \mathbb{1}_{i\neq j}\mathrm{d}\left[M^i\right]_s, \\ &= \mathbb{1}_{i\neq j}\mathbb{E}\left[\lambda_s^i\right]\mathrm{d}s.\end{aligned}$$

Since the expected value of λ_s^i is bounded by a constant independent from T and since all of $\big(B - A\,\mathrm{diag}(m^1,\cdots,m^d)\big)$'s eigen-values are positive we obtain the result. □

Lemma 4.5. *Assume that Assumptions 3 and 1 are is force. For $i,j \in [\![1,d]\!]$ let I_T be the quantity defined in 9, then we have*

$$\frac{1}{T}I_T \leq K.$$

Proof. Let $m \leq n$ be two integers in $[\![1,p]\!]$ and i,j in $[\![1,d]\!]$. We start by expanding 9

$$\begin{aligned}I_T &= \mathbb{E}\left[\mathbb{E}_X\left[\left(\int_0^{v_mT} X\lambda_t^i \hat{M}_{v_nT}^{j,i,t,X}\mathrm{d}t\right)^2\right]\right], \\ &= 2\mathbb{E}\left[\mathbb{E}_X\left[\int_0^{v_mT}\int_0^t X^2\lambda_t^i\hat{M}_{v_nT}^{j,i,t,X}\lambda_s^i\hat{M}_{v_nT}^{j,i,s,X}\mathrm{d}s\mathrm{d}t\right]\right], \\ &= 2\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\mathbb{E}\left[\lambda_t^i\lambda_s^i\mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\hat{M}_{v_nT}^{j,i,s,x}\right]\right]\mathrm{d}s\mathrm{d}t\nu^i\mathrm{d}x,\end{aligned}$$

where we recall that $\hat{M}_{v_nT}^{j,i,t,x} = \hat{L}_{v_nT}^{j,i,t,x} - m^j\int_t^{v_nT}\hat{\lambda}_s^{j,i,t,x}ds$ is a martingale. Thus the product's expectation is

$$\mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\hat{M}_{v_nT}^{j,i,s,x}\right] = \mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\left(\hat{M}_{v_nT}^{j,i,s,x} - \hat{M}_t^{j,i,s,x}\right)\right] + \mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\hat{M}_t^{j,i,s,x}\right],$$

the last term vanishes since for any $t \leq v_mT \leq v_nT$, $\mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\hat{M}_t^{j,i,s,x}\right] = \hat{M}_t^{j,i,s,x}\mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\right] = 0$. Thus

$$\begin{aligned}\mathbb{E}_t\left[\hat{M}_{v_nT}^{j,i,t,x}\hat{M}_{v_nT}^{j,i,s,x}\right] &= \mathbb{E}_t\left[\int_t^{v_nT}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} y^2\mathbb{1}_{\theta\leq\hat{\lambda}_u^{j,i,t,x}}\mathbb{1}_{\theta\leq\hat{\lambda}_u^{j,i,s,x}}\mathrm{d}\theta\nu^j(\mathrm{d}y)\mathrm{d}u\right], \\ &= \mathbb{E}_t\left[\int_t^{v_nT}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} y^2\mathbb{1}_{\theta\leq\min(\hat{\lambda}_u^{j,i,t,x},\hat{\lambda}_u^{j,i,s,x})}\mathrm{d}\theta\nu^j(\mathrm{d}y)\mathrm{d}u\right], \\ &= \int_{\mathbb{R}_+} y^2 d\nu^j(y)\int_t^{v_nT}\mathbb{E}_t\left[\min(\hat{\lambda}_u^{j,i,t,x},\hat{\lambda}_u^{j,i,s,x})\right]\mathrm{d}u, \\ &\leq K\int_t^{v_nT}\mathbb{E}_t\left[\hat{\lambda}_u^{j,i,s,x}\right]\mathrm{d}u.\end{aligned}$$

As shown in remark 5, $\hat{\boldsymbol{\lambda}}_u^{i,s,x}$ has the same dynamics as a Hawkes process $\tilde{\boldsymbol{\lambda}}_u^{i,s,x}$ that satisfies -if we follow the lines of the proof of Lemma 4.2- the SDE

$$\mathrm{d}\tilde{\boldsymbol{\lambda}}_u^{i,s,x} = -B\tilde{\boldsymbol{\lambda}}_u^{i,s,x}\mathrm{d}u + A\mathrm{d}\tilde{\boldsymbol{L}}_u^{i,s,x} \tag{14}$$

whose solution yields

$$\mathbb{E}_t\left[\tilde{\boldsymbol{\lambda}}_u^{i,s,x}\right] = e^{-V(u-t)}\tilde{\boldsymbol{\lambda}}_t^{i,s,x},$$

where $V = B - A\,\mathrm{diag}(m^1, \cdots, m^d)$ whose eigen-values $\rho_1, \cdots, \rho_d$ are positive. Hence

$$\begin{aligned}\mathbb{E}_t\left[\hat{M}^{j,i,t,x}_{v_nT}\hat{M}^{j,i,s,x}_{v_nT}\right] &\leq K\left[\int_t^{v_nT} e^{-V(u-t)}\tilde{\boldsymbol{\lambda}}^{i,s,x}_t \mathrm{d}u\right]^j,\\ &= K\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]^j.\end{aligned}$$

By plugging this inequality in I_T's expression we get

$$\begin{aligned}I_1 &\leq K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\mathbb{E}\left[\lambda^i_t\lambda^i_s\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]^j\right]\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x),\\ &= K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\mathbb{E}\left[\lambda^i_t\lambda^i_s\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]\right]^j\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x),\\ &= K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\mathbb{E}\left[\lambda^i_s\mathbb{E}_s\left[\lambda^i_t\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]\right]\right]^j\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x),\\ &= K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\mathbb{E}\left[\lambda^i_s\mathbb{E}_s\left[\lambda^i_t\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]\right]\right]^j\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x),\end{aligned}$$

since $\tilde{\boldsymbol{\lambda}}^{i,s,x}_t$ starts at s, $\mathbb{E}_s\left[\lambda^i_t\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right] = \mathbb{E}_s\left[\lambda^i_t\right]\mathbb{E}\left[\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right]$. By solving the expectation value of the SDE 14 with the initial condition $\mathbb{E}\left[\tilde{\boldsymbol{\lambda}}^{i,s,x}_s\right] = \boldsymbol{A}_{.i}x$ we get

$$\mathbb{E}\left[\tilde{\boldsymbol{\lambda}}^{i,s,x}_t\right] = e^{-V(t-s)}\boldsymbol{A}_{.i}x,$$

which yields after being plugged in the last inequality

$$\begin{aligned}I_T \leq& K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^2\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)e^{-V(t-s)}\boldsymbol{A}_{.i}x\right]^j\mathbb{E}\left[\lambda^i_s\mathbb{E}_s\left[\lambda^i_t\right]\right]\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x),\\ \leq& K\int_{\mathbb{R}_+}\int_0^{v_mT}\int_0^t x^3\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)e^{-V(t-s)}\boldsymbol{A}_{.i}\right]^j\mathbb{E}\left[\lambda^i_s\lambda^i_t\right]\mathrm{d}s\mathrm{d}t\nu^i(\mathrm{d}x).\end{aligned}$$

By combining Cauchy-Schwarz's inequality and Lemma 4.4 we have that

$$\mathbb{E}\left[\lambda^i_s\lambda^i_t\right] \leq \mathbb{E}\left[|\lambda^i_s|^2\right]^{1/2}\mathbb{E}\left[|\lambda^i_t|^2\right]^{1/2} \leq K,$$

thus

$$\begin{aligned}I_T \leq& K\int_{\mathbb{R}_+}x^3\nu^i(\mathrm{d}x)\int_0^{v_mT}\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)\int_0^t e^{-V(t-s)}\mathrm{d}s\boldsymbol{A}_{.i}\right]^j\mathrm{d}t,\\ \leq& K\int_{\mathbb{R}_+}x^3\nu^i(\mathrm{d}x)\int_0^{v_mT}\left[V^{-1}\left(I_d - e^{-V(v_nT-t)}\right)V^{-1}\left(I_d - e^{-Vt}\right)\boldsymbol{A}_{.i}\right]^j\mathrm{d}t,\\ \leq& K\int_0^{v_mT}\left[V^{-2}\left(I_d - e^{-V(v_nT-t)}\right)\left(I_d - e^{-Vt}\right)\boldsymbol{A}_{.i}\right]^j\mathrm{d}t, \quad \text{because } V \text{ commutes with its exponential},\\ =& K\int_0^{v_mT}\left[V^{-2}\left(I_d - e^{-V(v_nT-t)} - e^{-Vt} + e^{-Vv_nT}\right)\boldsymbol{A}_{.i}\right]^j\mathrm{d}t,\end{aligned}$$

and since $v_m T \leq v_n T$,

$$\begin{aligned} I_T \leq & K \int_0^{v_n T} \|I_d\|_{op} + \left\| e^{-V(v_n T - t)} \right\|_{op} + \left\| e^{-Vt} \right\|_{op} + \left\| e^{-VT} \right\|_{op} \mathrm{d}t, \\ \leq & K \int_0^{v_n T} 1 + (1 + (v_n T - t)^{d-1}) e^{-(\beta - \rho_d)(v_n T - t)} + (1 + t^{d-1}) e^{-(\beta - \rho_d)t} + (1 + v_n T^{d-1}) e^{-(\beta - \rho_d) v_n T} \mathrm{d}t, \\ \leq & KT. \end{aligned}$$

□

Lemma 4.6. *Set* $\boldsymbol{F}_T = \frac{\boldsymbol{L}_T - \mathrm{diag}(m^1, \cdots, m^d) \int_0^T \boldsymbol{\lambda}_t \mathrm{d}t}{\sqrt{T}}$ *and* $\boldsymbol{Y}_T = \frac{\boldsymbol{L}_T - \mathrm{diag}(m^1, \cdots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] \mathrm{d}t}{\sqrt{T}}$. *Then we have the equality*

$$\boldsymbol{Y}_T = J \boldsymbol{F}_T + \boldsymbol{R}_T, \tag{15}$$

where

$$J = \left(I_d - \mathrm{diag}(m^1, \cdots, m^d) B^{-1} A \right)^{-1}$$

and

$$\boldsymbol{R}_T = \mathrm{diag}(m^1, \cdots, m^d) \left(B - A \, \mathrm{diag}(m^1, \cdots, m^d) \right)^{-1} \frac{\mathbb{E}[\boldsymbol{\lambda}_T] - \boldsymbol{\lambda}_T}{\sqrt{T}}.$$

Proof. By taking the expected value of SDE 12 we have the system

$$\begin{cases} \mathrm{d}\boldsymbol{\lambda}_t = B(\boldsymbol{\mu} - \boldsymbol{\lambda}_t)\mathrm{d}t + A \mathrm{d}\boldsymbol{L}_t, \\ \mathrm{d}\mathbb{E}\left[\boldsymbol{\lambda}_t\right] = B(\boldsymbol{\mu} - \mathbb{E}[\boldsymbol{\lambda}_t])\mathrm{d}t + A \, \mathrm{diag}(m^1, \cdots, m^d) \mathbb{E}[\boldsymbol{\lambda}_t] \mathrm{d}t, \end{cases}$$

which yields after integrating with respect to time

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu} T - B \int_0^T \boldsymbol{\lambda}_t \mathrm{d}t + A \boldsymbol{L}_T, \\ \mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\mu} = B\boldsymbol{\mu} T - \left(B - A \, \mathrm{diag}(m^1, \cdots, m^d) \right) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] \mathrm{d}t. \end{cases}$$

In order to involve the quantities of interest $\boldsymbol{F}_T$ and $\boldsymbol{Y}_T$, we state the fact that $\mathrm{diag}(m^1, \cdots, m^d)$ is invertible (since $\nu^1, \cdots, \nu^d$ are suppoted by $\mathbb{R}_+^*$), hence

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu} T - B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \, \mathrm{diag}(m^1, \cdots, m^d) \int_0^T \boldsymbol{\lambda}_t \mathrm{d}t + A \mathrm{d}\boldsymbol{L}_T, \\ \mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\mu} = B\boldsymbol{\mu} T - \left(B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} - A \right) \mathrm{diag}(m^1, \cdots, m^d) \int_0^T \mathbb{E}[\boldsymbol{\lambda}_t] \mathrm{d}t, \end{cases}$$

which yields by adding and subtracting $\boldsymbol{L}_T$

$$\begin{cases} \boldsymbol{\lambda}_T - \boldsymbol{\mu} = B\boldsymbol{\mu} T + B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \left(\sqrt{T} \boldsymbol{F}_T - \boldsymbol{L}_T \right) + A \mathrm{d}\boldsymbol{L}_T, \\ \mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\mu} = B\boldsymbol{\mu} T + \left(B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} - A \right) \left(\sqrt{T} \boldsymbol{Y}_T - \boldsymbol{L}_T \right). \end{cases}$$

Subtracting the first equation from the second yields

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\lambda}_T &= \left(B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} - A \right) \sqrt{T} \boldsymbol{Y}_T - B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \sqrt{T} \boldsymbol{F}_T, \\ &= \left(B - A \, \mathrm{diag}(m^1, \cdots, m^d) \right) \mathrm{diag}(m^1, \cdots, m^d)^{-1} \sqrt{T} \boldsymbol{Y}_T - B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \sqrt{T} \boldsymbol{F}_T. \end{aligned}$$

Since $\left(B - \mathrm{diag}(m^1, \cdots, m^d) A \right)$ is invertible, we have

$$\begin{aligned} \left(B - A \, \mathrm{diag}(m^1, \cdots, m^d) \right)^{-1} \frac{\mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\lambda}_T}{\sqrt{T}} = & \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \boldsymbol{Y}_T \\ & - \left(B - A \, \mathrm{diag}(m^1, \cdots, m^d) \right)^{-1} B \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \boldsymbol{F}_T, \\ = & \, \mathrm{diag}(m^1, \cdots, m^d)^{-1} \boldsymbol{Y}_T \\ & - \left(\mathrm{diag}(m^1, \cdots, m^d) - \mathrm{diag}(m^1, \cdots, m^d) B^{-1} A \, \mathrm{diag}(m^1, \cdots, m^d) \right)^{-1} \boldsymbol{F}_T. \end{aligned}$$

Multiplying to the left by $\mathrm{diag}(m^1, \cdots, m^d)$ we get

$$\mathrm{diag}(m^1, \cdots, m^d)\left(B - A\,\mathrm{diag}(m^1, \cdots, m^d)\right)^{-1} \frac{\mathbb{E}\left[\boldsymbol{\lambda}_T\right] - \boldsymbol{\lambda}_T}{\sqrt{T}} = \boldsymbol{Y}_T - \left(I_d - \mathrm{diag}(m^1, \cdots, m^d)B^{-1}A\right)^{-1} \boldsymbol{F}_T.$$

□

Acknowledgement

I would like to thank Prof. Anthony Réveillac for his guidance and his useful critiques of this paper.

References

[1] Massil Achab, Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, and Jean-François Muzy. Uncovering causality from multivariate hawkes integrated cumulants. In *International Conference on Machine Learning*, pages 1–10. PMLR, 2017.

[2] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative finance*, 13(1):65–77, 2013.

[3] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Some limit theorems for hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

[4] Yannick Bessy-Roland, Alexandre Boumezoued, and Caroline Hillairet. Multivariate Hawkes process for cyber insurance. working paper or preprint, April 2020.

[5] Paul Embrechts, Thomas Liniger, and Lu Lin. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48(A):367–378, 2011.

[6] Eymen Errais, Kay Giesecke, and Lisa R Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665, 2010.

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

[8] Caroline Hillairet, Lorick Huang, Mahmoud Khabou, and Anthony Réveillac. The Malliavin-Stein method for Hawkes functionals, 2021.

[9] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, May 2012.

[10] Ivan Nourdin, Giovanni Peccati, and Anthony Réveillac. Multivariate normal approximation using stein's method and malliavin calculus. In *Annales de l'IHP Probabilités et statistiques*, volume 46, pages 45–58, 2010.

[11] Eva Löcherbach. Large deviations for cascades of diffusions arising in oscillating systems of interacting hawkes processes, 2017.

[12] 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*, Austin, United States, December 2013.

[13] Giovanni Peccati , Cengbo Zheng, et al. Multi-dimensional gaussian fluctuations on the poisson space. *Electronic Journal of Probability*, 15:1487–1527, 2010.

[14] Jean Picard. Formules de dualité sur l'espace de poisson. In *Annales de l'IHP Probabilités et statistiques*, volume 32, pages 509–548, 1996.