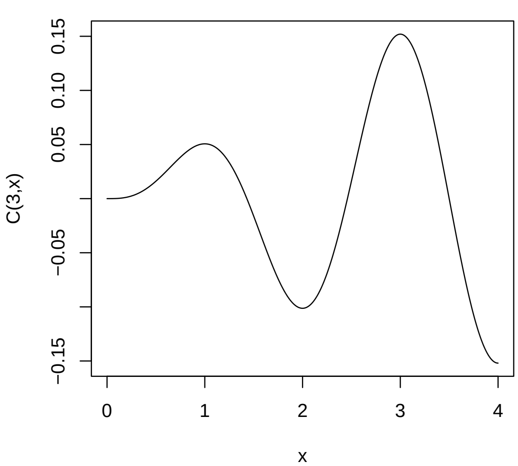
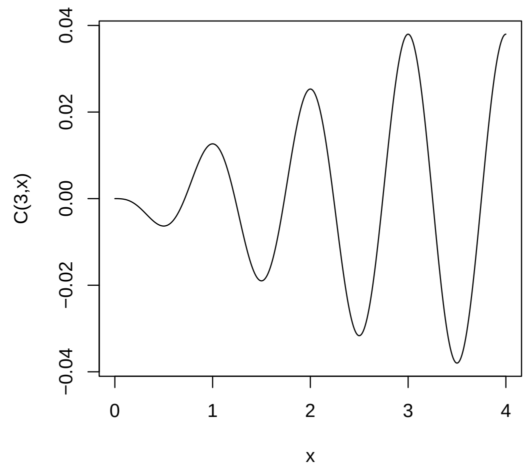
Online Supplement to: Efficient Modeling of Quasi-Periodic Data with Seasonal Gaussian Process

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

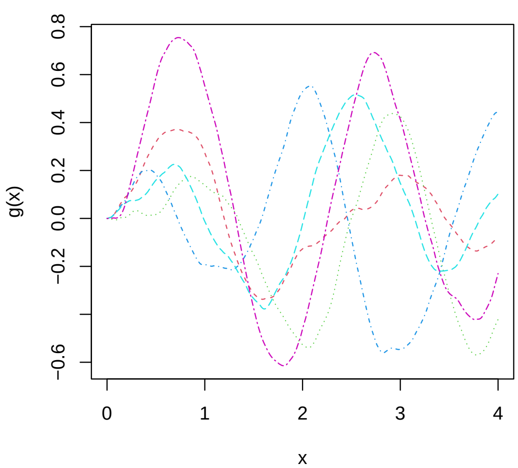
S1. Additional figures and table



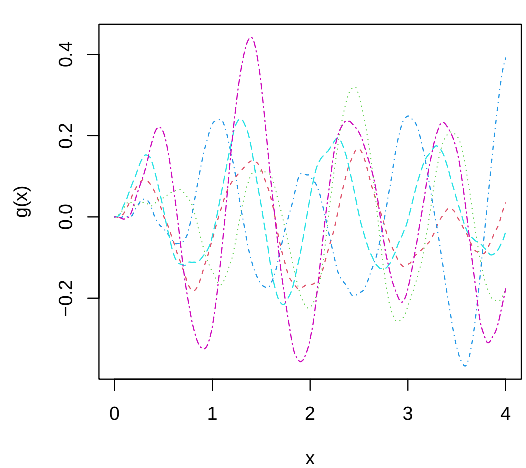
(a) $\alpha = \pi$: Covariance



(b) $\alpha = 2\pi$: Covariance



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



(d) $\alpha = 2\pi$: Samples

Figure S1: (a,b) display the covariance functions of two sGPs at 3, and (c,d) display five sample paths from the two sGPs. The frequency parameter α equals to π in (a,c) and 2π in (b,d), and the SD parameter $\sigma = 1$ in both sGPs. Both the covariance functions and the sample paths exhibit quasi-periodic behavior with amplitudes varying overtime.

	1st Quantile	Median	3rd Quantile
$\sigma_{tr}(10)$	0.4371 (102.05)	0.6446 (120.81)	0.9462 (139.49)
$\sigma_1(10)$	1.8179 (–)	2.4542 (–)	3.1709 (–)
$\sigma_{\frac{1}{2}}(10)$	1.8179 (–)	2.4542 (–)	3.1709 (–)
$\sigma_{\frac{44}{12}}(10)$	0.4004 (–)	0.4514 (–)	0.5019 (–)
$\sigma_{9.1}(10)$	0.004248 (–)	0.01045 (–)	0.02445 (–)
$\sigma_{10.4}(10)$	0.01090 (–)	0.02035 (–)	0.03335 (–)
σ_ϵ	0.5870 (0.6099)	0.5928 (0.6159)	0.5988 (0.6220)

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

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 $x_1, x_2 \in \mathbb{R}^+$ 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:

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

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

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

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 $\boldsymbol{s} = \{s_1, ..., s_n\} \subset \mathbb{R}^+$ denotes a set of n sorted locations and spacing $d_1 = s_1$ and $d_i = s_i - s_{i-1}$ for $i \in \{2, .., n\}$. Then the augmented vector $\boldsymbol{g}_{aug}(s_i) = [g(s_i), g'(s_i)]^T$ 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 $\mathbf{F}^{2k} = (-a^2)^k\mathbf{I}$ and $\mathbf{F}^{2k+1} = (-a^2)^k\mathbf{F}$, then apply the Taylor series expansion for both components in the integral above. It then can be rewritten as:

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

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:

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

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

$$Lg(x) = \sigma\xi(x),$$

where $L = a^2 + \frac{d^2}{dx^2}$ 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 $\mathbb{B}_k := \{\varphi_i, i \in [k]\}$ 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 $\tilde{g}(.) = \sum_{i=1}^{k} w_i\varphi_i(.)$. 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 $\mathbb{T}_k := \{\phi_i = L\varphi_i, i \in [k]\}$, 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

$$\langle L\tilde{g}(x), \phi_i(x)\rangle \overset{d}{=} \sigma\langle\xi(x), \phi_i(x)\rangle, \tag{10}$$

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

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

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

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:

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

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

S5. Proof of the Convergence Result

Theorem (Convergence of B-spline Approximation)**.** *Let $\Omega = [a, b]$ where $a, b \in \mathbb{R}^+$ and let $g \sim sGP(\alpha, \sigma)$. 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, then:*

$$\lim_{k \to \infty} \mathcal{C}_k(x_1, x_2) = \mathcal{C}(x_1, x_2),$$

for any $x_1, x_2 \in \Omega$, where $\mathcal{C}(x_1, x_2) = \mathrm{Cov}[g(x_1), g(x_2)]$, $\mathcal{C}_k(x_1, x_2) = \mathrm{Cov}[\tilde{g}_k(x_1), \tilde{g}_k(x_2)]$.

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 paramter of the sGP $\sigma = 1$ and $\Omega = [0, 1]$. We denote the 3rd order Sobolev space as $H^3(\Omega) = \{f \in \mathcal{L}^2(\Omega) : D^q f \in \mathcal{L}^2(\Omega)\ \forall |q| \leq 3\}$. We then define the constrained Sobolev space $\mathcal{H}$ as:

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

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 $f, h \in \mathcal{H}$.

Let $b > 1$ and $\tilde{\Omega} = [0, b) \supset \Omega$, then for each $f \in H^3(\Omega)$, there exists a zero-extension $f_0 \in H^3(\tilde{\Omega})$ such that $f_0(x) = f(x)$ for $x \in \Omega$, and $f_0(b) = f_0'(b) = 0$. Hence, we assume without the loss of generality that for each $f \in \mathcal{H}$, $f(1) = f'(1) = 0$ from now. This implies that the differential operator D^q in $\mathcal{H}$ has adjoint operator $(D^q)^* = (-1)^q D^q$ for each $q \in \mathbb{Z}$.

Define $\mathcal{H}_k = \text{span}\{\mathbb{B}_k\}$. Note $\mathcal{H}_k \subset \mathcal{H}$ by our construction of the B-spline basis. This implies if $f(x) \in \mathcal{H}$, then there exists a projection $\tilde{f}(x) = \sum_{i=1}^{k} w_i \varphi_i(x) \in \mathcal{H}_k$ which satisfies:

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

To prove $\lim_{k\to\infty} \mathcal{C}_k(x_1, x_2) = \mathcal{C}(x_1, x_2)$, note that for any $f, h \in \mathcal{H}$,

$$\begin{aligned} \text{Cov}[\langle g, f \rangle_{\mathcal{H}}, \langle g, h \rangle_{\mathcal{H}}] &:= \text{Cov}[\langle Lg, Lf \rangle_{\Omega}, \langle Lg, Lh \rangle_{\Omega}] \\ &= \text{Cov}[\langle \xi, Lf \rangle_{\Omega}, \langle \xi, Lh \rangle_{\Omega}] \\ &= \langle Lf, Lh \rangle_{\Omega}, \\ &= \langle f, h \rangle_{\mathcal{H}}. \end{aligned} \tag{14}$$

The second equality follows from the definition of the sGP, and the third equality follows from the property of Gaussian white noise (Harvey, 1990).

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

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

Using this result and the fact that the B-spline approximation $\tilde{g}_k$ is a least square solution, we have

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

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. Its projection into $\mathcal{H}_k$ is denoted as $\tilde{\mathcal{C}}_s(x)$

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

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

Proof. The first part directly follows from the proof in Proposition S1. The second part can be proved using the fact that L is self-adjoint and $L\mathcal{C}_{x_1}(x) = \frac{1}{a}\sin[a(x_1 - x)^+]$ is the Green function, which implies $\left\langle \varphi_i, \tilde{\mathcal{C}}_{x_1} \right\rangle_{\mathcal{H}} = \varphi_i(x_1)$ for each $i \in [k]$. The detailed proof proceeds as the following.

By construction of the B-spline approximation,

$$\begin{aligned} \mathcal{C}_k(x_1, x_2) &= \mathrm{Cov}\left[\sum_i w_i \varphi_i(x_1), \sum_i w_i \varphi_i(x_2) \right] \\ &= \mathrm{Cov}\left[\mathbf{\Phi}(x_1)^T \boldsymbol{w}, \mathbf{\Phi}(x_2)^T \boldsymbol{w} \right] \\ &= \mathbf{\Phi}(x_1)^T \Sigma_{\boldsymbol{w}} \mathbf{\Phi}(x_2), \end{aligned} \tag{17}$$

where $\mathbf{\Phi}(x) = [\varphi_1(x), ..., \varphi_k(x)]^T$.

Since $\tilde{\mathcal{C}}_{x_1}(x)$ is the projection of $\mathcal{C}_{x_1}(x)$ to $\mathcal{H}_k$, $\tilde{\mathcal{C}}_{x_1}(x) = \sum_i w_{x_1,i}\varphi_i(x)$ 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

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

Solving $\left\langle \tilde{\mathcal{C}}_{x_1}, \varphi_i \right\rangle_{\mathcal{H}} = \langle \mathcal{C}_{x_1}, \varphi_i \rangle_{\mathcal{H}}$ for each $i \in [k]$ yields that $\boldsymbol{w}_{x_1} = \Sigma_{\boldsymbol{w}} \boldsymbol{\omega}_{x_1}$, where $\boldsymbol{\omega}_{x_1} \in \mathbb{R}^k$ with ith element $\omega_{x_1,i} = \langle \mathcal{C}_{x_1}, \varphi_i \rangle_{\mathcal{H}}$. Hence it only remains to show $\boldsymbol{\omega}_{x_1} = \boldsymbol{\Phi}(x_1)$, which holds because

$$\begin{aligned} \omega_{x_1,i} &= \langle \mathcal{C}_{x_1}, \varphi_i \rangle_{\mathcal{H}} \\ &= \langle L\mathcal{C}_{x_1}, L\varphi_i \rangle_{\Omega} \\ &= \langle L^* L\mathcal{C}_{x_1}, \varphi_i \rangle_{\Omega} \\ &= \varphi_i(x_1), \ \forall i \in [k]. \end{aligned} \tag{19}$$

The last equality holds because L is a self-adjoint operator and $L\mathcal{C}_{x_1}$ is the Green function. This lemma is hence proved. □

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

$$\left\langle \tilde{\mathcal{C}}_{x_1}, \tilde{\mathcal{C}}_{x_2} \right\rangle_{\mathcal{H}} \to \langle \mathcal{C}_{x_1}, \mathcal{C}_{x_2} \rangle_{\mathcal{H}}, \tag{20}$$

as $k \to \infty$. 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). 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

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

for each $0 \leq q \leq 2$, where c_q is a constant that depends on $||\mathcal{C}_s||_{H^3(\Omega)}$ but does not depend on k. Note that

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

where the second equality holds since $D^*=-D$. The lemma is proved since $||.||_{\mathcal{H}}$ and $||.||_{H^2(\Omega)}$ are equivalent norm.

□

Using Lemma 2, Eq. (20) can be proved with Cauchy Schwarz inequality and the fact that the sequence $||\tilde{\mathcal{C}}_{x_2}||_{\mathcal{H}}$ is bounded,

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

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.