Technical Supplement for 'Multiplicative component models for replicated point processes'

Daniel Gervini
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

June 13, 2018

1 Estimation

In Gervini (2017) we assume (X, Λ) is a doubly stochastic process with $X|\Lambda = \lambda \sim \mathscr{P}(\lambda)$ and Λ a random function that follows the model

$$\log \Lambda(t) = \mu(t) + \sum_{k=1}^{p} U_k \phi_k(t) \tag{1}$$

with $\mu \in L^2(B)$ and $\{\phi_k\}$ orthonormal in $L^2(B)$, where B is a fixed bounded region of $\mathscr{S}$ and $\mathscr{S}$ is $\mathbb{R}$ or $\mathbb{R}^2$. The U_ks are assumed independent $N(0, \sigma_k^2)$ random variables. Model (1) translates into a multiplicative model for $\Lambda(t)$,

$$\Lambda(t) = \lambda_0(t) \prod_{k=1}^{p} \xi_k(t)^{U_k}, \tag{2}$$

with $\lambda_0 = \exp \mu$ and $\xi_k = \exp \phi_k$. The functions μ and ϕ_ks are modeled in terms of basis functions $\{\beta_1, \ldots, \beta_q\}$. In the paper we use B-splines for temporal processes and radial Gaussian kernels for spatial processes, but there are other possibilities. Then we have $\mu(t) = \mathbf{c}_0^T \boldsymbol{\beta}(t)$ and $\phi_k(t) = \mathbf{c}_k^T \boldsymbol{\beta}(t)$, where $\boldsymbol{\beta}$ is the vector of the β_ks. From (1) we can express

$$\log \Lambda(t) = (\mathbf{c}_0 + \mathbf{C}\mathbf{U})^T \boldsymbol{\beta}(t)$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_p]$ and $\mathbf{U} = (U_1, \ldots, U_p)^T$. The orthonormality of the ϕ_ks can be expressed as $\mathbf{C}^T \mathbf{J}_0 \mathbf{C} = \mathbf{I}$, with $\mathbf{J}_0 = \int \boldsymbol{\beta}\boldsymbol{\beta}^T$.

The parameters $\mathbf{c}_0$ and $\mathbf{c}_k$s, along with the variances σ_k^2s of the U_ks, are estimated by penalized maximum likelihood. For notational simplicity, we collect all these parameters into a single vector $\boldsymbol{\theta}$.

Let x denote a realization of $X_B = X \cap B$, $x = \{t_1, \ldots, t_m\}$. Then the marginal density of X_B at x is

$$\begin{aligned} f(x; \boldsymbol{\theta}) &= \int\int f(x, \mathbf{u}) \, d\mathbf{u} \qquad (3) \\ &= \int\int f(x \mid \mathbf{u}) f(\mathbf{u}) \, d\mathbf{u} \end{aligned}$$

where

$$f(x \mid \mathbf{u}) = \exp\left\{-\int_B \lambda_{\mathbf{u}}(t)dt\right\} \frac{1}{m!} \prod_{j=1}^{m} \lambda_{\mathbf{u}}(t_j)$$

with $\lambda_{\mathbf{u}}$ denoting a λ that follows model (1) for a given $\mathbf{u}$. Then

$$\begin{aligned}\log f(x \mid \mathbf{u}) &= -\int_B \lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m} \log \lambda_{\mathbf{u}}(t_j) - \log m! \\ &= -\int_B \exp\{(\mathbf{c}_0 + \mathbf{C}\mathbf{u})^T \boldsymbol{\beta}(t)\}dt + (\mathbf{c}_0 + \mathbf{C}\mathbf{u})^T \sum_{j=1}^{m} \boldsymbol{\beta}(t_j) - \log m!.\end{aligned}$$

Note that this expression depends on the t_js only through

$$\mathbf{b} := \sum_{j=1}^{m} \boldsymbol{\beta}(t_j).$$

For $f(\mathbf{u})$ we have

$$f(\mathbf{u}) = \prod_{k=1}^{p} \frac{1}{(2\pi\sigma_k^2)^{1/2}} \exp\left(-\frac{u_k^2}{2\sigma_k^2}\right),$$

so

$$\log f(\mathbf{u}) = \sum_{k=1}^{p} \left(-\frac{1}{2}\log 2\pi\sigma_k^2 - \frac{u_k^2}{2\sigma_k^2}\right).$$

The penalized maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ based on n independent realizations $x_1, \ldots, x_n$ is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \rho_n(\boldsymbol{\theta})$$

where

$$\rho_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i; \boldsymbol{\theta}) - \nu_1 P(\mu) - \nu_2 \sum_{k=1}^{p} P(\phi_k) \tag{4}$$

for smoothing parameters ν_1 and ν_2. The roughness penalty functions $P(\mu)$ and $P(\phi_k)$ are quadratic forms $P(\mu) = \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0$ and $P(\phi_k) = \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k$ for a matrix $\boldsymbol{\Omega}$ that only depends on $\boldsymbol{\beta}$, as shown in Section 1.4. Two different νs are used because μ and the ϕ_ks may have different orders of magnitude (the ϕ_ks are comparable among themselves because they all have unit norms).

1.1 Derivatives

The derivatives of $\rho_n(\boldsymbol{\theta})$ are obtained as follows. First, note that if $\mathbf{c}$ generically represents $\mathbf{c}_0$ or a $\mathbf{c}_k$, then (omitting $\boldsymbol{\theta}$ from the notation of f)

$$\begin{aligned}
\nabla_{\mathbf{c}} \log f(x) &= \frac{1}{f(x)} \int\int \{\nabla_{\mathbf{c}} f(x \mid \mathbf{u})\} f(\mathbf{u})\, d\mathbf{u} \\
&= \frac{1}{f(x)} \int\int \{\nabla_{\mathbf{c}} \log f(x \mid \mathbf{u})\} f(x \mid \mathbf{u}) f(\mathbf{u})\, d\mathbf{u} \\
&= \int\int \{\nabla_{\mathbf{c}} \log f(x \mid \mathbf{u})\} f(\mathbf{u} \mid x)\, d\mathbf{u} \\
&= \mathbb{E}\{\nabla_{\mathbf{c}} \log f(x \mid \mathbf{U}) \mid x\}.
\end{aligned}$$

Since

$$\nabla_{\mathbf{c}} \log f(x \mid \mathbf{u}) = -\nabla_{\mathbf{c}} \int_B \lambda_{\mathbf{u}}(t) dt + \sum_{j=1}^{m} \nabla_{\mathbf{c}} \log \lambda_{\mathbf{u}}(t_j)$$

and

$$\begin{aligned}
\nabla_{\mathbf{c}} \int_B \lambda_{\mathbf{u}}(t) dt &= \int_B \nabla_{\mathbf{c}} \lambda_{\mathbf{u}}(t) dt \\
&= \int_B \{\nabla_{\mathbf{c}} \log \lambda_{\mathbf{u}}(t)\} \lambda_{\mathbf{u}}(t) dt,
\end{aligned}$$

we have

$$\nabla_{\mathbf{c}} \log f(x \mid \mathbf{u}) = -\int_B \{\nabla_{\mathbf{c}} \log \lambda_{\mathbf{u}}(t)\} \lambda_{\mathbf{u}}(t) dt + \sum_{j=1}^{m} \nabla_{\mathbf{c}} \log \lambda_{\mathbf{u}}(t_j).$$

Then

$$\nabla_{\mathbf{c}} \log f(x) = -\int_B \mathbb{E}\left[\{\nabla_{\mathbf{c}} \log \lambda_{\mathbf{U}}(t)\} \lambda_{\mathbf{U}}(t) \mid x\right] dt + \sum_{j=1}^{m} \mathbb{E}\left[\nabla_{\mathbf{c}} \log \lambda_{\mathbf{U}}(t_j) \mid x\right].$$

For the σ_k^2s the procedure is analogous, only that $f(\mathbf{u})$ is differentiated:

$$\begin{aligned}
\frac{\partial}{\partial \sigma_k^2} \log f(x) &= \frac{1}{f(x)} \int\int f(x \mid \mathbf{u}) \{ \frac{\partial}{\partial \sigma_k^2} f(\mathbf{u}) \} \, d\mathbf{u} \\
&= \frac{1}{f(x)} \int\int f(x \mid \mathbf{u}) \{ \frac{\partial}{\partial \sigma_k^2} \log f(\mathbf{u}) \} f(\mathbf{u}) \, d\mathbf{u} \\
&= \int\int \{ \frac{\partial}{\partial \sigma_k^2} \log f(\mathbf{u}) \} f(\mathbf{u} \mid x) \, d\mathbf{u} \\
&= \mathbb{E} \left\{ \frac{\partial}{\partial \sigma_k^2} \log f(\mathbf{U}) \mid x \right\}.
\end{aligned}$$

Since

$$\frac{\partial}{\partial \sigma_k^2} \log f(\mathbf{u}) = -\frac{1}{2\sigma_k^2} + \frac{u_k^2}{2(\sigma_k^2)^2},$$

we get

$$\frac{\partial}{\partial \sigma_k^2} \log f(x) = -\frac{1}{2\sigma_k^2} + \frac{\mathbb{E}\,(U_k^2 \mid x)}{2(\sigma_k^2)^2}.$$

1.2 Maximizing the penalized likelihood

We will use a Newton–Raphson algorithm on separated variables to maximize $\rho_n(\boldsymbol{\theta})$; that is, we will update the parameters $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_p, \sigma_1^2, \ldots, \sigma_p^2$ one at a time and in that order. So we discuss each of the three cases separately.

Updating $\mathbf{c}_0$: From above we have

$$\nabla_{\mathbf{c}_0} \rho_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \nabla_{\mathbf{c}_0} \log f(x_i) - \nu_1 2 \boldsymbol{\Omega} \mathbf{c}_0.$$

Since

$$\nabla_{\mathbf{c}_0} \log \lambda_{\mathbf{u}}(t) = \boldsymbol{\beta}(t),$$

then

$$\begin{aligned}
\nabla_{\mathbf{c}_0} \rho_n(\boldsymbol{\theta}) &= -\frac{1}{n} \sum_{i=1}^n \int_B \boldsymbol{\beta}(t) \mathbb{E} \{ \lambda_{\mathbf{U}}(t) \mid x_i \} \, dt + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \boldsymbol{\beta}(t_{ij}) - \nu_1 2 \boldsymbol{\Omega} \mathbf{c}_0 \\
&= -\frac{1}{n} \sum_{i=1}^n \int_B \boldsymbol{\beta}(t) \mathbb{E} \{ \lambda_{\mathbf{U}}(t) \mid x_i \} \, dt + \frac{1}{n} \sum_{i=1}^n \mathbf{b}_i - \nu_1 2 \boldsymbol{\Omega} \mathbf{c}_0
\end{aligned}$$

with

$$\mathbf{b}_i = \sum_{j=1}^{m_i} \boldsymbol{\beta}(t_{ij}).$$

For the second derivatives, it is best to use the approximation (which ignores the fact that $\mathbb{E}$ depends on $\boldsymbol{\theta}$)

$$\begin{aligned} \nabla_{\mathbf{c}} \mathbb{E}\left\{\lambda_{\mathbf{U}}(t) \mid x\right\} &\approx \mathbb{E}\left\{\nabla_{\mathbf{c}} \lambda_{\mathbf{U}}(t) \mid x\right\} \\ &= \mathbb{E}\left[\{\nabla_{\mathbf{c}} \log \lambda_{\mathbf{U}}(t)\} \lambda_{\mathbf{U}}(t) \mid x\right]. \end{aligned}$$

So, for the particular case of $\mathbf{c}_0$ we have

$$\mathsf{H}_{\mathbf{c}_0} \rho_n(\boldsymbol{\theta}) \approx -\frac{1}{n} \sum_{i=1}^{n} \int_B \boldsymbol{\beta}(t) \boldsymbol{\beta}(t)^T \mathbb{E}\left\{\lambda_{\mathbf{U}}(t) \mid x_i\right\} dt - \nu_1 2\boldsymbol{\Omega}.$$

Since $\mathbf{c}_0$ is unconstrained, we use an ordinary Newton update:

$$\mathbf{c}_0^{\text{new}} = \mathbf{c}_0^{\text{old}} - \{\mathsf{H}_{\mathbf{c}_0} \rho_n(\boldsymbol{\theta}^{\text{old}})\}^{-1} \nabla_{\mathbf{c}_0} \rho_n(\boldsymbol{\theta}^{\text{old}}).$$

<u>Updating the $\mathbf{c}_k$s:</u> Now we have

$$\nabla_{\mathbf{c}_k} \rho_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \nabla_{\mathbf{c}_k} \log f(x_i) - \nu_2 2\boldsymbol{\Omega} \mathbf{c}_k.$$

Since

$$\nabla_{\mathbf{c}_k} \log \lambda_{\mathbf{u}}(t) = u_k \boldsymbol{\beta}(t),$$

it follows that

$$\begin{aligned} \nabla_{\mathbf{c}_k} \rho_n(\boldsymbol{\theta}) &= -\frac{1}{n} \sum_{i=1}^{n} \int_B \boldsymbol{\beta}(t) \mathbb{E}\left\{U_k \lambda_{\mathbf{U}}(t) \mid x_i\right\} dt + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbb{E}\left(U_k \mid x_i\right) \boldsymbol{\beta}(t_{ij}) - \nu_2 2\boldsymbol{\Omega} \mathbf{c}_k \\ &= -\frac{1}{n} \sum_{i=1}^{n} \int_B \boldsymbol{\beta}(t) \mathbb{E}\left\{U_k \lambda_{\mathbf{U}}(t) \mid x_i\right\} dt + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(U_k \mid x_i\right) \mathbf{b}_i - \nu_2 2\boldsymbol{\Omega} \mathbf{c}_k \end{aligned}$$

For the second derivative we proceed as before:

$$\begin{aligned} \nabla_{\mathbf{c}_k} \mathbb{E}\left\{U_k \lambda_{\mathbf{U}}(t) \mid x_i\right\} &\approx \mathbb{E}\left\{U_k \nabla_{\mathbf{c}_k} \lambda_{\mathbf{U}}(t) \mid x_i\right\} \\ &= \mathbb{E}\left[U_k\{\nabla_{\mathbf{c}_k} \log \lambda_{\mathbf{U}}(t)\} \lambda_{\mathbf{U}}(t) \mid x_i\right] \\ &= \mathbb{E}\left[U_k^2 \boldsymbol{\beta}(t) \lambda_{\mathbf{U}}(t) \mid x_i\right], \end{aligned}$$

so

$$\mathsf{H}_{\mathbf{c}_k} \rho_n(\boldsymbol{\theta}) \approx -\frac{1}{n} \sum_{i=1}^n \int_B \boldsymbol{\beta}(t) \boldsymbol{\beta}(t)^T \mathbb{E}\left\{U_k^2 \lambda_{\mathbf{U}}(t) \mid x_i\right\} dt - \nu_2 2 \boldsymbol{\Omega}.$$

To update the $\mathbf{c}_k$s we use projected Newton steps due to the constraints. This is done sequentially as follows: given that $\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}$ have already been updated, let $\boldsymbol{\Gamma}$ be a $q \times \{q-(k-1)\}$ orthonormal basis of $\{\mathbf{J}_0 \mathbf{c}_1^{\text{new}}, \ldots, \mathbf{J}_0 \mathbf{c}_{k-1}^{\text{new}}\}^{\perp}$; then

$$\begin{aligned} \tilde{\mathbf{c}}_k^{\text{new}} &= \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T \mathbf{c}_k^{\text{old}} - \boldsymbol{\Gamma}\{\boldsymbol{\Gamma}^T \mathsf{H}_{\mathbf{c}_k} \rho_n(\boldsymbol{\theta}^{\text{old}}) \boldsymbol{\Gamma}\}^{-1} \boldsymbol{\Gamma}^T \nabla_{\mathbf{c}_k} \rho_n(\boldsymbol{\theta}^{\text{old}}), \\ \mathbf{c}_k^{\text{new}} &= \tilde{\mathbf{c}}_k^{\text{new}} / \{(\tilde{\mathbf{c}}_k^{\text{new}})^T \mathbf{J}_0 \tilde{\mathbf{c}}_k^{\text{new}}\}^{1/2}. \end{aligned} \quad (5)$$

Updating the σ_k^2s: Now

$$\begin{aligned} \frac{\partial}{\partial \sigma_k^2} \rho_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \sigma_k^2} \log f(x_i) \\ &= -\frac{1}{2\sigma_k^2} + \frac{1}{2(\sigma_k^2)^2} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left(U_k^2 \mid x_i\right) \end{aligned}$$

and then we get, explicitly,

$$(\sigma_k^2)^{\text{new}} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left(U_k^2 \mid x_i\right).$$

1.3 Initialization

To initialize the mean we take a constant vector $\mathbf{c}_0 = c\mathbf{1}$ as initial estimator of $\mathbf{c}_0$. Since B-splines and normalized Gaussian kernel bases satisfy $\boldsymbol{\beta}(t)^T \mathbf{1} = 1$, this is basically like assuming that $\mu(t) \equiv c$ and then (for a mean-only model) that $\lambda_0(t) \equiv e^c$. Then $\int_B \lambda_0(t) dt = e^c |B|$. Since, for a mean-only model, the m_is are i.i.d. $\mathscr{P}(\int_B \lambda_0)$, we can estimate $e^c |B|$ by $\overline{m}$ and then we can take $\hat{c} = \log(\overline{m}/|B|)$.

Similarly, to initialize the first component we take $\mathbf{c}_1 = \alpha \mathbf{1}$, which implies $\phi_1(t) \equiv$

α and then $\|\phi_1\|^2 = \int_B \alpha^2 = \alpha^2 |B| = 1$, so $\alpha = 1/|B|^{1/2}$. For the initial one-component model we have $\lambda_i(t) = \lambda_0(t)e^{u_{i1}\alpha}$, so $\int \lambda_i = e^{u_{i1}\alpha} \int \lambda_0$. Now, since m_i estimates $\int \lambda_i$, we can take $\hat{u}_{i1} = \log(m_i / \int \lambda_0)/\alpha$. The variance σ_1^2 is then estimated as the sample variance of the $\hat{u}_{i1}$s.

For subsequent components we simply take $\hat{\sigma}_k^2 = \hat{\sigma}_{k-1}^2/2$ and $\hat{\mathbf{c}}_k = \tilde{\mathbf{c}}_k/(\tilde{\mathbf{c}}_k^T \mathbf{J}_0 \tilde{\mathbf{c}}_k)^{1/2}$ with $\tilde{\mathbf{c}}_k$ the orthogonal projection of $\mathbf{1}$ on the space $\{\mathbf{J}_0\mathbf{c}_1, \ldots, \mathbf{J}_0\mathbf{c}_{k-1}\}^\perp$. That is, if $\mathbf{\Gamma}$ is a $q \times \{q-(k-1)\}$ orthonormal basis of $\{\mathbf{J}_0\mathbf{c}_1, \ldots, \mathbf{J}_0\mathbf{c}_{k-1}\}^\perp$, then $\tilde{\mathbf{c}}_k = \mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{1}$.

1.4 Roughness penalty functions

In (4) we use roughness penalties of the form $P(g) = \int_B \|\mathsf{H}g(t)\|_F^2 \; dt$, where H denotes the Hessian and $\|\cdot\|_F$ the Frobenius matrix norm. Then $P(g) = \int_B (g'')^2$ for a temporal process and $P(g) = \int_B \{(\frac{\partial^2 g}{\partial t_1^2})^2 + 2(\frac{\partial^2 g}{\partial t_1 \partial t_2})^2 + (\frac{\partial^2 g}{\partial t_2^2})^2\}$ for a spatial process. In either case, when $g(t) = \mathbf{c}^T\boldsymbol{\beta}(t)$ for some basis $\boldsymbol{\beta}(t)$, the function $P(g)$ is quadratic in $\mathbf{c}$: $P(g) = \mathbf{c}^T\mathbf{\Omega}\mathbf{c}$.

For the temporal case we simply have $\mathbf{\Omega} = \int_B \boldsymbol{\beta}''(t)\boldsymbol{\beta}''(t)^T dt$, where the second derivatives are taken in a component-wise way.

For the spatial case we can decompose

$$\mathbf{\Omega} = \mathbf{\Omega}_{11} + 2\mathbf{\Omega}_{12} + \mathbf{\Omega}_{22}$$

where

$$\mathbf{\Omega}_{ij} = \int_B \boldsymbol{\beta}^{ij}(t)\boldsymbol{\beta}^{ij}(t)^T dt$$

and $\boldsymbol{\beta}^{ij}(t)$ denotes the vector of second derivatives of $\boldsymbol{\beta}^{ij}(t)$ with respect to t_i and t_j.

1.5 Laplace approximation of integrals

The marginal densities $f(x)$ are computed by Laplace approximation. We have

$$\begin{aligned} f(x) &= \iint f(x \mid \mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= \iint \exp g(\mathbf{u}) d\mathbf{u} \end{aligned}$$

with

$$g(\mathbf{u}) = \log f(x \mid \mathbf{u}) + \log f(\mathbf{u}).$$

If $\hat{\mathbf{u}} = \arg\max g(\mathbf{u})$ then $g(\mathbf{u}) \approx g(\hat{\mathbf{u}}) + .5(\mathbf{u} - \hat{\mathbf{u}})^T \mathsf{H} g(\hat{\mathbf{u}})(\mathbf{u} - \hat{\mathbf{u}})$ and

$$f(x) \approx \exp\{g(\hat{\mathbf{u}})\}(2\pi)^{p/2} \det(\mathbf{S})^{1/2}$$

with

$$\mathbf{S} = \{-\mathsf{H} g(\hat{\mathbf{u}})\}^{-1}.$$

In effect, we are approximating

$$f(x \mid \mathbf{u}) f(\mathbf{u}) \approx \exp\{g(\hat{\mathbf{u}})\}(2\pi)^{p/2} \det(\mathbf{S})^{1/2} \varphi_{(\hat{\mathbf{u}},\mathbf{S})}(\mathbf{u})$$

where $\varphi_{(\hat{\mathbf{u}},\mathbf{S})}(\mathbf{u})$ denotes the pdf of a $N_p(\hat{\mathbf{u}}, \mathbf{S})$, so $\mathbf{U} \mid x \approx N_p(\hat{\mathbf{u}}, \mathbf{S})$. Then we can also approximate the moments:

$$\begin{aligned} E(\mathbf{U} \mid x) &\approx \hat{\mathbf{u}}, \\ E(\mathbf{U}\mathbf{U}^T \mid x) &\approx \mathbf{S} + \hat{\mathbf{u}}\hat{\mathbf{u}}^T. \end{aligned}$$

We find $\hat{\mathbf{u}}$ by (a few steps of) Newton–Raphson for each x_i. Since

$$g(\mathbf{u}) = -\int \lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m} \log \lambda_{\mathbf{u}}(t_j) - \log m! + \sum_{k=1}^{p} \left(-\frac{1}{2} \log 2\pi\sigma_k^2 - \frac{u_k^2}{2\sigma_k^2} \right)$$

the derivatives with respect to $\mathbf{u}$ are, with $\mathbf{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma}^2)$,

$$\nabla g(\mathbf{u}) = -\int \lambda_{\mathbf{u}}(t)\boldsymbol{\phi}(t)dt + \sum_{j=1}^{m} \boldsymbol{\phi}(t_j) - \mathbf{\Sigma}^{-1}\mathbf{u}$$

and

$$\mathsf{H} g(\mathbf{u}) = -\int \lambda_{\mathbf{u}}(t)\boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^T dt - \mathbf{\Sigma}^{-1}.$$

1.6 Cyclic border conditions

For temporal processes on an interval $[a, b]$, e.g. $[0, 24]$ for daily processes, it is often natural to require that $\lambda_i(a) = \lambda_i(b)$. A sufficient condition for this is that $\mu(a) = \mu(b)$ and $\phi_k(a) = \phi_k(b)$ for all k. In terms of the basis coefficients we have $\mathbf{a}^T \mathbf{c}_k = \mathbf{0}$, $k = 0, \ldots, p$, with $\mathbf{a} = \boldsymbol{\beta}(a) - \boldsymbol{\beta}(b)$. So we modify the N–R algorithms for the $\hat{\mathbf{c}}_k$s by projecting on $\mathbf{a}^\perp$. That is, the N–R updates for $\hat{\mathbf{c}}_0$ have the form

$$\mathbf{c}_0^{\text{new}} = \mathbf{c}_0^{\text{old}} - \mathbf{P}\{\mathbf{P}^T \mathsf{H}_{\mathbf{c}_0}\rho_n(\boldsymbol{\theta}^{\text{old}})\mathbf{P}\}^{-1}\mathbf{P}^T \nabla_{\mathbf{c}_0}\rho_n(\boldsymbol{\theta}^{\text{old}})$$

with $\mathbf{P}$ a $q \times (q-1)$ matrix whose columns form an orthonormal basis of $\mathbf{a}^\perp$. For the $\hat{\mathbf{c}}_k$s we use formula (5) with $\boldsymbol{\Gamma}$ an orthonormal basis of $\{\mathbf{a}, \mathbf{J}_0\mathbf{c}_1^{\text{new}}, \ldots, \mathbf{J}_0\mathbf{c}_{k-1}^{\text{new}}\}^\perp$.

Acknowledgement

This research was partly supported by US National Science Foundation grant DMS 1505780.

References

Gervini, D. (2017). Multiplicative component models for replicated point processes. *ArXiv* 1705.09693.