Additional File 3 of "Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis": adaptive properties of the Lasso estimate for Hawkes processes

Patricia Reynaud-Bouret∗1 and Vincent Rivoirard2 and Franck Grammont1 and Christine Tuleau-Malot1

1Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France
2CEREMADE UMR CNRS 7534, Université Paris Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 PARIS Cedex 16, France

Email: Patricia Reynaud-Bouret - reynaudb@unice.fr; Vincent Rivoirard - rivoirard@ceremade.dauphine.fr; Franck Grammont - grammont@unice.fr; Christine Tuleau-Malot - malot@unice.fr;

*Corresponding author

Theorem 1 of [1] does not explicit particular basis on which the interaction functions are expanded. It is stated in its general form as follows. Note that [1] is a proceedings and therefore no proof has been given of the result in [1]. If \( n \) i.i.d. trials are recorded, each trial \( i \) corresponds to the observation of \( N_i = (N_i^{(1)}, ..., N_i^{(M)}) \), the multivariate Hawkes process whose intensity is given by the predictable transformation denoted \( \psi_i \).

Furthermore, to each trial \( i \), we can associate an intensity \( \lambda_i \) and a contrast \( \gamma(i) \). The global least-squares contrast over the \( n \) trials can also be seen as

\[
\gamma_n(f) = \sum_{i=1}^n \gamma(i)(f). \tag{1}
\]

We use the following notation: for any predictable processes \( H = (H_i^{(1)}, ..., H_i^{(M)})_{i=1,...,n} \), \( K = (K_i^{(1)}, ..., K_i^{(M)})_{i=1,...,n} \), set

\[
H \bullet N = \sum_{i=1}^n \sum_{m=1}^M \int_{T_1}^{T_2} H_i^{(m)}(t) dN_i^{(m)}(t), \tag{2}
\]

\[
H \circ K = \sum_{i=1}^n \sum_{m=1}^M \int_{T_1}^{T_2} H_i^{(m)}(t) K_i^{(m)}(t) dt, \tag{3}
\]

and \( H^{\circ 2} = H \circ H \).

In general, we use a dictionary \( \Phi \) of known functions of \( \mathcal{H} \) and we only consider linear combinations of functions of \( \Phi \) for estimating \( f^\ast \):

\[
f_a = \sum_{\varphi \in \Phi} a_\varphi \varphi, \text{ for } a \in \mathbb{R}^\Phi. \tag{4}
\]
Then, by linearity of $\psi$, one can rewrite (1) as
\[ \gamma_n(f_a) = -2a'b_n + a'G_na, \]  
where for any $\varphi$ and $\tilde{\varphi}$ in $\Phi$,
\[ (b_n)_\varphi = \psi(\varphi) \bullet N \quad \text{and} \quad (G_n)_{\varphi, \tilde{\varphi}} = \psi(\varphi) \circ \psi(\tilde{\varphi}). \]

Given a vector of positive weights $d$, the Lasso estimate of $f^*$ is $\tilde{f}_n := \tilde{a}_n$ where $\tilde{a}_n$ is a minimizer of the following $\ell_1$-penalized least-square contrast:
\[ \tilde{a}_n \in \arg \min_{a \in \mathbb{R}^\Phi} \{-2a'b_n + a'G_na + 2d'|a|\}. \]

Then Theorem 1 of [1] is stated as follows:

**Theorem 1.** We introduce the following two events:
\[ \Omega_{V,B} = \{ \forall \varphi \in \Phi, \sup_{t \in [T_1, T_2], m, i} |\psi^{(m)}_i(\varphi)| \leq B_\varphi \text{ and } (\psi(\varphi))^2 \bullet N \leq V_\varphi \}, \]
for positive deterministic constants $B_\varphi$ and $V_\varphi$ and
\[ \Omega_c = \{ \forall a \in \mathbb{R}^\Phi, \quad a'G_na \geq c \ a' \}, \]
for a positive constant $c$. Let $x$ and $\varepsilon$ be strictly positive constants and for all $\varphi \in \Phi$,
\[ d_\varphi = \sqrt{2(1 + \varepsilon) \bar{V}_\varphi^\mu x + \frac{B_\varphi^2 x}{3}}, \]
with
\[ \bar{V}_\varphi^\mu = \frac{\mu}{\mu - \phi(\mu)} (\psi(\varphi))^2 \bullet N + \frac{B_\varphi^2 x}{\mu - \phi(\mu)} \]
for a real number $\mu$ such that $\mu > \phi(\mu)$, where $\phi(\mu) = \exp(\mu) - \mu - 1$. Then, with probability larger than
\[ 1 - 4 \sum_{\varphi \in \Phi} \left( \frac{\log \left( \frac{1 + \frac{\mu \bar{V}_\varphi^\mu x}{B_\varphi^2 x}}{\log(1 + \varepsilon)} \right)}{\log(1 + \varepsilon)} + 1 \right) e^{-x} - \mathbb{P}((\Omega_{V,B} \cup \Omega_c)^c), \]
the following inequality holds
\[ |\psi(\tilde{f}_n) - \lambda|_{\psi^2} \leq C \inf_{a \in \mathbb{R}^\Phi} \left\{ |\psi(f_a) - \lambda|_{\psi^2} + \frac{1}{c} \sum_{\varphi \in S(a)} d_\varphi^2 \right\}, \]
where $C$ is an absolute positive constant and where $S(a)$ is the support of $a$, i.e. its coordinates with non-zero coefficients.
Proof. We use the notation of [2] and transposition of this notation. First by scaling the data, it is always possible to assume that $A = 1$. We have at hand $n \times M$ point processes $N_{m,i}^{(i)}$. In the more general case, we need to model each $\lambda^{(m,i)}$, intensity of $N_{m,i}^{(i)}$, by a

$$\psi_{f_n}^{(m,i)}(t) = \mu^{(m,i)} + \sum_{\ell,j} \int_{-\infty}^{t-} g_{\ell,j}^{(m,i)}(t-u) dN_{\ell,j}^{(i)}(u),$$

where $f_n$ belongs to $H_n$ which replaces the space $H$:

$$H_n = (\mathbb{R} \times L^2((0,1]))^{nM} = \left\{ f_n = \left( (\mu^{(m,i)}, g_{\ell,j}^{(m,i)})_{\ell=1,...,M, j=1,...,n} \right)_{m=1,...,M, i=1,...,n} : g_{\ell,j}^{(m,i)} \text{ with support in } (0,1) \text{ and } \|f_n\|^2 = \sum_{m,i} (\mu^{(m,i)})^2 + \sum_{m,i} \sum_{\ell,j} \int_{0}^{1} g_{\ell,j}^{(m,i)}(t)^2 dt < \infty \right\}.$$  

For every $f_n = \left( (\mu^{(m,i)}, g_{\ell,j}^{(m,i)})_{\ell=1,...,M, j=1,...,n} \right)_{m=1,...,M, i=1,...,n}$ in $H_n$, we denote for each $m$ and $i$,

$$f_n^{(m,i)} = (\mu^{(m,i)}, g_{\ell,j}^{(m,i)})_{\ell=1,...,M, j=1,...,n}.$$ 

In the same way for every $f = \left( (\mu^{(m)}, g_{\ell,j}^{(m)})_{\ell=1,...,M, j=1,...,n} \right)_{m=1,...,M}$ in $H$, we denote for each $m$ and $i$,

$$f^{(m)} = (\mu^{(m)}, g_{\ell,j}^{(m)})_{\ell=1,...,M, j=1,...,n}.$$ 

Now our dictionary $\Phi$ of $H$ can be transformed into a dictionary $\Phi_n$ of $H_n$ by stating that for any $\varphi$ in $\Phi$ we associate a $\varphi_n$ in $H_n$ such that for all $i,m$, $\varphi_n^{(m,i)} = \varphi^{(m)}$. Therefore it is easy to see that the vector $b$ of [2] associated to $f_n$ is actually our vector $b_n$ and that the matrix $G$ of [2] is our matrix $G_n$. The $V$ and $B$ are in the same way translated and the present result is a pure application of Theorem 2 of [2].

References
