SIAM J. NUMER. ANAL.
Vol. 62, No. 3, pp. 1191–1211

MEAN DIMENSION OF RADIAL BASIS FUNCTIONS*

CHRISTOPHER HOYT† AND ART B. OWEN‡

Abstract. We show that generalized multiquadric radial basis functions (RBFs) on $\mathbb{R}^d$ have a mean dimension that is $1 + O(1/d)$ as $d \to \infty$ with an explicit bound for the implied constant, under moment conditions on their inputs. Under weaker moment conditions the mean dimension still approaches 1. As a consequence, these RBFs become essentially additive as their dimension increases. Gaussian RBFs by contrast can attain any mean dimension between 1 and d. We also find that a test integrand due to Keister that has been influential in quasi-Monte Carlo theory has a mean dimension that oscillates between approximately 1 and approximately 2 as the nominal dimension d increases.

Key words. Jansen identity, Keister function, mesh-free approximation, multiquadrics

MSC codes. 33F05, 66M99

DOI. 10.1137/23M1614833

1. Introduction.

For high-dimensional functions it is very useful to find parameterizations in terms of some vectors of the same dimension as the input space. Two such parameterizations are ridge functions $\phi(\boldsymbol{x}^\mathsf{T}\theta)$ and radial basis functions (RBFs) $\phi(\|\boldsymbol{x}-\boldsymbol{c}\|)$ for vectors $\boldsymbol{c}, \theta$ of the same dimension as $\boldsymbol{x}$ and appropriate functions $\phi(\cdot)$. In this paper we study RBFs. We are interested in them because they serve as building blocks for numerical methods used in the mathematical sciences: interpolation, machine learning, Gaussian process regression (kriging), and multidimensional integration.

There are some results based on concentration of measure wherein high-dimensional Lipschitz functions become essentially constant as the dimension d of their domain tends to infinity. See, for example, [3]. In this paper we study the way in which some of these high-dimensional functions $\phi(\cdot)$ fluctuate around their nearly constant value. Our main results are that for certain RBFs these fluctuations must become essentially additive as $d \to \infty$, while others are not so constrained. Our techniques are based on the functional ANOVA decomposition of [9, 26, 4]. A function of d independent variables has $2^d - 1$ nontrivial variance components σ_u^2 for nonempty subsets u of variables. The mean dimension is the weighted average of cardinalities $|u|$ with weights proportional to σ_u^2. It can take values between 1 and d. A mean dimension near one means that the function is nearly additive in a least squares sense. If a function has mean dimension $1 + \epsilon$, then it has an additive approximation that explains at least $1 - \epsilon$ of its variance.

Our main result concerns generalized multiquadric RBFs. These take the form $f(\boldsymbol{x}) = (a + \sum_{j=1}^d (x_j - c_j)^2)^p$ for a point $\boldsymbol{c} \in \mathbb{R}^d$, intercept $a \geqslant 0$, and power $p < 1$. We show that under mild assumptions, their mean dimension is at most

*Received by the editors November 6, 2023; accepted for publication (in revised form) February 21, 2024; published electronically May 21, 2024.
https://doi.org/10.1137/23M1614833
Funding: This work was supported by the National Science Foundation, through grants IIS-1837931 and DMS-2152780, and by Hitachi, Ltd.
†Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305 USA (crhoyt@stanford.edu).
‡Department of Statistics, Stanford University, Stanford, CA 94305 USA (owen@stanford.edu).

$1 + (p-1)^2V/(2M^2d) + O(1/d^2)$, where M and V are certain average moments presented later. As a result these functions become essentially additive in high dimensions.

The well-known Gaussian RBF (that we define below) need not be of low mean dimension. We show that it can be parameterized to attain any mean dimension in the interval $(1, d)$ when the variables in it have continuous distributions. This RBF is known as the Gaussian RBF in machine learning, and also as the squared exponential RBF elsewhere in the literature.

To fix ideas, suppose that we have measured values $f(\boldsymbol{x}_i)$ for $\boldsymbol{x}_i \in \mathbb{R}^d$ and $i = 1, \dots, n$. We seek an interpolant $\tilde{f}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$. We might then use

$$\tilde{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \beta_i \phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$$

after solving n equations in n unknowns to compute $\beta = (\beta_1, \dots, \beta_n)^\mathsf{T} \in \mathbb{R}^n$. Only certain special functions ϕ are good choices for this usage. We describe some of those in a later section based on material from [6]. For now we mention generalized multiquadrics $\phi(\|\boldsymbol{x}\|) = (a + \|\boldsymbol{x}\|\vartheta^2)^p$ and Gaussians, $\phi(\|\boldsymbol{x}\|) = \exp(-\|\boldsymbol{x}\|^2\vartheta^2)$ for parameters $p \in \mathbb{R} \setminus \mathbb{N}_0$ and $a \geqslant 0$ and $\vartheta > 0$.

Now suppose that $f(\boldsymbol{x})$ is not nearly additive but all of the $\phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$ are nearly additive. It would still be possible to interpolate if the β_i took values of large magnitude with opposite signs that mostly canceled the additive parts in $\phi(\cdot)$. We would, however, expect serious numerical conditioning difficulties in that setting. RBF approximation is often ill-conditioned even with functions that are not nearly additive. Fitting a nonadditive function by nearly additive basis functions can only make things worse.

The covariance functions used in Gaussian process regression often take the RBF form, especially in geoscience. An additive covariance function implies additive realizations of the random field, a potentially serious limitation. This may be why covariances of product form are more popular than covariance models of the RBF form in high-dimensional Gaussian process models such as those used in computer experiments [23].

An important test function for quasi-Monte Carlo (QMC) integration is the Keister function from [14]. This is a radial basis function. Although it is expressed as a sinusoidal function of $\|\boldsymbol{x}\|^2$ for Gaussian $\boldsymbol{x}$, making all d variable equally important, we will see that it is generally of low mean dimension.

The asymptotic mean dimension of ridge functions was studied in [10] for $\boldsymbol{x} \sim \mathcal{N}(0, I)$. If the ridge function $\phi(\cdot)$ is Lipschitz continuous, then the mean dimension of $f(\boldsymbol{x}) = \phi(\boldsymbol{x}^\mathsf{T}\theta)$ for a unit vector $\theta \in \mathbb{R}^d$ remains bounded as the nominal dimension $d \to \infty$. Some discontinuous ridge functions can have mean dimensions that grow proportionally to $\sqrt{d}$. A form of conditional QMC known as preintegration (see [7]) can convert them into Lipschitz continuous ridge functions, greatly reducing their asymptotic mean dimension, which then makes them easier to integrate numerically.

An outline of this paper is as follows. Section 2 introduces our notation, gives some properties of RBFs, and presents the functional ANOVA and related material for mean dimension. Section 3 shows that, under some moment conditions, generalized multiquadric RBFs have mean dimension $1 + O(1/d)$ as $d \to \infty$ with an explicit upper bound on the implied constant in the $O(1/d)$ term. Under much weaker moment conditions, the mean dimension still approaches 1 as $d \to \infty$. Section 4 shows that the Gaussian RBFs can attain any mean dimension in the interval $(1, d)$ when the inputs have continuous distributions with bounded densities having support near $\boldsymbol{c}$.

Section 5 shows that the mean dimension of the Keister function oscillates between nearly 1 and nearly 2 as the nominal dimension d increases. Section 6 discusses how mean dimension varies among alternative methods. Finally, there are appendices for the lengthier proofs.

2. Notation and elementary results.

We study functions $f : \mathbb{R}^d \to \mathbb{R}$. The argument to f is denoted by $\boldsymbol{x} = (x_1, \ldots, x_d)$. The components of $\boldsymbol{x}$ are independent random variables. We use $\boldsymbol{x}'$ to denote another variable with the same distribution as $\boldsymbol{x}$, which is independent of $\boldsymbol{x}$. We will use hybrid points $\boldsymbol{x}_{-j}{:}\boldsymbol{x}'_j \in \mathbb{R}^d$ that combine inputs from both $\boldsymbol{x}$ and $\boldsymbol{x}'$. If $\boldsymbol{y} = \boldsymbol{x}_{-j}{:}\boldsymbol{x}'_j$, then $y_j = x'_j$ and $y_\ell = x_\ell$ for $\ell \neq j$. We use $[d]$ to denote the set $\{1, 2, \ldots, d\}$. For $u \subseteq [d]$ we use $|u|$ for the cardinality of u. The point $\boldsymbol{x}_u \in \mathbb{R}^{|u|}$ is comprised of x_j for $j \in u$. The complement $[d] \setminus u$ is denoted by $-u$, and $\boldsymbol{x}_{-u}$ consists of those x_j with $j \notin u$.

2.1. Radial basis functions.

The description here is based on [6]. Radial basis functions are used for scattered data interpolation, also known as mesh-free approximation, meaning that the sample points are not necessarily in a regular structure like a grid. One strong motivation for them is that polynomial interpolation is not necessarily well defined for an arbitrary set of points $\boldsymbol{x}_i \in \mathbb{R}^d$ for $d \geqslant 2$, but some RBFs can interpolate at any distinct points.

The RBF interpolant is of the form $\sum_{i=1}^n \beta_i \phi(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$. Fasshauer [6] also considers the more general form

$$\sum_{i=1}^n \beta_i \tilde{\phi}(\boldsymbol{x} - \boldsymbol{x}_i),$$

where $\tilde{\phi}(\cdot)$ is not necessarily "radial," i.e., not necessarily a function of the norm of its argument. To interpolate in this more general setting we must solve

$$K\beta = y \tag{2.1}$$

for $\beta \in \mathbb{R}^n$, where $K \in \mathbb{R}^{n\times n}$ has ij entry $\tilde{\phi}(\boldsymbol{x}_i - \boldsymbol{x}_j)$ and $y \in \mathbb{R}^n$ has ith entry $f(\boldsymbol{x}_i)$. The function $\tilde{\phi}$ is "radial" if $\tilde{\phi}(\boldsymbol{x}_i - \boldsymbol{x}_j) = \phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|)$ for a function $\phi : [0, \infty) \to \mathbb{R}$. The function $\tilde{\phi}$ is called positive definite if K is always positive semidefinite for any $n \geqslant 1$ and any distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$. If this K is always positive definite, then $\tilde{\phi}$ is strictly positive definite.

Strictly positive definite functions can be used to interpolate any values $f(\boldsymbol{x}_i)$ at distinct $\boldsymbol{x}_i$. Chapter 6 of [6] describes conditionally positive definite functions of order m that can be used to interpolate functions that are orthogonal to all multivariate polynomials of order less than or equal to $m-1$. To use them, one interpolates with a suitable polynomial plus a conditionally positive definite RBF.

Chapter 3 of [6] provides numerous properties and characterizations of positive definite functions and strictly positive definite functions. If $\tilde{\phi}$ is positive definite, then $|\tilde{\phi}(\boldsymbol{x})| \leqslant \tilde{\phi}(0)$. A real valued continuous and positive definite function must be even.

Our main interest here is in (strictly) positive definite radial functions. If $\tilde{\phi}(\cdot) = \phi(\|\cdot\|)$ is (strictly) positive definite for dimension d, then the same holds (strictly or not) for all dimensions $d' \leqslant d$. Because we want to study the limit as $d \to \infty$ we are interested in ϕ that provide strictly positive definite functions for all $d \geqslant 1$. By Theorem 3.8 of [6], due to Schoenberg, the function $\phi : [0, \infty) \to \mathbb{R}$ with

$$\phi(r) = \int_0^\infty e^{-r^2 t^2} \mu(\mathrm{d}t) \tag{2.2}$$

provides a strictly positive definite radial function for all dimensions $d \geqslant 1$ if and only if μ is a finite positive Borel measure not concentrated on $\{0\}$. It follows that these desirable functions ϕ can take no negative values, must be strictly decreasing, and cannot have compact support.

It is clear from (2.2) that the Gaussian RBF $\phi(r) = e^{-r^2\vartheta^2}$ is a strictly positive definite radial function for $\vartheta > 0$ in all dimensions $d \geqslant 1$. So are generalized inverse multiquadrics

$$(1 + \|\boldsymbol{x}\|^2)^p, \quad p < 0,$$

from [6, p. 42].

Table 3.1 of [5] names some of the more widely used generalized multiquadric RBFs $\phi(r)$:

$$\begin{aligned} (1+\vartheta^2 r^2)^{-1} &\quad \text{Inverse quadratic,}\\ (1+\vartheta^2 r^2)^{-1/2} &\quad \text{Inverse multiquadric,}\\ (1+\vartheta^2 r^2)^{1/2} &\quad \text{Multiquadric,}\end{aligned}$$

with a parameter $\vartheta > 0$. The last one is the one in the influential paper of Hardy [8].

2.2. ANOVA and Sobol' indices. The analysis of variance (ANOVA) decomposition of $f \in L_2([0,1]^d)$ is in [9, 26, 4]. We use its generalization to measurable functions f of $\boldsymbol{x} \in \mathbb{R}^d$ where the components x_j are independent random variables and $\mathbb{E}(f(\boldsymbol{x})^2) < \infty$. This decomposition writes as

$$f(\boldsymbol{x}) = \sum_{u \subseteq [d]} f_u(\boldsymbol{x}),$$

where the ANOVA effect f_u is a function that only depends on $\boldsymbol{x}$ through components x_j for $j \in u$. In this decomposition, $\mathbb{E}(f_u(\boldsymbol{x})f_v(\boldsymbol{x})) = 0$ for $u \neq v$ and $f_\varnothing$ is the constant function everywhere equal to $\mathbb{E}(f(\boldsymbol{x}))$. The quantities

$$\sigma_u^2 = \mathrm{Var}(f_u(\boldsymbol{x})) = \begin{cases} \mathbb{E}(f_u(\boldsymbol{x})^2), & u \neq \varnothing,\\ 0, & u = \varnothing,\end{cases}$$

are known as the variance components of f. They satisfy $\sigma^2 = \sum_{u \subseteq [d]} \sigma_u^2$, where $\sigma^2 = \mathrm{Var}(f(\boldsymbol{x}))$.

When $0 < \sigma^2 < \infty$, we define the mean dimension of f as

$$\nu(f) = \frac{1}{\sigma^2} \sum_{u \subseteq [d]} |u| \sigma_u^2.$$

The closest additive function to f in mean square is

$$f_{\text{add}}(\boldsymbol{x}) = f_\varnothing(\boldsymbol{x}) + \sum_{j=1}^{d} f_{\{j\}}(\boldsymbol{x}).$$

If $\nu(f)$ is close to one, then f is nearly additive in an L^2 sense. More precisely

$$\nu(f) \leqslant 1 + \epsilon \implies \frac{\mathrm{Var}(f(\boldsymbol{x}) - f_{\text{add}}(\boldsymbol{x}))}{\mathrm{Var}(f(\boldsymbol{x}))} \leqslant \epsilon.$$

We can get a good computational and theoretical handle on the mean dimension by using Sobol' indices as follows. The unnormalized Sobol' indices of f for $u \subseteq [d]$ are

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_u^2 \quad \text{and} \quad \overline{\tau}_u^2 = \sum_{v : v \cap u \neq \varnothing} \sigma_u^2,$$

respectively. Normalized versions $\underline{\tau}_u^2/\sigma^2$ and $\overline{\tau}_u^2/\sigma^2$ are widely used in global sensitivity analysis. See [22] for context and an extensive bibliography. We will use the identity

$$\overline{\tau}_u^2 = \mathbb{E}(\mathrm{Var}(f(\boldsymbol{x}) \,|\, \boldsymbol{x}_{-u})). \tag{2.3}$$

Our greatest need is for $\overline{\tau}_{\{j\}}^2$, which we abbreviate to $\overline{\tau}_j^2$.

An elementary result from [15] is that

$$\nu(f) = \frac{1}{\sigma^2} \sum_{j=1}^{d} \overline{\tau}_j^2. \tag{2.4}$$

Jansen [12] has the useful identity

$$\overline{\tau}_j^2 = \frac{1}{2} \mathbb{E}\big((f(\boldsymbol{x}_{-j}{:}\boldsymbol{x}'_j) - f(\boldsymbol{x}))^2\big), \tag{2.5}$$

which allows sampling-based estimates of $\overline{\tau}_j^2$. This identity underlies our theoretical analysis along with the more familiar identity $\sigma^2 = \mathbb{E}((f(\boldsymbol{x}) - f(\boldsymbol{x}'))^2)/2$.

3. Generalized multiquadrics.

These functions take the form $(a + \vartheta \|\boldsymbol{x}\|^2)^p$. We can rewrite them as $(a + \|\boldsymbol{x}\|^2)^p$ after replacing a by a/ϑ and rescaling the coefficients β_i by a factor of ϑ^p. The cases that interest us most have nonzero $p < 1$ because those get the most use. The case $p = 1$ is obviously of mean dimension one. We will include cases with $a = 0$ and $p < 0$. As [6] notes, these are not well suited to interpolation due to their singularities, but they are of interest as generalized Coulomb potentials.

3.1. Parametrization of generalized multiquadrics.

A radial basis function uses the inputs $\boldsymbol{x}$ only through $\sum_{j=1}^d (x_j - c_j)^2$. Here x_j is the jth component of $\boldsymbol{x}$ and c_j is the jth component of a center point such as $\boldsymbol{x}_i$. We let

$$z_j = \begin{cases} a + (x_1 - c_1)^2, & j = 1, \\ (x_j - c_j)^2 & \text{otherwise}, \end{cases}$$

and then we study mean dimension in terms of random $\boldsymbol{z} = (z_1, \dots, z_d)$. We have folded any $a > 0$ into z_1 to remove a from further expressions. The case of $a = 0$ is the most challenging because it can produce a singularity at $\boldsymbol{z} = 0$ that we don't have to consider when $a > 0$.

The RBFs we study are functions of $\boldsymbol{z}$, where $\boldsymbol{z}$ is defined componentwise from $\boldsymbol{x}$. If we use f^* to represent the RBF in terms of $\boldsymbol{z}$, then we find the same mean and variance components and mean dimension for f^* as we get for f. For simplicity, we will use f also for the RBF written in terms of $\boldsymbol{z} \in [0, \infty)^d$. We retain the distinction between $\boldsymbol{x}$ and $\boldsymbol{z}$ because that makes our input assumptions easier to interpret. We will study the mean dimension of $(\sum_{j=1}^d z_j)^p$ for independent not necessarily identically distributed random $z_j \geqslant 0$ and nonzero $p \leqslant 1$.

3.2. Assumptions on $\boldsymbol{z}$. We study a collection of independent nonnegative random variables z_j for $j=1,\ldots,d$. We write $\mu_j=\mathbb{E}(z_j)$ and $\sigma_j^2=\mathrm{Var}(z_j)$. Some higher moments are denoted by $\mu_j^{(k)}=\mathbb{E}((z_j-\mu_j)^k)$ for positive integers k. For certain sums we write

$$z_{1:d}:=\sum_{j=1}^d z_j,\quad \mu_{1:d}:=\sum_{j=1}^d \mu_j,\quad \sigma^2_{1:d}:=\sum_{j=1}^d \sigma_j^2,\quad \text{and}\quad \mu^{(k)}_{1:d}:=\sum_{j=1}^d \mu_j^{(k)}.$$

We want to bound the mean dimension of $(z_{1:d})^p$. It is convenient to define

$$f(\boldsymbol{z})=\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^p. \tag{3.1}$$

This function of $\boldsymbol{z}$ has the same mean dimension as if we had not scaled the input by $\mu_{1:d}$, and it has the same mean dimension as the original function of $\boldsymbol{x}$.

We will use a bounded mean assumption

$$0<\underline{\mu}\leqslant\mu_j\leqslant\overline{\mu}<\infty,\quad 1\leqslant j\leqslant d, \tag{3.2}$$

and a bounded variance assumption

$$0<\sigma_j^2\leqslant\overline{\sigma}^2<\infty,\quad 1\leqslant j\leqslant d, \tag{3.3}$$

and for some $\alpha>0$ a negative moment assumption

$$\mathbb{E}(z_j^{-\alpha})\leqslant M_\alpha<\infty,\quad 1\leqslant j\leqslant d. \tag{3.4}$$

For some of our sharper result we will require that

$$|\mathbb{E}((z_j-\mu_j)^k)|\leqslant\lambda\quad\text{for } 2\leqslant k\leqslant 6 \tag{3.5}$$

hold for some $\lambda<\infty$.

We do not lose much generality requiring $\sigma_j^2>0$ because $\sigma_j^2=0$ implies that z_j is redundant. Our main results will still hold if some $\sigma_j^2=0$ so long as $\sigma^2_{1:d}>0$.

One very important case has $x_j\sim\mathcal{N}(0,1)$. Then z_j has a noncentral chi-squared distribution with one degree of freedom and noncentrality parameter c_j^2. This distribution satisfies the bounded mean and variance assumptions, provided that c_j^2 is bounded. It satisfies the negative moment assumption if $\alpha<2$ because the central $\chi^2_{(1)}$ satisfies that condition and the noncentral distribution is a mixture of central χ^2 distributions with odd numbers of degrees of freedom. For the case with finite $a>0$, z_1 satisfies (3.2), (3.3), and (3.4) if $(x_1-c_1)^2$ does.

3.3. Main result for generalized multiquadrics. Here we present our main result for mean dimension of generalized multiquadric RBFs. We give moment conditions on z_j under which

$$\nu(f)\leqslant 1+\frac{(p-1)^2}{2}\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+O\left(\frac{1}{d^2}\right). \tag{3.6}$$

Most of the proof details are in Appendix B. We assume throughout that independent $z_j\geqslant 0$ satisfy the bounded mean condition (3.2), the bounded variance condition (3.3), the negative moment condition (3.4), and the sixth moment condition (3.5). The results in Appendix B depend on some results in Appendix A about positive and negative moments of sums of z_j.

THEOREM 3.1. *Let $z_j \geqslant 0$ be independent random variables satisfying assumptions* (3.2) *through* (3.5). *Let $f(\boldsymbol{z}) = (z_{1:d}/\mu_{1:d})^p$ for nonzero $p < 1$. Then*

$$\nu(f) \leqslant 1 + \frac{(p-1)^2}{2} \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + O\left(\frac{1}{d^2}\right)$$

as $d \to \infty$.

Proof. The mean dimension equals $\sum_{j=1}^d \overline{\tau}_j^2/\sigma^2$, and we use asymptotic expressions for the numerator and denominator of this ratio. For the numerator, Proposition B.2 from Appendix B shows that

$$\sum_{j=1}^d \overline{\tau}_j^2 \leqslant \frac{p^2\sigma_{1:d}^2}{(\mu_{1:d})^2}\left(1 + (p-1)(2p-3)\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + (p-1)\frac{\mu_{1:d}^{(3)}}{\mu_{1:d}\sigma_{1:d}^2} + O(d^{-2})\right).$$

For the denominator, Proposition A.5 in Appendix A shows that $\mathbb{E}((z_{1:p}/\mu_{1:p})^p)$ equals

$$1 + \left(\frac{(p)_2}{2!} \cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}\right) + \left(\frac{(p)_3}{3!} \cdot \frac{\mu_{1:d}^{(3)}}{(\mu_{1:d})^3} + \frac{(p)_4}{4!} \cdot \frac{3(\sigma_{1:d}^2)^2}{(\mu_{1:d})^4}\right) + O(d^{-3})$$

for $p < 6$ as $d \to \infty$. Here $(p)_k = p(p-1)\cdots(p-k+1)$. Using this result for first and second moments of $(z_{1:d}/\mu_{1:d})^p$ for $p < 1$, Corollary B.1 shows that $\sigma^2 = \mathrm{Var}\left((z_{1:d}/\mu_{1:d})^p\right)$ equals

$$\frac{p^2\sigma_{1:d}^2}{(\mu_{1:d})^2}\left(1 + (p-1) \cdot \frac{\mu_{1:d}^{(3)}}{\sigma_{1:d}^2\mu_{1:d}} + \frac{1}{2}(p-1)(3p-5) \cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + O(d^{-2})\right).$$

Combining the upper bound from Proposition B.2 with the asymptotic variance in Corollary B.1 we get

$$\begin{aligned}\nu(f) &\leqslant \frac{1 + (p-1)(2p-3)\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + (p-1)\frac{\mu_{1:d}^{(3)}}{\mu_{1:d}\sigma_{1:d}^2} + O(d^{-2})}{1 + \frac{1}{2}(p-1)(3p-5)\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + (p-1)\frac{\mu_{1:d}^{(3)}}{\sigma_{1:d}^2\mu_{1:d}} + O(d^{-2})} \\ &= 1 + \frac{(p-1)^2}{2}\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2} + O\left(\frac{1}{d^2}\right). \end{aligned}$$ ☐

Remark 3.2. Under the assumptions we have made, $\sigma_{1:d}^2/(\mu_{1:d})^2 = \Theta(1/d)$.

Remark 3.3. We notice that the bound in Theorem 3.1 can be evaluated for the degenerate case $p = 0$. We conjecture that this might be the rate for $f(\boldsymbol{z}) = \log(z_{1:d})$. Our reasoning is that the mean dimension of $(z_{1:d})^p$ is the same as that of $((z_{1:d})^p - 1)/p$, which approaches $\log(z_{1:d})$ as $p \to 0$.

3.4. Weaker conditions. Theorem 3.1 relies on a sixth moment assumption in order to get an expression for the coefficient of $1/d$ in the bound on $\nu(f)$. This section shows that the mean dimension of generalized multiquadric RBFs tends to 1 as $d \to \infty$ under very mild moment conditions: means and variances of z_j bounded uniformly from 0 and ∞ and a finite negative moment. Under these conditions, Lemmas C.4 and C.5 in Appendix C show that

$$\limsup_{d\to\infty} \frac{\sum_{j=1}^d \overline{\tau}_j^2}{p^2 \cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}} \leqslant 1 \quad \text{and} \quad \liminf_{d\to\infty} \frac{\mathrm{Var}(f(\boldsymbol{z}))}{p^2 \cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}} \geqslant 1,$$

respectively.

THEOREM 3.4. *Let independent random $z_j \geqslant 0$ for $j = 1, \dots, d$ satisfy assumptions* (3.2) *through* (3.4) *and let $f(\boldsymbol{z}) = (z_{1:d}/\mu_{1:d})^p$ for nonzero $p < 1$. Then the mean dimension of f satisfies*

$$\lim_{d\to\infty} \nu(f) = 1.$$

Proof. By definition $\nu(f) \geqslant 1$. Next

$$\lim_{d\to\infty} \nu(f) = \lim_{d\to\infty} \frac{\sum_{j=1}^d \overline{\tau}_j^2}{\mathrm{Var}(f(\boldsymbol{z}))} \leqslant \frac{\limsup_{d\to\infty} \sum_{j=1}^d \overline{\tau}_j^2/[p^2\sigma_{1:d}^2/(\mu_{1:d})^2]}{\liminf_{d\to\infty} \mathrm{Var}(f(\boldsymbol{z}))/[p^2\sigma_{1:d}^2/(\mu_{1:d})^2]},$$

which equals 1 by Lemmas C.4 and C.5. □

4. The Gaussian RBF.

Here we show how the Gaussian RBF is not limited to low mean dimension as $d \to \infty$ because the scale parameter can be chosen to control mean dimension. This makes it very different from multiquadric and related RBFs where the asymptotic mean dimension must converge to one. The Gaussian RBF is special in that it can be parameterized as a product:

$$f(\boldsymbol{x}) = \prod_{j=1}^d \exp(-(x_j - c_j)^2/\vartheta^2)$$

for $\vartheta > 0$. We have changed the scaling from $(x_j - c_j)^2\vartheta^2$ to $(x_j - c_j)/\vartheta^2$ to give ϑ^2 an interpretation as twice the variance of a Gaussian random variable. We assume that x_j are independent with a continuous distribution. Without loss of generality we assume that x_j have mean zero.

We use three propositions. The product form of the Gaussian RBF allows for a simplification of the mean dimension. Proposition 4.1 below applies to general products, not just Gaussian RBFs.

PROPOSITION 4.1. *Let $f(\boldsymbol{x}) = \prod_{j=1}^d g_j(x_j)$, where x_j are independent random variables with $\mathrm{Var}(g_j(x_j)) < \infty$ and $\min_{1\leqslant j\leqslant d} \mathrm{Var}(g_j(x_j)) > 0$. Then f has mean dimension*

$$\nu(f) = \frac{\sum_{j=1}^d \rho_j}{1 - \prod_{j=1}^d (1 - \rho_j)}, \tag{4.1}$$

where

$$\rho_j = \frac{\mathrm{Var}(z_j)}{\mathbb{E}(z_j^2)} \in [0, 1].$$

Proof. This is Proposition 1 of [18]. □

PROPOSITION 4.2. *Under the conditions of Proposition* 4.1*,*

$$\frac{\partial}{\partial \rho_k} \nu(f) \geqslant 0.$$

Proof. The result is trivial for $d = 1$, so we assume that $d \geqslant 2$. The partial derivative is

$$\frac{1 - \prod_{j\neq k}(1 - \rho_j)[1 + \sum_{j\neq k} \rho_j]}{[1 - \prod_{j=1}^d (1 - \rho_j)]^2}.$$

The denominator above is positive. Letting $\bar\rho_{-k}=(d-1)^{-1}\sum_{j\ne k}\rho_j$, the numerator is at least

$$1-(1-\bar\rho_{-k})^{d-1}(1+(d-1)\bar\rho_{-k}) \tag{4.2}$$

because the geometric mean of $1-\rho_j$ for $j\ne k$ is no larger than their arithmetic mean. The expression in (4.2) is increasing in $\bar\rho_{-k}$ over $\bar\rho_{-k}\in(0,1)$ and it equals zero for $\bar\rho_{-k}=0$. ∎

Proposition 4.3. *Let x be a random variable with probability density function h on $\mathbb{R}$. Assume that $h(x)\leqslant M$ and that $c\in\mathbb{R}$ belongs to an interval I of length at least $\ell>0$ on which $h(x)\geqslant h_0>0$. Then*

$$\lim_{\vartheta\to\infty}\frac{\mathbb{E}\left(e^{-2(x-c)^2/\vartheta^2}\right)}{\mathbb{E}\left(e^{-(x-c)^2/\vartheta^2}\right)^2}=1$$

and

$$\lim_{\vartheta\downarrow 0}\frac{\mathbb{E}\left(e^{-2(x-c)^2/\vartheta^2}\right)}{\mathbb{E}\left(e^{-(x-c)^2/\vartheta^2}\right)^2}=\infty. \tag{4.3}$$

Remark 4.4. The first limit has a mean square over a squared mean approaching 1. Then the variance becomes negligible, so $\rho\to 0$ in the above notation. The second limit has a mean square divided by a squared mean approaching infinity, so $\rho\to 1$ in the above notation.

Proof. The first claim is easy as both numerator and denominator approach 1 as $\vartheta\to\infty$. For the second claim

$$\mathbb{E}(e^{-(x-c)^2/\vartheta^2})=\int_{-\infty}^{\infty}e^{-(x-c)^2/\vartheta^2}h(x)\,\mathrm{d}x\leqslant M\sqrt{\pi}\vartheta.$$

We let $I=(a,b)$ with $b-a=\ell$. Next by change of variable

$$\begin{aligned}\mathbb{E}(e^{-2(x-c)^2/\vartheta^2})&=\frac{\vartheta}{2}\int_{-\infty}^{\infty}e^{-y^2/2}h(c+\vartheta y/2)\,\mathrm{d}y\\&\geqslant\frac{\vartheta h_0}{2}\int_{-\infty}^{\infty}e^{-y^2/2}\mathbf{1}_{c+\vartheta y/2\in I}\,\mathrm{d}y\\&=\frac{\vartheta h_0}{2}\int_{2(a-c)/\vartheta}^{2(b-c)/\vartheta}e^{-y^2/2}\,\mathrm{d}y\\&=\vartheta h_0\sqrt{\frac{\pi}{2}}\left(\Phi\left(\frac{2(b-c)}{\vartheta}\right)-\Phi\left(\frac{2(a-c)}{\vartheta}\right)\right)\\&\geqslant\vartheta h_0\sqrt{\frac{\pi}{2}}\left(\Phi\left(\frac{2(b-a)}{\vartheta}\right)-\frac{1}{2}\right).\end{aligned}$$

For any $\ell=b-a>0$ we can choose ϑ small enough to make $\Phi(2\ell/\vartheta)\geqslant 3/4$ and then $\mathbb{E}(e^{-2(x-c)^2/\vartheta^2})\geqslant\vartheta h_0\sqrt{\pi/32}$. Now the numerator in (4.3) is $\Omega(\vartheta)$, while the denominator is $O(\vartheta^2)$ both as $\vartheta\to\infty$. The result follows. ∎

In the Gaussian setting, $\rho_j>0$ and ruling out uninteresting variables with $\mathrm{Var}(x_j)=0$ we also have $\rho_j<1$. The mean dimension of f is continuous and nondecreasing in each ρ_j by Proposition 4.2. By Proposition 4.3, each $\rho_j\to 1$ as $\vartheta\to 0$, when x_j has a continuous distribution and so $\nu(f)\to d$. Conversely as $\vartheta\to\infty$, each $\rho_j\to 0$ and then $\nu(f)\to 1$. Therefore any mean dimension in $(1,d)$ can be attained at some value of ϑ.

5. Keister's function. The Keister function was used by [14] and also [2], to compare multidimensional quadrature methods. They use $\int_{\mathbb{R}^d} e^{-\|\boldsymbol{x}\|^2}\cos(\|\boldsymbol{x}\|)\,\mathrm{d}\boldsymbol{x}$ as an example of the sort of integration problem arising in atomic, nuclear, and particle physics. We make a change of variable and consider

$$f(\boldsymbol{x}) = \cos(\|\boldsymbol{x}\|/2)$$

for $\boldsymbol{x} \sim \mathcal{N}(0, I)$. This f is an RBF but not one of those commonly used for approximation. References [2] and [14] give precise values for $\mathbb{E}(f(\boldsymbol{x}))$ at certain values of d, and [11] gives a recursion for this expectation.

Keister's function has become a test function for QMC since [20]. The success of QMC on some integrands from finance could possibly be explained by the unequal importance of the variables in those integrands. Perhaps many of them were quite unimportant, leaving an integrand that depends on only a few variables. All d variables enter Keister's function symmetrically so there would need to be another explanation for QMC's successes there. The explanation is that it is dominated by its low-dimensional ANOVA components. Computations in [18] show that for $d = 25$ (the dimension considered by [20]) and $d = 80$, over 99% of the variance of the Keister function comes from variance components σ_u^2 with $|u| \leqslant 3$ making it of effective dimension 3 in the sense of [1]. Here we study the Keister function's mean dimension for $2 \leqslant d \leqslant 1000$.

By symmetry, $\overline{\tau}_1^2 = \overline{\tau}_2^2 = \cdots = \overline{\tau}_d^2$ for the Keister function and so its mean dimension is $\nu(f) = d\overline{\tau}_1^2/\sigma^2$. The variance σ^2 can easily be approximated by sampling because $\|\boldsymbol{x}\|^2 \sim \chi^2_{(d)}$. For this paper, we used a midpoint rule on $n = 2^{14} = 16{,}384$ points in $(0,1)$, transformed them to $\chi^2_{(d)}$ quantiles, took the square root to get a sample value for $\|\boldsymbol{x}\|$, and then computed the sample variance of the $\cos(\|\boldsymbol{x}\|/2)$ values.

To estimate $\overline{\tau}_1^2$, we find using the Jansen identity (2.5) that

$$\overline{\tau}_1^2 = \frac{1}{2}\mathbb{E}\big((f(z_1+z_2) - f(z_1+z_3))^2\big)$$

for $z_1 = \sum_{j=2}^d x_j^2$, $z_2 = x_1^2$, and $z_3 = {x_1'}^2$. Now $z_1 \sim \chi^2_{(d-1)}$, $z_2 \sim \chi^2_{(1)}$, and $z_3 \sim \chi^2_{(2)}$ are independent random variables. We then estimated $\overline{\tau}_1^2$ by using randomized Sobol' points in $(0,1)^3$, transforming them to the needed χ^2 values by inversion of their cumulative distribution functions and applied the Jansen formula. For this integral we used a Sobol' sequence [25] with direction numbers from [13] and a nested uniform scramble of [17] with $n = 2^{14} = 16{,}384$ points.

The above computation was replicated five times independently. With a bit of foresight, we plot the mean dimension of Keister's function in dimension d versus $\sqrt{d}$ in Figure 1. The plot shows all five replicates, but they overlap each other in the figure. The mean dimension is not monotone in d. Instead for $d \geqslant 2$, the mean dimension oscillates regularly from just over 1 to peaks that are eventually just over 2.

From Figure 1 it becomes clear what is going on. The random variable $\|\boldsymbol{x}\|^2$ has a $\chi^2_{(d)}$ distribution. For large d, this is approximately $\mathcal{N}(d, 2d)$. Then by the delta method (Taylor approximation about the mean), $\|\boldsymbol{x}\|/2$ has approximately the $\mathcal{N}(\sqrt{d}/2, 1/4)$ distribution. The central 99.9% of $\mathcal{N}(\alpha, 1/4)$ values belong to the range $\alpha \pm \Phi^{-1}(0.9995)/\sqrt{4}$ or about $\alpha \pm 1.65$. Then $\cos(\|\boldsymbol{x}\|/2)$ primarily uses the cosine function over an interval of length about 3.3, roughly half of the period 2π of the cosine function. When $\sqrt{d}/2$ is nearly an integer multiple of π, then the cosine function is being sampled predominantly in a region where it is nearly quadratic and we find that the mean dimension is close to 2. If instead $\sqrt{d}/2$ is nearly $\pi/2$ plus an integer

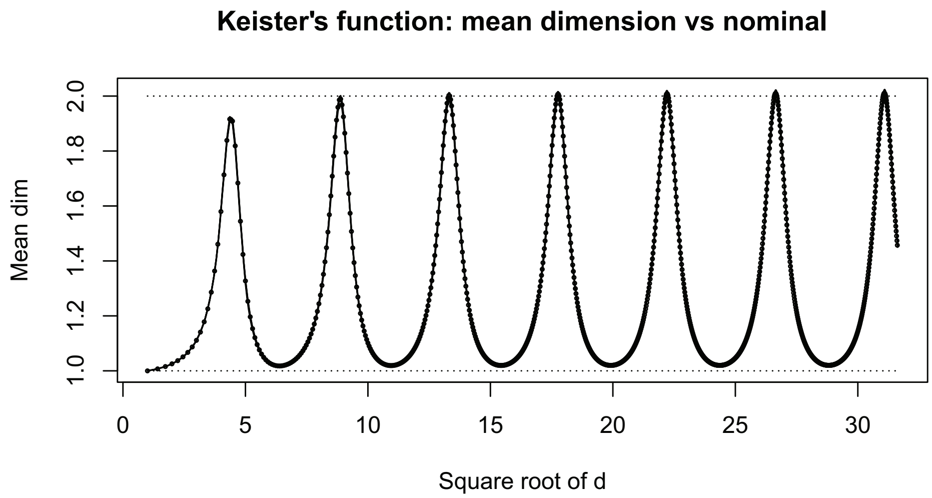

FIG. 1. *The horizontal axis has* $\sqrt{d}$ *for* $1 \leqslant d \leqslant 1000$. *The vertical axis plots five lines, each an independent randomized QMC estimate of mean dimension versus* $\sqrt{d}$. *Points mark the average of the five values. There are dotted horizontal reference lines at levels* 1 *and* 2.

multiple of π, then the cosine is being sampled over a nearly linear range and the mean dimension is close to 1.

6. Discussion. Much success in high-dimensional numerical methods comes from the target function having less complexity than we might expect given its nominal dimension. See [1] or [16, Chapter 9] or [24], among other references. In that literature, tractability results provide sets of assumptions under which there is no curse of dimensionality for integration. Reference [19] and references therein show that some weighted Hilbert spaces for which dimension-independent tractability has been established have very low effective dimension in a superposition sense (e.g., 3 or less from the $\eta = 1$ column of Table 1 in [19]). In the definition of that paper, an effective dimension of 3 implies that there are only negligible contributions to f from variance components σ_u^2 with $|u| > 3$.

The effective dimension is hard to measure empirically. The mean dimension provides a measure that can easily be estimated through Sobol' indices and Jansen's identity, while also quantifying the extent to which a function is dominated by its low-dimensional ANOVA effects. For instance, a mean dimension below 1.01 implies an effective dimension of one in the superposition sense using the definition in [1].

We have given conditions under which generalized multiquadric RBFs have a mean dimension of $1 + O(1/d)$. This makes it very difficult for them to be used in approximations of functions involving even two or three factor interactions. We believe that this explains why Gaussian RBFs and ridge functions are more commonly used in machine learning. Gaussian RBFs can attain almost any mean dimension if their parameters are well chosen, so they do not have the limitation that multiquadric RBFs do. Ridge functions $\phi(\theta^{\mathsf{T}}\boldsymbol{x})$ with Lipschitz continuous $\phi(\cdot)$ and a unit vector θ attain an $O(1)$ mean dimension automatically, under Gaussian sampling [10], but are not limited to $1 + O(1/d)$.

Appendix A. Moments of some sums. Here we provide some moment formulas needed later. We begin by working out some expressions for central moments of sums of our random variables. For integers $k \geqslant 2$ we use $\mu_j^{(k)} = \mathbb{E}((z_j - \mu_j)^k)$ to denote the kth central moments and

$$\mu_{1:d}^{(k)} \equiv \sum_{j=1}^{d} \mu_j^{(k)}.$$

For $k = 1$ we use $\mu_j = \mathbb{E}(z_j)$ and $\mu_{1:d} = \sum_{j=1}^d \mu_j$, and for $k = 2$ we use σ_j^2 and $\sigma_{1:d}^2 = \sum_{j=1}^d \sigma_j^2$. The following theorem simplifies some derivations.

Theorem A.1. *For $d \geqslant 1$, let $x_1, \dots, x_d$ be independent random variables with $\mathbb{E}(|x_j|^k) < \infty$ for $j = 1, \dots, d$ and some integer $k \geqslant 2$. Set $x_{1:d} = \sum_{j=1}^d x_j$. If $\mathbb{E}(x_j) = 0$ for $j = 1, \dots, d$, then*

$$\mathbb{E}[|x_{1:d}|^k] \leqslant c(p) d^{k/2-1} \sum_{j=1}^{d} \mathbb{E}[|x_j|^k]$$

for some $c(k) < \infty$.

Proof. This is in [21]. ∎

Proposition A.2. *For $j = 1, \dots, d$, let z_j be independent random variables with means μ_j and variances σ_j^2. Let $z_1, \dots, z_d$ satisfy the sixth moment bounds* (3.5) *for some $\lambda < \infty$. Then*

$$\mathbb{E}\left[(z_{1:d}/\mu_{1:d} - 1)^2\right] = \sigma_{1:d}^2/(\mu_{1:d})^2, \tag{A.1}$$
$$\mathbb{E}\left[(z_{1:d}/\mu_{1:d} - 1)^3\right] = \mu_{1:d}^{(3)}/(\mu_{1:d})^3, \tag{A.2}$$
$$\mathbb{E}\left[(z_{1:d}/\mu_{1:d} - 1)^4\right] = 3(\sigma_{1:d}^2)^2/(\mu_{1:d})^4 + O(d^{-3}), \tag{A.3}$$
$$\mathbb{E}\left[(z_{1:d}/\mu_{1:d} - 1)^5\right] = O(d^{-3}), \quad \textit{and} \tag{A.4}$$
$$\mathbb{E}\left[(z_{1:d}/\mu_{1:d} - 1)^6\right] = O(d^{-3}). \tag{A.5}$$

Proof. The results for exponents $k = 2, 3, 4$ are elementary. Theorem A.1 (Petrov) yields

$$\mathbb{E}\left[\left(\frac{z_{1:d}}{\mu_{1:d}} - 1\right)^k\right] \leqslant c(k) d^{k/2-1} \frac{\mu_{1:d}^{(k)}}{(\mu_{1:d})^k} \leqslant \frac{\lambda c(k) d^{k/2}}{(d\underline{\mu})^k} = O(d^{-k/2}).$$

Taking $k = 6$ above provides the result (A.5) for the sixth moment.

The case of $k = 5$ remains. Petrov's theorem would only give us $O(d^{-5/2})$. The difference is that Petrov's theorem is about an absolute moment, and our requirement is for just for an expected fifth power. For $k = 5$ we get

$$\mathbb{E}\left[(z_{1:d} - \mu_{1:d})^5\right] = 10 \sum_{\substack{j_1, j_2 \in [d] \\ \text{distinct}}} \mathbb{E}\left[(z_{j_1} - \mu_{j_1})^3 (z_{j_2} - \mu_{j_2})^2\right] + \sum_{j=1}^{d} \mathbb{E}\left[(z_j - \mu_j)^5\right]$$
$$= 10\mu_{1:d}^{(3)} \sigma_{1:d}^2 - 10 \sum_{j=1}^{d} \mu_j^{(3)} \sigma_j^2 + \mu_{1:d}^{(5)}, \tag{A.6}$$

where the factor 10 comes from there being 10 partitions like $j_1 = j_2 \neq j_3 = j_4 = j_5$. The quantity in (A.6) is then $O(d^2)$ establishing (A.4).

The implied constant in the fourth degree term can be taken as 3λ. The implied constant in the fifth degree term can be taken as $10\lambda^2/\underline{\mu}^5 + \epsilon$ for any $\epsilon > 0$. The implied constant in the sixth degree term can be taken as $\lambda c(6)/\underline{\mu}^6$. ∎

For the next result, we prove an upper bound on negative moments. We use the quantity

$$\beta := \frac{1}{\underline{\mu}\sqrt[\alpha]{M_\alpha}}, \tag{A.7}$$

recalling that $\mathbb{E}(z_j^{-\alpha}) \leqslant M_\alpha < \infty$ from (3.4). This β is useful in providing constant upper bounds for negative moments.

Proposition A.3. *Let $z_1, z_2, \ldots, z_d$ be independent nonnegative random variables that satisfy the mean bounds* (3.2) *and the negative moment assumption* (3.4) *for some $\alpha > 0$ and $M_\alpha < \infty$, and choose an exponent $p < 0$. Then $\mathbb{E}((z_{1:d}/\mu_{1:d})^p) \leqslant \beta^p$ for all $d \geqslant -p/\alpha$.*

Proof. For $d \geqslant -p/\alpha$, we find that $\phi(x) = x^{-\alpha d/p}$ is a convex function. Then using the mean lower bound (3.2), the arithmetic-geometric mean identity, and Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^p\right] &= \left(\frac{d}{\mu_{1:d}}\right)^p \cdot \mathbb{E}\left[\left(\frac{z_{1:d}}{d}\right)^p\right] \\
&\leqslant \left(\frac{d}{d\underline{\mu}}\right)^p \cdot \mathbb{E}\left[z_1^{p/d} z_2^{p/d} \cdots z_d^{p/d}\right] \\
&\leqslant \underline{\mu}^{-p} \cdot \mathbb{E}\left[\left(z_1^{p/d} z_2^{p/d} \cdots z_d^{p/d}\right)^{-\alpha d/p}\right]^{-p/\alpha d} \\
&\leqslant \underline{\mu}^{-p} \cdot \mathbb{E}\left[z_1^{-\alpha} z_2^{-\alpha} \cdots z_d^{-\alpha}\right]^{-p/\alpha d} \\
&\leqslant \underline{\mu}^{-p} \cdot (M_\alpha^d)^{-p/\alpha d} \\
&\leqslant \beta^p.
\end{aligned}$$
∎

Remark A.4. This result shows that any negative moment of the sample average is $O(1)$ as $d \to \infty$, under the given conditions.

This next result is used to control the Lagrange error term in some Taylor approximations.

Proposition A.5. *Let independent $z_j \geqslant 0$ satisfy the mean bounds* (3.2) *as well as condition* (3.5) *on their first six central moments and the negative moment condition* (3.4)*. Then for $p < 6$, $\mathbb{E}((z_{1:d}/\mu_{1:d})^p)$ equals*

$$1 + \left(\frac{(p)_2}{2!} \cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}\right) + \left(\frac{(p)_3}{3!} \cdot \frac{\mu_{1:d}^{(3)}}{(\mu_{1:d})^3} + \frac{(p)_4}{4!} \cdot \frac{3(\sigma_{1:d}^2)^2}{(\mu_{1:d})^4}\right) + O(d^{-3})$$

as $d \to \infty$.

Proof. Using a fifth order Taylor expansion, we get

$$\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^p = \sum_{k=0}^{5} \frac{(p)_k}{k!} \left(\frac{z_{1:d}}{\mu_{1:d}} - 1\right)^k + \frac{(p)_6}{6!} \left(\frac{z_{1:d}}{\mu_{1:d}} - 1\right)^6 \cdot \theta^{p-6}$$

for some θ between 1 and $z_{1:d}/\mu_{1:d}$. Using the results in Proposition A.2 we find that the expected value of the sum for $0 \leqslant k \leqslant 5$ is

$$1 + \left(\frac{(p)_2}{2!} \cdot \frac{\sigma^2_{1:d}}{(\mu_{1:d})^2} \right) + \left(\frac{(p)_3}{3!} \cdot \frac{\mu^{(3)}_{1:d}}{(\mu_{1:d})^3} + \frac{(p)_4}{4!} \cdot \frac{3(\sigma^2_{1:d})^2}{(\mu_{1:d})^4} \right) + O(d^{-3}).$$

It remains to show that the remainder term with $k = 6$ is $O(d^{-3})$.

We can assume that $d \geqslant 2$ and then define

$$A_d = \frac{\sum_{j\in[d],\text{even}} z_j}{\sum_{j\in[d],\text{even}} \mu_j}, \quad B_d = \frac{\sum_{j\in[d],\text{odd}} z_j}{\sum_{j\in[d],\text{odd}} \mu_j}, \quad \text{and} \quad t_d = \frac{\sum_{j\in[d],\text{even}} \mu_j}{\mu_{1:d}}.$$

Here, A_d and B_d are independent random variables, $t_d \in (0, 1)$ is nonrandom, and $z_{1:d}/\mu_{1:d} = t_d A_d + (1 - t_d)B_d$. Because $\phi(x) = (x - 1)^6$ is a convex function,

$$\left(\frac{z_{1:d}}{\mu_{1:d}} - 1 \right)^6 \leqslant t_d(A_d - 1)^6 + (1 - t_d)(B_d - 1)^6 \leqslant (A_d - 1)^6 + (B_d - 1)^6.$$

Next, θ is between 1 and $z_{1:d}/\mu_{1:d}$, and so $\theta^{p-6} \leqslant 1 + (z_{1:d}/\mu_{1:d})^{p-6}$. Since $t_d A_d$ and $(1 - t_d)B_d$ are both lower bounds for $z_{1:d}/\mu_{1:d}$, we can take either $1 + t_d^{p-6} A_d^{p-6}$ or $1 + (1 - t_d)^{p-6} B_d^{p-6}$ as an upper bound for θ^{p-6}.

Because the exponent $p - 6$ is negative we will need to bound t_d away from zero below. Using upper and lower bounds on μ_j we know that

$$t_d \geqslant \frac{((d/2) - 1)\underline{\mu}}{d\overline{\mu}}.$$

That lower bound is strictly positive for $d = 3$ and it increases with d, so $t_d^{p-6} = O(1)$. A similar argument shows that $(1 - t_d)^{p-6} = O(1)$ too, and so $\max(t_d^{p-6}, (1 - t_d)^{p-6}) \leqslant C$ for some $C < \infty$ and all $d \geqslant 3$. Therefore, we find that for d large enough

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{z_{1:d}}{\mu_{1:d}} - 1 \right)^6 \theta^{p-6} \right] \\
&\quad \leqslant \mathbb{E}[((A_d - 1)^6 + (B_d - 1)^6)\theta^{p-6}] \\
&\quad \leqslant \mathbb{E}[(A_d - 1)^6 \cdot (1 + t_d^{p-6} B_d^{p-6}) + (B_d - 1)^6 \cdot (1 + (1 - t_d)^{p-6} A_d^{p-6})] \\
&\quad \leqslant \mathbb{E}[(A_d - 1)^6] \cdot (1 + C \cdot \mathbb{E}[B_d^{p-6}]) + \mathbb{E}[(B_d - 1)^6] \cdot (1 + C \cdot \mathbb{E}[A_d^{p-6}]).
\end{aligned}$$

Now both $\mathbb{E}((A_d - 1)^6)$ and $\mathbb{E}((B_d - 1)^6))$ are $O(d^{-3})$ by (A.5) of Proposition A.2 and $\max(\mathbb{E}[A_d^{p-6}], \mathbb{E}[B_d^{p-6}]) \leqslant \beta^{p-6} = O(1)$ by Proposition A.3. We also note that $(z_{1:d}/\mu_{1:d} - 1)^6 \theta^{p-6}$ is nonnegative, so the expectation is bounded below by zero. Therefore, $\mathbb{E}((z_{1:d}/\mu_{1:d} - 1)^6 \theta^{p-6}) = O(d^{-3})$, as required. ∎

Remark A.6. The implied constant in the $O(d^{-3})$ error term depends only on the constants in bounds (3.2), (3.4), and (3.5).

Appendix B. Convergence rates for multiquadrics. In this section we have the main background results to support our finding that $\nu(f) = 1 + O(1/d)$ for generalized multiquadric RBFs under moment conditions.

For the next result we use falling factorial notation $(p)_k = p(p - 1) \cdots (p - k + 1)$, where p need not be an integer and $k \geqslant 0$ is an integer.

COROLLARY B.1. *Suppose $p<1$ and that independent random variables $z_j \geqslant 0$ satisfy assumptions* (3.2) *through* (3.5). *Then* $\mathrm{Var}((z_{1:d}/\mu_{1:d})^p)$ *is*

$$\frac{p^2\sigma^2_{1:d}}{(\mu_{1:d})^2}\left(1+(p-1)\cdot\frac{\mu^{(3)}_{1:d}}{\sigma^2_{1:d}\mu_{1:d}}+\frac{1}{2}(p-1)(3p-5)\cdot\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+O(d^{-2})\right)$$

as $d\to\infty$.

Proof. Because $p<1$ we have both $p<6$ and $2p<6$. So we can use Proposition A.5 with exponents p and $2p$ to write $\mathrm{Var}((z_{1:d}/\mu_{1:d})^p)$ as

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^{2p}\right]-\mathbb{E}\left[\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^{p}\right]^2\\
&\quad=\left(1+\frac{(2p)_2}{2!}\cdot\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+\frac{(2p)_3}{3!}\cdot\frac{\mu^{(3)}_{1:d}}{(\mu_{1:d})^3}+\frac{(2p)_4}{4!}\cdot\frac{3(\sigma^2_{1:d})^2}{(\mu_{1:d})^4}+O(d^{-3})\right)\\
&\qquad-\left(1+\frac{(p)_2}{2!}\cdot\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+\frac{(p)_3}{3!}\cdot\frac{\mu^{(3)}_{1:d}}{(\mu_{1:d})^3}+\frac{(p)_4}{4!}\cdot\frac{3(\sigma^2_{1:d})^2}{(\mu_{1:d})^4}+O(d^{-3})\right)^2\\
&\quad=\frac{p^2\sigma^2_{1:d}}{(\mu_{1:d})^2}\left(1+(p-1)\cdot\frac{\mu^{(3)}_{1:d}}{\sigma^2_{1:d}\mu_{1:d}}+\frac{1}{2}(p-1)(3p-5)\cdot\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+O(d^{-2})\right)
\end{aligned}$$

after some algebra. ∎

PROPOSITION B.2. *Let independent random variables $z_j \geqslant 0$ satisfy assumptions* (3.2) *through* (3.5). *Then for* $p<1$

$$\sum_{j=1}^d \underline{\tau}_j^2 \leqslant \frac{p^2\sigma^2_{1:d}}{(\mu_{1:d})^2}\left(1+(p-1)(2p-3)\frac{\sigma^2_{1:d}}{(\mu_{1:d})^2}+(p-1)\frac{\mu^{(3)}_{1:d}}{\mu_{1:d}\sigma^2_{1:d}}+O(d^{-2})\right)$$

as $d\to\infty$.

Proof. For each $j\in[d]$ we form a Taylor expansion of $(z_{1:d}/\mu_{1:d})^p$ in powers of $z_j-\mu_j$ as follows:

(B.1)
$$\begin{aligned}
&S_j^p+\frac{p}{\mu_{1:d}}S_j^{p-1}(z_j-\mu_j)+\frac{(p)_2}{2(\mu_{1:d})^2}S_j^{p-2}(z_j-\mu_j)^2+\frac{(p)_3}{6(\mu_{1:d})^3}(S'_j)^{p-3}(z_j-\mu_j)^3\\
&\quad=: T_0+T_1+T_2+T_3,
\end{aligned}$$

where

$$S_j=\frac{(z_{1:d}-z_j)+\mu_j}{\mu_{1:d}}\quad\text{and}\quad S'_j=\frac{(z_{1:d}-z_j)+\theta}{\mu_{1:d}}$$

for some θ between μ_j and z_j.

Now $\underline{\tau}_j^2=\mathbb{E}(\mathrm{Var}((z_{1:d}/\mu_{1:d})^p\mid \boldsymbol{x}_{-j})$, so we begin by bounding the conditional variances of the terms T_k defined at (B.1). Because S_j is a function of $\boldsymbol{z}_{-j}$, $\mathrm{Var}(T_0\mid \boldsymbol{x}_{-j})=\mathrm{Var}(S_j\mid\boldsymbol{z}_{-j})=0$. Similarly

$$\mathrm{Var}(T_1\mid\boldsymbol{z}_{-j})=\frac{p^2}{(\mu_{1:d})^2}S_j^{2p-2}\sigma_j^2.$$

Next, noting that $\mathrm{Var}((z_j-\mu_j)^2\,|\,\boldsymbol{z}_{-j})=\mathrm{Var}((z_j-\mu_j)^2)\leqslant\mu_j^{(4)}\leqslant\lambda$,

$$\mathrm{Var}(T_2\,|\,\boldsymbol{x}_{-j})=\frac{(p)_2^2}{4(\mu_{1:d})^4}S_j^{2p-2}\mathrm{Var}((z_j-\mu_j)^2\,|\,\boldsymbol{z}_{-j})\leqslant\frac{(p)_2^2}{4(\mu_{1:d})^4}S_j^{2p-2}\lambda.$$

Turning to the one term with S_j',

$$\begin{aligned}\mathrm{Var}(T_3\,|\,\boldsymbol{z}_{-j})&\leqslant\frac{(p)_3^2}{36(\mu_{1:d})^6}\mathbb{E}((S_j')^{2p-6}(z_j-\mu_j)^6\,|\,\boldsymbol{z}_{-j})\\&\leqslant\frac{(p)_3^2}{36(\mu_{1:d})^6}\mathbb{E}\Big(\Big(\frac{z_{1:d}-z_j}{\mu_{1:d}}\Big)^{2p-6}(z_j-\mu_j)^6\,|\,\boldsymbol{z}_{-j}\Big)\\&\leqslant\frac{(p)_3^2}{36(\mu_{1:d})^6}\Big(\frac{z_{1:d}-z_j}{\mu_{1:d}}\Big)^{2p-6}\lambda.\end{aligned}$$

With the above decomposition, we write

$$\begin{aligned}\overline{\tau}_j^2\leqslant{}&\mathbb{E}(\mathrm{Var}(T_1\,|\,\boldsymbol{z}_{-j}))+\mathbb{E}(\mathrm{Var}(T_2\,|\,\boldsymbol{z}_{-j}))+\mathbb{E}(\mathrm{Var}(T_3\,|\,\boldsymbol{z}_{-j}))\\&+2\mathbb{E}(\mathrm{Cov}(T_1,T_2\,|\,\boldsymbol{z}_{-j}))+2\mathbb{E}(\mathrm{Cov}(T_1,T_3\,|\,\boldsymbol{z}_{-j}))+2\mathbb{E}(\mathrm{Cov}(T_2,T_3\,|\,\boldsymbol{z}_{-j})).\end{aligned}$$

Proposition A.5 shows that for $q<0$ and large enough d

$$\begin{aligned}\mathbb{E}[S_j^q]=\mathbb{E}\left[\left(\frac{z_{1:d}-z_j+\mu_j}{\mu_{1:d}}\right)^q\right]&=1+\frac{q(q-1)}{2}\frac{\sigma_{1:d}^2-\sigma_j^2}{(\mu_{1:d})^2}+O\left(\frac{1}{d^2}\right)\\&=1+\frac{q(q-1)}{2}\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}+O\left(\frac{1}{d^2}\right).\end{aligned}$$

In this application of Proposition A.5, the variable z_j with variance σ_j^2 is replaced by μ_j with variance 0. That proposition does not assume strictly positive σ_j^2. Note that the implied constant within $O(1/d^2)$ depends only on moment conditions from Remark A.6 and can be bounded independently of j.

Also

$$0\leqslant\mathbb{E}\left[S_j^{p-1}\left(\frac{z_{1:d}-z_j}{\mu_{1:d}}\right)^{p-3}\right]\leqslant\mathbb{E}[S_j^{2p-4}]\left(1+\frac{\overline{\mu}}{(d-1)\underline{\mu}}\right)^{3-p}=1+O\Big(\frac{1}{d}\Big).$$

Then the expected variances are

$$\begin{aligned}\mathbb{E}(\mathrm{Var}(T_1\,|\,\boldsymbol{z}_{-j}))&=\left(\frac{p}{\mu_{1:d}}\right)^2\sigma_j^2\mathbb{E}(S_j^{2p-2})\\&\leqslant\left(\frac{p}{\mu_{1:d}}\right)^2\sigma_j^2\left(1+\frac{(2p-2)(2p-3)}{2}\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}+O\left(\frac{1}{d^2}\right)\right),\\\mathbb{E}(\mathrm{Var}(T_2\,|\,\boldsymbol{z}_{-j}))&\leqslant\left(\frac{(p)_2}{2(\mu_{1:d})^2}\right)^2\mathbb{E}(S_j^{2p-2})\lambda=O\left(\frac{1}{d^4}\right),\quad\text{and}\\\mathbb{E}(\mathrm{Var}(T_3\,|\,\boldsymbol{z}_{-j}))&\leqslant\frac{(p)_3^2}{36(\mu_{1:d})^6}\mathbb{E}\left(\left(\frac{z_{1:d}-z_j}{\mu_{1:d}-\mu_j}\right)^{2p-6}\right)\lambda=O\left(\frac{1}{d^6}\right).\end{aligned}$$

Because S_j is a function of $\boldsymbol{z}_{-j}$,

$$\begin{aligned}\mathrm{Cov}(T_1,T_2\,|\,\boldsymbol{z}_{-j})&=\frac{p(p)_2}{2(\mu_{1:d})^3}S_j^{2p-3}\mathrm{Cov}\left(z_j-\mu_j,(z_j-\mu_j)^2\right)\\&=\frac{p(p)_2}{2(\mu_{1:d})^3}S_j^{2p-3}\mu_j^{(3)},\end{aligned}$$

and so

$$\begin{aligned}\mathbb{E}(\mathrm{Cov}(T_1,T_2\,|\,\boldsymbol{z}_{-j})) &= \frac{p(p)_2}{2(\mu_{1:d})^3}\mathbb{E}\left(S_j^{2p-3}\right)\mu_j^{(3)}\\ &= \frac{p(p)_2}{2(\mu_{1:d})^3}\left(1+O\left(\frac{1}{d}\right)\right)\mu_j^{(3)}.\end{aligned}$$

Similarly

$$\begin{aligned}\mathrm{Cov}(T_1,T_3\,|\,\boldsymbol{z}_{-j}) &\leqslant \frac{p(p)_3}{6(\mu_{1:4})^4}S_j^{p-1}\Big(\frac{z_{1:d}-z_j}{\mu_{1:d}}\Big)^{p-3}\lambda \quad \text{and}\\ \mathrm{Cov}(T_2,T_3\,|\,\boldsymbol{z}_{-j}) &\leqslant \frac{p(p)_2}{12(\mu_{1:4})^5}S_j^{p-1}\Big(\frac{z_{1:d}-z_j}{\mu_{1:d}}\Big)^{p-3}\lambda.\end{aligned}$$

so that

$$\mathbb{E}[\mathrm{Cov}(T_1,T_3\,|\,\boldsymbol{z}_{-j})] = O\left(\frac{1}{d^4}\right) \quad \text{and} \quad \mathbb{E}[\mathrm{Cov}(T_2,T_3\,|\,\boldsymbol{z}_{-j})] = O\left(\frac{1}{d^5}\right).$$

Combining all of our bounds, we get

$$\begin{aligned}\overline{\tau}_j^2 \leqslant &\left(\frac{p}{\mu_{1:d}}\right)^2\sigma_j^2\left(1+(2p-1)(p-3)\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}+O\Big(\frac{1}{d^2}\Big)\right)\\ &+\frac{p(p)_2}{2(\mu_{1:d})^3}\mu_j^{(3)}+O\left(\frac{1}{d^4}\right).\end{aligned}$$

The implied constants in both $O(\cdot)$ expressions above can be chosen independently of j from Remark A.6. Then summing over j yields

$$\sum_{j=1}^d \overline{\tau}_j^2 \leqslant \frac{p^2\sigma_{1:d}^2}{(\mu_{1:d})^2}\left(1+(2p-1)(p-3)\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}+\frac{p-1}{2}\frac{\mu_{1:d}^{(3)}}{\mu_{1:d}\sigma_{1:d}^2}\right)+O\left(\frac{1}{d^3}\right). \qquad \square$$

Appendix C. Mean dimension approaching one. Here we prove the lemmas needed for Theorem 3.4. We have a subsection to prove upper bounds on Sobol' indices and another for lower bounds on the variance.

C.1. Sobol' index upper bounds. Here we find upper bounds for the Sobol' indices $\overline{\tau}_j^2$ that form the numerator of $\nu(f)$. We will need some properties of

$$T_d := \left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^p, \tag{C.1}$$

where $J \in [d]$ is a random index with

$$\Pr(J=j) = \frac{\sigma_j^2}{\sigma_{1:d}^2}, \quad 1 \leqslant j \leqslant d, \tag{C.2}$$

chosen independently of $\boldsymbol{z}$. In particular, we need to show that $\mathbb{E}(|T_d-1|) \to 0$ as $d \to \infty$.

Proposition C.1. *Let independent $z_j \geqslant 0$ satisfy the lower bound condition* (3.2) *for some $\underline{\mu} > 0$, the variance bounds* (3.3), *and the negative moment condition* (3.4) *for some $\alpha > 0$ and $M_\alpha < \infty$ for all $j = 1,\dots,d$. If the random index J satisfies* (C.2) *and is independent of $\boldsymbol{z}$, then for $d > 1$*

$$\mathbb{E}\left(\left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^{-\alpha(d-1)}\right) \leqslant \beta^{-\alpha(d-1)}e^{\alpha},$$

where β is given in (A.7).

Proof. Directly, we find that

$$\begin{aligned}\mathbb{E}\left[\left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^{-\alpha(d-1)}\right] &= \mathbb{E}\left[\left(\frac{z_{1:d}-z_J}{d-1}\right)^{-\alpha(d-1)}\right]\cdot\left(\frac{d-1}{\mu_{1:d}}\right)^{-\alpha(d-1)} \\ &\leqslant \mathbb{E}\left[\prod_{j\in[d]\setminus\{J\}} z_j^{-\alpha}\right]\cdot\left(\frac{d-1}{d\underline{\mu}}\right)^{-\alpha(d-1)} \\ &\leqslant (M_\alpha)^{d-1}\cdot\underline{\mu}^{\alpha(d-1)}\left(\frac{d}{d-1}\right)^{\alpha(d-1)} \\ &\leqslant \beta^{-\alpha(d-1)}e^{\alpha}.\end{aligned}$$

∎

LEMMA C.2. *Let independent random variables $z_j \geqslant 0$ satisfy the upper and lower bound mean conditions in* (3.2)*, the variance bounds in* (3.3)*, and the negative moment condition* (3.4) *for some $\alpha > 0$ and $M_\alpha < \infty$. Let the index J be chosen according to* (C.2) *independently of $\boldsymbol{z}$. If T_d is defined by* (C.1) *with $p < 0$, then $\mathbb{E}(|T_d - 1|) \to 0$ as $d \to \infty$.*

Proof. We will show that T_d converges to 1 in probability and that T_d is uniformly integrable for large enough d. Then the result follows by the Vitali convergence theorem.

Writing

$$\frac{z_{1:d}-z_J}{\mu_{1:d}} = \frac{z_{1:d}}{\mu_{1:d}} - \frac{z_J}{\mu_{1:d}}$$

we see that the first term converges to one in probability (by our variance assumptions), and the second term converges to zero in probability by our assumptions on μ_j. Therefore $(z_{1:d}-z_J)/\mu_{1:d}$ converges to one in probability and then, by continuity T_d, converges to one in probability as $d \to \infty$.

Now, we prove that T_d is uniformly integrable for all $d > \lceil 1 - 2p/\alpha \rceil$, so that $1 + \alpha(d-1)/p < -1$. Consider any $\epsilon > 0$, and select any value $M > \min(\beta^p, \beta^{2p}e^{\alpha}/\epsilon)$. Noting that $x \mapsto x^{-\alpha(d-1)/p}$ is a monotonically increasing function and then using Proposition C.1, we get

$$\begin{aligned}\int_M^\infty \Pr(|T_d| \geqslant z)\,\mathrm{d}z &= \int_M^\infty \Pr\left(\left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^{p} \geqslant z\right)\mathrm{d}z \\ &= \int_M^\infty \Pr\left(\left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^{-\alpha(d-1)} \geqslant z^{-\alpha(d-1)/p}\right)\mathrm{d}z \\ &\leqslant \int_M^\infty \mathbb{E}\left[\left(\frac{z_{1:d}-z_J}{\mu_{1:d}}\right)^{-\alpha(d-1)}\right] z^{\alpha(d-1)/p}\,\mathrm{d}z \\ &\leqslant \beta^{-\alpha(d-1)}e^{\alpha}\int_M^\infty z^{\alpha(d-1)/p}\,\mathrm{d}z \\ &= \beta^{-\alpha(d-1)}e^{\alpha}\cdot\frac{M^{1+\alpha(d-1)/p}}{-(1+\alpha(d-1)/p)} \\ &\leqslant \beta^{p}e^{\alpha}\cdot(M\beta^{-p})^{1+\alpha(d-1)/p}\end{aligned}$$

(as $M\ \beta^{-p} \geqslant 1$ and $1 + \alpha(d-1)/p \leqslant -1$)

$$\begin{aligned}&\leqslant \beta^{p}e^{\alpha}(M\beta^{-p})^{-1} \\ &= \beta^{2p}e^{\alpha}/M \\ &\leqslant \epsilon\end{aligned}$$

because $M \geqslant \beta^{2p}e^{\alpha}/\epsilon$.

Therefore $\int_M^\infty \Pr(T_d| \geqslant z)\,\mathrm{d}z \leqslant \epsilon$. It follows that T_d is uniformly integrable for all $d \geqslant \lceil 1-2p/\alpha \rceil$, which completes our claim. ◻

PROPOSITION C.3. *Let $h:\mathbb{R}^2 \to \mathbb{R}$ be a function where $h(\cdot, z)$ is an $M(z)$-Lipschitz function for every z. If x and z are independent random variables, then*

$$\mathbb{E}[\mathrm{Var}(h(x,z)\,|\,z)] \leqslant \mathbb{E}[M(z)^2]\cdot \mathrm{Var}(x).$$

Proof. First $\mathrm{Var}(h(x,z)\,|\,z) \leqslant M(z)^2\mathrm{Var}(x\,|\,z) = M(z)^2\mathrm{Var}(x)$ by independence of x and z. The result follows by taking the expectation over z. ◻

LEMMA C.4. *Let independent $z_j \geqslant 0$ satisfy the upper and lower mean bounds in* (3.2) *and the upper and lower variance bounds in* (3.3). *Let $f(\boldsymbol{z}) = (z_{1:d}/\mu_{1:d})^p$ for nonzero $p<1$. Then*

$$\limsup_{d\to\infty} \frac{\sum_{j=1}^d \overline{\tau}_j^2}{p^2\cdot \frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}} \leqslant 1.$$

Proof. For all $j \in [d]$,

$$0 \leqslant \frac{\partial}{\partial z_j} f(\boldsymbol{z}) = \frac{p}{\mu_{1:d}}\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^{p-1} \leqslant \frac{p}{\mu_{1:d}}\left(\frac{z_{1:d}-z_j}{\mu_{1:d}}\right)^{p-1},$$

which we can use as a conditional Lipschitz bound independent of z_j. Then using the identity (2.3) and Proposition C.3

$$\overline{\tau}_j^2 = \mathbb{E}\left(\mathrm{Var}\left(\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^p \,|\, z_{-j}\right)\right) \leqslant \frac{\sigma_j^2 p^2}{(\mu_{1:d})^2}\mathbb{E}\left(\left(\frac{z_{1:d}-z_j}{\mu_{1:d}}\right)^{2p-2}\right).$$

Now

$$\frac{1}{p^2\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}}\sum_{j=1}^d \overline{\tau}_j^2 \leqslant \sum_{j=1}^d \frac{\sigma_j^2}{\sigma_{1:d}^2}\mathbb{E}\left(\left(\frac{z_{1:d}-z_j}{\mu_{1:d}}\right)^{2p-2}\right),$$

which we recognize as $\mathbb{E}(T_d)$ defining T_d as in (C.1) but with exponent $2p-2<0$. Then Lemma C.2 finishes the proof. ◻

C.2. Variance lower bounds. In section C.1 we found an upper bound for a normalized upper bound of Sobol' indices. Here we get a lower bound for the variance of the RBFs.

We will use the following inequality. If Y_d for $d \geqslant 1$ are random variables that have finite variances and converge in distribution to a random variable Y, then

$$\liminf_{d\to\infty} \mathrm{Var}(Y_d) \geqslant \mathrm{Var}(Y). \tag{C.3}$$

LEMMA C.5. *Let independent $z_j \geqslant 0$ satisfy the upper and lower mean bounds in* (3.2) *and the upper and lower variance bounds in* (3.3). *Let $f(\boldsymbol{z}) = (z_{1:d}/\mu_{1:d})^p$ for nonzero $p<1$. Then*

$$\liminf_{d\to\infty} \mathrm{Var}(f(\boldsymbol{z}))\cdot\left(p^2\cdot\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}\right)^{-1} \geqslant 1.$$

Proof. From the mean value theorem

$$\frac{\mu_{1:d}}{\sigma_{1:d}^2}(f(\boldsymbol{z})-1)=\frac{\mu_{1:d}}{\sigma_{1:d}^2}\left[\left(\frac{z_{1:d}}{\mu_{1:d}}\right)^p-1\right]=p\theta^{p-1}\cdot\frac{z_{1:d}-\mu_{1:d}}{\sigma_{1:d}^2}$$

for some θ between 1 and $z_{1:d}/\mu_{1:d}$. That ratio converges to 1 in probability as $d\to\infty$ and so $\theta\to 1$ in probability. Then by the continuous mapping theorem, $p\theta^{p-1}$ converges to p in probability.

Next $(z_{1:d}-\mu_{1:d})/\sigma_{1:d}^2 \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ by the central limit theorem, and so, using Slutsky's theorem,

$$\frac{\mu_{1:d}}{\sigma_{1:d}^2}(f(\boldsymbol{z})-1)\xrightarrow{\mathrm{d}}\mathcal{N}(0,p^2).$$

Finally, from (C.3)

$$\liminf_{d\to\infty}\frac{\mathrm{Var}(f(\boldsymbol{z}))}{p^2\cdot\frac{\sigma_{1:d}^2}{(\mu_{1:d})^2}}=\liminf_{d\to\infty}\frac{1}{p^2}\mathrm{Var}\left(\frac{\mu_{1:d}}{\sigma_{1:d}}[f(z_{1:d})-1]\right)$$
$$\geqslant\frac{1}{p^2}\cdot\mathrm{Var}(\mathcal{N}(0,p))=1.\qquad\square$$

Acknowledgments. We thank Naofumi Hama for comments on the role of RBFs in machine learning. We also thank an anonymous reviewer for helpful comments.

REFERENCES

[1] R. E. Caflisch, W. Morokoff, and A. B. Owen, *Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension*, J. Comput. Finance, 1 (1997), pp. 27–46.

[2] S. Capstick and B. D. Keister, *Multidimensional quadrature algorithms at higher degree and/or dimension*, J. Comput. Phys., 123 (1996), pp. 267–273.

[3] D. L. Donoho, *High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality*, AMS Lecture: Mathematical Challenges of the 21st Century, 2000, pp. 1–32.

[4] B. Efron and C. Stein, *The jackknife estimate of variance*, Ann. Statist., 9 (1981), pp. 586–596.

[5] G. Fasshauer and M. J. McCourt, *Kernel-Based Approximation Methods Using Matlab*, World Scientific, Singapore, 2015.

[6] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB*, Vol. 6, World Scientific, Singapore, 2007.

[7] A. Griewank, F. Y. Kuo, H. Leövey, and I. H. Sloan, *High dimensional integration of kinks and jumps—Smoothing by preintegration*, J. Comput. Appl. Math., 344 (2018), pp. 259–274.

[8] R. L. Hardy, *Multiquadric equations of topography and other irregular surfaces*, J. Geophys. Res., 76 (1971), pp. 1905–1915.

[9] W. Hoeffding, *A class of statistics with asymptotically normal distribution*, Ann. Math. Stat., 19 (1948), pp. 293–325.

[10] C. R. Hoyt and A. B. Owen, *Mean dimension of ridge functions*, SIAM J. Numer. Anal., 58 (2020), pp. 1195–1216, https://doi.org/10.1137/19M127149X.

[11] R. Jagadeeswaran and F. J. Hickernell, *Fast automatic Bayesian cubature using lattice sampling*, Stat. Comput., 29 (2019), pp. 1215–1229.

[12] M. J. W. Jansen, *Analysis of variance designs for model output*, Comput. Phys. Commun., 117 (1999), pp. 35–43.

[13] S. Joe and F. Y. Kuo, *Constructing Sobol' sequences with better two-dimensional projections*, SIAM J. Sci. Comput., 30 (2008), pp. 2635–2654, https://doi.org/10.1137/070709359.

[14] B. D. Keister, *Multidimensional quadrature algorithms*, Comput. Phys., 10 (1996), pp. 119–122.

[15] R. Liu and A. B. Owen, *Estimating mean dimensionality of analysis of variance decompositions*, J. Amer. Stat. Assoc., 101 (2006), pp. 712–721.

[16] E. Novak and H. Wozniakowski, *Tractability of Multivariate Problems* II*: Standard Information for Functionals*, European Mathematical Society, Zurich, 2010.

[17] A. B. Owen, *Randomly permuted* (t, m, s)*-nets and* (t, s)*-sequences*, in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, H. Niederreiter and P. J.-S. Shiue, eds., Springer-Verlag, New York, 1995, pp. 299–317.

[18] A. B. Owen, *The dimension distribution and quadrature test functions*, Stat. Sinica, 13 (2003), pp. 1–17.

[19] A. B. Owen, *Effective dimension of some weighted pre-Sobolev spaces with dominating mixed partial derivatives*, SIAM J. Numer. Anal., 57 (2019), pp. 547–562, https://doi.org/10.1137/17M1158975.

[20] A. Papageorgiou and J. F. Traub, *Faster evaluation of multidimensional integrals*, Comput. Phys., (1997), pp. 574–578.

[21] V. V. Petrov, *Moments of sums of independent random variables*, J. Sov. Math., 61 (1992), pp. 1905–1906.

[22] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. L. Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. A. Guillaume, J. Jakeman, H. Gupta, N. Milillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. R. Maier, *The future of sensitivity analysis: An essential discipline for systems modeling and policy support*, Environ. Model. Softw., 137 (2021), 104954.

[23] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Stat. Sci., 4 (1989), pp. 409–423.

[24] I. H. Sloan and H. Wozniakowski, *When are quasi-Monte Carlo algorithms efficient for high dimensional integration?*, J. Complexity, 14 (1998), pp. 1–33.

[25] I. M. Sobol', *The distribution of points in a cube and the accurate evaluation of integrals*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 86–112.

[26] I. M. Sobol', *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).