Web Appendix: Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets

Rajarshi Guhaniyogi and Sudipto Banerjee

January 8, 2018

1 Web Appendix A: Posterior Concentration of Meta-Posteriors

We study the concentration of the meta posterior corresponding to two special instances of equation (4) in Section 2.2. The first is a full Gaussian process model and the second is a Gaussian process model with compactly supported correlation function (CSC) (Kaufman et al., 2008). The former arises from equation (4) (in Section 2.2) by taking $D(\theta) = \tau^2 I$ and $C(\theta)$ as a customary spatial covariance matrix whose (i, j)-th element is given by the Matérn covariance function

$$\kappa(s_i, s_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{2\nu}\|s_i - s_j\|\phi\right)^{\nu} \mathcal{K}_\nu(\sqrt{2\nu}\|s_i - s_j\|\phi) , \quad (1)$$

where ϕ (range parameter) and ν (smoothness parameter) control rate of decay in spatial correlation and the degree of process smoothness respectively. Γ is the usual Gamma function and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind with order ν (Stein, 2012). σ^2 is referred to as the spatial variance parameter. Use of Matérn covariance function results in a dense $C(\theta)$ and the computational benefits accrue from fitting SMK. The Gaussian process with compactly supported correlation function results by assuming that $C(\theta) = H(\theta) \odot T$, where $H(\theta) = ((\kappa(s_i, s_j)))_{i,j=1}^N$ is a spatial covariance matrix constructed using the Matérn covariance function and $T = ((\kappa_\delta(s_i, s_j))_{i,j=1}^N$ is a sparse covariance matrix constructed from a compactly supported spatial covariance function κ_δ (Wendland, 2004). $\odot$ refers to the Hadamard product (Kaufman et al., 2008) between two matrices. Compactly supported correlation functions set correlation between $w(s_i)$ and $w(s_j)$ to 0 if $||s_i - s_j|| > \delta$. Thus, T (and hence $C(\theta)$) under CSC is sparse. The readers should note that CSC as above can also be visualized as a Gaussian process with covariance kernel $\kappa(s_i, s_j)\kappa_\delta(s_i, s_j)$.

Posterior concentration of the meta posterior is studied in two steps. In the first step, we carefully investigate concentration of each subset posterior around the true data generating function. While such a concentration result is readily available for Gaussian processes (Vaart and Zanten, 2011), we carefully adapt their ideas to derive similar results for the CSC. The second step invokes Theorem 2.1 in Minsker et al. (2014) to establish the improved concentration of meta posteriors compared to individual subset posteriors around the true function.

1.0.1 Notations

Without loss of generality, we assume that the spatial domain of interest is $[0, 1]^2$. For our exposition we will make several simplifications in the problem as specified below.

(a) X in Section 2.2 is assumed to be $N \times 1$, i.e. Gaussian process model is fitted with the intercept only.

(b) τ^2, ϕ are both known. Without loss of generality, assume $\tau^2 = 1$ and $\phi = 1$.

(a) and (b) are assumed to substantially ease mathematical calculations and in no way to detract our attention from the general problem. For example, van der Vaart and van Zanten (2009) show that concentration results on Gaussian processes with a fixed ϕ can readily be extended to settings with random ϕ. Similarly, the result with a fixed τ^2 can be extended to unknown τ^2 with some additional calculations, as seen in various articles, see e.g. Ghosal et al. (2007).

Smoothness for the a-priori surface and the data generating surface:

Let the true data generating spatial model be $y(s) = w_0(s) + \epsilon(s)$, where $w_0(s)$ is the unknown data generating spatial surface. Under assumptions (a) and (b), the Gaussian process regression fitted to each data subset is given by $y(s) = w(s) + \epsilon(s)$. We intend to study the posterior concentration of $w(s)$ around the true data generating function $w_0(s)$ under the Gaussian (or CSC) prior on $w(s)$ on each subset.

To study convergence properties of a nonparametric Gaussian process regression, it is important to specify the degree of "smoothness" for the a-priori class of fitted spatial functions and the smoothness of $w_0(s)$. "Smoothness" of a function roughly means how many higher order derivatives exist for a function. In this context, we focus upon two classical spaces for smooth functions, Holder-α on $[0,1]^2$, denoted by $H^\alpha[0,1]^2$ and Sobolev-α space on $[0,1]^2$, denoted by $S^\alpha[0,1]^2$ (Vaart and Zanten, 2011). They include functions which are $[\alpha]$ times continuously differentiable ($[\alpha]$ is the greatest integer $\leq \alpha$).

Normed distance between probability measures:

Let $C_b([0,1]^2)$ be the space of bounded continuous functions on $[0,1]^2$ and $||w||_\infty = \sup_{s\in[0,1]^2}|w(s)|$ be the supremum norm of w in $C_b([0,1]^2)$. When s's are drawn randomly from $[0,1]^2$ with a density $h(s)$, define L_2 norm of a function w by $\|w\|_2 = \left(\int w(s)^2 h(s) ds\right)^{1/2}$.

To assess concentration of the meta posterior of w around w_0, we employ *Wasserstein distance* between the meta posterior $\pi^*(w)$ and δ_{w_0}, δ_{w_0} being the Dirac-delta function at w_0. Let $\zeta(w_1,w_2) = \|w_1 - w_2\|_2$. The Wasserstein distance between two probability measures π_1 and π_2, denoted by $d_{W_{1,\zeta}}(\pi_1,\pi_2)$, is expressed by $\inf\{E\zeta(w_1,w_2) : \mathcal{L}(w_1) = \pi_1, \mathcal{L}_2 = \pi_2\}$, where infimum is taken over all joint distributions of (w_1,w_2). Using this representation, Wasserstein distance $\pi^*(w)$ and δ_{w_0} is given by, $d_{W_{1,\zeta}}(\pi^*(w),\delta_{w_0}) = \int \zeta(w,w_0)\pi^*(w)$. Similarly, the Wasserstein distance between subset posteriors $p_k(w)$ and δ_{w_0} is given by $d_{W_{1,\zeta}}(p_k(w),\delta_{w_0}) = \int \zeta(w,w_0)p_k(w)$.

Denote the reproducing kernel Hilbert space (RKHS) of the Gaussian process prior with the associated RKHS norm by $(\mathcal{H}, ||\cdot||_{\mathcal{H}})$. The key element to derive posterior concentration under GP prior is the concentration function defined as

$$\phi_{w_0}(\gamma) = \inf_{h\in\mathcal{H}:||h-w_0||_\infty<\gamma} ||h||^2_{\mathcal{H}} - \log p(w : ||w||_\infty < \gamma).$$

It has been shown in van der Vaart and van Zanten (2008) that $\phi_{w_0}(\gamma)$ equals $-\log p(w : ||w - w_0||_\infty < \gamma)$ upto some constants, so that $\phi_{w_0}(\cdot)$ provides prior concentration of w around the true w_0. Finally we define the function $\psi_{w_0}(\gamma) = \frac{\phi_{w_0}(\gamma)}{\gamma^2}$ and its inverse $\psi^{-1}_{w_0}(\eta) =$

$\inf\{\gamma > 0 : \psi_{w_0}(\gamma) \geq \eta\}$. All concentration results will be based on $\psi_{w_0}^{-1}(n)$.

Framework:

The posterior concentration for meta posterior of $w(s)$ around $w_0(s)$ is assessed by $P_{w_0}\left[d_{W_{1,\varsigma}}(\pi^*(w), \delta_{w_0}) > \epsilon\right]$. This signifies the probability that the Wasserstein distance between the meta posterior $\pi^*(w)$ and δ_{w_0} is greater than ϵ. In the first step, a bound on $P_{w_0}\left[d_{W_{1,\varsigma}}(p_k(w), \delta_{w_0}) > \epsilon\right]$ is established for the kth subset, $k = 1, ..., K$. Theorem 1.2 and 1.3 provide such results for Gaussian process and CSC models respectively. In the second step, these upper bounds on subset posteriors are used in Theorem 1.4 to develop an upper bound on $P_{w_0}\left[d_{W_{1,\varsigma}}(\pi^*(w), \delta_{w_0}) > \epsilon\right]$. For simplicity, we assume $N = nK$, where n is the number of observations per subset.

1.1 Results

We begin by stating the fact that for any distribution $\tilde{\pi}$ of w, $d_{W_{1,\varsigma}}(\tilde{\pi}(w), \delta_{w_0})^2 \leq \int ||w - w_0||_2^2 \tilde{\pi}(w)$ which follows from the Jensens inequality. This fact leads to the straightforward extension of Theorem 2 of Vaart and Zanten (2011).

Theorem 1.1 *Suppose for some $\alpha > 0$, a Gaussian process prior $\tilde{\pi}$ on w assigns probability 1 to a Holder-α space on $[0,1]^2$. Then for a constant C_s depending on $h(\cdot)$, the density of s, if $\psi_{w_0}^{-1}(n) \leq n^{-1/(2\alpha+2)}$*

$$E_{w_0}[d_{W_{1,\varsigma}}(\tilde{\pi}(w), \delta_{w_0})^2] \leq C_s \psi_{w_0}^{-1}(n)^2. \tag{2}$$

Otherwise, the same assertion holds with the upper bound $C_s n \psi_{w_0}^{-1}(n)^{(2\alpha+4)}$.

In what follows, Theorem 1.1 is suitably used to show bounds on the concentration of subset posteriors for both Gaussian process and CSC priors on w.

It is well known that GP prior with matern kernel (1) assigns probability one to $H^{\nu}[0,1]^2$. Therefore, Theorem 1.1 in Vaart and Zanten (2011) combined with Theorem 1.1 in this article lead to the following result.

Theorem 1.2 *Suppose $\{(y_{k_1}, s_{k_1}), ..., (y_{k_n}, s_{k_n})\}$ are data points in the k-th subsample $\mathcal{S}_k$ for every $k = 1, .., K$, where $(s_{k_1}, ..., s_{k_n})$ emerges from a random design with density h. Further assume*

1. Covariance kernel $\kappa(\cdot,\cdot)$ for the GP prior is a Matern kernel (1).

2. $w_0 \in H^{\chi}[0,1]^2 \cap S^{\chi}[0,1]^2$ with $\min(\nu,\chi) > 1$.

Then for every $\epsilon > 0$, $P_{w_0}\left[d_{W_{1,\varsigma}}(p_k(w), \delta_{w_0}) > \epsilon\right] \leq \frac{C_k}{\epsilon^2}\left(\frac{1}{n}\right)^{2\min(\nu,\chi)/(2\nu+2)}$, for some constant $C_k > 0$.

Remark: Theorem 1.2 presents a probabilistic upper bound on the distance between the subset posteriors and Dirac delta function at w_0. The upper bound is a decreasing function of n. In other words, as the sample size in each subset increases, the probability that the distance between $p_k(w)$ and δ_{w_0} is greater than ϵ converges to 0, i.e. $p_k(w)$ lies within a distance ϵ from δ_{w_0} with a high probability. Clearly, this probability depends on the smoothness parameters ν of the Gaussian process and χ of the true data generating function.

We will next turn our attention to proving a result similar to Theorem 1.2, when $w(s)$ is assigned a CSC prior (Kaufman et al., 2008). To begin with, recall that the correlation kernel for the CSC is given by $\kappa(s,t)\kappa_\delta(s,t)$, where $\kappa(s,t)$ is the Matérn correlation kernel as defined in (1). According to Bochner's theorem, any stationary covariance kernel has a one to one correspondence with its *spectral density* or the Fourier transform. Let $f_M : \mathbb{R}^2 \to \mathbb{R}$ be the spectral density of $\kappa(s,t)$. The spectral density f_M is related to $\kappa(s,t)$ by $\kappa(s,t) = \sigma^2 \int e^{i\lambda'(s-t)} f_M(\lambda) d\lambda$. It is well known that the spectral density of a Matérn covariance kernel of the form (1) is given by $f_M(\lambda) = \frac{1}{[1+||\lambda||^2]^{\nu+1}}$.

Assume that the compactly supported correlation function κ_δ is also stationary with the spectral density $f_{tap}(\lambda)$. Using the well known result of convolution of fourier transforms, one readily obtains the spectral density for the CSC as $f_{tGP}(\lambda) = f_{tap} * f_M(\lambda) = \int f_M(x) f_{tap}(\lambda - x)dx$.

Condition on the spectral density of tapering kernel:

To prove concentration results of $p_k(w)$ under the CSC prior, it is essential to impose some restrictions on the tapering kernel κ_δ or equivalently on its spectral density function f_{tap}. Assume that $f_{tap}(\lambda)$ satisfies the *taper condition*, i.e. for some $\mu > 0$ and $M_\delta < \infty$ depending upon the taper range δ, $0 < f_{tap}(\lambda) < \frac{M_\delta}{(1+||\lambda||^2)^{\nu+1+\mu}}$. Proposition 1 in Furrer et al. (2012) confirms the taper condition being equivalent to assuming $\lim\limits_{\lambda\to\infty} \frac{f_{tGP}(\lambda)}{f_M(\lambda)}$ is bounded away from 0 and ∞. The taper condition simply ensures that f_M and f_{tGP} exhibit the same tail behavior. In particular, the Wendland tapering kernel (see Section 3.4.2) used in simulation studies is shown to satisfy the taper condition (see Furrer et al. (2012)).

To show a result similar to Theorem 1.2 for the CSC, we proceed in a few steps. In step 1, we argue that the realizations from the CSC lie in $H_\nu[0,1]^2$ with probability 1. Though showing such a result for general ν is difficult, we prove the result for $0 < \nu < 1$. Clearly, the popular exponential kernel is encompassed under this assumption. Lemma 1.5 in the Appendix states and proves such a result. Theorem 1.1 is then readily applicable for the CSC. Finally, $\psi_{w_0}^{-1}(n)$ is calculated for the CSC in the Appendix (see Lemma 1.6, Lemma 1.7 and 1.8) to arrive at the following result.

Theorem 1.3 *Suppose $\{(y_{k_1}, s_{k_1}), ..., (y_{k_n}, s_{k_n})\}$ are data points in the k-th subsample $\mathcal{S}_k$ for every $k = 1, .., K$, where $(s_{k_1}, ..., s_{k_n})$ emerges from a random design with density h. Further assume*

1. *Covariance kernel for CSC is $\kappa(\cdot,\cdot)\kappa_\delta(\cdot,\cdot)$, κ is a Matérn kernel with smoothness parameter ν.*

2. *$w_0 \in H^\chi[0,1]^2 \cap S^\chi[0,1]^2$ with $\min(\nu,\chi) > 1$.*

Then for every $\epsilon > 0$, $P_{w_0}\left[d_{W_{1,\varsigma}}(p_k(w), \delta_{w_0}) > \epsilon\right] \leq \frac{H_k}{\epsilon^2}\left(\frac{1}{n}\right)^{2\min(\nu,\chi)/(2\nu+2)}$, for some constant $H_k > 0$.

Remark: Several authors have shown that CSC yield similar asymptotic mean squared error as a Gaussian process with Matérn correlation (e.g Furrer et al. (2012), Kaufman et al. (2008)). Theorem 1.3 reveals a stronger property of the CSC as an approximation to the Gaussian process. It shows that under Wasserstein distance, posterior for the unknown random function $w(s)$ concentrates around $w_0(s)$ at a similar rate for CSC and GP.

It remains to state the results on concentration of the meta posterior π^*. In this regard, we invoke Theorem 2.1 in Minsker et al. (2014).

Theorem 1.4 *Let $\alpha \in (0, 1/2)$ and $C_\alpha = (1-\alpha)\sqrt{\frac{1}{1-2\alpha}}$. Further assume that $P(d_{W_{1,\varsigma}}(p_k(w), \delta_{w_0}) \geq \epsilon) < p < \alpha$, for all $k = 1, ..., K$. Then*

$$P_{w_0}(d_{W_{1,\varsigma}}(\pi^*(w), \delta_{w_0}) > C_\alpha \epsilon) \leq \exp\left(-K\left[(1-\alpha)\log\left(\frac{1-\alpha}{1-p}\right) + \alpha\log\left(\frac{\alpha}{p}\right)\right]\right). \quad (3)$$

Remark: Theorem 1.4 has several implications. It shows improved concentration of the meta posterior over subset posteriors around δ_{w_0}. It is also easy to see that keeping K finite and $n \to \infty$, $P_{w_0}(d_{W_{1,\zeta}}(\pi^*(w), \delta_{w_0}) > C_\alpha \epsilon) \to 0$, for any α, i.e. theoretical convergence of meta posterior is achieved. On the other hand, when K varies slowly with N, i.e. $K = N^r$, for $0 < r < 1$, Theorem 1.4, together with earlier results, shows convergence of the meta posterior. To elaborate further, in this case $n = N^{1-r}$, so that for any $\epsilon > 0$, one is able to choose p s.t. $\max\{\frac{H_k}{\epsilon^2}\left(\frac{1}{N^{1-r}}\right)^{2\min(\nu,\chi)/(2\nu+2)}\} < p$ by Theorem 1.3. Considering the fact that $K = N^r$, the R.H.S of (3) converges to 0 as $N \to \infty$. The convergence of the meta posterior now follows by applying Theorem 1.4.

Web Appendix B: Proofs of the results

Lemma 1.5 *When $0 < \nu < 1$, CSC belongs to Holder-ν space.*

Proof Since CSC is a centered isotropic random field, we will make use of the results by Adler (2010). Note that for $0 < \nu < 1$ and $s \to 0$, $\kappa(s) - 1 = B||s||^{2\nu} + o(||s||^{2\nu})$, for some constant $B > 0$. Using the fact that tapering kernel κ_δ is a finite powered polynomial starting with nonzero coefficient for $||s||^0$, as $s \to 0$, $\kappa(s)\kappa_\delta(s) - 1 = B_1||s||^{2\nu} + o(||s||^{2\nu})$, for some constant $B_1 > 0$. Therefore, $\sqrt{\kappa(s)\kappa_\delta(s) - 1} = o(||s||^\nu)$, when $s \to 0$. According to Adler (2010), an isotropic random field satisfying this is Holder-ν continuous.

Lemma 1.5 together with Theorem 1.1 prove that (2) holds for CSC. We need to calculate $\psi_{w_0}(\epsilon)$. To this end, we would like to state a result (van der Vaart and van Zanten, 2009) that characterizes RKHS $\mathcal{H}$ of a Gaussian process.

Lemma 1.6 *RKHS $\mathcal{H}$ consists of real parts of the functions of the following form $f_m(s) = \int e^{i\lambda' s} m(\lambda) G(\lambda) d\lambda$, for m a function which is square integrable w.r.t the density G and $G(\lambda)$ is the spectral density for the correlation function of the Gaussian process. In this case the squared RKHS norm of f_m is given by*

$$||f_m||^2_{\mathcal{H}} = \min_{g: f_m = f_g} \int |g(\lambda)|^2 G(\lambda) d\lambda. \quad (4)$$

Denote $\mathcal{H}_{tGP}$ as the reproducing kernel Hilbert space for the CSC. Let $\hat{f}$ be the Fourier transform of a function f. Using the formula for Inverse Fourier transformation for any function f, $f(t) = \int e^{i\lambda' t}\hat{f}(\lambda)d\lambda$. By Lemma 1.6, $f \in \mathcal{H}_{tGP}$ if $\frac{\hat{f}}{f_{tGP}}$ is square integrable under the measure f_{tGP}. As a corollary, the RKHS norm of f is bounded by the L_2 norm $||\frac{\hat{f}}{f_{tGP}}||_2$ under the density f_{tGP}, i.e. $\int \frac{|\hat{f}(\lambda)|^2}{f_{tGP}(\lambda)}d\lambda$. We will use this fact to bound the RKHS norm of an element in the RKHS of CSC that is contained in an ϵ ball around w_0. We begin by bounding the prior probability of w in an open ball around 0.

Lemma 1.7 *For C_∞ a constant depending on ϵ, $-\log(||w||_\infty < \epsilon) \leq C_\infty\left(\frac{1}{\epsilon}\right)^{2/\nu}$, when w follows a CSC.*

Proof We will use Lemma 1.6 to prove this result. Recall that $f_{tGP} = f_{tap} * f_M(\lambda)$ and note that the Fourier transform of any $f_m \in \mathcal{H}_{tGP}$ is given by $\hat{f}_m(\lambda) = m(\lambda)f_{tap} * f_M(\lambda)$. Clearly for g to be the minimal choice as in (4), we have

$$\int |g(\lambda)|^2 \frac{1}{f_{tap} * f_M(\lambda)} d\lambda = ||f_m||^2_{\mathcal{H}_{tGP}}. \tag{5}$$

Use the fact that the *taper condition* implies, $\lim_{||\lambda||\to\infty} \frac{f_M(\lambda)}{f_{tGP}(\lambda)} = \gamma_1$, $0 < \gamma_1 < \infty$. Therefore, we have $\inf\limits_{||\lambda||} \frac{f_M(\lambda)}{f_{tap}*f_M(\lambda)} = \min\left\{\inf\limits_{||\lambda||>||\lambda_0||} \frac{f_M}{f_{tap}*f_M(\lambda)}, \inf\limits_{||\lambda||<||\lambda_0||} \frac{f_M}{f_{tap}*f_M(\lambda)}\right\}$, where $||\lambda_0||$ is chosen so that $\left|\frac{f_M(\lambda)}{f_{tap}*f_M(\lambda)} - \gamma_1\right| < \gamma_1/2$ for all $||\lambda|| > ||\lambda_0||$. Note that $[0, ||\lambda_0||]$ is a compact interval, so that $\frac{f_M(\lambda)}{f_{tap}*f_M(\lambda)} > 0$ for all $||\lambda|| \in [0, ||\lambda_0||]$ implies $\inf\limits_{||\lambda||<||\lambda_0||} \frac{f_M(\lambda)}{f_{tap}*f_M(\lambda)} > 0$. Therefore, $\inf\limits_{||\lambda||} \frac{f_M(\lambda)}{f_{tap}*f_M(\lambda)} = C_2 > 0$. Using this fact and (5), we obtain

$$||f_m||^2_{\mathcal{H}_{tGP}} = \int |g(\lambda)|^2 \frac{1}{f_{tap} * f_M(\lambda)} d\lambda \geq C_2 \int |g(\lambda)|^2 \frac{1}{f_M(\lambda)} d\lambda = \int |g(\lambda)|^2 [1 + ||\lambda||^2]^{\nu+1} d\lambda.$$

Thus unit ball in RKHS is contained in some Sobolev ball of order $\nu + 1$. Applying Theorem 3.3.2 of Edmunds and Triebel (1996) gives us $\log N(\epsilon, \mathcal{H}^1_{tGP}, ||\cdot||_\infty) \leq C_\infty(\frac{1}{\epsilon})^{2/(1+2\nu)}$, for some C_∞ constant, where $\mathcal{H}^1_{tGP}$ is the unit RKHS ball for CSC and $N(\epsilon, \mathcal{H}^1_{tGP}, ||\cdot||_\infty)$ is its metric entropy defined as the minimum number of balls of radius ϵ needed to cover $\mathcal{H}^1_{tGP}$ under $||\cdot||_\infty$ or sup-norm. Now applying Theorem 2 and Proposition 3 of Kuelbs and Li (1993) and Theorem 1.1 and 1.2 of Li et al. (1999) which connect $-\log(||w||_\infty < \epsilon)$ with $\log N(\epsilon, \mathcal{H}^1_{tGP}, ||\cdot||_\infty)$, the result follows.

Next we provide bound on $\inf\limits_{h\in\mathcal{H}_{tGP}:||h-w_0||_\infty<\gamma} ||h||^2_{\mathcal{H}_{tGP}}$. In fact

Lemma 1.8 *If* $w_0 \in S^{\chi}[0,1]^2 \cap H^{\chi}[0,1]^2$*, with* $\chi \leq \nu$*, then* $\inf_{h \in \mathcal{H}_{tGP}: ||h-w_0||_{\infty} < \gamma} ||h||^2_{\mathcal{H}_{tGP}} \leq \left(\frac{1}{\epsilon}\right)^{\frac{2\nu+2-2\chi}{2\chi}}$.

Proof Let $\hat{\kappa}_1$ be the Fourier transform of a function $\kappa_1 : \mathbb{R} \to \mathbb{R}$. Let the Fourier transform equals $\frac{1}{2\pi}$ in a nbd. of 0 and has compact support. Then $\frac{1}{2\pi} = \frac{1}{2\pi}\int e^{i\lambda t}\kappa_1(t)dt$. This gives rise to $\int \kappa_1(t)dt = 1$ and $\int (it)^k \kappa_1(t)dt = 0$ for $k \geq 1$. Thus κ_1 is a density which has vanishing moments. Define $\phi(t) = \kappa_1(t_1)\kappa_1(t_2)$. Clearly, ϕ also integrates to 1 and vanishing moments of all order. Further, for $\eta_1 > 0$, if we define $\phi_{\eta_1}(t) = \frac{1}{\eta_1^2}\phi(t/\eta_1)$ and $h = \phi_{\eta_1} * w_0$, using kernel estimation theory it can be shown that $||w_0 - \phi_{\eta_1} * w_0||_{\infty} \leq C_4 \eta_1^{\chi}$ (Van der Vaart and Van Zanten, 2009).

It must also be noted here that $\phi_{\eta_1} * w_0 \in \mathcal{H}_{tGP}$. To show this, we start with the fact that the Fourier transform of $\phi_{\eta_1} * w_0$ is $\hat{\phi}_{\eta_1}\hat{w}_0$. Now use the fact that the function is the inverse Fourier transform of its Fourier transform, i.e.

$$\phi_{\eta_1} * w_0(t) = (2\pi)^{-2}\int e^{i\lambda' t}\hat{w}_0(-\lambda)\hat{\phi}(\lambda\eta_1)d\lambda = (2\pi)^{-2}\int e^{i\lambda' t}\frac{\hat{w}_0(-\lambda)\hat{\phi}(\lambda\eta_1)}{f_{tGP}(\lambda)}d\lambda.$$

By construction of $\hat{w}_0$ and square integrability of $w(s)$, $\frac{\hat{w}_0(-\lambda)\hat{\phi}(\lambda\eta_1)}{f_{tGP}(\lambda)}$ is square integrable w.r.t f_{tGP}. $\phi_{\eta_1} * w_0 \in \mathcal{H}_{tGP}$ now follows by invoking Lemma 1.6. From the discussion preceding Lemma 1.6, for some constant $K > 0$,

$$\begin{aligned} ||h||^2_{\mathcal{H}_{tGP}} &\leq K\int |\hat{\phi}(\eta_1\lambda)\hat{w}_0(\lambda)|^2 \frac{1}{f_{tGP}(\lambda)}d\lambda \\ &\leq K \sup\left[\frac{1}{f_{tGP}(\lambda)(1+||\lambda||^2)^{\chi}}|\hat{\phi}(\eta_1\lambda)|^2\right]\int (1+||\lambda||^2)^{\chi}|\hat{w}_0(\lambda)|^2 d\lambda. \end{aligned}$$

Choosing the same $||\lambda_0||$ as in the proof of the Lemma 1.7 yields, $\sup_{\lambda} \frac{f_M(\lambda)}{f_{tGP}(\lambda)} \leq \max\{\sup_{||\lambda||<||\lambda_0||} \frac{f_M(\lambda)}{f_{tGP}(\lambda)}, 3\gamma_1/2\} = K_2$. Therefore, with $||w_0||_{\chi|2} = \int (1+||\lambda||^2)^{\chi}|\hat{w}_0(\lambda)|^2 d\lambda$,

$$\begin{aligned} ||h||^2_{\mathcal{H}_{tGP}} &\leq K_4 \sup\left[\frac{1}{f_M(\lambda)(1+||\lambda||^2)^{\chi}}|\hat{\phi}(\eta_1\lambda)|^2\right]||w_0||_{\chi|2} \leq \sup[(1+||\lambda||^2)^{\nu+1-\chi}|\hat{\phi}(\eta_1\lambda)|^2]||w_0||_{\chi|2} \\ &\leq K_5\frac{(1+||\lambda||^2)^{\nu+1-\chi}}{(1+||\eta_1\lambda||^2)^{\nu+1-\chi}} \leq K_6\left(\frac{1}{\eta_1}\right)^{\nu+1-\chi}, \end{aligned}$$

for constants $K_4, K_5, K_6 > 0$. Choosing $\eta_1 = \epsilon^{1/\chi}$, the result follows.

Lemma 1.7 and Lemma 1.8 yield, for $\chi \leq \nu$ and $w_0 \in S^\chi[0,1]^2 \cap H^\chi[0,1]^2$

$$\psi_{w_0}(\epsilon) \leq K \left[\frac{1}{\epsilon^{1/\nu}} + \frac{1}{\epsilon^{(2\nu+2-2\chi)/\chi}} \right]. \tag{6}$$

References

Adler, R. J. (2010). *The geometry of random fields*, Volume 62. Siam.

Edmunds, D. E. and H. Triebel (1996). Function spaces. *Entropy numbers, Differential operators 120*.

Furrer, R., M. G. Genton, and D. Nychka (2012). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*.

Ghosal, S., A. Van Der Vaart, et al. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics 35*(1), 192–223.

Kaufman, C. G., M. J. Schervish, and D. W. Nychka (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association 103*(484), 1545–1555.

Kuelbs, J. and W. V. Li (1993). Small ball estimates for brownian motion and the brownian sheet. *Journal of Theoretical Probability 6*(3), 547–577.

Li, W. V., W. Linde, et al. (1999). Approximation, metric entropy and small ball estimates for gaussian measures. *The Annals of Probability 27*(3), 1556–1578.

Minsker, S., S. Srivastava, L. Lin, and D. B. Dunson (2014). Robust and scalable bayes via a median of subset posterior measures. *arXiv preprint arXiv:1403.2660*.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging.* Springer Science & Business Media.

Vaart, A. v. d. and H. v. Zanten (2011). Information rates of nonparametric gaussian process methods. *Journal of Machine Learning Research 12* (Jun), 2095–2119.

van der Vaart, A. W. and J. H. van Zanten (2008). Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, 1435–1463.

van der Vaart, A. W. and J. H. van Zanten (2009). Adaptive bayesian estimation using a gaussian random field with inverse gamma bandwidth. *The Annals of Statistics*, 2655–2675.

Wendland, H. (2004). *Scattered data approximation*, Volume 17. Cambridge university press.