Statistica Sinica **29** (2019), 1007-1024
doi:https://doi.org/10.5705/ss.202016.0440

ASYMPTOTIC DISTRIBUTION FOR REGRESSION IN A SYMMETRIC PERIODIC GAUSSIAN KERNEL HILBERT SPACE

Xianli Zeng[1] and Yingcun Xia[1,2]

[1] *National University of Singapore and*
[2] *University of Electronic Science and Technology, China*

Abstract: The reproducing kernel Hilbert space (RKHS) method is arguably the most popular approach in machine learning to deal with nonlinearity in data. However, it has not been as widely adopted in statistical analyses as might be expected. One reason is that its statistical properties have not yet been adequately understood and, in particular, its asymptotic inference. In this paper, we introduce a symmetric periodic Gaussian kernel and show, in the generic regression setting where the regression function is in the Sobolev spaces, that the method under consideration is asymptotically normal. This asymptotic distribution also provides an explanation to the estimation efficiency of regularization method using the Gaussian reproducing kernel. We include simulation results to illustrate the finite sample properties of the method.

Key words and phrases: Asymptotic normality, estimation efficiency, Gaussian reproducing kernel, nonparametric estimation, sobolev space.

1. Introduction

The method of regularization with reproducing kernel has gained tremendous popularity in machine learning literature to deal with nonlinearity in the data; see, for example, Wahba (1999), Schölkopf and Smola (2002) and Hofmann, Schölkopf and Smola (2008). The key reason for such popularity is the computational efficiency of the method that has led to successful applications in many areas of research. In machine learning literature, much attention has been paid to the consistency of the method, while, in statistical analysis, asymptotic inference is of more interest. As far as we know, little research has been undertaken on the asymptotic distribution of the estimated function. Our asymptotic results developed in this article deal with this statistical interest.

In statistical analysis, many problems can be considered as the nonparametric regression model

$$y_i = g_0(x_i) + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where (x_i, y_i), $i = 1, \ldots, n$, are IID samples of (X, Y), g_0 is an unknown regression function in a Hilbert space (denoted by $\mathcal{H}$, hereafter), and ϵ is the random error, which is independent of X with $E(\epsilon_i) = 0$ and $E(\epsilon_i^2) = \sigma^2$. Denote by f the density function of X, supported on $[0, \pi]$ without loss of generality. The goal is to estimate the unknown regression function g_0 based on data $(x_i, y_i)_{i=1}^n$. The method of regularization estimates the true regression function g_0 by minimizing the penalized loss function

$$\hat{g} = \text{argmin}_{g \in \mathcal{H}}[L(g|\text{data}) + \lambda J(g)]. \tag{1.2}$$

In (1.2), the first term $L(g|\text{data})$ assesses the fitness of function g to observed random samples, while $J(g)$ penalizes the complexity of g, with the parameter λ that balances the fitness and the complexity. Compared with other regression methods, the method of regularization with reproducing kernel enjoys several advantages in computation. First, the solution of this method is guaranteed to exist and is unique. Second, the solution can be represent by the linear combinations of the kernel function evaluated at sample points. Third, it is convenient to extend to multivariate cases. Furthermore, for some specific kernels, the solutions of this RKHS method can be expressed by a series of orthogonal basis functions, which facilitates the theoretic analysis of its asymptotic performance.

For any Sobolev space in which the regression function g_0 is considered, there is a corresponding inherent kernel that generates the space. Even though statistical properties of using those kernels have been extensively studied, for example by Yuan and Cai (2010) and Shang and Cheng (2013), only a few practitioners really use those kernels because they are very complicated and do not have explicit forms for multivariate spaces. Hence, in existing studies, other reproducing kernels such as Gaussian reproducing kernel and polynomial kernel are often used instead. Using standard empirical processes and other techniques for deriving consistency, convergence rates for kernel based learning, especially for support vector machines(SVMs), have been obtained. See for example Williamson, Smola and Schölkopf (2001), Mendelson (2002), Kosorok (2008) and Steinwart and Christmann (2008).

Among all the reproducing kernels, the Gaussian reproducing kernel is the most popular in practice. Some theories for using the Gaussian reproducing kernel have been studied in machine learning literature. Keerthi and Lin (2003) proved the consistency of SVMs with the Gaussian reproducing kernel. Steinwart, Hush and Scovel (2006) provided explicit descriptions for the RKHS corresponding to the Gaussian reproducing kernel and discussed the applications of their results in analyzing the learning performance of SVMs. Steinwart and Scovel (2007) generated a fast convergence rate for SVMs using the Gaussian reproducing kernel by proposing a new geometric noise condition. More recently, Eberts and Steinwart (2013) derived the optimal rates for SVMs using a Gaussian reproducing kernel under certain smoothness conditions. Nonetheless, asymptotic normality has not yet been studied, which impedes its application in making statistical inference.

Because the eigen-decomposition for the Gaussian reproducing kernel is very complicated and the eigen-functions are unbounded, theoretical analysis for asymptotic properties is very difficult. To gain insights into the performance of the Gaussian reproducing kernel, Lin and Brown (2004) systematically analyzed the asymptotic properties for the method that uses the periodic Gaussian kernel. Their results only explained the consistency of the method. By deriving the asymptotic distributions, our work gives a clearer explanation of the estimation efficiency of the method that uses the Gaussian reproducing kernel.

In this paper, we introduce a symmetric periodic Gaussian kernel that possesses a number of advantages. First, this kernel has a much closer approximation to a Gaussian reproducing kernel than the periodic Gaussian kernel. In this sense, our results can provide deeper insights on the statistical properties of the method of regularization with the Gaussian reproducing kernel. Second, the method is applicable to regression functions that are not necessarily periodic on their support. In addition, it also allows us to implement the eigen-decomposition and simultaneous diagonalization to derive asymptotic theory. Our estimator achieves the same consistency rate as in Lin and Brown (2004), which is optimal in estimating functions in any finite order of Sobolev spaces, and is asymptotically minimax in estimating functions in the infinite order Sobolev Space. Moreover, the so-called functional Bahadur representation(FBR, Shang and Cheng (2013)) is derived for our estimated function, and is applied to derive the asymptotic normality for our method and for that of Lin and Brown (2004).

The paper is organized as follows. In Section 2, we make a brief review of the RKHS and relevant notations, and introduce our symmetric periodic Gaussian kernel. Estimation procedures are discussed in Section 3. Our main asymptotic results are shown in Section 4. In Section 5, we derive the asymptotic bias and variance for our method and that of Lin and Brown (2004), and compare the two methods using three examples. Simulation studies are contained in Section 6. Proofs are presented in the online Supplementary Material.

2. Symmetric Periodic Gaussian Kernel Hilbert Space

2.1. Reproducing kernel Hilbert space (RKHS)

The reproducing kernel Hilbert space (RKHS), developed by Aronszajn (1950), is a Hilbert space of functions in which all the evaluation functionals are bounded. Let $\mathcal{X}$ be an arbitrary set of real values, $\mathcal{H}$ be a Hilbert space of functions supported on $\mathcal{X}$, and $||\cdot||_{\mathcal{H}}$ be the norm on $\mathcal{H}$. $\mathcal{H}$ is a RKHS if, for all $x \in \mathcal{X}$, there exists a positive number M such that,

$$L_x(f) = |f(x)| \leq M||f||_{\mathcal{H}} \quad \text{for all} \;\; f \in \mathcal{H},$$

where L_x is called evaluation functional.

According to this definition, if f and g are two functions in RKHS with small $||f-g||_{\mathcal{H}}$, then f and g are also close pointwise. A more intuitive definition of RKHS is provided by using a reproducing kernel(RK) as a representer of an evaluation functional.

A function $K(\cdot,\cdot)$ denfined on $\mathcal{X} \times \mathcal{X}$ is a RK if it is symmetric positive definite: for any $t_1, t_2, \ldots, t_N \in \mathcal{X}$ and $a_1, a_2, \ldots, a_N \in \mathcal{R}, N = 1, 2, \ldots,$

$$\sum_{i,j=1}^{N} a_i a_j K(t_i, t_j) \geq 0.$$

As stated in Moore-Aronszajn Theorem (Aronszajn (1950)), every symmetric positive definite function corresponds to a unique RKHS of real-valued functions. More concisely, the RK generates a Hilbert space of functions as

$$\mathcal{K} = \{g(\cdot) = \sum_{i=1}^{m} \alpha_i K(t_i, \cdot) : m \in \boldsymbol{N}, t_i \in \mathcal{X}, \alpha_i \in \boldsymbol{R}\},$$

with inner product defined such that

$$\langle K(t_i, \cdot), K(t_j, \cdot) \rangle_{\mathcal{K}} = K(t_i, t_j). \tag{2.1}$$

It can be proved that $\mathcal{K}$ is a RKHS. In addition, functions in $\mathcal{K}$ satisfy the reproducing property:

$$\langle K(t, \cdot), g(\cdot) \rangle_{\mathcal{K}} = g(t) \quad \text{for every } g \in \mathcal{K}.$$

In this discussion, we see how a RK generates a RKHS. Conversely, a RKHS defines a unique RK as follows. Let $\mathcal{K}$ be a RKHS endowed with inner product $\langle \cdot, \cdot \rangle_{\mathcal{K}}$. By Riesz representing theorem, for each $t \in \mathcal{X}$, there exists a function $K_t(\cdot)$ such that $\langle K_t(\cdot), g(\cdot) \rangle_{\mathcal{K}} = g(t)$. If $K(s,t) = \langle K_s(\cdot), K_t(\cdot) \rangle_{\mathcal{K}}$, then $K(\cdot, \cdot)$ is a reproducing kernel.

By applying Mercer's theorem (1909) on the integral operator defined by the reproducing kernel, we present another view on reproducing kernel Hilbert space. Let

$$L_K g(\cdot) = \int_{\mathcal{X}} K(\cdot, t) g(t) dt.$$

Mercer (1909) states that any integral operator defined by a reproducing kernel has a spectral decomposition. And the reproducing kernel can be represented by

$$K(s,t) = \sum_{k=0}^{\infty} \lambda_k \phi_k(s) \phi_k(t).$$

Here λ_k's are the eigen-values of $K(\cdot,\cdot)$ with the $\phi_k(\cdot)$ as the corresponding eigen-function,

$$K\phi_k(\cdot) = \int_{\mathcal{X}} K(\cdot, t)\phi_k(t) dt = \lambda_k \phi_k(\cdot).$$

Furthermore, $\phi_k(\cdot)$'s satisfy

$$\langle \phi_i, \phi_j \rangle_{\mathcal{K}} = \frac{\delta_{i,j}}{\lambda_i},$$

where $\delta_{i,j}$ is the Kronecker delta, and the eigen-functions form a sequence of orthogonal basic functions. Since RKHS and RK are one-to-one correspondence, the RKHS can be defined alternatively by these eigen-values and eigen-functions,

$$\mathcal{K} = \left\{ g : g(t) = \sum_{k=1}^{\infty} g_k \phi_k(t), \ \sum_{k=1}^{\infty} \frac{g_k^2}{\lambda_k} < \infty \right\},$$

where $g_k = \int_{\mathcal{X}} g(t)\phi_k(t)dt$. We then call the λ_k's and $\phi_k(\cdot)$'s the eigen-values and eigen-functions for RKHS $\mathcal{K}$. As any function in RKHS can be represented by a series of orthogonal basic functions, the asymptotic behaviour of derived estimator can thus be analyzed more conveniently.

2.2. Symmetric periodic gaussian kernel Hilbert space

We first introduce the periodic Gaussian kernel and its corresponding RKHS. Smola, Schölkopf and Müller (1998) proposed the periodic Gaussian kernel to estimate periodic functions on a compact interval, say $[0,\pi]$ without loss of generality.

$$K_{\omega,0,\pi}(s,t) = \sum_{k=-\infty}^{\infty} K_\omega(s - t - k\pi, 0) \qquad \text{with } s,t \in [0,\pi] \tag{2.2}$$

where

$$K_\omega(s,t) = \frac{1}{\sqrt{2\pi}\omega} e^{-(s-t)^2/\omega^2} \tag{2.3}$$

is the well-known Gaussian reproducing kernel.

To study the asymptotic performance of their estimator, Lin and Brown (2004) introduced two RKHSs: an infinite order Sobolev space with periodic functions,

$$\mathcal{S}^{\infty}_{\omega[a,b]} = \{g \in L^2(a,b) : g \text{ is } (b-a) - \text{periodic with} \\ \sum_{m=0}^{\infty} \frac{\omega^{2m}}{m!2^m} \int_a^b [g^{(m)}(t)]^2 dt < \infty\}.$$

and an m-th order Sobolev space with periodic functions,

$$\mathcal{S}^{m}_{[a,b]} = \{g \in L^2(a,b) : g \text{ is } (b-a) - \text{periodic with} \\ \int_a^b [g(t)]^2 + [g^{(m)}(t)]^2 dt < \infty\};$$

The norm of the estimated function is a natural penalty function for $J(g)$ in (1.2); see for example Evgeniou, Pontil and Poggio (2000). In some cases, the explicit form of this norm or penalty can be derived with the help of Green functions and Fourier transforms. For example, the norms or penalties with respect to the period Gaussian kernel can be written as,

$$\sum_{m=0}^{\infty} \frac{\omega^{2m}}{2^m m!} \int_0^{\pi} [g^{(m)}(t)]^2 dt.$$

The RKHS generated by periodic Gaussian kernel (2.2) is the infinite order Sobolev space $\mathcal{S}^{\infty}_{\omega[0,\pi]}$.

We introduce the symmetric periodic Gaussian kernel

$$H_{\omega}(s,t) = K_{\omega,-\pi,\pi}(s,t) + K_{\omega,-\pi,\pi}(s,-t). \tag{2.4}$$

Here, $K_{\omega,-\pi,\pi}(s,t) = \sum_{k=-\infty}^{\infty} K_{\omega}(s-t-2k\pi, 0)$ is the periodic Gaussian kernel with period 2π. Denote by $\mathcal{H}^{\infty}_{\omega[-\pi,\pi]}$ the RKHS corresponding to $H_{\omega}(s,t)$. This RKHS consists of symmetric functions on $[-\pi,\pi]$, and is a subspace of infinite order Sobolev space. As we can see, $\mathcal{H}^{\infty}_{\omega[-\pi,\pi]}$ is an infinite order Sobolev space with symmetric functions.

Proposition 1. *Let $\mathcal{H}^{\infty}_{\omega[-\pi,\pi]}$ be the RKHS corresponding to kernel H_{ω}, then*

$$\mathcal{H}^{\infty}_{\omega[-\pi,\pi]} = \left\{ g : \; g(t) = \sum_{k=0}^{\infty} g_k \xi_k(t), \; \sum_{k=0}^{\infty} \frac{g_k^2}{\lambda_{k,\omega}} < \infty \right\} \\ = \left\{ g : g(-t) = g(t), \; g \in \mathcal{S}^{\infty}_{\omega[-\pi,\pi]} \right\}$$

with $\lambda_{k,\omega} = e^{-(k^2\omega^2)/2}$, $\xi_0(t) = \pi^{-(1/2)}$, $\xi_k(t) = \sqrt{2/\pi}\cos(kt)$.

For the m-th order Sobolev space with symmetric functions,

$$\mathcal{H}^m_{[-\pi,\pi]} = \left\{ g : g(-t) = g(t), g \in \mathcal{S}^m_{[-\pi,\pi]} \right\}.$$

Similar to Proposition 1, we have

$$\mathcal{H}^m_{[-\pi,\pi]} = \left\{ g : \ g(t) = \sum_{k=0}^{\infty} g_k \xi_k(t), \sum_{k=0}^{\infty} \frac{g_k^2}{\rho_k} < \infty \right\}$$

with $\rho_0 = 1$ and $\rho_k = k^{2m} + 1$.

Theorem 1. *Suppose $K_\omega(s,t)$ and $H_\omega(s,t)$ are defined in* (2.3) *and* (2.4) *respectively, and $x_1, x_2, \ldots, x_n$ is any set of values on* $[0,\pi]$. *Let $\omega = n^{-a}$ for $a > 0$. Then, for any $\delta > 0$ and $M > 0$,*

$$\sup_{\max_i |\alpha_i| < M, x \in [\delta, \pi-\delta]} \left| \sum_{i=1}^{n} [H_\omega(x_i, x)\alpha_i - K_\omega(x_i, x)\alpha_i] \right| \to 0 \quad as \quad n \to \infty. \tag{2.5}$$

Thus, the symmetric periodic Gaussian kernel is close to the Gaussian reproducing kernel when $\omega \to 0$. The results in Section 4 facilitate the understanding of the asymptotic performance of estimators that are based on the Gaussian reproducing kernel.

3. The method of regularization with a symmetric periodic Gaussian kernel

Suppose (x_i, y_i) are IID samples from model (1.1). The method of regularization with a symmetric periodic Gaussian kernel estimates g_0 as follows. Consider a loss function $L_n(g)$ with a penalty function $J(g)$,

$$L_n(g) = \frac{1}{n} \sum_{i=1}^{n} (y_i - g(x_i))^2, \quad \text{and} \quad J(g) = ||g||^2_{\mathcal{H}_\omega} = \langle g, g \rangle_{\mathcal{H}_\omega},$$

where $\langle g, g \rangle_{\mathcal{H}_\omega}$ is defined in the same way as $\langle g, g \rangle_{\mathcal{K}}$ in (2.1). Then, the estimator is the solution of

$$\hat{g} = \text{argmin}_g \{ L_n(g) + \lambda J(g) \}. \tag{3.1}$$

Proposition 2 (representer theorem,Wahba (1990))**.** *If $g_0 \in \mathcal{H}^\infty_{\omega_0[-\pi,\pi]}$ for $\omega_0 > \omega$ or $g_0 \in \mathcal{H}^m_{[-\pi,\pi]}$, then there exist constants $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that*

$$\hat{g}(t) = \sum_{i=1}^{n} H_\omega(x_i, t)\alpha_i. \tag{3.2}$$

The representer theorem indicates that the solution of (3.1) can be found in the space generated by basis functions $H_\omega(x_i, \cdot), i = 1, 2, \ldots, n$. Denote this function space by $\mathcal{H}_n = \{ g(t) = \sum_{i=1}^n H_\omega(x_i, t)\alpha_i \}$. This result is important to both numerical implementation and theoretical study. It follows that for any function $g \in \mathcal{H}_n$,

$$L_n(g) = \frac{1}{n}(Y - H\alpha)^\top (Y - H\alpha) \quad \text{and} \quad J(g) = \alpha^\top H\alpha,$$

where $Y = (y_1, y_2, \dots, y_n)^\top$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^\top$ and $H = (H_{ij})$ is a $n \times n$ matrix with $H_{ij} = H_\omega(x_i, x_j)$. Then

$$L_n(g) + \lambda J(g) = \frac{1}{n}(Y - H\alpha)^\top (Y - H\alpha) + \lambda\alpha^\top H\alpha.$$

Thus, estimating g in (3.1) is equivalent to estimating α. The solution is,

$$\hat{\alpha} = (H + n\lambda)^{-1} Y.$$

4. Asymptotic Analysis

To obtain the asymptotic performance of our method, we need some assumptions.

Assumption 1. *$\{x_i\}_{i=1}^n$ are IID samples of variable X. The density function $f(x)$ of X is supported on $[0, \pi]$ and satisfies $0 < c < f(x) < C < \infty$ for some constants c and C;*

Assumption 2. *$\{\epsilon_i\}_{i=1}^n$ is a sequence of IID random variables that are independent of X; $E(\epsilon_i) = 0$ and $E(\epsilon_i^2) = \sigma^2$;*

Assumption 3.a. $g_0 \in \mathcal{H}^\infty_{\omega_0[-\pi,\pi]}$;

Assumption 3.b. $g_0 \in \mathcal{H}^m_{[-\pi,\pi]}$.

To facilitate the theoretical calculation, we standardize our symmetric periodic Gaussian kernel as

$$\tilde{H}_\omega(s, t) = \frac{H_\omega(s, t)}{\sqrt{f(s)f(t)}}.$$

We also denote this kernel by $H_\omega(s, t)$ for simplicity.

Proposition 3. *Let $H_\omega(s, t)$ be the standardized symmetric periodic Gaussian kernel, then*

$$H_\omega(s, t) = \sum_{k=0}^\infty \lambda_{k,\omega} \phi_k(s) \phi_k(t)$$

with $\lambda_{k,\omega}$'s and $\phi_k(t)$'s satisfying

(i) $\lambda_{k,\omega} = e^{-(k^2\omega^2)/2}$;

(ii) $\phi_0(t) = 1/\sqrt{\pi f(t)}$ *and* $\phi_k(t) = (\sqrt{2}\cos(kt))/\sqrt{\pi f(t)}$;

(iii) $\int_0^\pi \phi_i(t)\phi_j(t)f(t)dt = \delta_{i,j}$;

(iv) $\langle \phi_i(t), \phi_j(t)\rangle_{\mathcal{H}_\omega} = \delta_{i,j}/\lambda_i$;

(v) $\sup_k \sup_t |\phi_k(t)| < \infty$.

Let $||g||^2_{\mathcal{H}_\omega} = \langle g, g\rangle_{\mathcal{H}_\omega}$ as defined above. For the asymptotic analysis, we take

$$\begin{aligned} ||g||_0 &= [E_f(g^2(X))]^{1/2} = \left[\int_0^\pi g^2(t)f(t)dt\right]^{1/2}, \\ ||g||_\lambda &= (||g||_0^2 + \lambda||g||^2_{\mathcal{H}_\omega})^{1/2}, \\ \langle g_1, g_2\rangle_0 &= \frac{1}{4}(||g_1 + g_2||_0^2 - ||g_1 - g_2||_0^2), \\ \langle g_1, g_2\rangle_\lambda &= \langle g_1, g_2\rangle_0 + \lambda\langle g_1, g_2\rangle_{\mathcal{H}_\omega}. \end{aligned}$$

Theorem 2 (simultaneous diagonalization)**.** *For any $g \in \mathcal{H}_\omega$, if $g_k = \langle g, \phi_k\rangle_0$ with the ϕ_k given in Proposition* 3, then

$$g(t) = \sum_{k=0}^{\infty} g_k\phi_k(t),$$

with

$$||g||_0^2 = \sum_{k=0}^{\infty} g_k^2, \quad J(g) = \sum_{k=0}^{\infty} \frac{g_k^2}{\lambda_{k,\omega}}.$$

4.1. Consistency

The consistency for RKHS-based learning has been established using general empirical processes; see for example Mendelson (2002) and Steinwart and Christmann (2008). Based on the technique in Silverman (1982), the consistency rates for regularization estimators of model (1.1) in $\mathcal{H}^\infty_{\omega[-\pi,\pi]}$ and $\mathcal{H}^m_{[-\pi,\pi]}$, can also be derived and are given below. It can been seen that, the consistency rate is minimax for the estimator in the infinite order Sobolev space, as shown in Lin and Brown (2004). The rate is also optimal for nonparametric regression in the mth order derivative function space; see, for example, Stone (1982).

Hereafter, for any two sequences a and b that depend on n, take $a \sim b$ if $\lim_{n\to\infty} a/b = c$ with $0 < c < \infty$.

Theorem 3. *Suppose Assumptions* 1, 2 and 3 hold. When $\omega = \omega_0$ and $\lambda \sim (\ln n)^{1/2}/n$, the regularization estimator $\hat{g}$ satisfies

$$||\hat{g} - g_0||_0^2 = O_p\left(\frac{(\ln n)^{1/2}}{n}\right).$$

When $\omega^2 < \omega_0^2/2$ and $\lambda \sim (\ln n)^{1/4}/(n\omega)^{1/2}$, the regularization estimator $\hat{g}$ satisfies

$$||\hat{g} - g_0||_0^2 = O_p\left(\frac{(\ln n)^{1/2}}{n\omega}\right).$$

Theorem 4. *Suppose Assumptions* 1, 2 and 3.b hold. The regularization estimator $\hat{g}$ derived with $\lambda = o(1)$ and $(-\ln\lambda)^{1/2}/\omega \sim n^{1/(2m+1)}$ satisfies

$$||\hat{g} - g_0||_0^2 = O_p\left(n^{-2m/(2m+1)}\right).$$

4.2. Asymptotic normality

Here, we present our main results for the asymptotic normality of the regularization estimator we introduced.

Let $H_{\omega_t} = H_\omega(t,\cdot)$. For any $\Delta g \in \mathcal{H}_\omega$, write

$$S_{n\lambda}(g) = -\frac{2}{n}\sum_{i=1}^{n}(y_i - g(x_i))H_{\omega_{x_i}} + 2\lambda g;$$

$$DS_{n\lambda}(g)\Delta g = \frac{2}{n}\sum_{i=1}^{n}\Delta g(x_i)H_{\omega_{x_i}} + 2\lambda\Delta g.$$

Let $S_\lambda(g) = E_f\{S_{n\lambda}(g)\}$ and $DS_\lambda(g)\Delta g = E_f\{DS_{n\lambda}(g)\Delta g\}$.

Lemma 1 (Functional Bahadur representation)**.** *Suppose Assumptions* 1, 2 and 3.a hold. When $\omega = \omega_0$ and $\lambda \sim (\ln n)^{1/2}/n = o(1)$ as $n \to \infty$, we have

$$||\hat{g} - g_0 + (DS_\lambda(g_0))^{-1}S_{n\lambda}(g_0)||_\lambda^2 = O_p\left(\frac{\ln n}{n^2}\right).$$

We apply this FBR to get point-wise asymptotic normality for the estimators in the two Sobolev spaces.

Theorem 5. *Suppose Assumptions* 1, 2 and 3.a hold, and that parameter λ and ω satisfy the conditions in Lemma 1. With $g_0(t) = \sum_{k=0}^{\infty} g_{0,k}\phi_k(t)$, for any $x_0 \in [-\pi,\pi]$, there exists a constant $\sigma_{x_0}^2 > 0$ such that

$$\lim_{n\to\infty}\frac{\sigma^2}{(\ln n)^{1/2}}\sum_{k=0}^{\infty}\left(1 + \frac{\lambda}{\lambda_{k,\omega}}\right)^{-2}\phi_k^2(x_0) = \sigma_{x_0}^2. \tag{4.1}$$

with $g^\star(t) = \sum_{k=0}^{\infty}(\lambda_{k,\omega}g_{0,k})/(\lambda + \lambda_{k,\omega})\phi_k(t)$,

$$\sqrt{\frac{n}{(\ln n)^{1/2}}}(\hat{g}(x_0) - g^\star(x_0)) \xrightarrow{d} N\left(0, \sigma_{x_0}^2\right). \tag{4.2}$$

Theorem 6. *Suppose Assumptions* 1, 2 and 3.b hold. When $\lambda = o(1)$ and

$(-\ln\lambda)^{1/2}/\omega \sim n^{1/(2m+1)}$ as $n \to \infty$, we have

$$||\hat{g} - g_0 + (DS_\lambda(g_0))^{-1}S_{n\lambda}(g_0)||_\lambda^2 = O_p\left(n^{-4m/(2m+1)}\right). \tag{4.3}$$

For any $x_0 \in [-\pi, \pi]$, there exists a constant $\tilde{\sigma}_{x_0}^2 > 0$ such that

$$\lim_{n\to\infty} \sigma^2 n^{-1/(2m+1)} \sum_{k=0}^{\infty} \left(1 + \frac{\lambda}{\lambda_{k,\omega}}\right)^{-2} \phi_k^2(x_0) = \tilde{\sigma}_{x_0}^2. \tag{4.4}$$

If $g_0(t) = \sum_{k=0}^\infty g_{0,k}\phi_k(t)$ and $g^\star(t) = \sum_{k=0}^\infty (\lambda_{k,\omega} g_{0,k})/(\lambda + \lambda_{k,\omega})\phi_k(t)$, then

$$n^{m/(2m+1)}(\hat{g}(x_0) - g^\star(x_0)) \xrightarrow{d} N\left(0, \tilde{\sigma}_{x_0}^2\right). \tag{4.5}$$

5. Comparison of Estimation Efficiency

We compare the periodic Gaussian kernel and the symmetric periodic Gaussian kernel by their estimation efficiencies and asymptotic distributions. Lin and Brown (2004) did not obtained the asymptotic distribution for their estimators. Here, we establish the asymptotic normality for the method in Lin and Brown (2004), and then we compare the asymptotic bias and variance with those of our method.

Theorem 7. *Under Assumptions* 1 and 2, suppose $g_0 \in \mathcal{S}_{\omega_0[0,\pi]}^\infty$ and ω_0 is a fixed value. Let the η_k's and ψ_k's be the eigen-values and eigen-functions of the standardized periodic Gaussian kernel $K_{\omega,0,\pi}$. Write $g_0(t) = \sum_{k=0}^\infty g_{0,k}\psi_k(t)$. If $\omega = \omega_0$ and $\lambda \sim (\ln n)^{1/2}/n = o(1)$, then for any $x_0 \in [0,\pi]$, there exists a constant $a_{x_0}^2 > 0$ such that

$$\lim_{n\to\infty} \frac{\sigma^2}{(\ln n)^{1/2}} \sum_{k=0}^{\infty} \left(1 + \frac{\lambda}{\eta_k}\right)^{-2} \psi_k^2(x_0) = a_{x_0}^2. \tag{5.1}$$

For $g^\star(t) = \sum_{k=0}^\infty (\eta_k g_{0,k})/(\lambda + \eta_k)\psi_k(t)$, we have

$$\sqrt{\frac{n}{(\ln n)^{1/2}}}(\hat{g}(x_0) - g^\star(x_0)) \xrightarrow{d} N\left(0, a_{x_0}^2\right). \tag{5.2}$$

Theorem 8. *Suppose Assumptions* 1 and 2 hold, and that $g_0 \in \mathcal{S}_{[0,\pi]}^m$. If $\lambda = o(1)$ and $(-\ln\lambda)^{1/2}/\omega \sim n^{1/(2m+1)}$ as $n \to \infty$, then for any $x_0 \in [0,\pi]$, there exists a constant $\tilde{a}_{x_0}^2 > 0$ such that

$$\lim_{n\to\infty} \sigma^2 n^{-1/(2m+1)} \sum_{k=0}^{\infty} \left(1 + \frac{\lambda}{\eta_k}\right)^{-2} \phi_k^2(x_0) = \tilde{a}_{x_0}^2. \tag{5.3}$$

If $g_0(t) = \sum_{k=0}^\infty g_{0,k}\psi_k(t)$, then for $g^\star(t) = \sum_{k=0}^\infty (\eta_k g_{0,k})/(\lambda + \eta_k)\psi_k(t)$, we have

$$n^{m/(2m+1)}(\hat{g}(x_0) - g^\star(x_0)) \xrightarrow{d} N\left(0, \tilde{a}_{x_0}^2\right). \tag{5.4}$$

For any kernel K, denote the asymptotic bias and variance of the regularization estimator at point x by $B_{g_0,K,\lambda}(x)$ and $V_{g_0,K,n,\lambda}(x)$, respectively. For $g_0(t) = \sum_{k=0}^{\infty} g_{0,k}\phi_k(t) = \sum_{k=0}^{\infty} \tilde{g}_{0,k}\psi_k(t)$, by Theorems 5 and 7 we have

$$B_{g_0,K_{\omega,0,\pi},\lambda}(x) = -\lambda \sum_{k=0}^{\infty} \frac{\tilde{g}_{0,k}}{\lambda + \eta_k} \psi_k(x),$$

$$V_{g_0,K_{\omega,0,\pi},n,\lambda}(x) = \frac{\sigma^2}{n} \sum_{k=0}^{\infty} \left(1 + \frac{\lambda}{\eta_k}\right)^{-2} \psi_k^2(x),$$

$$B_{g_0,H_\omega,\lambda}(x) = -\lambda \sum_{k=0}^{\infty} \frac{g_{0,k}}{\lambda + \lambda_{k,\omega}} \phi_k(x),$$

$$V_{g_0,H_\omega,n,\lambda}(x) = \frac{\sigma^2}{n} \sum_{k=0}^{\infty} \left(1 + \frac{\lambda}{\lambda_{k,\omega}}\right)^{-2} \phi_k^2(x).$$

The asymptotic variance here is not associated with the true regression functions. For simplicity, we assume that the predictor is uniformly distributed on $[0, \pi]$, and $\sigma^2 = 1$. Under these assumptions we have

(i) $\eta_0 = 1, \eta_{2k-1} = \eta_{2k} = e^{-2k^2\omega^2}$;

(ii) $\psi_0(t) = 1, \psi_{2k-1}(t) = \sqrt{2}\sin(2kt), \psi_{2k}(t) = \sqrt{2}\cos(2kt)$;

(iii) $\lambda_{k,\omega} = e^{-(k^2\omega^2)/2}$;

(iv) $\phi_0(t) = 1, \phi_k(t) = \sqrt{2}\cos(kt)$.

The asymptotic variances of the two methods are, respectively,

$$V_{g_0,K_{\omega,0,\pi},n,\lambda}(x) = \frac{1}{n(1+\lambda)^2} + \frac{2}{n} \sum_{k=1}^{\infty} \left(1 + \lambda e^{2k^2\omega^2}\right)^{-2};$$

$$V_{g_0,H_\omega,n,\lambda}(x) = \frac{1}{n(1+\lambda)^2} + \frac{2}{n} \sum_{k=1}^{\infty} \left(1 + \lambda e^{k^2\omega^2/2}\right)^{-2} \cos^2(kx).$$

With a common λ, the variance of using H_ω is smaller than that of using $K_{\omega/2,0,\pi}$. However, as the values of λ depend on the trade-off mentioned in Section 1, the variances of using the two kernels are not easy to compare in general. Instead, we use three examples to show their differences. The calculations are given at the end of the supplementary material.

Example 1 ($g_0(t) = C$). The constant function is a smooth periodic function and satisfies all the conditions in both methods. For any λ and ω, we have

$$B_{g_0,K_{\omega,0,\pi},\lambda}(x) = B_{g_0,H_\omega,\lambda}(x) = -\frac{\lambda C}{\lambda + 1},$$

$$V_{g_0,K_{\omega/2,0,\pi},n,\lambda}(x) > V_{g_0,H_\omega,n,\lambda}(x).$$

Thus, our method has the same bias as that of Lin and Brown (2004), but with a smaller variance when the true regression function is constant.

Example 2 ($g_0(t) = \sin(t)$). This function is periodic on $[0, \pi]$ and satisfies the conditions for both methods. We have

$$\begin{aligned} B_{g_0,K_{\omega,0,\pi},\lambda}(x) &= B_{g_0,H_\omega,\lambda}(x) \\ &= \sum_{k=1}^{\infty} \frac{4\lambda}{\pi(4k^2-1)(\lambda+e^{-2k^2\omega^2})} \cos(2kx) - \frac{2\lambda}{\pi(\lambda+1)}. \end{aligned}$$

However, neither has uniformly smaller variances than the other. As a consequence, neither has smaller mean squared error (MSE) than the other when the optimal λ is used.

Example 3 ($g_0(t) = \cos(t)$). This function is not periodic on $[0, \pi]$. We have

$$B_{g_0,K_{\omega,0,\pi},\lambda}(x) = -\sum_{k=1}^{\infty} \frac{8k\lambda}{\pi(4k^2-1)(\lambda+e^{-2k^2\omega^2})} \sin(2kx);$$

$$B_{g_0,H_\omega,\lambda}(x) = -\sum_{k=1}^{\infty} \frac{8k\lambda}{\pi(4k^2-1)(\lambda+e^{-\omega^2/2})} \sin(2kx).$$

It follows that

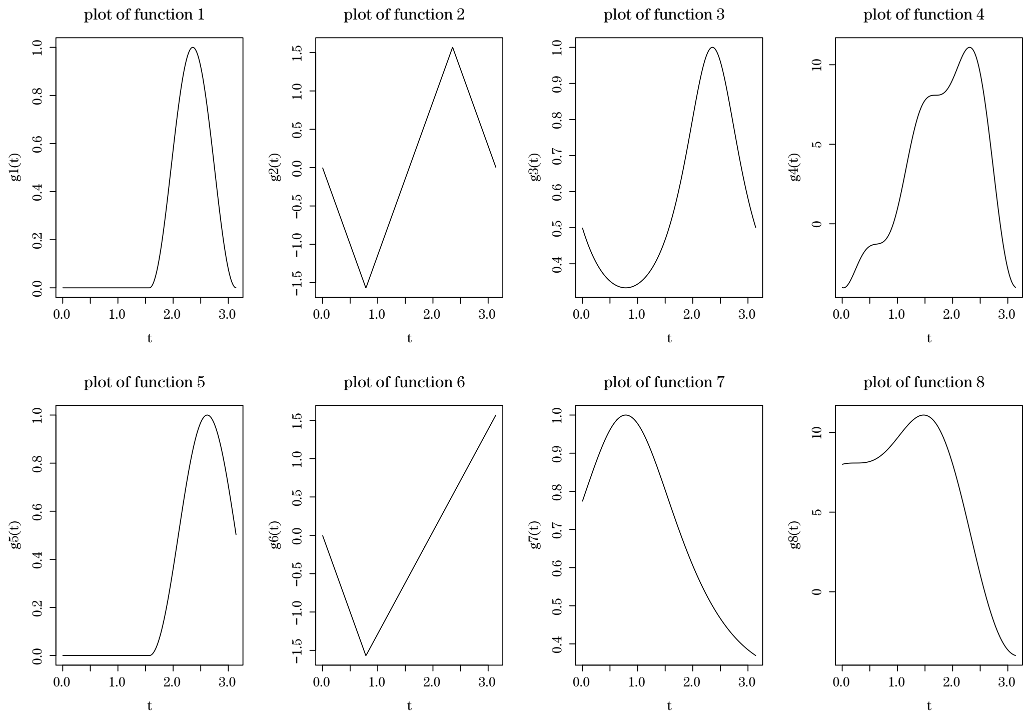
$$\int_0^\pi B^2_{g_0,K_{\omega/2,0,\pi},\lambda}(x)dx > \int_0^\pi B^2_{g_0,H_\omega,\lambda}(x)dx,$$

$$V_{g_0,K_{\omega/2,0,\pi},n,\lambda}(x) > V_{g_0,H_\omega,n,\lambda}(x).$$

In this case, our symmetric periodic Gaussian regularization achieves smaller integrated MSE than the periodic Gaussian regularization of Lin and Brown (2004).

6. Simulation

Lin and Brown (2004) did simulations which suggested that their method is comparable to, or even better than, the other such nonparametric smoothing methods as smoothing splines. In our simulation, we only need to compare their finite sample performances, but we also compare with the method of the Gaussian regularization to show its similarity with our method. Data are generated from the model (1.1) with noise $N(0, 1)$ and the design variable uniformly distributed on $[0, \pi]$. Consider eight regression functions, the first four of which were used in Lin and Brown (2004), and are periodic on $[0, \pi]$; the others are not periodic.



plot of function 1

plot of function 2

plot of function 3

plot of function 4

plot of function 5

plot of function 6

plot of function 7

plot of function 8

Figure 1. Plots of the eight regression functions

$$
\begin{aligned}
g_1(t) &= \sin^2(2t-\pi)I_{t\geq\pi/2},\\
g_2(t) &= -2t+(4t-\pi)I_{t\geq\pi/4}+(-4t+3\pi)I_{t\geq 3\pi/4},\\
g_3(t) &= \frac{1}{2-\sin(2t-\pi)},\\
g_4(t) &= 2+\sin(2t-\pi)+2\cos(2t-\pi)+3\sin^2(2t-\pi)\\
&\quad +4\cos^2(2t-\pi)+5\sin^3(2t-\pi),\\
g_5(t) &= \sin^2\left(\frac{3t}{2}-\frac{3\pi}{4}\right)I_{t\geq\pi/2},\\
g_6(t) &= -2t+\left(\frac{10t}{3}-\frac{5\pi}{6}\right)I_{t\geq\pi/4},\\
g_7(t) &= \frac{1}{2-\cos(t-\pi/4)},\\
g_8(t) &= 2+\sin(t)+2\cos(t)+3\sin^2(t)+4\cos^2(t)+5\sin^3(t).
\end{aligned}
$$

Plots of these functions are shown in figure 1.

For periodic Gaussian regularization, following Lin and Brown (2004), we approximate $K_{\omega,0,\pi}$ by $K^J_{\omega,0,\pi} = \sum_{k=-J}^{J} K_\omega(s-t-k\pi, 0)$. In fact

$$0 < K_{\omega,0,\pi}(s,t) - K^J_{\omega,0,\pi}(s,t) < 10^{-20} \qquad \forall s,t \in [0,\pi] \text{ for } 2J+1 > 3\omega.$$

Here we choose $J = 4$. In practice, $J = 1$ is enough. For the selection of λ and ω, we did a grid search over points $(\omega, \lambda) : \omega = 0.3k_1 - 0.1, \lambda = \exp(-0.4^{k_2} + 7)$, for $k_1 = 1, \ldots, 10$ and for $k_2 = 1, \ldots, 50$ so as to minimize Mallows' C_p; see Lin and Brown (2004) for the details.

For the symmetric periodic Gaussian regularization, the kernel H_ω is approximated by

$$H^J_\omega(s,t) = \sum_{k=-J}^{J} [K_\omega(s-t-2k\pi, 0) + K_\omega(s+t-2k\pi, 0)],$$

where J is the same as in the periodic kernel above. Similar to periodic Gaussian regularization, the parameter (ω, λ) is searched to minimize Mallows' C_p over the same collection of possible values.

For each model with sample sizes n=50, 100, 200, we repeated the simulations 100 times. For each replication, k, the mean squared error (MSE) was calculated as $MSE_k = (1/n)\sum_{i=1}^{n}(\hat{g}(x_{k_i}) - g_0(x_{k_i}))^2$, n the sample size and $\{x_{k_i}\}_{i=1}^{n}$ the sample points. The averaged mean squared error (AMSE) over the 100 replications was then calculated as

$$AMSE = \frac{1}{100}\sum_{k=1}^{100} MSE_k.$$

Table 1 lists the AMSEs for different combinations of regression functions and sample sizes. We have some observations. First, the AMSE decreases as the sample size grows, which supports our theoretical consistency. Second, our symmetric periodic Gaussian regularization has almost identical performance to Gaussian regularization in all simulations which reflects the approximation in Theorem 1. Third, for periodic functions (functions 1-4), our method is comparable to, or slightly better than, that of Lin and Brown (2004). Our method can achieve much better performance when the true regression function is not periodic, as with functions 5-8.

As Lin and Brown (2004) demonstrated that the periodic Gaussian regularization is comparable to, or better than, the smoothing splines in their simulations, our simulations indicate that our method can do even better.

In conclusion, our newly introduced reproducing kernel performs satisfactorily for nonparametric estimation problem, and is almost identical to that of the popularly used Gaussian regularization. Based on Theorem 1 and the identical

Table 1. AMSEs for the Gaussian regularization, the symmetric periodic Gaussian regularization, and the periodic Gaussian regularization on eight different regression functions

				Symmetric	Periodic
Function	Regression	Sample	Gaussian	Periodic	Gaussian
Type	Functions	Size	kernel	Gaussian kernel	kernel
Periodic Functions on $[0, \pi]$	$g_1(t)$	50	0.120	0.128	0.173
		100	0.064	0.069	0.084
		200	0.038	0.036	0.041
	$g_2(t)$	50	0.145	0.152	0.135
		100	0.068	0.077	0.084
		200	0.042	0.043	0.036
	$g_3(t)$	50	0.099	0.099	0.126
		100	0.052	0.053	0.057
		200	0.030	0.031	0.029
	$g_4(t)$	50	0.225	0.209	0.183
		100	0.108	0.097	0.081
		200	0.054	0.050	0.039
Non-periodic Functions on $[0, \pi]$	$g_5(t)$	50	0.110	0.112	0.123
		100	0.055	0.056	0.070
		200	0.033	0.034	0.039
	$g_6(t)$	50	0.126	0.137	0.159
		100	0.057	0.070	0.103
		200	0.039	0.038	0.080
	$g_7(t)$	50	0.093	0.092	0.096
		100	0.045	0.046	0.057
		200	0.022	0.022	0.030
	$g_8(t)$	50	0.166	0.143	0.698
		100	0.073	0.065	0.514
		200	0.040	0.035	0.376

numerical performance in Table 1, the kernel we introduced and our associated results collectively shed some light on the success of the Gaussian reproducing kernel.

Supplementary Materials

The supplementary materials include all the proofs of the our theoretical results in the present paper.

Acknowledgment

We thank an associate editor and two referees for their very constructive comments which lead to substantial improvement of the paper. YC Xia's research is partially supported by NUS Grant: R-155-000-162-112, R-155-000-193-114, MOE grant of Singapore: MOE2014-T2-1-072, and National Natural Science Foundation of China Grant 71371095, 11771066.

References

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society* **68**, 337–404.

Eberts, M. and Steinwart, I. (2013). Optimal regression rates for SVMs using Gaussian kernels. *Electronic Journal of Statistics* **7**, 1–42.

Evgeniou, T., Pontil, M. and Poggio, T. (2000). Regularization networks and support vector machines. *Advances in Computational Mathematics* **13**, 1–50.

Hofmann, T., Schölkopf, B. and Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 1171–1220.

Keerthi, S. S. and Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* **15**, 1667–1689.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lin, Y. and Brown, L. D. (2004). Statistical properties of the method of regularization with periodic Gaussian reproducing kernel. *Annals of Statistics*, 1723–1743.

Mendelson, S. (2002). Geometric parameters of kernel machines. In: *International Conference on Computational Learning Theory*. Springer Berlin Heidelberg, pp. 29–43.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **209**, 415–446.

Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press.

Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. *The Annals of Statistics* **41**, 2608–2638.

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*, 795–810.

Smola, A. J., Schölkopf, B. and Müller, K. R. (1998). The connection between regularization operators and support vector kernels. *Neural Networks* **11**, 637–649.

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer Science and Business Media.

Steinwart, I., Hush, D. and Scovel, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Transactions on Information Theory* **52**, 4635–4643.

Steinwart, I. and Scovel, C. (2007). Fast rates for support vector machines using Gaussian kernels. *The Annals of Statistics*, 575–607.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 1040–1053.

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

Wahba, G. (1999). Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. *Advances in Kernel Methods-Support Vector Learning* **6**, 69–87.

Williamson, R. C., Smola, A. J. and Schölkopf, B. (2001). Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators. *Information Theory, IEEE Transactions on* **47**, 2516–2532.

Yuan, M. and Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics* **38**, 3412–3444.

National University of Singapore, Singapore.

E-mail: A0123862@u.nus.edu

University of Electronic Science and Technology, China.

E-mail: staxyc@nus.edu.sg

(Received September 2016; accepted November 2017)