Supplement to "On Prediction Properties of Kriging: Uniform Error Bounds and Robustness"

Wenjia Wang*[1], Rui Tuo†[2] and C. F. Jeff Wu‡[3]

[1]The Statistical and Applied Mathematical Sciences Institute, Durham, NC 27709, USA
[2]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA
[3]The H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

1 Auxiliary tools

In this section, we review some mathematical tools which are used in the proofs presented in Appendix.

1.1 Reproducing kernel Hilbert spaces

In this subsection we introduce the reproducing kernel Hilbert spaces and several results from literature. Let Ω be a subset of $\mathbf{R}^d$. Assume that $K : \Omega \times \Omega \rightarrow \mathbf{R}$ is a symmetric

*Wenjia Wang is a postdoctoral fellow in the Statistical and Applied Mathematical Sciences Institute, Durham, NC 27709, USA (Email: wenjia.wang234@duke.edu);

†Rui Tuo is Assistant Professor in Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA (Email: ruituo@tamu.edu); Tuo's work is supported by NSF grant DMS 1564438 and NSFC grants 11501551, 11271355 and 11671386.

‡C. F. Jeff Wu is Professor and Coca-Cola Chair in Engineering Statistics at School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA (Email: jeff.wu@isye.gatech.edu). Wu's work is supported by NSF grant DMS 1564438.

The authors are grateful to an AE and referees for very helpful comments.

positive definite kernel. Define the linear space

$$F_K(\Omega) = \left\{ \sum_{i=1}^{n} \beta_i K(\cdot, x_i) : \beta_i \in \mathbf{R}, x_i \in \Omega, n \in \mathbb{N} \right\}, \tag{1.1}$$

and equip this space with the bilinear form

$$\left\langle \sum_{i=1}^{n} \beta_i K(\cdot, x_i), \sum_{j=1}^{m} \gamma_j K(\cdot, x_j') \right\rangle_K := \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_i \gamma_j K(x_i, x_j'). \tag{1.2}$$

Then the *reproducing kernel Hilbert space* $\mathcal{N}_K(\Omega)$ generated by the kernel function K is defined as the closure of $F_K(\Omega)$ under the inner product $\langle \cdot, \cdot \rangle_K$, and the norm of $\mathcal{N}_K(\Omega)$ is $\|f\|_{\mathcal{N}_K(\Omega)} = \sqrt{\langle f, f \rangle_{\mathcal{N}_K(\Omega)}}$, where $\langle \cdot, \cdot \rangle_{\mathcal{N}_K(\Omega)}$ is induced by $\langle \cdot, \cdot \rangle_K$. More detail about reproducing kernel Hilbert space can be found in Wendland (2004) and Wahba (1990). In particular, we have the following theorem, which gives another characterization of the reproducing kernel Hilbert space when K is defined by a stationary kernel function Φ, via the Fourier transform of Φ.

Theorem 1 (Theorem 10.12 of Wendland (2004))**.** *Let Φ be a positive definite kernel function which is continuous and integrable in $\mathbf{R}^d$. Define*

$$\mathcal{G} := \{ f \in L_2(\mathbf{R}^d) \cap C(\mathbf{R}^d) : \tilde{f} / \sqrt{\tilde{\Phi}} \in L_2(\mathbf{R}^d) \},$$

with the inner product

$$\langle f, g \rangle_{\mathcal{N}_\Phi(\mathbf{R}^d)} = (2\pi)^{-d} \int_{\mathbf{R}^d} \frac{\tilde{f}(\boldsymbol{\omega}) \overline{\tilde{g}(\boldsymbol{\omega})}}{\tilde{\Phi}(\boldsymbol{\omega})} d\boldsymbol{\omega}.$$

Then $\mathcal{G} = \mathcal{N}_\Phi(\mathbf{R}^d)$, and both inner products coincide.

For $f \in \mathcal{N}_\Phi(\Omega)$, a pointwise error bound for the radial basis function interpolation is given by (Wendland (2004), Theorem 11.4):

$$|f(\boldsymbol{x}) - \mathcal{I}_{\Phi,\mathbf{X}} f(\boldsymbol{x})| \leq P_{\Phi,\mathbf{X}}(\boldsymbol{x}) \|f\|_{\mathcal{N}_\Phi(\Omega)}. \tag{1.3}$$

In addition, it can be shown that the interpolant $\mathcal{I}_{\Phi,\mathbf{X}} f(\boldsymbol{x})$ satisfies the following properties (Corollary 10.25, Wendland (2004)):

$$\|\mathcal{I}_{\Phi,\mathbf{X}} f(\boldsymbol{x})\|_{\mathcal{N}_\Phi(\Omega)} \leq \|f\|_{\mathcal{N}_\Phi(\Omega)}. \tag{1.4}$$

In addition, if $\mathbf{X}' \subset \mathbf{X}$,

$$\|\mathcal{I}_{\Phi,\mathbf{X}'} h\|_{\mathcal{N}_\Phi(\Omega)} \leq \|\mathcal{I}_{\Phi,\mathbf{X}} h\|_{\mathcal{N}_\Phi(\Omega)}. \tag{1.5}$$

1.2 A Maximum inequality for Gaussian processes

The theory of bounding the maximum value of a Gaussian process is well-established in the literature. The main step of finding an upper bound is to calculate the *covering number* of the index space. Here we review the main results. Detailed discussions can be found in Adler and Taylor (2009).

Let Z_t be a Gaussian process indexed by $t \in T$. Here T can be an arbitrary set. The Gaussian process Z_t induces a metric on T, defined by

$$\mathfrak{d}(t_1, t_2) = \sqrt{\mathbb{E}(Z_{t_1} - Z_{t_2})^2}. \tag{1.6}$$

The ϵ-covering number of the metric space $(T, \mathfrak{d})$, denoted as $N(\epsilon, T, \mathfrak{d})$, is the minimum integer N so that there exist N distinct balls in $(T, \mathfrak{d})$ with radius ϵ, and the union of these balls covers T. Let D be the diameter of T. The supremum of a Gaussian process is closely tied to a quantity called the *entropy integral*, defined as

$$\int_0^{D/2} \sqrt{\log N(\epsilon, T, \mathfrak{d})} d\epsilon. \tag{1.7}$$

Lemma 1 gives a maximum inequality for Gaussian processes, which is a direct consequence of Theorems 1.3.3 and 2.1.1 of Adler and Taylor (2009).

Lemma 1. *Let Z_t be a centered separable Gaussian process on a $\mathfrak{d}$-compact T, $\mathfrak{d}$ the metric, and N the ϵ-covering number. Then there exists a universal constant K such that for all $u > 0$,*

$$\mathbb{P}(\sup_{t \in T} |Z_t| > K \int_0^{D/2} \sqrt{\log N(\epsilon, T, \mathfrak{d})} d\epsilon + u) \leq 2e^{-u^2/2\sigma_T^2}, \tag{1.8}$$

where $\sigma_T^2 = \sup_{t \in T} \mathbb{E} Z_t^2$.

2 Additional figure related to Table 2

Figure 1 shows the relationship between the logarithm of the fill distance (i.e., $\log h_{\mathbf{X}}$) and the logarithm of the average prediction error (i.e., $\log \mathcal{E}$) in scatter plots for the four cases given in Table 2. The solid line in each panel shows the linear regression fit calculated from the data. Each of the regression lines in Figure 1 fits the data very well, which

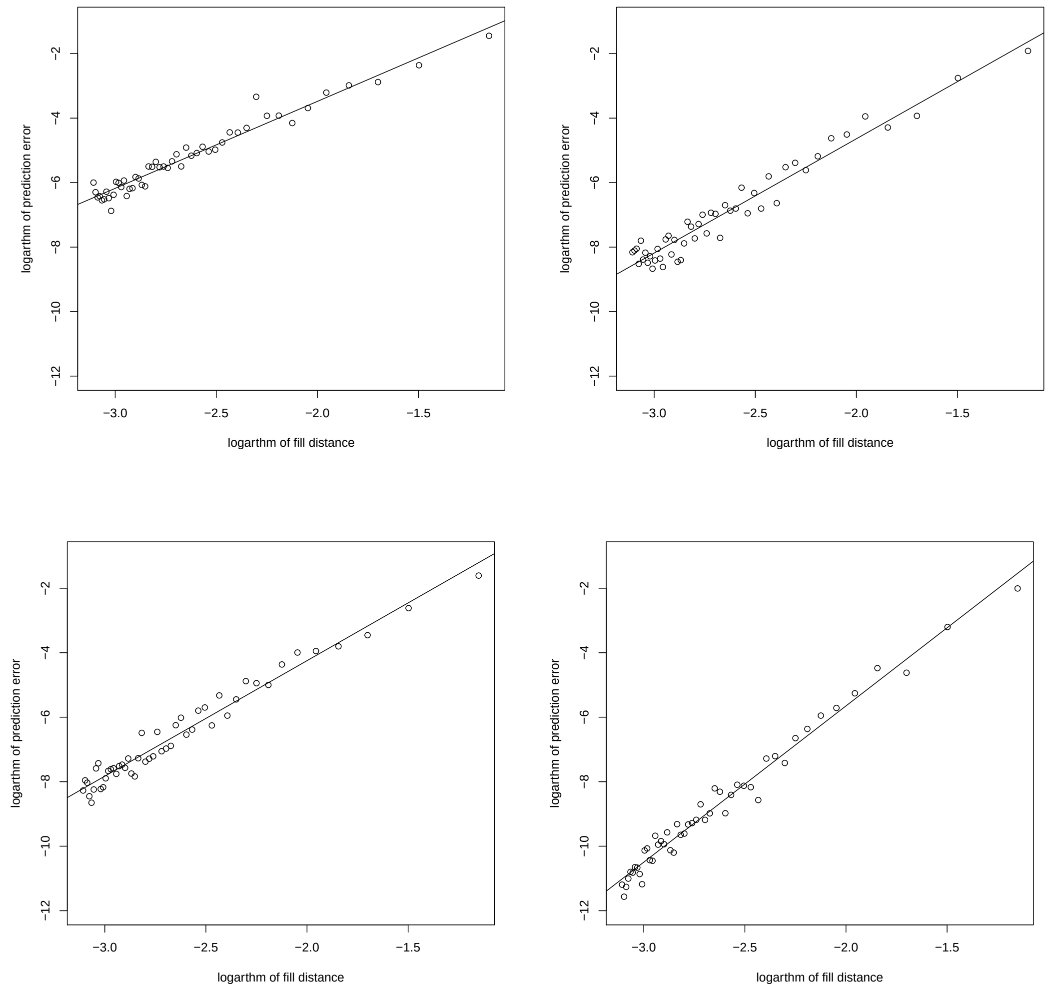

Figure 1: The regression line of $\log \sup_{\boldsymbol{x} \in \Omega} \epsilon(\boldsymbol{x})$ on $\log h_{\mathbf{X}}$. Each point denotes one average prediction error for each n. **Panel 1:** $\nu_0 = 3$, $\nu = 2.5$. **Panel 2:** $\nu_0 = 5$, $\nu = 3.5$. **Panel 3:** $\nu_0 = \nu = 3.5$. **Panel 4:** $\nu_0 = \nu = 5$.

gives an empirical confirmation of the approximation in (4.2) in the main text. It is also observed from Figure 1 that, as the fill distance decreases, the maximum prediction error also decreases.

References

Adler, R. J. and J. E. Taylor (2009). *Random Fields and Geometry.* Springer Science & Business Media.

Wahba, G. (1990). *Spline Models for Observational Data*, Volume 59. SIAM.

Wendland, H. (2004). *Scattered Data Approximation*, Volume 17. Cambridge University Press.