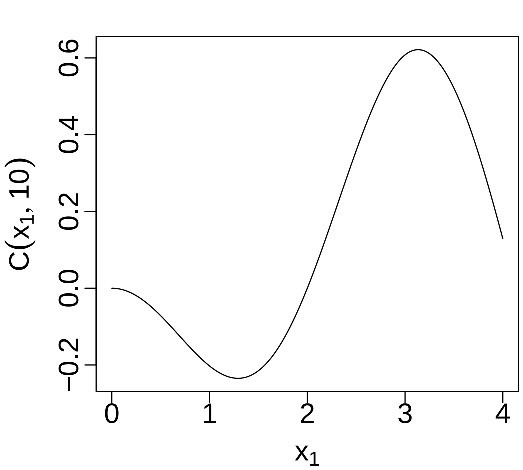
Online Supplement to: Efficient Modeling of Quasi-Periodic Data with Seasonal Gaussian Process

Ziang Zhang
Department of Statistical Sciences, University of Toronto
and
Patrick Brown
Department of Statistical Sciences, University of Toronto
Centre for Global Health Research, St Michael's Hospital
and
Jamie Stafford
Department of Statistical Sciences, University of Toronto

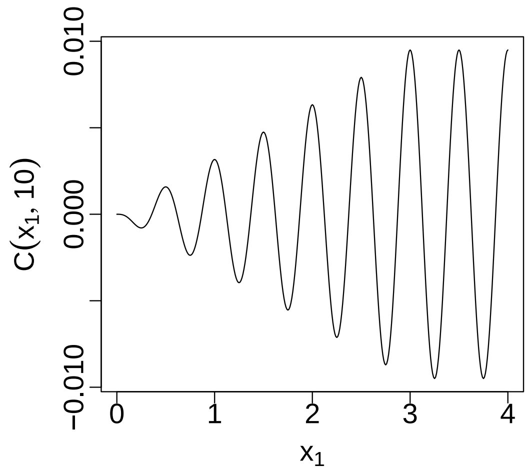
S1. Additional figures and table

	0.025 Quantile	Median	0.975 Quantile
$\sigma_{tr}(10)$	0.040 (15.336)	0.960 (37.655)	9.897 (106.646)
$\sigma_1(10)$	0.126 (–)	0.270 (–)	0.509 (–)
$\sigma_{\frac{1}{2}}(10)$	0.004 (–)	0.054 (–)	0.146 (–)
$\sigma_{\frac{44}{12}}(10)$	0.162 (–)	0.518 (–)	0.889 (–)
$\sigma_{9.1}(10)$	0.025 (–)	0.391 (–)	1.429 (–)
$\sigma_{10.4}(10)$	0.022 (–)	0.430 (–)	1.568 (–)
σ_{ϵ}	0.553 (0.582)	0.584 (0.615)	0.620 (0.651)

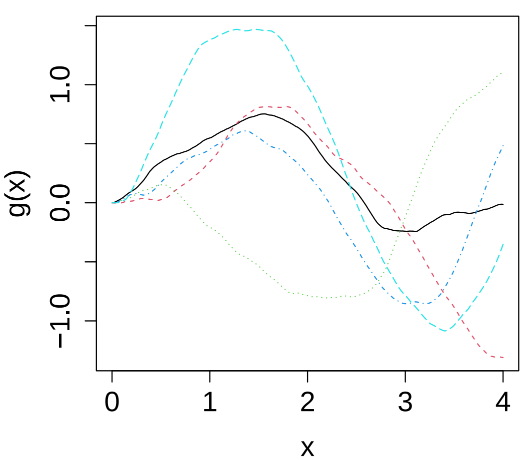
Table S1: Posterior summary of standard deviation parameters for the CO2 example in Section 5.4. Results from M2 are shown in parenthesis.



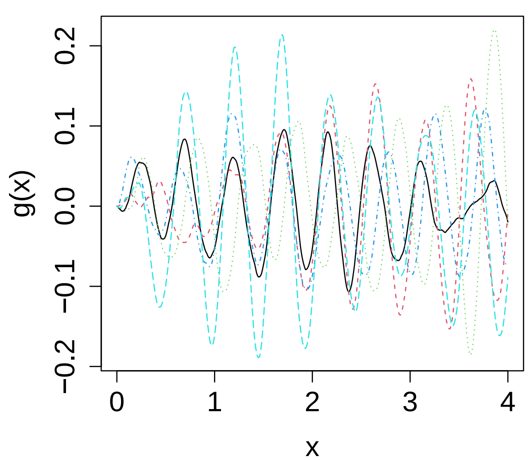
(a) $\alpha = \pi/2$: Covariance



(b) $\alpha = 4\pi$: Covariance

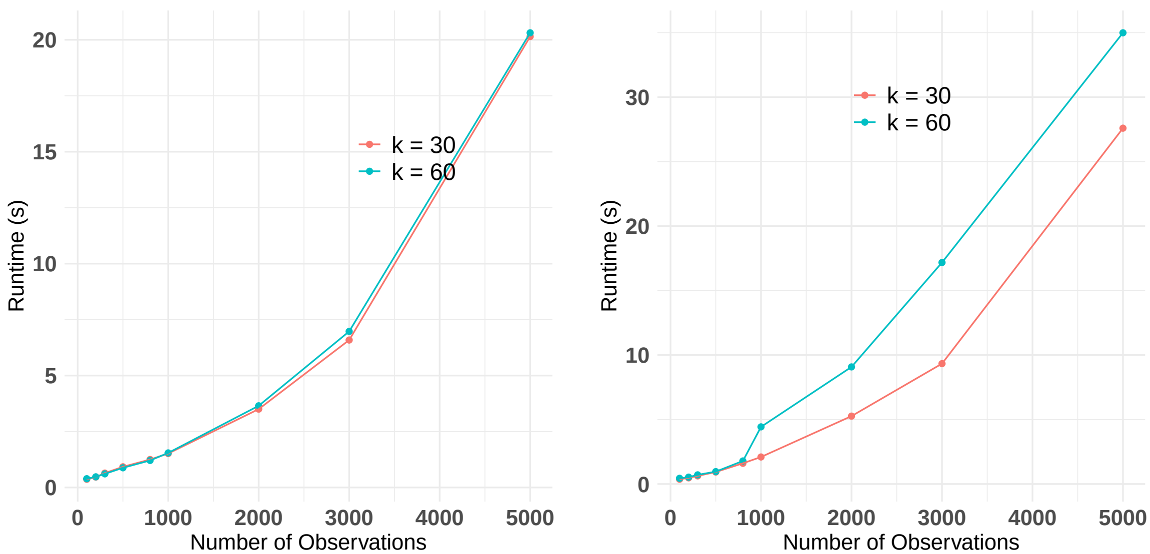


(c) $\alpha = \pi/2$: Samples



(d) $\alpha = 4\pi$: Samples

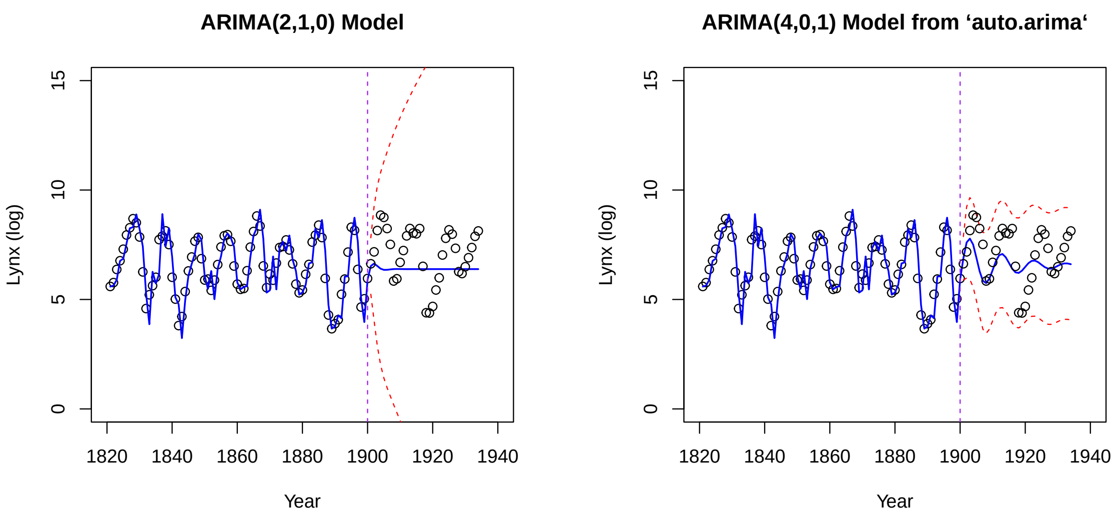
Figure S1: Figures (a,b) show covariance functions of the sGP with different α, where the first argument in the covariance function is fixed at 3. Figures (c,d) display five sample paths from the two sGPs. The frequency parameter α equals to $\pi/2$ in (a,c) and 4π in (b,d), and the SD parameter $\sigma = 1$ in both sGPs.



(a) Easy Setting (A)

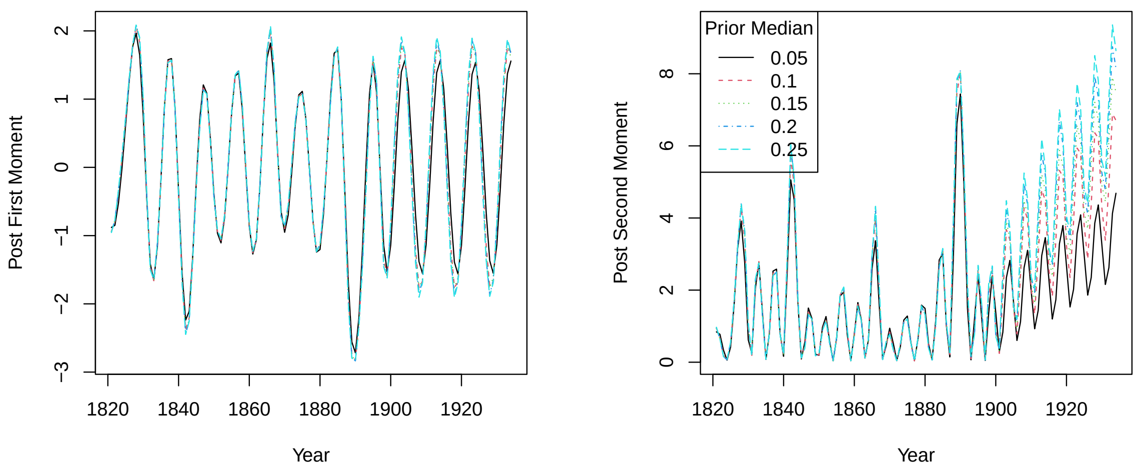
(b) Hard Setting (B)

Figure S2: Average runtimes using the sB-spline approximation with $k = 30$ (red) and $k = 60$ (blue) for the two simulation settings in Section 4.1, when the sample size n varies. Each average is computed from ten replications.



(a) (b)

Figure S3: Additional comparisons for Section 4.3. Frequentist ARIMA models fitted using maximum likelihood estimation: (a) ARIMA with order fixed at (2,1,0); (b) ARIMA with optimal order (4,0,1) selected based on AIC.



(a) First Moment (b) Second Moment

Figure S4: Additional sensitivity analysis for Section 4.3. Posterior moments obtained from the sGP model, with the median in the PSD prior varying over a range of possible values.

S2. Derivation of the sGP covariance

Proposition S1 (Covariance Function of the Seasonal Gaussian Process)**.** *Let* $g \sim sGP_\alpha(\sigma)$*. Then* g *has a covariance function:*

$$\begin{aligned} C(x_1, x_2) &= \left(\frac{\sigma}{\alpha}\right)^2 \left[\frac{x_1}{2}\cos(\alpha(x_2 - x_1)) - \frac{\cos(\alpha x_2)\sin(\alpha x_1)}{2\alpha}\right] \\ &= \left(\frac{\sigma}{\alpha}\right)^2 \left[\frac{\cos(\alpha x_2)x_1}{2}\cos(\alpha x_1) + \left(\frac{\sin(\alpha x_2)x_1}{2} - \frac{\cos(\alpha x_2)}{2\alpha}\right)\sin(\alpha x_1)\right], \end{aligned} \tag{1}$$

for any $x_1, x_2 \in \Omega$ *such that* $x_1 \leq x_2$*.*

Proof. It is obvious that the differential operator L is linear. Define $\boldsymbol{g}_{aug}(x) = (g(x), g'(x))^T$ and therefore $\boldsymbol{g}'_{aug}(x) = (g'(x), g''(x))^T$, then the SDE can be rewritten in the vector form:

$$\boldsymbol{g}'_{aug} = \mathbf{F}\boldsymbol{g}_{aug} + \mathbf{J}W, \tag{2}$$

where $\mathbf{F} = \begin{bmatrix} 0 & 1 \\ -\alpha^2 & 0 \end{bmatrix}$ and $\mathbf{J} = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}$.

Using the result from Särkkä and Solin (2019) (section 4.3), the solution of the linear SDE can be written as:

$$\begin{aligned} \boldsymbol{g}_{aug}(x) &= \exp(\mathbf{F}x)\boldsymbol{g}_{aug}(0) + \int_0^x \exp(\mathbf{F}(x-\tau))\mathbf{J}W(\tau)d\tau \\ &= \int_0^x \exp(\mathbf{F}(x-\tau))\mathbf{J}W(\tau)d\tau, \end{aligned} \tag{3}$$

where $\exp(\mathbf{F}x)$ denotes the matrix exponential defined as $\exp(\mathbf{F}x) = \sum_k \frac{\mathbf{F}^k x^k}{k!}$.

Note that $\mathbf{F}^{2k} = (-\alpha^2)^k \mathbf{I}$ and $\mathbf{F}^{2k} = (-\alpha^2)^k \mathbf{F}$. With Taylor series, the first component of $\boldsymbol{g}_{aug}(x)$ can be therefore written as:

$$g(x) = \int_0^x \frac{\sigma}{\alpha}\sin(\alpha(x-\tau))W(\tau)d\tau. \tag{4}$$

Assume arbitrary $0 < x_1 \leq x_2$, the covariance function can be computed for g as:

$$\begin{aligned} C(x_1, x_2) &= \int_0^{x_1} \frac{\sigma}{\alpha}\sin(\alpha(x_1-\tau))\frac{\sigma}{\alpha}\sin(\alpha(x_2-\tau))d\tau \\ &= \left(\frac{\sigma}{\alpha}\right)^2 \left[\frac{x_1}{2}\cos(\alpha(x_2 - x_1)) - \frac{\cos(\alpha x_2)\sin(\alpha x_1)}{2\alpha}\right], \end{aligned} \tag{5}$$

using properties of Gaussian white noise (Harvey, 1990).

□

S3. Proof of the State-Space Representation

Theorem S1 (State Space Representation of the sGP). *Consider $g \sim sGP_\alpha(\sigma)$, and let $\boldsymbol{s} = \{s_1, ..., s_n\} \subset \Omega$ be sorted with spacing $d_1 = s_1$ and $d_i = s_i - s_{i-1}$ for $i \in \{2, .., n\}$. Then $\boldsymbol{g}_{aug}(s_i) = [g(s_i), g'(s_i)]^T$ can be written as a Markov model:*

$$\boldsymbol{g}_{aug}(s_{i+1}) = \boldsymbol{R}_{i+1}\boldsymbol{g}_{aug}(s_i) + \boldsymbol{\epsilon}_{i+1}, \tag{6}$$

where $\boldsymbol{\epsilon}_i \overset{ind}{\sim} N(0, \boldsymbol{\Sigma}_i)$. The 2×2 matrices $\boldsymbol{R}_i$ and $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i^{-1}$ are respectively defined as:

$$\boldsymbol{R}_i = \begin{bmatrix} \cos(\alpha d_i) & \frac{1}{\alpha}\sin(\alpha d_i) \\ -\alpha\sin(\alpha d_i) & \cos(\alpha d_i) \end{bmatrix}, \ \boldsymbol{\Sigma}_i = \sigma^2 \begin{bmatrix} \frac{1}{\alpha^2}\left(\frac{d_i}{2} - \frac{\sin(2\alpha d_i)}{4\alpha}\right) & \frac{\sin^2(\alpha d_i)}{2\alpha^2} \\ \frac{\sin^2(\alpha d_i)}{2\alpha^2} & \frac{2\alpha d_i + \sin(2\alpha d_i)}{4\alpha} \end{bmatrix}. \tag{7}$$

Proof. To show the above Markov representation, note that the value of $g(s_{i+1})$ given $g(s_i)$ can be written similarly as (Särkkä and Solin, 2019):

$$\boldsymbol{g}_{aug}(s_{i+1}) = \exp(\mathbf{F}d_{i+1})\boldsymbol{g}_{aug}(s_i) + \int_{s_i}^{s_{i+1}} \exp(\mathbf{F}(s_{i+1} - \tau))\mathbf{J}W(\tau)d\tau.$$

Recall that $\mathbf{F}^{2k} = (-\alpha^2)^k\mathbf{I}$ and $\mathbf{F}^{2k+1} = (-\alpha^2)^k\mathbf{F}$, then apply the Taylor series expansion for both components in the integral above. It then can be rewritten as:

$$\begin{aligned} \boldsymbol{g}_{aug}(s_{i+1}) &= \exp(\mathbf{F}d_{i+1})\boldsymbol{g}_{aug}(s_i) + \int_{s_i}^{s_{i+1}} \exp(\mathbf{F}(s_{i+1} - \tau))\mathbf{J}W(\tau)d\tau \\ &= \boldsymbol{R}_{i+1}\boldsymbol{g}_{aug}(s_i) + \int_{s_i}^{s_{i+1}} \begin{bmatrix} \frac{1}{\alpha}\sin\left(\alpha(s_{i+1} - \tau)\right) \\ \cos\left(\alpha(s_{i+1} - \tau)\right) \end{bmatrix} \sigma W(\tau)d\tau \\ &:= \boldsymbol{R}_{i+1}\boldsymbol{g}_{aug}(s_i) + \boldsymbol{\epsilon}_{i+1}. \end{aligned} \tag{8}$$

Note that since each $\boldsymbol{\epsilon}_{i+1}$ involves integration at disjoint intervals, their independence follows from the property of Gaussian white noise (Harvey, 1990). To check its covariance matrix $\boldsymbol{\Sigma_{i+1}}$, note that:

$$\begin{aligned} \boldsymbol{\Sigma_{i+1}} &= \sigma^2 \begin{bmatrix} \int_{s_i}^{s_{i+1}} \frac{1}{\alpha^2}\sin^2\left(\alpha(s_{i+1} - \tau)\right)d\tau & \frac{1}{\alpha}\int_{s_i}^{s_{i+1}} \sin\left(\alpha(s_{i+1} - \tau)\right)\cos\left(\alpha(s_{i+1} - \tau)\right)d\tau \\ \frac{1}{\alpha}\int_{s_i}^{s_{i+1}} \sin\left(\alpha(s_{i+1} - \tau)\right)\cos\left(\alpha(s_{i+1} - \tau)\right)d\tau & \int_{s_i}^{s_{i+1}} \cos^2\left(\alpha(s_{i+1} - \tau)\right) \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} \frac{1}{\alpha^2}\left(\frac{d_{i+1}}{2} - \frac{\sin(2\alpha d_{i+1})}{4\alpha}\right) & \frac{\sin^2(\alpha d_{i+1})}{2\alpha^2} \\ \frac{\sin^2(\alpha d_{i+1})}{2\alpha^2} & \frac{2\alpha d_{i+1} + \sin(2\alpha d_{i+1})}{4\alpha} \end{bmatrix}, \end{aligned} \tag{9}$$

which completes the proof.

□

S4. Details of the Finite Element Method

The Finite Element Method (FEM) used to construct the finite-dimensional approximation can be understood as the following procedures.

Given the stochastic differential equation (SDE) that defines the (standard) sGP model:

$$LW(x) = \xi(x),$$

where $L = \alpha^2 + \frac{d^2}{dx^2}$ is a linear differential operator and $\xi(x)$ is the standard Gaussian white noise process. Let $\Omega \subset \mathbb{R}^+$ denotes a bounded interval of interest. Let $\mathbb{B}_k := \{\psi_i, i \in [k]\}$ denote the set of k pre-specified basis functions, and let $\mathbb{T}_q := \{\phi_i, i \in [q]\}$ denote the set of q pre-specified test functions. We consider finite dimensional approximation with form $\widetilde{W}_k(.) = \sum_{i=1}^k w_i\psi_i(.)$. The weights $\boldsymbol{w} := [w_1, ..., w_k]^T \in \mathbb{R}^k$ is a set of random weights to be determined.

In our FEM construction, we used the sB-splines defined over Ω as the basis functions, and chose the test functions by $\mathbb{T}_k := \{\phi_i = L\psi_i, i \in [k]\}$, which is called a least squares approximation in Lindgren et al. (2011). The distribution of the unknown weight vector can be found by fulfilling the weak formulation at the test function spaces $\mathbb{T}_k$, such that

$$\langle L\widetilde{W}_k(x), \phi_i(x)\rangle \stackrel{d}{=} \langle \xi(x), \phi_i(x)\rangle, \tag{10}$$

for any test function $\phi_i \in \mathbb{T}_k$. This equation can also be vectorized as:

$$\langle L\widetilde{W}_k(x), \phi_i(x)\rangle_{i=1}^k = H\boldsymbol{w},$$

where the ij component of the $k \times k$ H matrix can be computed as $H_{ij} = \langle L\psi_j(x), L\psi_i(x)\rangle_{i=1}^k$.

The inner product on the right $\langle \xi(x), \phi_i(x)\rangle_{i=1}^k$ will have Gaussian distribution with zero mean vector and covariance matrix H by properties of Gaussian white noise (Harvey, 1990). Therefore, the basis coefficients $\boldsymbol{w}$ will be multivariate Gaussian with zero mean and covariance $H^{-1}HH^{-1} = H^{-1}$. Each element of the matrix H can be written as:

$$\begin{aligned} H_{ij} &= \langle L\psi_j, L\psi_i \rangle \\ &= \langle a^2\psi_j + \frac{d^2\psi_j}{dx^2}, a^2\psi_i + \frac{d^2\psi_i}{dx^2} \rangle \\ &= a^4\langle \psi_j, \psi_i \rangle + a^2\langle \frac{d^2\psi_j}{dx^2}, \psi_i \rangle + a^2\langle \psi_j, \frac{d^2\psi_i}{dx^2} \rangle + \langle \frac{d^2\psi_j}{dx^2}, \frac{d^2\psi_i}{dx^2} \rangle, \end{aligned} \tag{11}$$

hence $H = a^4G + C + a^2M$ with $G_{ij} = \langle \psi_i, \psi_j \rangle$, $C_{ij} = \langle \frac{d^2\psi_i}{dx^2}, \frac{d^2\psi_j}{dx^2} \rangle$ and $M_{ij} = \langle \psi_i, \frac{d^2\psi_j}{dx^2} \rangle + \langle \frac{d^2\psi_i}{dx^2}, \psi_j \rangle$ for each element of the matrices.

S5. Proof of the Convergence Result

Theorem (Covariance Convergence of B-spline Approximation)**.** *Assume* $\mathbb{B}_k$ *is a set of* k *cubic B-splines constructed with equally spaced knots over* Ω*, and* $\tilde{g}_k$ *denotes the corresponding FEM approximation for* $sGP_\alpha(\sigma)$*, then for any* $x_1, x_2 \in \Omega$*:*

$$|\mathcal{C}_k(x_1, x_2) - \mathcal{C}(x_1, x_2)| = O(1/k),$$

where $\mathcal{C}(x_1, x_2)$ *is the covariance in Proposition S1 and* $\mathcal{C}_k(x_1, x_2) = \mathrm{Cov}[\tilde{g}_k(x_1), \tilde{g}_k(x_2)]$.

Proof. The proof of this theorem starts with a similar strategy as in Lindgren et al. (2011). Without the loss of generality, we assume the variance parameter of the sGP $\sigma = 1$, $\Omega = [0, 1]$ and the initial conditions of the sGP are zero. We denote the 3rd order Sobolev space as $H^3(\Omega) = \{f \in \mathcal{L}^2(\Omega) : D^q f \in \mathcal{L}^2(\Omega)\ \forall |q| \leq 3\}$ and the constrained Sobolev space $\mathcal{H}$ as:

$$\mathcal{H} = \{f \in H^3(\Omega) : f(0) = f'(0) = 0\} \subset H^3(\Omega).$$

Since $Lf = 0$ implies $f \in \mathrm{span}\{\cos(\alpha x), \sin(\alpha x)\}$, it is clear that

$$\langle f, h \rangle_{\mathcal{H}} := \langle Lf, Lh \rangle_{\Omega} = \int_{\Omega} Lf(x) Lh(x) dx \tag{12}$$

defines an inner product for $f, h \in \mathcal{H}$.

Define $\mathcal{H}_k = \mathrm{span}\{\mathbb{B}_k\}$, and note $\mathcal{H}_k \subset \mathcal{H}$ by our construction of the B-spline basis. Since $\mathcal{H}_k$ is a finite-dimensional subspace, for each $f(x) \in \mathcal{H}$ there exists an unique projection $\tilde{f}(x) = \sum_{i=1}^{k} w_i \psi_i(x) \in \mathcal{H}_k$ which satisfies:

$$\left\langle f - \tilde{f}, \tilde{h} \right\rangle_{\mathcal{H}} = \left\langle f, \tilde{h} \right\rangle_{\mathcal{H}} - \left\langle \tilde{f}, \tilde{h} \right\rangle_{\mathcal{H}} = 0,\ \forall \tilde{h} \in \mathcal{H}_k. \tag{13}$$

Based on the property of Gaussian white noise (Harvey, 1990), for any $f, h \in \mathcal{H}$ we have:

$$\mathrm{Cov}[\langle \xi, Lf\rangle_\Omega, \langle \xi, Lh\rangle_\Omega] = \langle Lf, Lh\rangle_\Omega = \langle f, h\rangle_{\mathcal{H}}. \tag{14}$$

Since the FEM approximation $\tilde{g}_k \in \mathcal{H}_k$, we know

$$\langle \tilde{g}_k, f\rangle_{\mathcal{H}} = \left\langle \tilde{g}_k, f - \tilde{f} + \tilde{f}\right\rangle_{\mathcal{H}} = \left\langle \tilde{g}_k, \tilde{f}\right\rangle_{\mathcal{H}} + \left\langle \tilde{g}_k, f - \tilde{f}\right\rangle_{\mathcal{H}} = \left\langle \tilde{g}_k, \tilde{f}\right\rangle_{\mathcal{H}},$$

where the last equality follows as $\tilde{f}$ is the projection of f. Using this result and the fact that the FEM approximation $\tilde{g}_k$ is a least square solution, we have

$$\begin{aligned}
\mathrm{Cov}\left[\langle \tilde{g}_k, f\rangle_{\mathcal{H}}, \langle \tilde{g}_k, h\rangle_{\mathcal{H}}\right] &= \mathrm{Cov}\left[\left\langle \tilde{g}_k, \tilde{f}\right\rangle_{\mathcal{H}}, \left\langle \tilde{g}_k, \tilde{h}\right\rangle_{\mathcal{H}}\right] \\
&= \mathrm{Cov}\left[\left\langle L\tilde{g}_k, L\tilde{f}\right\rangle_{\Omega}, \left\langle L\tilde{g}_k, L\tilde{h}\right\rangle_{\Omega}\right] \\
&= \mathrm{Cov}\left[\left\langle \xi, L\tilde{f}\right\rangle_{\Omega}, \left\langle \xi, L\tilde{h}\right\rangle_{\Omega}\right] \\
&= \left\langle L\tilde{f}, L\tilde{h}\right\rangle_{\Omega} = \left\langle \tilde{f}, \tilde{h}\right\rangle_{\mathcal{H}}.
\end{aligned} \tag{15}$$

Let $\mathcal{C}_s(x) = \mathcal{C}(s, x)$ denote the covariance function of the sGP defined at any $s \in \Omega$. Based on the previous result in Proposition S1, we know $\mathcal{C}_s(x) \in \mathcal{H}$ and $L\mathcal{C}_s(x) = \frac{1}{\alpha}\sin[\alpha(s-x)^+]$ is the Green function of L. The projection of $\mathcal{C}_s(x)$ into $\mathcal{H}_k$ is denoted as $\tilde{\mathcal{C}}_s(x)$.

Lemma 1. *Given the same setting in the main theorem*

$$\begin{aligned}
\mathcal{C}(x_1, x_2) &= \langle C_{x_1}, C_{x_2}\rangle_{\mathcal{H}} \\
\mathcal{C}_k(x_1, x_2) &= \left\langle \tilde{\mathcal{C}}_{x_1}, \tilde{\mathcal{C}}_{x_2}\right\rangle_{\mathcal{H}}.
\end{aligned} \tag{16}$$

Proof. The first part directly follows from the proof in Proposition S1. The second part can be proved using the fact that $L\mathcal{C}_{x_1}(x)$ is the Green function, which implies

$$\left\langle \psi_i, \tilde{\mathcal{C}}_{x_1}\right\rangle_{\mathcal{H}} = \langle \psi_i, \mathcal{C}_{x_1}\rangle_{\mathcal{H}} = \langle L\psi_i, L\mathcal{C}_{x_1}\rangle_\Omega = \psi_i(x_1),$$

for each $\psi_i \in \mathbb{B}_k$. The detailed proof proceeds as follow.

By construction of the B-spline approximation,

$$\begin{aligned}\mathcal{C}_k(x_1, x_2) &= \mathrm{Cov}\Bigg[\sum_i w_i\psi_i(x_1), \sum_i w_i\psi_i(x_2)\Bigg] \\ &= \mathrm{Cov}\Big[\boldsymbol{\Phi}(x_1)^T\boldsymbol{w}, \boldsymbol{\Phi}(x_2)^T\boldsymbol{w}\Big] \\ &= \boldsymbol{\Phi}(x_1)^T\Sigma_{\boldsymbol{w}}\boldsymbol{\Phi}(x_2) \\ &= \boldsymbol{\gamma}_{x_1}^T\Sigma_{\boldsymbol{w}}^{-1}\boldsymbol{\gamma}_{x_2},\end{aligned} \tag{17}$$

where $\boldsymbol{\Phi}(x) = [\psi_1(x), ..., \psi_k(x)]^T$, and $\Sigma_{\boldsymbol{w}}$ is defined in Section 3.2, and $\boldsymbol{\gamma}_{x_1} = \Sigma_{\boldsymbol{w}}\boldsymbol{\Phi}(x_1)$ and $\boldsymbol{\gamma}_{x_2} = \Sigma_{\boldsymbol{w}}\boldsymbol{\Phi}(x_2)$.

Since $\tilde{\mathcal{C}}_{x_1}(x)$ is the projection of $\mathcal{C}_{x_1}(x)$ to $\mathcal{H}_k$, $\tilde{\mathcal{C}}_{x_1}(x) = \sum_i w_{x_1,i}\psi_i(x)$ for some weights $\boldsymbol{w}_{x_1} = [w_{x_1,1}, ..., w_{x_1,k}]^T$. The same argument can be used for $\tilde{\mathcal{C}}_{x_2}$. Therefore

$$\begin{aligned}\mathrm{Cov}\Big[\left\langle \tilde{g}_k, \tilde{\mathcal{C}}_{x_1}\right\rangle_{\mathcal{H}}, \left\langle \tilde{g}_k, \tilde{\mathcal{C}}_{x_2}\right\rangle_{\mathcal{H}}\Big] &= \left\langle \tilde{\mathcal{C}}_{x_1}, \tilde{\mathcal{C}}_{x_2}\right\rangle_{\mathcal{H}} \\ &= \left\langle \boldsymbol{\Phi}(x)^T\boldsymbol{w}_{x_1}, \boldsymbol{\Phi}(x)^T\boldsymbol{w}_{x_2}\right\rangle_{\mathcal{H}} \\ &= \left\langle L\boldsymbol{\Phi}(x)^T\boldsymbol{w}_{x_1}, L\boldsymbol{\Phi}(x)^T\boldsymbol{w}_{x_2}\right\rangle_{\Omega} \\ &= \boldsymbol{w}_{x_1}^T\Sigma_{\boldsymbol{w}}^{-1}\boldsymbol{w}_{x_2},\end{aligned} \tag{18}$$

since $[\Sigma_{\boldsymbol{w}}^{-1}]_{ij} = \langle L\psi_i, L\psi_j\rangle_{\Omega}$, hence it only remains to show $\boldsymbol{\gamma}_{x_1} = \boldsymbol{w}_{x_1}$. Since

$$\begin{aligned}\left\langle \tilde{C}_{x_1}, \psi_i\right\rangle_{\mathcal{H}} &= \langle C_{x_1}, \psi_i\rangle_{\mathcal{H}} = \psi_i(x_1), \\ \left\langle \tilde{C}_{x_1}, \psi_i\right\rangle_{\mathcal{H}} &= \left\langle \boldsymbol{\Phi}(x)^T\boldsymbol{w}_{x_1}, \psi_i\right\rangle_{\mathcal{H}} = \sum_{j=1}^{k} w_{x_1,j}\left\langle L\psi_j, L\psi_i\right\rangle_{\Omega},\end{aligned} \tag{19}$$

for each $\psi_i \in \mathbb{B}_k$, we get $\Sigma_{\boldsymbol{w}}^{-1}\boldsymbol{w}_{x_1} = \boldsymbol{\Phi}(x_1)$. This lemma is hence proved. □

Using the above result and Lemma 1, it suffices to prove that

$$\left|\left\langle \tilde{\mathcal{C}}_{x_1}, \tilde{\mathcal{C}}_{x_2}\right\rangle_{\mathcal{H}} - \langle \mathcal{C}_{x_1}, \mathcal{C}_{x_2}\rangle_{\mathcal{H}}\right| = O(1/k). \tag{20}$$

For this step, we will use the following lemma on the spline approximation:

Lemma 2. *Given the same setting in the main theorem, define the norm* $||f||_{\mathcal{H}} = \langle f, f \rangle_{\mathcal{H}}^{1/2}$ *for* $f \in \mathcal{H}$ *then*

$$||\mathcal{C}_s - \tilde{\mathcal{C}}_s||_{\mathcal{H}} = O(1/k), \tag{21}$$

for each $s \in \Omega$.

Proof. The proof of this lemma mostly follows from the result in Schultz (1969). First, since $D^2\mathcal{C}_s(x)$ is a continuous function on Ω and differentiable everywhere except at $x = s$, $C_s(x)$ has week derivatives up to order 3. As the derivative of $D^2\mathcal{C}_s(x)$ is bounded and continuous for $x < s$ and $x > s$, we can conclude $\mathcal{C}_s \in H^3(\Omega)$. Given $\mathcal{H}_k$ is a spline space with degree 3 and mesh size $1/k$ and $\mathcal{C}_s \in H^3(\Omega)$, by theorem 3.3 in Schultz (1969) we have

$$||D^q(\mathcal{C}_s - \tilde{\mathcal{C}}_s)||_{\mathcal{L}^2} \leq c_q \left(\frac{1}{k}\right)^{3-q} \tag{22}$$

for each $0 \leq q \leq 2$, where c_q is a constant that only depends on $||\mathcal{C}_s||_{H^3(\Omega)}$. Note that

$$\begin{aligned} ||\mathcal{C}_s - \tilde{\mathcal{C}}_s||_{\mathcal{H}}^2 &= \alpha^4 ||\mathcal{C}_s - \tilde{\mathcal{C}}_s||_{\mathcal{L}^2}^2 + ||D^2(\mathcal{C}_s - \tilde{\mathcal{C}}_s)||_{\mathcal{L}^2}^2 - 2\alpha \left\langle (\mathcal{C}_s - \tilde{\mathcal{C}}_s), D^2(\mathcal{C}_s - \tilde{\mathcal{C}}_s) \right\rangle_{\Omega} \\ &\leq \alpha^4 ||\mathcal{C}_s - \tilde{\mathcal{C}}_s||_{\mathcal{L}^2}^2 + ||D^2(\mathcal{C}_s - \tilde{\mathcal{C}}_s)||_{\mathcal{L}^2}^2 + 2\alpha ||D^2(\mathcal{C}_s - \tilde{\mathcal{C}}_s)||_{\mathcal{L}^2} ||(\mathcal{C}_s - \tilde{\mathcal{C}}_s)||_{\mathcal{L}^2} \end{aligned} \tag{23}$$

where the second equality holds by Cauchy-Schwarz Inequality. The lemma hence proved. □

Using Lemma 2, Eq. (20) can be proved with an application of Triangle Inequality followed with a use of Cauchy-Schwarz Inequality and the fact that the sequence $||\tilde{\mathcal{C}}_{x_2}||_{\mathcal{H}}$ is bounded,

$$\begin{aligned} \left| \left\langle \tilde{\mathcal{C}}_{x_1}, \tilde{\mathcal{C}}_{x_2} \right\rangle_{\mathcal{H}} - \langle \mathcal{C}_{x_1}, \mathcal{C}_{x_2} \rangle_{\mathcal{H}} \right| &= \left| \left\langle \tilde{\mathcal{C}}_{x_1} - \mathcal{C}_{x_1}, \tilde{\mathcal{C}}_{x_2} \right\rangle_{\mathcal{H}} - \left\langle \mathcal{C}_{x_1}, \mathcal{C}_{x_2} - \tilde{\mathcal{C}}_{x_2} \right\rangle_{\mathcal{H}} \right| \\ &\leq ||\tilde{\mathcal{C}}_{x_1} - \mathcal{C}_{x_1}||_{\mathcal{H}} ||\tilde{\mathcal{C}}_{x_2}||_{\mathcal{H}} + ||\tilde{\mathcal{C}}_{x_2} - \mathcal{C}_{x_2}||_{\mathcal{H}} ||\mathcal{C}_{x_1}||_{\mathcal{H}} \\ &\leq c/k, \end{aligned} \tag{24}$$

where c is some constant independent of k. The theorem is hence proved. □

References

Harvey, A. C. (1990). Forecasting, structural time series models and the kalman filter.

Lindgren, F., H. Rue, and J. Lindstrom (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 73*(4), 423–498.

Särkkä, S. and A. Solin (2019). *Applied stochastic differential equations*, Volume 10. Cambridge University Press.

Schultz, M. H. (1969). Approximation theory of multivariate spline functions in sobolev spaces. *SIAM Journal on Numerical Analysis 6*(4), 570–582.