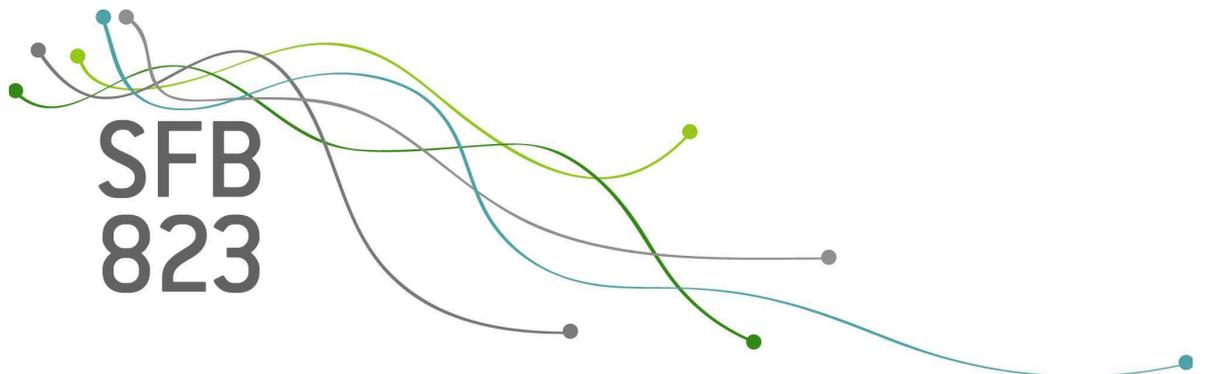


SFB
823

Reproducing kernel Hilbert spaces, polynomials and the classical moment problems

Holger Dette, Anatoly Zhigljavsky

Nr. 4/2021



Discussion Paper

Reproducing kernel Hilbert spaces, polynomials and the classical moment problems

Holger Dette

Anatoly Zhigljavsky

Ruhr-Universität Bochum

School of Mathematics

Fakultät für Mathematik

Cardiff University

44780 Bochum, Germany

Cardiff, CF24 4AG, UK

Abstract

We show that polynomials do not belong to the reproducing kernel Hilbert space of infinitely differentiable translation-invariant kernels whose spectral measures have moments corresponding to a determinate moment problem. Our proof is based on relating this question to the problem of best linear estimation in continuous time one-parameter regression models with a stationary error process defined by the kernel. In particular, we show that the existence of a sequence of estimators with variances converging to 0 implies that the regression function cannot be an element of the reproducing kernel Hilbert space. This question is then related to the determinacy of the Hamburger moment problem for the spectral measure corresponding to the kernel.

In the literature it was observed that a non-vanishing constant function does not belong to the reproducing kernel Hilbert space associated with the Gaussian kernel (see Corollary 4.44 in Steinwart and Christmann, 2008). Our results provide a unifying view of this phenomenon and show that the mentioned result can be extended for arbitrary polynomials and a broad class of translation-invariant kernels.

AMS Subject Classification: 46E22, 62J05, 44A60

Keywords and Phrases: Reproducing kernel Hilbert spaces, classical moment problem, best linear estimation, continuous time regression model

1 Introduction

1.1 Main results

Let $X \subseteq \mathbb{R}^d$, $d \geq 1$, $K : X \times X \rightarrow \mathbb{R}$ be a positive definite kernel on X and define $H(K)$ as the corresponding Reproducing Kernel Hilbert Space (RKHS). We assume that X has a non-empty interior and K is an infinitely differentiable (on the diagonal) translation-invariant kernel so that $K(x, y) = k(x - y)$, where $k : \mathbb{R}^d \rightarrow \mathbb{R}$ is a non-constant positive definite function infinitely differentiable at the point 0. Without loss of generality, we suppose $k(0) = 1$.

It is well-known, see e.g. Corollary 4.44 in Steinwart and Christmann (2008), that in the case of the squared exponential (Gaussian) kernel

$$k(x) = \exp\{-\lambda\|x\|^2\} \quad \text{with } \lambda > 0, \quad (1.1)$$

the constant function does not belong to $H(K)$. The purpose of this paper is to significantly extend this result by considering:

- (a) a substantially larger class of kernels, for which this statement remains true, and
- (b) arbitrary polynomials rather than simply constant functions.

The results of this paper have definite consequences for the methodologies of function approximation, Bayesian global optimization and support vector machines (SVM) and other kernel-based machine learning, see Section 1.2 for a short discussion.

In the main part of the paper, we consider the case $X \subset \mathbb{R}$. In this case, by Bochner's theorem (Bochner and Chandrasekharan, 1949), there exists a measure α , such that the kernel k can be represented in the form

$$k(x) = \int_{-\infty}^{\infty} e^{itx} \alpha(dt) \quad \text{for all } x \in X. \quad (1.2)$$

The measure α is called *spectral measure*. Because $k(x) = k(-x)$ for all x and $k(0) = 1$, α is a probability measure symmetric around the point 0. We denote by

$$c_k = \int_{-\infty}^{\infty} t^k \alpha(dt), \quad k = 1, 2, \dots \quad (1.3)$$

the moments of this measure (in the case of their existence). The classical *Hamburger*

moment problem is to give necessary and sufficient conditions such that a given real sequence $(c_k)_{k \in \mathbb{N}}$ is in fact a sequence of moments of a distribution α defined on the Borel sets of $\mathbb{R} = (-\infty, \infty)$ such that (1.3) holds. In particular, the sequence $(c_k)_{k \in \mathbb{N}}$ is a sequence of moments of some distribution if and only if the *Hankel* matrices $(c_{i+j})_{i,j=0,\dots,n}$ are positive semidefinite for all $n \in \mathbb{N}$ (see e.g. Shohat and Tamarkin, 1943; Schmüdgen, 2017, among many others). The Hamburger moment problem is called *determinate* if the sequence of moments $(c_k)_{k \in \mathbb{N}}$ determines the measure $\alpha(dt)$ uniquely.

The main results of this paper are Theorems 1.1 and 1.2 formulated below. These theorems and provide sufficient conditions ensuring that the polynomials do not belong to the RKHS $H(K)$. The proofs are given in Section 2.5.

Theorem 1.1 *Let $X \subset \mathbb{R}$ and assume that the spectral measure $\alpha(dt)$ in (1.2) has infinite support and no mass at the point 0. If the Hamburger moment problem for this measure is determinate, then the non-zero constant functions do not belong to the RKHS $H(K)$.*

Theorem 1.2 *Let $X \subset \mathbb{R}$, m be a positive integer and assume that the spectral measure $\alpha(dt)$ in (1.2) has infinite support. If the Hamburger moment problem for the measure $\alpha_m(dt) = t^{2m}\alpha(dt)/c_{2m}$ is determinate, then the RKHS $H(K)$ does not contain polynomials on X of degree m .*

Note that under the assumptions of Theorem 1.1 and 1.2 all functions of the form $f+g \notin H(K)$ with $g \in H(K)$ and f being either a non-zero constant or a polynomial do not belong to $H(K)$.

Note also that in Theorem 1.2, the spectral measure $\alpha(dt)$ is allowed having a positive mass at the point 0. Combining Theorems 1.1 and 1.2 with their variations in the cases when the spectral measure $\alpha(dt)$ has finite support (see Section 3.1) and when this measure has positive mass at 0 (see Theorem 3.1), we obtain the following corollary.

Corollary 1.1 *Let $X \subset \mathbb{R}$ and the Hamburger moment problem for the spectral measure $\alpha(dt)$ is determinate. Then we have the following:*

- (a) *The constant functions $f(x) = \text{const} \neq 0, \forall x \in X$, belong to $H(K)$ if and only if $\alpha(dt)$ has a positive mass at the point 0;*
- (b) *If the Hamburger moment problems for the measures $t^{2m}\alpha(dt)/c_{2m}$ are determinate for all $m \in \mathbb{N}$, then $H(K)$ does not contain non-constant polynomials on X .*

Theorems 1.1, 1.2 and Corollary 1.1 can be easily extended to the multivariate case, see Section 3.2 for a more detailed discussion.

1.2 Implications and related results

Many methods of function approximation, interpolation and prediction (see Stein, 1999; Wendland, 2004), Bayesian methods of global optimization (see Zhigljavsky and Zilinskas, 2007) and machine learning (see Rasmussen and Williams, 2006; Steinwart and Christmann, 2008) are kernel-based. Properties of all these algorithms depend on whether an unknown function of interest belongs to the corresponding RKHS, see Steinwart et al. (2006), Section 4.4 in Steinwart and Christmann (2008) for a discussion on the importance of this issue for the learning performance of SVM in the case of the squared exponential kernel. The condition that the non-zero constant functions do not belong to the RKHS $H(K)$ is a convenient simplifying condition which allows to avoid function centering, see, for example, Assumption 2 in Lee et al. (2016).

Let us briefly consider two approaches for the problem of approximation of a function $f : X \rightarrow \mathbb{R}$. First assume that $f \in H = H(K)$ with $\|f\|_H < \infty$, where $\|\cdot\|_H$ denotes the norm on $H(K)$. Let $\hat{f}_n(x) = \sum_{i=1}^n w_i f(x_i)$ be a linear predictor of $f(x)$ (for an $x \in X$) based on evaluations of f at an n -point design $X_n = \{x_1, \dots, x_n\} \subset X$; note that the set of weights $w = (w_1, \dots, w_n)$ may depend on x . Define

$$\rho_n(x, w) = \left\| K(x, \cdot) - \sum_{i=1}^n w_i K(x_i, \cdot) \right\|_H,$$

which is a quantity depending on x, w and design X_n . The Cauchy-Schwarz inequality implies (see Sect. 2.1 in Pronzato and Zhigljavsky, 2020, for details)

$$|f(x) - \hat{f}_n(x)| \leq \|f\|_H \cdot \rho_n(x, w). \quad (1.4)$$

The inequality (1.4) yields that if $f \notin H(K)$ then the best linear predictors of $f(x)$, constructed under the assumption $f \in H(K)$, can be poor. Moreover, in view of Corollary 3.3 of Section 3.4, $f \in H(K)$ if and only if $\lim_{n \rightarrow \infty} \widehat{\sigma}_n^2 < \infty$, where $\widehat{\sigma}_n^2$ is the MLE of σ^2 constructed from n observations of f under the assumption that f is a realization of a Gaussian random field with covariance kernel $R(x, x') = \sigma^2 K(x, x')$. Mean squared error of prediction of $f(x)$ at any $x \in X$ is proportional to σ^2 (and $\widehat{\sigma}_n^2$ in the empirical versions) and therefore prediction error of Bayesian predictors of f can only be accurate when $f \in H(K)$.

For a given set X , kernel K and function f , the problem of deciding whether $f \in H(K)$ is very important and often difficult to resolve; this is discussed e.g. in Steinwart et al.

(2006). In the case where the spectral measure has a positive density, say φ , on $X = \mathbb{R}$, there is an easy condition for deciding if f belongs to the RKHS. In fact, $f \in H(K)$ if and only if the squared modulus of the Fourier transform is integrable with respect to the measure $dt/\varphi(t)$ (see formula (2.4) in Berline and Thomas-Agnan, 2011). This condition, however, is not very helpful when $X \subset \mathbb{R}^d$ like $X = [0, 1]$. A necessary condition for an m times continuously differentiable kernel k on X can be found in Corollary 4.36 of Steinwart and Christmann (2008) saying that any function $f \in H(K)$ has to be at least m times differentiable. The statements like “*roughly speaking, a function f belongs to a RKHS only if it is at least as smooth as the kernel*” (see e.g. p. 12 in Berline and Thomas-Agnan, 2011) may give an impression that the smoothness of f guarantees that $f \in H(K)$. As Theorems 1.1, 1.2 and Corollary 1.1 show this impression is misleading: polynomials are infinitely smooth but for many infinitely differentiable translation-invariant kernels they do not belong to $H(K)$ for any $X \subset \mathbb{R}^d$ with non-empty interior.

The case of the squared exponential kernel (1.1) has attracted a lot of attention in the literature. As mentioned earlier, the result of Theorem 1.1 for the kernel (1.1) is known and clearly formulated (along with its important implications) in Steinwart et al. (2006), see also Steinwart and Christmann (2008). Bearing in mind the interpretation discussed in Section 3.5, this result (for the kernel (1.1)) also follows from Theorem 2.3 in Xu and Stein (2017). Moreover, Theorem 2.4 in Xu and Stein (2017) yields the result of Theorem 1.2 for squared exponential kernel (1.1) and the function $f(x) = x$, $x \in [0, 1]$.

1.3 Sufficient conditions for moment determinacy

There exists a vast amount of literature on sufficient conditions for the determinacy of Hamburger and Stieltjes moment problems, see e.g. Lin (2017), Schmüdgen (2017) and Stoyanov (2013, Chapter 11). The following two sufficient conditions for moment determinacy of a measure $\alpha(dt)$ with moments c_k in Hamburger moment problem are commonly used:

$$\sum_{n=1}^{\infty} c_{2n}^{-1/(2n)} = \infty, \tag{1.5}$$

$$\limsup_{n \rightarrow \infty} \frac{1}{2n} c_{2n}^{1/(2n)} < \infty. \tag{1.6}$$

Condition (1.5) is called *Carleman condition* and is slightly weaker than the condition (1.6), which is often very easy to verify. Note that all measures $\alpha_m(dt)$ ($m = 0, 1, 2, \dots$) in Theorem 1.2 are moment-determinant if the original spectral measure $\alpha(dt)$ satisfies (1.6). A well-known sufficient condition for indeterminacy in Hamburger moment problem is the so-called *Krein condition*

$$\int_{-\infty}^{\infty} \frac{-\log \varphi(t)}{1+t^2} < \infty \tag{1.7}$$

applicable for absolutely continuous measures $\alpha(dt) = \varphi(t)dt$, for which all moments exist. Two important examples of kernels satisfying conditions of Theorems 1.1 and 1.2 are the squared exponential (Gaussian) kernel defined by (1.1) with spectral density

$$\varphi(t) = \frac{1}{2\sqrt{\pi\lambda}} \exp\{-t^2/(4\lambda)\},$$

and the Cauchy kernel $k(x) = 1/(1+x^2/\lambda^2)$ with spectral density

$$\varphi(t) = \frac{\lambda}{2} \exp\{-\lambda|t|\}, \quad t \in \mathbb{R}.$$

In both cases one can use condition (1.6) to prove that the moment problem is determinate. Further examples of kernels are given in Section 4.2 of this paper and many other examples can be constructed using moment-determinate spectral measures.

1.4 Main steps in the proofs and structure of the remaining part of the paper

Section 2 is devoted to the proofs of Theorems 1.1 and 1.2, which are given in several steps. The main idea in our approach is to relate the problem of interest to properties of the best linear unbiased estimate (BLUE) in linear regression models, which will be worked out in Sections 2.1 and 2.2. Sections 2.3 and 2.4 provide different characterizations of the moment determinacy of spectral measures and finally the proofs will be completed in Section 2.5. We now explain the different steps in more detail.

In Section 2.1 we consider the BLUE in a one-parameter linear regression model with a regression function $f \in H(K)$ and show that in this case the BLUE exists and its variance is strictly positive, see Lemma 2.1. We also show that in the case $f \notin H(K)$, the BLUE does not exist and establish in Lemma 2.2 that for proving $f \notin H(K)$, it is sufficient to

construct a sequence of linear unbiased estimators $\hat{\theta}_n$ of the unknown parameter with variances tending to 0. Such a sequence is constructed in Section 2.2 for the location scale model and an explicit expression for the variance of these estimators in terms of the ratio of determinants

$$\text{var}(\hat{\theta}_n) = \frac{\det(c_{2(i+j)})_{i,j=0}^n}{\det(c_{2(i+j)})_{i,j=1}^n} \quad (1.8)$$

of Hankel-type matrices of the moments of the spectral measure is derived in Lemma 2.4. In Section 2.3 we establish several properties of moment-determinant symmetric measures which we use in Section 2.4 for building up an equivalence between the moment determinacy of the spectral measures and the statement that the sequence (1.8) converges to zero. This is arguably the most important step in the proof of both theorems (see Lemma 2.7). Finally, these results are combined in the proofs of Theorem 1.1 and 1.2 in Section 2.5. In Section 3 we consider several extensions and interpretations of the main results. In Section 3.1 we consider spectral measures with finite support, while Section 3.2 discusses the multivariate case. This discussion is continued in Sections 3.3–3.5 where we also consider general metric spaces. In Section 3.3 we explain a technique of characterizing the fact $f \in H(K)$ via suitable discretization of the set X and show that $1/\|f\|_{H(K)}$ is the limit of variances of the related discrete BLUEs. These results are used in Section 3.4 where we show that the constant function belongs to $H(K)$ if and only if the spectral measure has positive mass at 0. In Section 3.5 we show that the problem of parameter estimation in a one-parameter regression model is equivalent to the problem of estimating the variance of a Gaussian process (field). Thus we are able to relate our findings to the estimation problems considered in Xu and Stein (2017). In Section 3.6 we return to the one-dimensional case and give an interpretation of Theorem 1.1 in terms of the L_2 -error of the best approximation of a constant function by polynomials of the form $a_1 t^2 + a_2 t^4 + \dots + a_n t^{2n}$. Finally, in Section 4, for two specific classes of kernels we derive explicit results on the rates of convergence to 0 of the variances of the main estimators $\hat{\theta}_n$. In the case of squared exponential kernel (1.1), we detail and improve one of the asymptotic expansions of Theorem 3.3 in Xu and Stein (2017).

2 Parameter estimation, moment determinacy and proofs of main results

2.1 BLUE in a one-parameter regression model

Consider a one-parameter regression model with stationary correlated errors:

$$y(x) = \theta f(x) + \varepsilon(x), \quad x \in X, \quad \mathbb{E}\varepsilon(x) = 0, \quad \mathbb{E}\varepsilon(x)\varepsilon(x') = k(x - x'). \quad (2.1)$$

Here θ is a scalar parameter, $f: X \rightarrow \mathbb{R}$ is a given regression function and $k(\cdot)$ is an infinitely differentiable positive definite function with $k(0) = 1$ making the kernel $K(\cdot, \cdot)$ defined by $K(x, y) = k(x - y)$ an infinitely differentiable correlation kernel. For constructing estimators of the parameter θ , the observations of the process $\{y(x) | x \in X\}$ along with observations of all of its derivatives $\{y^{(k)}(x) | x \in X\}$, $k = 1, 2, \dots$, can be used.

The (continuous) best linear unbiased estimator (BLUE) of θ is defined as an unbiased estimator $\hat{\theta}_{BLUE}$ such that $\text{var}(\hat{\theta}_{BLUE}) \leq \text{var}(\hat{\theta})$, where $\hat{\theta}$ is any linear unbiased estimator of θ . If the kernel K is differentiable and the BLUE exists, then for its computation all available derivatives of $y(x)$ are used, see Dette et al. (2019). In general, the BLUE may not exist but the next lemma shows that it does exist when $f \in H(K)$.

Lemma 2.1 *If $f \in H(K)$, then the BLUE $\hat{\theta}_{BLUE}$ in model (2.1) exists and*

$$\text{var}(\hat{\theta}_{BLUE}) = 1/\|f\|_{H(K)}^2 > 0.$$

The statement of lemma follows from Theorem 6C (p. 975) of Parzen (1961). Formally, only the case $X = [0, 1]$ is considered in Parzen (1961), but Parzen's proof does not use the structure of X and is therefore valid for a general metric space X .

Lemma 2.2 *If there exists a sequence of linear unbiased estimators $(\hat{\theta}_n)_{n \in \mathbb{N}}$ of θ in model (2.1) such that $\text{var}(\hat{\theta}_n) \rightarrow 0$ as $n \rightarrow \infty$, then $f \notin H(K)$.*

Proof. Assume that $f \in H(K)$. By Lemma 2.1, the continuous BLUE $\hat{\theta}_{BLUE}$ exists and $\text{var}(\hat{\theta}_{BLUE}) = 1/\|f\|_{H(K)}^2 > 0$. From the definition of the BLUE, $\text{var}(\hat{\theta}_n) \geq \text{var}(\hat{\theta}_{BLUE}) > 0$ for all $n \in \mathbb{N}$. We have arrived at a contradiction and hence $f \notin H(K)$. \square

2.2 A family of estimators $\hat{\theta}_n$ in the location scale model

Consider the location scale model

$$y(x) = \theta + \varepsilon(x), \quad x \in X \subset \mathbb{R}, \quad \mathbb{E}\varepsilon(x) = 0, \quad \mathbb{E}\varepsilon(x)\varepsilon(x') = k(x - x'), \quad (2.2)$$

where $k(\cdot)$ is an infinitely differentiable at 0 positive definite function. Choose any interior point $x_0 \in X$ and set $\varepsilon_0 = \varepsilon(x_0)$. For construction of the estimator $\hat{\theta}_n$, which we will apply in Lemma 2.2, we use the following $n + 1$ observations: the observation $y(x_0) = \theta + \varepsilon_0$ at the point x_0 and n mean-square derivatives of the process y at the point x_0 :

$$\varepsilon_j = y^{(j)}(x_0) = \left. \frac{d^j y(x)}{dx^j} \right|_{x=x_0} = \left. \frac{d^j \varepsilon(x)}{dx^j} \right|_{x=x_0}, \quad j = 2, 4, \dots, 2n. \quad (2.3)$$

The following result provides a necessary and sufficient condition for the existence of the derivatives. For a proof, see page 164 (Section 12) in Yaglom (1986).

Lemma 2.3 *Let x_0 be an interior point of X . The mean-square derivative $\varepsilon_j = d^j \varepsilon(x)/dx^j|_{x=x_0}$ of the stationary process $\{\varepsilon(x)|x \in X\}$ in (2.2) at the point x_0 exists if and only if $c_{2j} < \infty$, where*

$$c_{2j} = \int_{-\infty}^{\infty} t^{2j} \alpha(dt) = (-1)^j \left. \frac{\partial^{2j}}{\partial u^{2j}} k(u) \right|_{u=0} \quad (2.4)$$

is the $2j$ -th moment of the spectral measure α corresponding to the kernel k in Bochner's theorem.

As we have assumed that the kernel $k(\cdot)$ is infinitely differentiable at 0, all moments c_j ($j = 0, 1, \dots$) exist. As an immediate consequence of the existence of all moments and the representation (1.2), for the random variables ε_j defined in (2.3), we obtain by Lemma 2.3 for all $i, j = 0, 1, \dots$

$$\mathbb{E}\varepsilon_i \varepsilon_j = \left. \frac{\partial^{i+j}}{\partial x^i \partial y^j} k(x - y) \right|_{x,y=x_0} = (-1)^{i+j} \left. \frac{\partial^{i+j}}{\partial u^{i+j}} k(u) \right|_{u=0} = c_{i+j}. \quad (2.5)$$

Note that all derivatives $\partial^m / \partial u^m k(u)|_{u=0}$ of odd order m vanish as the function $k(\cdot)$ is symmetric around the point 0.

Next, we introduce the random variables $\delta_i = (-1)^i \varepsilon_{2i}$, $i = 0, 1, \dots$. The observations (2.3) used for constructing the discrete BLUE in model (2.2) can then be rewritten as

$$y_0 = y(x_0) = \theta + \varepsilon_0, \quad y_1 = y^{(2)}(x_0) = \delta_1, \quad \dots, \quad y_n = y^{(2n)}(x_0) = \delta_n.$$

Moreover, the covariance matrix of the vector $(\delta_0, \delta_1, \dots, \delta_n)^\top$ is the Hankel matrix

$$C_n = (\mathbb{E}\delta_i\delta_j)_{i,j=0}^n = (c_{2(i+j)})_{i,j=0}^n, \quad (2.6)$$

where c_2, \dots, c_{2n} are the moments defined in (2.4).

Assume that the spectral measure $\alpha(dt)$ has infinite support. In this case, the matrices C_n are positive definite for all $n = 0, 1, \dots$ (see, for example, Proposition 3.11 in Schmüdgen, 2017) and the discrete BLUE is obtained as

$$\hat{\theta}_n = \frac{e_{0,n}^\top C_n^{-1} Y_n}{e_{0,n}^\top C_n^{-1} e_{0,n}}, \quad (2.7)$$

where $Y_n = (y_0, y_1, \dots, y_n)^\top$ and $e_{0,n} = (1, 0, \dots, 0)^\top \in \mathbb{R}^{n+1}$ denotes the first coordinate vector in \mathbb{R}^{n+1} .

Lemma 2.4 *The variance of the estimator (2.7) is*

$$\text{var}(\hat{\theta}_n) = \frac{1}{e_{0,n}^\top C_n^{-1} e_{0,n}} = \frac{H_n}{G_n}, \quad (2.8)$$

where H_n and G_n are the determinants

$$H_n = \det(C_n) = \det \left[(c_{2(i+j)})_{i,j=0}^n \right], \quad G_n = \det \left[(c_{2(i+j)})_{i,j=1}^n \right]. \quad (2.9)$$

Proof. The expression (2.8) follows from the standard formula $\text{var}(\hat{\theta}_n) = 1/(e_{0,n}^\top C_n^{-1} e_{0,n})$ for the variance of the BLUE and Cramér's rule for computing elements of a matrix inverse; in our case, $e_{0,n}^\top C_n^{-1} e_{0,n}$ coincides with the top-left element of the matrix C_n^{-1} . \square

Observing Lemma 2.2 we conclude that a non-vanishing constant function does not belong to $H(K)$ if $\lim_{n \rightarrow \infty} H_n/G_n = 0$. In the following sections we relate this condition to the moment determinacy of the spectral measure.

Remark 2.1 Let us briefly consider the case where the spectral measure has a positive mass at the point 0. Consider the location scale model (2.2) and let

$$\alpha_\gamma(dt) = (1 - \gamma)\alpha(dt) + \gamma\delta_0(dt) \quad (2.10)$$

denote the spectral measure corresponding to a nonnegative definite and symmetric kernel k_γ , where $0 < \gamma < 1$, δ_0 is the Dirac measure at the point 0 and $\alpha(dt)$ is a symmetric probability measure on \mathbb{R} with no mass at 0. The measure $\alpha_\gamma(dt)$ is symmetric around the point 0 with even moments $\tilde{c}_0 = 1$ and

$$\tilde{c}_{2j} = (1 - \gamma)c_{2j}, \quad j = 1, 2, \dots$$

Recall the definition of the matrix C_n in (2.6) and define the matrices

$$\tilde{C}_n = (\tilde{c}_{2(i+j)})_{i,j=0}^n = \gamma e_{0,n} e_{0,n}^\top + (1 - \gamma)C_n$$

and the corresponding determinants

$$\tilde{H}_n = \det \tilde{C}_n, \quad \tilde{G}_n = \det \left[(\tilde{c}_{2(i+j)})_{i,j=1}^n \right] = (1 - \gamma)^n G_n,$$

where G_n is defined in (2.9). Using standard formulas of linear algebra we obtain

$$\tilde{H}_n = \det [\gamma e_{0,n} e_{0,n}^\top + (1 - \gamma)C_n] = (1 - \gamma)^n [(1 - \gamma) + \gamma e_{0,n}^\top C_n^{-1} e_{0,n}] H_n.$$

In accordance with (2.8), the variance of $\tilde{\theta}_n$, the BLUE of θ constructed similarly to $\hat{\theta}_n$ but for the spectral measure $\alpha_\gamma(dt)$, is given by

$$\begin{aligned} \text{var}(\tilde{\theta}_n) &= \frac{1}{e_{0,n}^\top \tilde{C}_n^{-1} e_{0,n}} = \frac{\tilde{H}_n}{\tilde{G}_n} = \frac{H_n}{G_n} [(1 - \gamma) + \gamma e_{0,n}^\top C_n^{-1} e_{0,n}] \\ &= \text{var}(\hat{\theta}_n) [(1 - \gamma) + \gamma / \text{var}(\hat{\theta}_n)] = (1 - \gamma) \text{var}(\hat{\theta}_n) + \gamma > 0. \end{aligned}$$

This implies that $\text{var}(\tilde{\theta}_n)$ cannot converge to 0 and Lemma 2.2 is not applicable if the spectral measure has a positive mass at the point 0.

In Theorem 3.1 of Section 3.4 we will prove that for any compact set $X \subset \mathbb{R}^d$ the constant functions indeed belong to $H(K)$, if the spectral measure has a positive mass at the point 0.

2.3 Moment-determinacy of the spectral measure

Consider the spectral measure α introduced in equation (1.2). As a spectral measure, α is a symmetric measure (around 0) on the real line and we have assumed that α does not have a positive mass at the point 0. Moreover, we have assumed $k(0) = 1$ making α a probability distribution. In the following we relate α to a (unique) measure on the nonnegative axis $[0, \infty)$. Loosely speaking, if a real valued random variable ξ has distribution $\alpha(dt)$, then $\alpha_+(dt)$ is the distribution of the random variable ξ^2 . In the opposite direction, if the nonnegative random variable η has distribution $\alpha_+(dt)$, then $\pm\sqrt{\eta}$ has distribution $\alpha(dt)$, where \pm denotes a random sign.

For a more formal construction we follow the arguments in Section 3.3 of Schmüdgen (2017) and denote by \mathcal{B} the Borel sigma field on \mathbb{R} , define $\tau : \mathbb{R} \rightarrow [0, \infty)$; $\tau(x) = x^2$ and $\kappa : [0, \infty) \rightarrow \mathbb{R}$, $\kappa(x) = \sqrt{x}$. Then for any symmetric (Radon) measure α on \mathcal{B} , the measure α_+ defined by

$$\alpha_+(B) = \alpha(\tau^{-1}(B)) \quad B \in \mathcal{B} \cap [0, \infty) \quad (2.11)$$

defines a measure on $\mathcal{B} \cap [0, \infty)$. Conversely, if α_+ is a measure on $\mathcal{B} \cap [0, \infty)$, then

$$\alpha(B) = \frac{1}{2}(\alpha_+(\kappa^{-1}(B)) + \alpha_+((-\kappa)^{-1}(B))) \quad (2.12)$$

defines a symmetric measure on \mathcal{B} . It now follows from Theorem 3.17 in Schmüdgen (2017) that the relations (2.11) and (2.12) define a bijection from the set of all symmetric measures on \mathbb{R} onto the set of all measures on $[0, \infty)$.

The even moments of a symmetric probability measure α on \mathcal{B} are related to the moments of the measure α_+ from (2.11) by

$$c_{2j} = \int_{-\infty}^{\infty} t^{2j} \alpha(dt) = 2 \int_0^{\infty} t^{2j} \alpha(dt) = \int_0^{\infty} t^j d\alpha_+(t) = b_j, \quad j \in \mathbb{N}, \quad (2.13)$$

and as a consequence the determinants H_n and G_n in (2.9) can be represented as

$$H_n = \det \left[(b_{i+j})_{i,j=0}^n \right], \quad G_n = \det \left[(b_{i+j})_{i,j=1}^n \right]. \quad (2.14)$$

Similarly to the case of the Hamburger moment problem, the *Stieltjes moment problem* is to give necessary and sufficient conditions such that a real sequence $(b_j)_{j \in \mathbb{N}}$ is in fact a sequence of moments of a measure $\alpha_+(dt)$ on the Borel sets of $[0, \infty)$; that is $b_j =$

$\int_0^\infty t^j d\alpha_+(t)$ for all $j \in \mathbb{N}_0$. The Stieltjes moment problem is determinate if the sequence of moments $(b_j)_{j \in \mathbb{N}}$ determines the measure $\alpha_+(dt)$ uniquely. For a proof of the following result, which relates the Hamburger and Stieltjes moment problem, see Heyde (1963, Lemma 1), Schmüdgen (2017, Proposition 3.19) and Stoyanov (2013, Sect. 11.10).

Lemma 2.5 *Let α be a symmetric probability measure on \mathcal{B} . The Hamburger moment problem for α is determinate if and only if the Stieltjes moment problems for the measure α_+ defined by (2.11) is determinate.*

Note that for the equivalence in Lemma 2.5 to hold, the assumption that α does not have mass at 0 is not required. This assumption, however, is needed in the next lemma.

Lemma 2.6 *Let α be a symmetric probability measure on \mathcal{B} with no mass at the point 0. The Hamburger moment problem for α is determinate if and only if the Hamburger moment problem for the measure α_+ defined by (2.11) is determinate.*

Proof. Using the result of Theorem A in Heyde (1963) (see also (Stoyanov, 2013, p.113) and (Schmüdgen, 2017, Remark 2.12)), if the Stieltjes moment problems for the measure α_+ is determinate and the measure α_+ has no mass at 0, then the Hamburger moment problems for this measure is also determinate. From Lemma 2.5, the required equivalence follows. \square

2.4 Relating moment-determinacy of the measure α_+ to $\text{var}(\hat{\theta}_n)$

Lemma 2.7 *Let α be a symmetric probability measure on \mathcal{B} with infinite support and no mass at the point 0. The Hamburger moment problem for the measure α_+ defined by (2.11) is determinate if and only if $H_n/G_n \rightarrow 0$ as $n \rightarrow \infty$, where the determinants H_n and G_n are defined in (2.14).*

Proof. (i) Assume that the moment problem for the measure α_+ is determinate. Let \mathcal{P}_n denote the class of all polynomials of degree n and define

$$\rho_n(t_0) = \min \left\{ \int_{\mathbb{R}} |P_n(t)|^2 \alpha_+(dt) \mid P_n \in \mathcal{P}_n, P_n(t_0) = 1 \right\}$$

for any $t_0 \in \mathbb{R}$, which is not a root of the n th orthogonal polynomial with respect to the measure α_+ (see equation (2.26) in Lemma 2.11 of Shohat and Tamarkin, 1943). Then

$$\lim_{n \rightarrow \infty} \rho_n(t_0) =: \rho(t_0)$$

exists, by Theorem 2.6 in Shohat and Tamarkin (1943). As the point 0 is not a support point of the measure α_+ and all roots of the orthogonal polynomials with respect to the measure α_+ are located in $\text{supp}(\alpha_+) \subset (0, \infty)$ we have from Corollary 2.6 in Shohat and Tamarkin (1943) that

$$\rho(0) = \lim_{n \rightarrow \infty} \rho_n(0) = 0.$$

Moreover, by the discussion on p. 72 (middle of the page) in Shohat and Tamarkin (1943) it follows that $\rho_n(0)$ is exactly the ratio H_n/G_n , where H_n and G_n are the determinants in (2.14). Hence the moment determinacy for the measure α_+ implies $H_n/G_n \rightarrow 0$ as $n \rightarrow \infty$.

(ii) To prove the converse, assume that $H_n/G_n \rightarrow 0$ as $n \rightarrow \infty$. Let λ_n be the smallest eigenvalue of the matrix C_n . Theorem 1.1 in Berg et al. (2002) states that the condition

$$\lim_{n \rightarrow \infty} \lambda_n = 0$$

is necessary and sufficient for the moment-determinacy of the measure α_+ .

From the definition of λ_n as the smallest eigenvalue of the matrix C_n and the representation (2.8) it follows

$$\lambda_n \leq \frac{1}{e_{0,n}^\top C_n^{-1} e_{0,n}} = \frac{H_n}{G_n} = \rho_n(0)$$

for all $n \in \mathbb{N}$ (see also a related discussion in Berg et al., 2002). Therefore, $H_n/G_n \rightarrow 0$ as $n \rightarrow \infty$ implies $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$ and this yields the moment determinacy of the measure α_+ . \square

2.5 Proof of Theorem 1.1 and 1.2

Proof of Theorem 1.1. Use Lemma 2.2 with the estimator defined in (2.7). By Lemma 2.4 the variance of this estimator is given by (2.8). From Lemma 2.7, the determinacy of the measure α_+ is equivalent to $\text{var}(\hat{\theta}_n) \rightarrow 0$ as $n \rightarrow \infty$. By Lemma 2.5, this is also equivalent to the moment determinacy of the spectral measure α . \square

Proof of Theorem 1.2. Assume that the function f in (2.1) is a polynomial of degree $m \geq 1$. Take m derivatives of both sides in (2.1). The model (2.1) thus reduces to

$$\tilde{y}(x) = \tilde{\theta} + \tilde{\varepsilon}(x), \quad x \in X, \quad (2.15)$$

where $\tilde{\theta}$ is the new parameter, $\tilde{y}(x) = y^{(m)}(x)$ are new observations and $\tilde{\varepsilon} = \varepsilon^{(m)}$ is the new error process.

From (Yaglom, 1986, (2.178)), the autocovariance function of the process $\{\varepsilon^{(m)}(x) | x \in X\}$ is given by

$$\mathbb{E}\varepsilon^{(m)}(x)\varepsilon^{(m)}(x') = k_m(x - x') \quad \text{with} \quad k_m(x) = (-1)^m k^{(2m)}(x).$$

From (1.2), the spectral measure associated with the kernel $k_m(x - x')$ is $\alpha_m(dt) = t^{2m}\alpha(dt)/c_{2m}$. Hence, the statement for the case when f is a polynomial of degree $m \geq 1$ is reduced to the case of the constant function proved in Theorem 1.1; this theorem is applicable as the measure $\alpha_m(dt)$ does not have mass at 0 for any $m \geq 1$. \square

3 Extensions of Theorems 1.1 and 1.2 and further discussion

In this section we discuss several extensions of the results derived in Sections 1 and 2. In particular, we consider spectral measures with positive mass at the point 0 and extends the results to the multivariate case. Moreover, we briefly indicate a relation of our results to the optimal approximation of a constant function by polynomials with no intercept.

3.1 Spectral measures with finite support

If the spectral measure $\alpha(dt)$ in (1.2) has finite support, say $\mathcal{T} = \{\pm t_1, \dots, \pm t_m\}$ with $m \geq 1$ and $0 < t_1 < \dots < t_m$, then the matrices C_n in (2.6) are invertible for $n \leq m - 1$ but

$$\det(C_n) = \det(c_{2(i+j)})_{i,j=0}^n = 0 \quad \text{for} \quad n \geq m. \quad (3.1)$$

Consequently, observing Lemma 2.4 we have in this case

$$\text{var}(\hat{\theta}_n) = 0 \quad \text{for } n = m, m + 1, \dots$$

Therefore, by Lemma 2.2 a non-vanishing constant function does not belong to $H(K)$ if the corresponding spectral measure has finite support.

The relation (3.1) follows, observing the representation

$$C_n = 2 \sum_{i=1}^m w_i g(t_i) g^\top(t_i) \in \mathbb{R}^{(n+1) \times (n+1)}$$

where $g(t) = (1, t^2, \dots, t^{2n})^\top$ and w_1, \dots, w_m are the masses of the measure α at the points t_1, \dots, t_m . As C_n is a sum of rank one matrices, it is singular whenever $n > m - 1$. On the other hand, in the case $m = n + 1$ we have by Vandermond's determinant formula

$$\det C_n = \prod_{i=1}^{n+1} (2w_i) \prod_{1 \leq i < j \leq n+1} (t_i^2 - t_j^2)^2 > 0,$$

which shows that C_n is nonsingular. Finally, if $m \geq n + 1$ we have (in the Loewner ordering)

$$C_n \geq 2 \sum_{i=1}^{n+1} w_i g(t_i) g^\top(t_i)$$

where the matrix on the right-hand side is positive definite.

3.2 Multivariate case

Consider the location scale model (2.2) but assume that X is a subset of \mathbb{R}^d with non-empty interior. Extensions of Theorems 1.1 and 1.2 to the multivariate case, when $d > 1$, essentially follow from the one-dimensional results because it is sufficient to use derivatives of the process $\{y(x); x \in X\}$ with respect to one variable for construction of estimators $(\hat{\theta}_n)_{n \in \mathbb{N}}$ and subsequent application of Lemma 2.2. In the following discussion we consider two cases for the kernel $K(x, x')$ using the notation $x = (x_1, \dots, x_d)^\top$, $x' = (x'_1, \dots, x'_d)^\top$ and $t = (t_1, \dots, t_d)^\top$. We also denote by $x_{(i)}, x'_{(i)}$ and $t_{(i)} \in \mathbb{R}^{d-1}$ the vectors x, x' and t with i -th component removed, respectively.

Case 1: Assume that K is a product kernel, that is

$$K(x, x') = \prod_{j=1}^d K_j(x_j, x'_j), \quad (3.2)$$

where for all $j = 1, \dots, d$ the kernel K_j (defined on a subset of \mathbb{R}^2) satisfies $K_j(x_j, x'_j) = k_j(x_j - x'_j)$ and k_j is a non-constant positive definite function infinitely differentiable at the point 0. Denote by $\alpha_j(dt_j)$ the spectral measure for k_j and define $\alpha(dt) = \alpha_1(dt_1) \cdots \alpha_d(dt_d)$. To construct the sequence of estimators $(\hat{\theta}_n)_{n \in \mathbb{N}}$ for the application of Lemma 2.2, we can use the derivatives with respect to the i -th coordinate for any i . Therefore, Corollary 1.1 can be generalized as follows.

Corollary 3.1 *Assume that $X \subset \mathbb{R}^d$ and the kernel K has the form (3.2). Then we have the following:*

- (a) *If the measure α has a positive mass at the point 0, then the constant functions belong to $H(K)$.*
- (b) *If for at least one $i \in \{1, \dots, d\}$ the Hamburger moment problem for the measure $\alpha_i(\cdot)$ is determinate and the measure α_i does not have a positive mass at the point 0, then any non-vanishing constant function does not belong to $H(K)$.*
- (c) *If for at least one $i \in \{1, \dots, d\}$ the Hamburger moment problem for the measures $t_i^{2m} \alpha_i(dt_i) / c_{2m}$ is determinate for all $m = 0, 1, \dots$, then $H(K)$ does not contain non-constant polynomials on X .*

Note that the set X in Corollary 3.1 does not have to be a product of one-dimensional sets. Moreover, we also point out that the assumption (3.2) can be generalized to kernels of the form

$$K(x, x') = k_i(x_i - x'_i) K_{(d-1)}(x_{(i)}, x'_{(i)}),$$

where $K_{(d-1)}(\cdot, \cdot)$ is a positive definite and suitably differentiable kernel on $\mathbb{R}^{d-1} \times \mathbb{R}^{d-1}$ and k_i is a non-constant positive definite function infinitely differentiable at the point 0.

Case 2: The kernel K satisfies

$$K(x, x') = k(x - x'),$$

where k is a positive definite function on \mathbb{R}^d . Consider the spectral measure $\alpha(dt)$ corresponding to k by Bochner's theorem, that is

$$k(x) = \int_{\mathbb{R}^d} e^{i(t_1x_1 + \dots + t_dx_d)} \alpha(dt) , \quad (3.3)$$

and denote by

$$\alpha_i(B) = \int_{\mathbb{R}^d} I_B(t_i) \alpha(dt) , \quad B \in \mathcal{B},$$

the i th the marginal distribution of the measure α ($i = 1, \dots, d$), where I_B denotes the indicator function of the set B . In this case, we can generalize Corollary 1.1 as follows.

Corollary 3.2 *If the spectral measure $\alpha(dt)$ does not have a positive mass at the point 0 and if for at least one $i \in \{1, \dots, d\}$ the Hamburger moment problems for the measures proportional to $t^{2m} \alpha_i(dt_i)$ are determinate for all $m = 0, 1, \dots$, then $H(K)$ does not contain non-vanishing polynomials.*

The case when the spectral measure has positive mass at the point 0 is treated similarly in one-dimensional and multi-dimensional cases, see Section 3.4.

3.3 Discretization of the space and the limit of discrete BLUEs

In Section 3.4 below we will prove that constant functions belong to $H(K)$ if the spectral measure has positive mass at the point 0. The proof requires an auxiliary result which is of own interest and shows that in the case $f \in H(K)$ the variance of the continuous BLUE is the limit of the variances of discrete BLUEs, after a suitable discretization of X has been performed.

Lemma 3.1 *Let X be a compact in \mathbb{R}^d , $(x_n)_{n \in \mathbb{N}}$ be a sequence of distinct points in X such that*

$$\sup_{x \in X} \min_{1 \leq i \leq n} \|x - x_i\| \rightarrow 0 \text{ as } n \rightarrow \infty . \quad (3.4)$$

Let $\hat{\theta}_{BLUE,n}$ be the BLUE of θ in model (2.1) from the observations of $y(x_1), \dots, y(x_n)$. Then $f \in H(K)$ if and only if $\text{var}(\hat{\theta}_{BLUE,n}) \rightarrow c > 0$ as $n \rightarrow \infty$.

Moreover, if $f \in H(K)$, the continuous BLUE $\hat{\theta}_{BLUE}$ of θ in model (2.1) exists and

$$c = 1/\|f\|_{H(K)} = \text{var}(\hat{\theta}_{BLUE}).$$

Proof. Let $X_n = \{x_1, \dots, x_n\}$, K_n denote the restriction of K on X_n , and define $H_n = H(K_n)$ as the RKHS corresponding to the kernel K_n . By Theorem 6 in Section 1.4.2 of Berlinet and Thomas-Agnan (2011) we have for the restriction f_n of f on X_n that $f_n \in H_n$ and

$$\|f_n\|_{H_n} \leq \|f_{n+1}\|_{H_{n+1}} \leq \|f\|_{H(K)}.$$

Consequently, the sequence of $(\text{var}(\hat{\theta}_{BLUE,n}))_{n \in \mathbb{N}} = (1/\|f_n\|_{H_n})_{n \in \mathbb{N}}$ is monotonously decreasing so that the limit $c = \lim_{n \rightarrow \infty} \text{var}(\hat{\theta}_{BLUE,n}) \geq 0$ exists for any f . Moreover, $\text{var}(\hat{\theta}_{BLUE,n}) \geq c$ for all $n \in \mathbb{N}$.

If $f \in H(K)$ we have by Proposition 4.12 in Paulsen (2016) that

$$\lim_{n \rightarrow \infty} \text{var}(\hat{\theta}_{BLUE,n}) = c = 1/\|f\|_{H(K)}.$$

Conversely, if $\text{var}(\hat{\theta}_{BLUE,n}) \rightarrow c$ as $n \rightarrow \infty$ for some $c > 0$, we can use the equivalence between (1) and (2) in Theorem 4.15 in Paulsen (2016) to deduce that $f \in H(K)$. \square

Recall that the explicit expression for the variance of the discrete BLUE $\hat{\theta}_{BLUE,n}$ of Lemma 3.1 is given by

$$\text{var}(\hat{\theta}_{BLUE,n}) = 1/F_n^\top W_n^{-1} F_n, \tag{3.5}$$

where

$$F_n = (f(x_1), \dots, f(x_n))^\top, \quad W_n = (K(x_i, x_j))_{i,j=1}^n. \tag{3.6}$$

3.4 Spectral measures with positive mass at the point 0

In this section, we investigate the case, where the spectral measure has a positive mass at the point 0 in more detail. In particular, we show that in this case the constant functions belong to $H(K)$. To be precise, assume that the covariance kernel of the error process

has the form

$$K_\gamma(x, x') = \gamma + (1 - \gamma)K(x, x'), \quad (3.7)$$

where $0 \leq \gamma < 1$ and $K(x, x')$ is a strictly positive definite kernel on a compact set $X \subset \mathbb{R}^d$. Note that in the particular case $d = 1$ and $K(x, x') = k(x - x')$ with k having the spectral measure $\alpha(dt)$, we obtain the representation (2.10) for the spectral measure α_γ .

Theorem 3.1 *Let $X \subset \mathbb{R}^d$ be a compact set and assume the kernel K_γ has the form (3.7) with $0 < \gamma < 1$. Then the constant functions belong to $H(K_\gamma)$.*

Proof. Consider the location scale model

$$y(x) = \theta + \varepsilon(x), \quad x \in X, \quad \mathbb{E}\varepsilon(x) = 0, \quad \mathbb{E}\varepsilon(x)\varepsilon(x') = K_\gamma(x, x'). \quad (3.8)$$

and let $(x_n)_{n \in \mathbb{N}}$ denote a sequence of distinct points in X such that (3.4) is satisfied. Let $\hat{\theta}_{m,\gamma}$ be the BLUE of θ in the model (3.8), constructed on the observations of $y(x_1), \dots, y(x_m)$. Define $W_{m,\gamma} = (K_\gamma(x_i, x_j))_{i,j=1}^m$, $Y_m = (y(x_1), \dots, y(x_m))^\top$ and $\mathbf{1}_m = (1, \dots, 1)^\top \in \mathbb{R}^m$. As the covariance kernel $K(x, x')$ is strictly positive definite, the matrix $W_{m,\gamma}$ is invertible for all $m \geq 1$, $0 \leq \gamma < 1$. Therefore, the BLUE is unique and given by

$$\hat{\theta}_{m,\gamma} = \mathbf{1}_m^\top W_{m,\gamma}^{-1} Y_m / \mathbf{1}_m^\top W_{m,\gamma}^{-1} \mathbf{1}_m.$$

Its variance is

$$\text{var}(\hat{\theta}_{m,\gamma}) = 1 / \mathbf{1}_m^\top W_{m,\gamma}^{-1} \mathbf{1}_m.$$

For simplicity of notation, denote $\kappa_{m,\gamma} = \mathbf{1}_m^\top W_{m,\gamma}^{-1} \mathbf{1}_m = 1 / \text{var}(\hat{\theta}_{m,\gamma})$. The same arguments as given in the proof of Lemma 3.1 show that for any $0 \leq \gamma < 1$, the sequence $(\kappa_{m,\gamma})_{m \in \mathbb{N}}$ is monotonously increasing with some limit $c_\gamma = \lim_{m \rightarrow \infty} \kappa_{m,\gamma} \in (0, \infty]$. Observing the representation

$$W_{m,\gamma} = (1 - \gamma)W_{m,0} + \gamma \mathbf{1}_m \mathbf{1}_m^\top$$

(for all $m = 1, 2, \dots$ and $0 < \gamma < 1$), we have

$$W_{m,\gamma}^{-1} = \frac{1}{1 - \gamma} \left[W_{m,0}^{-1} - \frac{\gamma}{1 - \gamma + \gamma \mathbf{1}_m^\top W_{m,0}^{-1} \mathbf{1}_m} W_{m,0}^{-1} \mathbf{1}_m \mathbf{1}_m^\top W_{m,0}^{-1} \right].$$

This implies

$$\kappa_{m,\gamma} = \frac{\kappa_{m,0}}{1-\gamma} \left[1 - \frac{\gamma\kappa_{m,0}}{1-\gamma+\gamma\kappa_{m,0}} \right] = \frac{\kappa_{m,0}}{1-\gamma+\gamma\kappa_{m,0}},$$

and therefore it follows that

$$\text{var}(\hat{\theta}_{m,\gamma}) = 1/\kappa_{m,\gamma} = \gamma + (1-\gamma)\text{var}(\hat{\theta}_{m,0}). \quad (3.9)$$

Taking the limit (as $m \rightarrow \infty$) in (3.9) we obtain for all $0 < \gamma < 1$:

$$\lim_{m \rightarrow \infty} \text{var}(\hat{\theta}_{m,\gamma}) = \gamma + (1-\gamma)/c_0 \geq \gamma > 0.$$

Lemma 3.1 now yields that the constant functions belong to $H(K_\gamma)$. □

3.5 Estimation of the variance of a Gaussian random field

Let $X \subset \mathbb{R}^d$ be a compact set, and let f denote of a Gaussian random process (field) on X with a strictly positive definite covariance kernel $R(x, x') = \sigma^2 K(x, x')$ on $X \times X$, where the kernel $K(x, x')$ is known but σ^2 is unknown. For estimating σ^2 we assume that one can observe f at n distinct points $x_1, \dots, x_n \in X$. Then it is easy to see (see p.140 in Xu and Stein, 2017)) that the corresponding log-likelihood function is given by

$$LL(\sigma^2) = \frac{1}{2} \left[-n \log(2\pi) - n \log(\sigma^2) - \log(\det(W_n)) - \frac{1}{\sigma^2} F_n^\top W_n^{-1} F_n \right], \quad (3.10)$$

where F_n and W_n are defined by (3.6) and a simple calculation shows that the maximum likelihood estimator (MLE) of σ^2 is given by

$$\widehat{\sigma}_n^2 = \frac{1}{n} F_n^\top W_n^{-1} F_n. \quad (3.11)$$

Comparing (3.11) with (3.5) we get

$$\widehat{\sigma}_n^2 = \frac{1}{n \text{var}(\widehat{\theta}_{BLUE,n})}, \quad (3.12)$$

and by Lemma 3.1 we obtain the following corollary.

Corollary 3.3 *Let X be a compact in \mathbb{R}^d , K be a strictly positive definite kernel on $X \times X$ and f denote a centred Gaussian random field with covariance kernel $R(x, x') = \sigma^2 K(x, x')$. If x_1, x_2, \dots is a sequence of distinct points in X satisfying (3.4) and $\widehat{\sigma}_n^2$ is the MLE of σ^2 from the observations $f(x_1), \dots, f(x_n)$, then the function f belongs to the reproducing kernel Hilbert space $H(K)$ if and only if $\lim_{n \rightarrow \infty} \widehat{\sigma}_n^2 < \infty$.*

3.6 Best polynomial approximation

Let $L_2(\alpha)$ denote the space of square integrable functions with respect to the measure $\alpha(dt)$ on the real line and define \mathcal{P}_{n-1} to be the space of polynomials of degree $n - 1$. For $p \in \mathcal{P}_{n-1}$ we consider the $L_2(\alpha)$ -distance

$$V(p) = \int_{-\infty}^{\infty} (1 - t^2 p(t^2))^2 \alpha(dt)$$

between the constant function $g(t) \equiv 1$ and the even polynomial $t^2 p(t^2)$ of degree $2n$ with no intercept. A well known result in approximation theory (see, for example, Achieser, 1956, p. 15-16) shows that

$$\min_{p_n \in \mathcal{P}_{n-1}} V(p) = \frac{\det(c_{2(i+j)})_{i,j=0}^n}{\det(c_{2(i+j)})_{i,j=1}^n} = \text{var}(\hat{\theta}_n), \quad (3.13)$$

where c_0, c_2, c_4, \dots are the (even) moments of the spectral measure α defined in (2.4) and the last equality is a consequence of Lemma 2.4.

From this representation it follows that $\text{var}(\hat{\theta}_n) \rightarrow 0$ as $n \rightarrow \infty$ if and only if non-zero constant functions can be approximated by polynomials of the form $\tilde{p}_n(t) = t^2 p_n(t^2)$ with arbitrary small error. Moreover, for any polynomial p on $(-\infty, \infty)$, we have

$$V(p) = \int_0^{\infty} (1 - tp(t))^2 \alpha_+(dt) = b_2 \int_0^{\infty} (1/t - p(t))^2 \alpha_{2,+}(dt),$$

where the measure $\alpha_+(dt)$ is defined by (2.11), $b_2 = \int_0^{\infty} t^2 d\alpha_+(t)$ and $\alpha_{2,+}(dt) = t^2 \tilde{\alpha}_+(dt)/b_2$. From Corollary 2.3.3 in Akhiezer (1965), it therefore follows that the set of all polynomials $\mathcal{P}_\infty = \cup_{n=0}^{\infty} \mathcal{P}_n$ is dense in the space $L_2([0, \infty), \nu)$ if the measure ν on $[0, \infty)$ is the (unique) solution of a determinate Hamburger moment problem. As the function $f(t) = 1/t$ belongs to $L_2((0, \infty), \alpha_{2,+})$ we thus obtain from (3.13) another proof of the

fact that if $\alpha(dt)$ has no mass at 0 and $\alpha_{2,+}(dt)$ is moment-determinate in the Hamburger sense then $\text{var}(\hat{\theta}_n) \rightarrow 0$. Note that this is almost equivalent to the ‘if’ statement in the important Lemma 2.7.

4 Rates of convergence

In this section, we derive for several specific classes of correlation kernels explicit results on the rate of convergence of the ratio $\text{var}(\hat{\theta}_n) = H_n/G_n$, see (2.8), where H_n and G_n are the determinants defined in (2.9).

4.1 Squared exponential (Gaussian) kernel

We first consider the case of the squared exponential kernel $K(x, x') = \exp\{-\lambda(x - x')^2\}$ with $X \subset \mathbb{R}$ and $\lambda > 0$. Assuming for simplicity $\lambda = 1/4$, we obtain that the spectral measure is absolute continuous with density

$$\varphi(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}, \quad -\infty < t < \infty.$$

The moments of even order of the measure α are given by

$$c_{2j} = \int_{-\infty}^{\infty} t^{2j} \varphi(t) dx = \int_0^{\infty} t^j g(t) dt = b_j = 2^j (2j - 1)!! \quad j = 0, 1, \dots,$$

where $g(y) = \frac{1}{\sqrt{\pi}} y^{-1/2} e^{-y}$, $y > 0$. Using (1.6) easy to see that the corresponding Hamburger moment problem is determinate and therefore non-vanishing constant functions (and all polynomials) do not belong to the corresponding RKHS. We now investigate the variance of the discrete BLUE defined in (2.7), which is given by the ratio of the determinants H_n and G_n .

It follows from results in Lau and Studden (1988) that the determinant of the Hankel matrix defined in (2.9) has the representation

$$H_n = |c_{2(i+j)}|_{i,j=0}^n = |b_{i+j}|_{i,j=0}^n = \prod_{i=1}^n \left(\tilde{d}_{2i-1} \tilde{d}_{2i} \right)^{n-i+1}, \quad (4.1)$$

where \tilde{d}_j are the coefficients of the three-term recurrence relation

$$P_{\ell+1}(t) = (t - \tilde{d}_{2\ell} - \tilde{d}_{2\ell+1})P_{\ell}(t) - \tilde{d}_{2\ell-1}\tilde{d}_{2\ell}P_{\ell-1}(t), \quad \ell = 0, 1, \dots \quad (4.2)$$

of the monic orthogonal polynomials with respect to measure $g(y)dy$ ($\tilde{d}_0 = 0$, $P_0(t) = 1$, $P_{-1}(t) = 0$). Observing the three-term recurrence relation

$$(\ell + 1)L_{\ell+1}^{(\alpha)}(t) = (-t + 2\ell + \alpha + 1)L_{\ell}^{(\alpha)}(t) - (\ell + \alpha)L_{\ell-1}^{(\alpha)}(t)$$

for the Laguerre polynomials $L_n^{(\alpha)}(t)$ (orthogonal with respect to $e^{-y}y^{\alpha}dy$, $y > 0$) we can identify the coefficients in (4.2). More precisely, the monic polynomials

$$\bar{L}_{\ell+1}^{(\alpha)}(t) = (-1)^{\ell+1}(\ell + 1)!L_{\ell+1}^{(\alpha)}(t)$$

satisfy a three-term recurrence relation of the form (4.2) with $\tilde{d}_{2k} = k$, $\tilde{d}_{2k-1} = k + \alpha$, see Dette and Studden (1992), Lemma 2.2 (b). As $P_{\ell}(t) = \bar{L}_{\ell}^{(-1/2)}(t)$ we have $\tilde{d}_{2k} = k$, $\tilde{d}_{2k-1} = k - 1/2$, and therefore obtain

$$H_n = \prod_{k=1}^n (k(2k-1))^{n-k+1} \prod_{k=1}^n \left(\frac{1}{2}\right)^{n-k+1} = \left(\frac{1}{2}\right)^{n(n+1)/2} \prod_{k=1}^n (k(2k-1))^{n-k+1}. \quad (4.3)$$

Now we move on to the determinant $G_n = |b_{i+j}|_{i,j=1}^n$. Note that we have

$$b_j = \frac{1}{\sqrt{\pi}} \int_0^{\infty} y^{j-2} y^{3/2} e^{-y} dy = \frac{3}{4} a_{j-2}$$

for $j \geq 2$, where $a_k = \int_0^{\infty} y^k \tilde{g}(y) dy$ and the density \tilde{g}_k is defined by $\tilde{g}(y) = \frac{4}{3\sqrt{\pi}} y^{3/2} e^{-y}$, $y > 0$. Therefore,

$$G_n = \left(\frac{3}{4}\right)^n |a_{i+j}|_{i,j=0}^{n-1} = \left(\frac{3}{4}\right)^n \prod_{l=1}^{n-1} (\bar{d}_{2l-1} \bar{d}_{2l})^{n-l},$$

where $\bar{d}_{2l-1} = l + 3/2$, $\bar{d}_{2l} = l$. Consequently,

$$G_n = \left(\frac{3}{4}\right)^n \left(\frac{1}{2}\right)^{n(n-1)/2} \prod_{k=1}^{n-1} (k(2k+3))^{n-k}$$

and it follows

$$\begin{aligned}\frac{H_n}{G_n} &= \left(\frac{4}{3}\right)^n \left(\frac{1}{2}\right)^n \left[\prod_{k=1}^{n-1} \frac{(k(2k-1))^{n-k+1}}{(k(2k+3))^{n-k}} \right] n(2n-1) \\ &= \left(\frac{2}{3}\right)^n n!(2n-1) \prod_{k=1}^{n-1} \frac{(2k-1)^{n-k+1}}{(2k+3)^{n-k}}.\end{aligned}$$

Since

$$\prod_{k=1}^{n-1} \frac{(2k-1)^{n-k+1}}{(2k+3)^{n-k}} = \frac{3^n}{(2n-1)(2n+1)!!}$$

we obtain

$$\frac{H_n}{G_n} = \frac{2^n n!}{(2n+1)!!} = \frac{\sqrt{\pi}}{2\sqrt{n}} \left[1 - \frac{3}{8n} + \frac{25}{128n^2} + O\left(\frac{1}{n^3}\right) \right], \quad n \rightarrow \infty. \quad (4.4)$$

The expansion (4.4) details the asymptotic relation formulated as Theorem 3.3 in Xu and Stein (2017) in the case $p = 0$. Note that formula (4.4) also corrects a minor mistake in this reference, which gives $\frac{\sqrt{\pi}}{\sqrt{2n}}$ as the leading term.

4.2 Spectral measure with Beta distribution

For measures with a compact support the determinants H_n and G_n can be conveniently evaluated using the theory of canonical moments, see e.g. Dette and Studden (1997). Exemplarily, we consider the symmetric Beta (α, α) distribution on the interval $[-1, 1]$ with density

$$\psi'_\alpha(t) = \frac{1}{2^{2\alpha+1} B(\alpha+1, \alpha+1)} (1-t^2)^\alpha, \quad -1 < t < 1, \quad (4.5)$$

where $\alpha > -1$ and $B(\alpha, \beta)$ denotes the Beta-function. For later purposes we also introduce the Beta (α, β) distribution on the interval $[0, 1]$ with density

$$\phi_{\alpha,\beta}(t) = \frac{1}{B(\beta+1, \alpha+1)} t^\beta (1-t)^\alpha, \quad 0 < t < 1, \quad (4.6)$$

where the $\alpha, \beta > -1$. The canonical moments of the Beta-distribution with density (4.6) are given by

$$p_{2j} = \frac{j}{2j+1+\alpha+\beta}, \quad p_{2j-1} = \frac{\beta+j}{2j+\alpha+\beta}; \quad (4.7)$$

see e.g. formula (1.3.11) in Dette and Studden (1997). It is easy to see that the distribution on the interval $[0, 1]$ related to the distribution ψ_α in (4.5) by the transformation (2.11) is a Beta $(\alpha, -\frac{1}{2})$ distribution. Therefore, it follows from (4.7) that the corresponding canonical moments are given by

$$p_{2j} = \frac{j}{2j+1/2+\alpha}, \quad p_{2j-1} = \frac{j-1/2}{2j-1/2+\alpha}. \quad (4.8)$$

Now Theorem 1.4.10 in Dette and Studden (1997) gives

$$H_n = |(b_{i+j})_{i,j=0}^n| = \prod_{i=1}^n (q_{2i-2}p_{2i-1}q_{2i-1}p_{2i})^{n+1-i}, \quad (4.9)$$

where $q_0 = 1, q_j = 1 - p_j$ ($j \geq 1$) and (observing (4.8))

$$q_{2i-2}p_{2i-1}q_{2i-1}p_{2i} = \frac{4i(i+\alpha)(2i-1+2\alpha)(2i-1)}{(4i+1+2\alpha)(4i-1+2\alpha)^2(4i-3+2\alpha)}, \quad i = 1, 2, \dots \quad (4.10)$$

For the calculation of the determinant $G_n = |(b_{i+j})_{i,j=1}^n|$ we note the relation

$$b_i = \frac{B(\frac{5}{2}, \alpha+1)}{B(\frac{1}{2}, \alpha+1)} \tilde{b}_{i-2} \quad i = 2, 3, \dots \quad (4.11)$$

where $\tilde{b}_0, \tilde{b}_1, \dots$ are the moments of the Beta $(\alpha, 3/2)$ distribution. Consequently, we obtain from Theorem 1.4.10 in Dette and Studden (1997) that

$$\begin{aligned} G_n &= |(b_{i+j})_{i,j=1}^n| = |(\tilde{b}_{i+j})_{i,j=0}^{n-1}| = \left[\frac{B(\frac{5}{2}, \alpha+1)}{B(\frac{1}{2}, \alpha+1)} \right]^n \times \prod_{i=1}^{n-1} (\tilde{q}_{2i-2}\tilde{p}_{2i-1}\tilde{q}_{2i-1}\tilde{p}_{2i})^{n-i} \\ &= \left[\frac{3}{(2\alpha+3)(2\alpha+5)} \right]^n \times \prod_{i=2}^n (\tilde{q}_{2i-4}\tilde{p}_{2i-3}\tilde{q}_{2i-3}\tilde{p}_{2i-2})^{n+1-i}, \end{aligned} \quad (4.12)$$

where \tilde{p}_1, \tilde{p}_2 are the canonical moments of Beta($\alpha, 3/2$) distribution, that is

$$\tilde{p}_{2i} = \frac{j}{2i + 5/2 + \alpha}, \quad \tilde{p}_{2i-1} = \frac{3/2 + i}{2i + 3/2 + \alpha},$$

and

$$\tilde{q}_{2i-2}\tilde{p}_{2i-1}\tilde{q}_{2i-1}\tilde{p}_{2i} = \frac{4i(i+\alpha)(2i+3+2\alpha)(2i+3)}{(4i+5+2\alpha)(4i+3+2\alpha)^2(4i+1+2\alpha)}, \quad i = 1, 2, \dots \quad (4.13)$$

Consequently, it follows from (4.13), (4.12) and (4.10)

$$\begin{aligned} \frac{H_n}{G_n} &= \left[\frac{(2\alpha+3)(2\alpha+5)}{3} \right]^n (q_0 p_1 q_1 p_2)^n \prod_{i=2}^n \left[\frac{q_{2i-2} p_{2i-1} q_{2i-1} p_{2i}}{\tilde{q}_{2i-4} \tilde{p}_{2i-3} \tilde{q}_{2i-3} \tilde{p}_{2i-2}} \right]^{n+1-i} \\ &= \left[\frac{4(1+\alpha)}{3(3+2\alpha)^2} \right]^n \prod_{i=2}^n \left[\frac{i(i+\alpha)(i-1/2)(i+\alpha-1/2)}{(i-1)(i-1+\alpha)(i+1/2)(i+\alpha+1/2)} \right]^{n+1-i}. \end{aligned}$$

Observing the relations

$$\begin{aligned} \prod_{i=2}^n \left[\frac{i}{i-1} \right]^{n+1-i} &= n!, \\ \prod_{i=2}^n \left[\frac{i-1/2}{i+1/2} \right]^{n+1-i} &= \frac{3^n}{(2n+1)!!}, \\ \prod_{i=2}^n \left[\frac{i+\alpha}{i-1+\alpha} \right]^{n+1-i} &= \frac{\Gamma(n+1+\alpha)}{(1+\alpha)^n \Gamma(1+\alpha)}, \\ \prod_{i=2}^n \left[\frac{i+\alpha-1/2}{i+\alpha+1/2} \right]^{n+1-i} &= \frac{(3+2\alpha)^n \Gamma(3/2+\alpha)}{2^n \Gamma(n+3/2+\alpha)}. \end{aligned}$$

we obtain

$$\begin{aligned} \frac{H_n}{G_n} &= \left[\frac{4(1+\alpha)}{3(3+2\alpha)} \right]^n \frac{n! 3^n \Gamma(n+1+\alpha) (3+2\alpha)^n \Gamma(3/2+\alpha)}{(2n+1)!! (1+\alpha)^n \Gamma(1+\alpha) 2^n \Gamma(n+3/2+\alpha)} \\ &= \frac{\sqrt{\pi}}{2^{2\alpha+1} B(\alpha+1, \alpha+1)} \times \frac{(2n)!!}{(2n+1)!!} \cdot \frac{\Gamma(n+1+\alpha)}{\Gamma(n+3/2+\alpha)} \\ &= \frac{\pi}{2^{2\alpha+2} B(\alpha+1, \alpha+1)} \times \frac{1}{n} \left(1 + O\left(\frac{1}{n}\right) \right), \quad n \rightarrow \infty \end{aligned} \quad (4.14)$$

where the expansion in the last line follows by straightforward but tedious calculation using Stirling’s formula.

We finally mention the special cases $\alpha = 0$ (the spectral measure is a uniform spectral density on the interval $[-1, 1]$ with corresponding kernel function $k(x) = \sin(x)/x$) and $\alpha = -1/2$ (the spectral measure is the arcsine distribution on $[-1, 1]$ and the corresponding kernel is $k(x) = 2J_1(x)/x$, where $J_\alpha(\cdot)$ is the Bessel function of the first kind) for which the expansions are given, respectively, by

$$\begin{aligned} \frac{H_n}{G_n} &= \left[\frac{(2n)!!}{(2n+1)!!} \right]^2 = \frac{\pi}{4n} + O\left(\frac{1}{n^2}\right), \\ \frac{H_n}{G_n} &= \left(\frac{8}{3}\right)^n \left[\frac{1}{8}\right]^n n! \frac{3^n}{(2n+1)!!} \frac{2^n \Gamma(n+1/2)}{\Gamma(1/2)} \frac{1}{n!} = \frac{1}{2n+1} = \frac{1}{2n} + O\left(\frac{1}{n^2}\right) \end{aligned} \tag{4.15}$$

as $n \rightarrow \infty$. Interestingly, the ratio H_n/G_n in (4.15) is the squared ratio H_n/G_n of (4.4).

Acknowledgements This research was partially supported by the Collaborative Research Center ‘Statistical modeling of nonlinear dynamic processes’ (*Sonderforschungsbereich 823, Teilprojekt C2*).

References

- Achieser, N. I. (1956). *Theory of Approximation*. Frederick Ungar Publishing Co., New York.
- Akhiezer, N. (1965). *The classical Moment Problem and some Related Questions in Analysis*. Oliver & Boyd.
- Berg, C., Chen, Y., and Ismail, M. E. (2002). Small eigenvalues of large Hankel matrices: The indeterminate case. *Mathematica Scandinavica*, 90:67–81.
- Berlinet, A. and Thomas-Agnan, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer.
- Bochner, S. and Chandrasekharan, K. (1949). *Fourier Transforms. (AM-19)*. University Press, Princeton, N.J.
- Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2019). The BLUE in continuous-time regression models with correlated errors. *Annals of Statistics*, 47(4):1928–1959.

- Detle, H. and Studden, W. (1992). On a new characterization of the classical orthogonal polynomials. *Journal of Approximation Theory*, 71(1):3–17.
- Detle, H. and Studden, W. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis*. John Wiley & Sons.
- Heyde, C. (1963). Some remarks on the moment problem (i). *The Quarterly Journal of Mathematics*, 14(1):91–96.
- Lau, T.-S. and Studden, W. (1988). On an extremal problem of Fejér. *Journal of Approximation Theory*, 53(2):184–194.
- Lee, K.-Y., Li, B., and Zhao, H. (2016). Variable selection via additive conditional independence. *Journal of the Royal Statistical Society B*, 78(5):1037–1055.
- Lin, G. D. (2017). Recent developments on the moment problem. *Journal of Statistical Distributions and Applications*, 4(1):1–17.
- Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, 32(4):951–989.
- Paulsen, V. (2016). *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge University Press.
- Pronzato, L. and Zhigljavsky, A. (2020). Bayesian quadrature, energy minimization, and space-filling design. *SIAM/ASA Journal on Uncertainty Quantification*, 8(3):959–1011.
- Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- Schmüdgen, K. (2017). *The Moment Problem*. Springer.
- Shohat, J. and Tamarkin, J. (1943). *The Problem of Moments*. American Math. Soc.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer.
- Steinwart, I., Hush, D., and Scovel, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. on Inform. Theory*, 52(10):4635–4643.
- Stoyanov, J. M. (2013). *Counterexamples in Probability*. Courier Corporation.
- Wendland, H. (2004). *Scattered Data Approximation*. Cambridge university press.
- Xu, W. and Stein, M. L. (2017). Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):138–175.
- Yaglom, A. (1986). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*. Springer.
- Zhigljavsky, A. and Zilinskas, A. (2007). *Stochastic Global Optimization*. Springer.

