SIAM J. SCI. COMPUT.
Vol. 42, No. 4, pp. A2262–A2280

MULTIVARIATE MONTE CARLO APPROXIMATION BASED ON SCATTERED DATA*

WENWU GAO†, XINGPING SUN‡, ZONGMIN WU§, AND XUAN ZHOU¶

Abstract. We propose and study a new multivariate stochastic scattered data quasi-interpolation scheme that is reminiscent of the classical Monte Carlo method for estimating integrals. We first employ a convolution operator to approximate (deterministically) Sobolev space functions and use a result of Cheney, Light, and Xu [*On kernels and approximation orders*, in Approximation Theory, Lecture Notes in Pure Appl. Math. 138, Dekker, 1992, pp. 227–242] and Cheney and Lei [*Quasi-interpolation on irregular points*, in Approximation and Computation, Internat. Ser. Numer. Math. 119, Birkhäuser Boston, 1994, pp. 121–135] to obtain an approximation error estimate in terms of moment conditions. We then approximate (stochastically) the convolution integral using a Monte Carlo method and derive the maximal mean squared error (M-MSE) estimate and mean L^p-error estimate on bounded domains which are in line with those obtained by the classical Monte Carlo method for estimating multivariate integrals. The introduction of convolution operators is solely for the purpose of facilitating error analysis. The implementation of this scheme does not require any numerical handling of the convolution integral involved. Our final approximant is in the form of scattered data quasi-interpolation. It enjoys a simple construction and optimal convergence rate, yet it provides an efficient tool in various computing environments. Asymptotic normality and confidence interval test results show that the scheme is computationally stable. Numerical simulation results show that the scheme is robust in the presence of noise.

Key words. moment condition, Monte Carlo approximation, uncertainty quantification, quasi-interpolation, scattered data, statistical integration

AMS subject classifications. 41A05, 41063, 41065, 65D05, 65D10, 65D15

DOI. 10.1137/19M1249138

1. Introduction.

For many real-world data mining problems [39], data come in large quantities, and some are corrupted or even outright false. This brings forth a challenging computing environment in which employing time-consuming and sophisticated algorithms is not feasible. For example, an inspired programmer may devote

*Submitted to the journal's Methods and Algorithms for Scientific Computing section March 11, 2019; accepted for publication (in revised form) April 3, 2020; published electronically July 27, 2020.
https://doi.org/10.1137/19M1249138

Funding: This work was supported by the NSFC through grants 11871074, 11501006, and 61672032; by the NSFC Key Projects 11631015 and 91330201; by SGST through grant 12DZ 2272800; by the Joint Research Fund of National Natural Science Foundation of China and Research Grants Council of Hong Kong through grant 11461161006; by the Fund of Introducing Leaders of Science and Technology of Anhui University through grant J10117700057; by the 4th Project of Cultivating Backbone of Young Teachers of Anhui University through grant J01005138; and by the Anhui Provincial Science and Technology Major Project 16030701091.

†Department of Statistics, School of Economics, Anhui University, Hefei, People's Republic of China; Shanghai Key Laboratory for Contemporary Applied Mathematics, School of Mathematical Sciences, Fudan University, Shanghai, People's Republic of China; and Anhui Engineering Laboratory of Agro-Ecological Big Data, Anhui University, Hefei, People's Republic of China (wenwugao528@163.com).

‡Corresponding author. College of Mathematics and Information Science, Henan Normal University, Xinxiang, People's Republic of China, and Department of Mathematics, Missouri State University, Springfield, MO 65897 (xsun@missouristate.edu).

§Shanghai Key Laboratory for Contemporary Applied Mathematics, School of Mathematical Sciences, Fudan University, Shanghai, People's Republic of China, and Shanghai Center for Mathematical Sciences, Shanghai, People's Republic of China (zmwu@fudan.edu.cn).

¶Shanghai Key Laboratory for Contemporary Applied Mathematics, School of Mathematical Sciences, Fudan University, Shanghai, People's Republic of China (15110180025@fudan.edu.cn).

time and other resources for establishing an interpolation model for the acquired data only to find that it does not depict the real-world situation at hand. Hard work devoted to finding such an interpolant is instantly rendered worthless. Algorithms that are suitable for this kind of harsh computing environment must be nimble and capable of quickly generating simple-structured approximants that are stable and robust. Providing such an algorithm is the motivation of our effort devoted to this research. A brief outline of our approaches and main results is in order.

Let d, l be two positive integers, and let Ω be a bounded domain with Lipschitz boundary in the Euclidean space $\mathbb{R}^d$. Let $1 \le p \le \infty$, and let $W_p^l(\mathbb{R}^d)$ denote the Sobolev space of functions f such that

$$D^\alpha f \in L^p(\mathbb{R}^d), \quad 0 \le |\alpha| \le l,$$

where $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multi-index, $|\alpha| = \sum_{j=1}^d \alpha_j$, and $D^\alpha f$ denote the distributional derivatives of f. Endowed with the norm

$$\|f\|_{W_p^l(\mathbb{R}^d)} = \sum_{0\le|\alpha|\le l} \|D^\alpha f\|_{L^p(\mathbb{R}^d)}, \tag{1.1}$$

$W_p^l(\mathbb{R}^d)$ is a Banach space. Suppose that $\Psi \in C(\mathbb{R}^d)$ satisfies the moment condition of order l ($l > d/2$) (Definition 2.1). Let $h > 0$ be a scale parameter. (The scale parameter is often referred to as "bandwidth.") Set $\Psi_h = h^{-d}\Psi(x/h)$. Let X be the random variable uniformly distributed in Ω, and let $(X_j)_{j\ge1}$ be independent copies of X. Let V be a bounded domain satisfying $V \subset \Omega$, and let $\operatorname{dist}(\partial\Omega, \partial V) \ge c$ for a fixed $0 < c < 1$. Here $\operatorname{dist}(\partial\Omega, \partial V)$ denotes the Hausdorff distance between $\partial\Omega$ and ∂V. Our final approximant is in the form of scattered data quasi-interpolation:

$$Q_{\Psi,h}(f)(x) = \frac{|\Omega|}{N} \sum_{j=1}^N f(X_j)\Psi_h(x - X_j), \quad x \in V. \tag{1.2}$$

Here $|\Omega|$ denotes the Lebesgue measure of Ω. Setting $h = N^{-1/(2l+d)}$, we establish the following two estimates:

(i) The maximal mean squared error (M-MSE) satisfies the inequality

$$\sup_{x\in V} \mathbb{E}\left(f(x) - Q_{\Psi,h}(f)(x)\right)^2 \le C\ N^{-\frac{2l}{2l+d}}, \quad f \in W_\infty^l(\mathbb{R}^d). \tag{1.3}$$

(ii) The mean L^p-convergence rate of our quasi-interpolants for $1 \le p \le 2$ satisfies the inequality

$$\mathbb{E}\left(||Q_{\Psi,h}f - f||^p_{L^p(V)}\right) \le C\ N^{\frac{-pl}{2l+d}}, \quad f \in W_p^l(\mathbb{R}^d), \quad 1 \le p \le 2. \tag{1.4}$$

Here in inequalities (1.3) and (1.4), C denotes a constant independent of h. We point out that inequality (1.4) is similar to a class of estimates obtained in the learning theory paradigm advocated by Cucker and Smale [21] and Cucker and Zhou [22].

In [7], it was shown that $h = N^{-1/(2l+d)}$ is an optimal choice of the bandwidth, which yields the optimal convergence rate $N^{-\frac{l}{2l+d}}$ from the stochastic perspective in the context of nonparametric kernel regression. Furthermore, we establish the asymptotic normality and conduct confidence interval tests for our quasi-interpolants. We also provide numerical simulations to substantiate the theoretic results. Our final approximant, $Q_{\Psi,h}$, is a finite linear combination of translates of a single function,

Ψ_h. If function values at scattered sites $(X_j)_{j=1}^N$ are available, such an approximant can be easily constructed, and the implementation of the scheme needs only minimal storage memory and computation time.

The classical Monte Carlo integration has long been a cherished method among engineers for estimating integrals. In this paper, we generalize this method to approximate convolution integrals. In deriving the convergence rates as shown in (1.3), we first utilize a convolution integral $\Psi_h \ * f$ to approximate f (deterministically). We use a result of Cheney and coauthors [19, 20, 44] to characterize the approximation order in terms of the moment condition. We then build a quasi-interpolant by stochastically discretizing the convolution integral, viewing the scattered points $X_1, \ldots, X_N$ as identically, independently, and uniformly distributed (iiud) random variables. This implies that for each fixed $x \in V$, $Q_{\Psi,h}f(x)$ is a sum of N random variables, which leads us to derive the M-MSE (maximal mean squared error) bound (1.3) and the mean L^p-convergence rate (1.4). In the execution stage, a smaller scale parameter h yields a more favorable deterministic approximation order of f by the convolution integral $\Psi_h * f$, while the opposite is true for the stochastic approximation order of $\Psi_h * f$ by $Q_{\Psi,h}$. The estimates as shown in (1.3) and (1.4) are results of an optimization procedure done on the scale parameter h.

We are keenly aware of the fact that data acquired in harsh computing environments often contain noise. Therefore, we have provided in this paper stability analysis of our quasi-interpolation scheme, culminating in the derivations of asymptotic normality results and confidence interval tests for the quasi-interpolants $Q_{\Psi,h}f$ (Theorems 3.1 and 3.3) from a stochastic viewpoint. Numerical simulation results show that the scattered data quasi-interpolation scheme given herein is robust in the presence of noise.

Literature abounds on the study of quasi-interpolation methods based on grid data; see [5, 8, 9, 11, 12, 14, 16, 17, 18, 20, 25, 33, 34, 37, 45, 46, 53, 54, 56] and the references therein. Various innovative quasi-interpolation schemes based on scattered data have been proposed and studied in the literature. Beatson and Powell [6] discussed three univariate multiquadric quasi-interpolation schemes defined on a bounded interval based on scattered data. Wu and Schaback [56] studied a new version of multiquadric quasi-interpolation and investigated its shape-preserving properties and convergence rates. Buhmann, Dyn, and Levin [13] constructed quasi-interpolation for multivariate approximation using radial basis functions with quasi-uniform centers. Dyn and Ron [26] introduced the "two-step construction" to deal with the mathematical difficulties stemming from quasi-interpolation at scattered sites. Specifically, they built a quasi-interpolant based on gridded data as a midway house to obtain quasi-interpolants based on scattered centers. Their approach was further generalized by Yoon [60]. For further reading about the interesting topic, we point out the following references: Aldroubi and Gröchenig [2]; Backus and Gilbert [4]; Maz'ya and Schmidt [47, 48, 49]; Lanzara, Maz'ya, and Schmidt [41]; Wu and Liu [57]; Fasshauer and Zhang [28, 29, 30, 31]; and Wu, Sun, and Ma [58].

In citing references, we strive for impartiality and comprehensiveness. However, unintended omission of important approaches and results is unavoidable, which we blame unequivocally on our own ignorance and offer our apology in advance. In this paper, we view the scattered sites as "footprints" of N independent copies of the random variable uniformly distributed in a bounded domain, and consequently we construct a numerical algorithm from a stochastic perspective. This viewpoint is different from all others mentioned in the references above, as are the error estimates we obtain. We hope this approach will open a door to broader and deeper interdisciplinary research between approximation theory and machine learning as advocated by Cucker and Smale [21] and Cucker and Zhou [22]. More importantly, we anticipate that synergetic results will be established in a timely fashion.

The paper is organized as follows. In section 2, we introduce notation and definitions, and review some important results which we use to establish our own in the current paper. These include *moment conditions* which determine the approximation order of Sobolev space functions by convolution operators. A more subliminal goal of section 2 is to clearly lay out the logistics in the derivation of the final approximant. In section 3, we state and prove the main results of the paper, which include various error estimates, asymptotic normality, and confidence interval tests of our quasi-interpolants. Section 4 is devoted to numerical examples from which the reader witnesses the simplicity and efficiency of our quasi-interpolation scheme. Some of the examples demonstrate the robustness of our method against noisy data.

2. Preliminaries.

Let $L^p(\mathbb{R}^d)$ denote the Banach spaces consisting of all Lebesgue measurable functions f on $\mathbb{R}^d$ for which $\|f\|_{L^p(\mathbb{R}^d)} < \infty$. Here,

$$\|f\|_{L^p(\mathbb{R}^d)} := \begin{cases} \left(\int_{\mathbb{R}^d} |f(x)|^p dx\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \inf\{C : |f(x)| \le C \text{ for almost all } x\} & \text{if } p = \infty. \end{cases}$$

Our target functions are from Sobolev spaces of integer orders. Fractional order Sobolev spaces have also been extensively studied in the literature; see, e.g., [1, 55]. In the current paper, we stick to integer order Sobolev spaces to comply with the requirement of the moment condition (Definition 2.1). We adopt the Sobolev norms as defined in (1.1). Let $\Psi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, and let Ψ_h be a scaled version of Ψ, that is, $\Psi_h(x) = h^{-d}\Psi(x/h)$, with h being a positive scale parameter. Consider the convolution operator

$$f \mapsto \Psi_h * f, \quad f \in W_p^l(\mathbb{R}^d),$$

where

$$\Psi_h * f(x) = \int_{\mathbb{R}^d} \Psi_h(x-y)f(y)dy.$$

We seek conditions on Ψ such that

$$\|\Psi_h * f - f\|_{L^p(\mathbb{R}^d)} \le C_{l,d}\, |f|_{l,p}\, h^l, \quad f \in W_p^l(\mathbb{R}^d), \quad 1 \le p \le \infty, \tag{2.1}$$

where $C_{l,d}$ is a constant depending only on d and l, and $|f|_{l,p}$ is a semi-Sobolev norm defined by

$$|f|_{l,p} = \sum_{|\alpha|=l} \|D^\alpha f\|_{L^p(\mathbb{R}^d)}.$$

One such condition is often referred to as the "moment condition," which we will discuss in detail in the next subsection.

2.1. Moment condition.

Moment conditions are the key ingredient in the so-called approximate approximation paradigm advocated by Maz'ya and his coauthors (see, e.g., [41, 42, 43, 47, 48, 49] and references therein). These have been formulated in several equivalent conditions by Cheney [19], Lei, Jia, and Cheney [44], Wu and Liu [57], and other authors.

Definition 2.1. *Let l be a positive integer. We say that a function $\Psi \in W_p^l(\mathbb{R}^d)$ satisfies the moment condition of order l if*

$$\int_{\mathbb{R}^d} |x|^l |\Psi(x)| dx < \infty,$$

and the following equations hold true:

$$\int_{\mathbb{R}^d} x^\alpha \Psi(x) dx = \delta_{\alpha,0},\ 0 \le |\alpha| \le l-1. \tag{2.2}$$

Here $\delta_{\alpha,0}$ is a Kronecker delta symbol. Precisely, we have

$$\delta_{\alpha,0} = \begin{cases} 1 & \text{if } |\alpha| = 0, \\ 0 & \text{if } |\alpha| > 0. \end{cases}$$

Let $x \cdot y$ denote the Euclidean dot product of $x, y \in \mathbb{R}^d$. The Fourier transform pair (Fourier transform and the inverse Fourier transform) $\hat{f}$ and $\check{f}$ for a function $f \in L^1(\mathbb{R}^d)$ is given by

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \xi \cdot x} dx, \quad \check{f}(x) = \int_{\mathbb{R}^d} f(x) e^{2\pi i \xi \cdot x} d\xi.$$

Here we take the liberty of assuming that the Fourier transform pair has been appropriately extended to include its application to Schwarz class distributions. If both f and $\hat{f} \in L^1(\mathbb{R}^d)$, then we can take the inverse Fourier transform to recover f pointwise:

$$f(x) = (\hat{f}(\xi))\check{}\,(x) = \int_{\mathbb{R}^d} \hat{f}(\xi) e^{2\pi i \xi \cdot x} d\xi.$$

In this case, both f and $\hat{f}$ are continuous on $\mathbb{R}^d$. If Ψ satisfies the moment condition of order l, then for every α with $|\alpha| \le l$, $D^\alpha \hat{\Psi}$ is continuous on $\mathbb{R}^d$, and equations (2.2) are equivalent to

$$D^{(\alpha)} \hat{\Psi}(0) = \delta_{\alpha,0},\ 0 \le |\alpha| \le l-1. \tag{2.3}$$

Using (2.3), one can show that large classes of functions satisfy moment conditions (2.2). For example, any appropriately normalized function Ψ satisfies the moment condition for $l = 1$. Some authors have applied various techniques to construct functions that satisfy higher order moment conditions [34, 48, 57].

The following result was proved in [44, Theorem 2.1].

Lemma 2.1. *Assume that Ψ satisfies the moment condition of order l. Then inequality* (2.1) *holds true.*

2.2. Monte Carlo method for numerical integration. We first briefly review the classical Monte Carlo method for estimating integrals [24, 27, 36]. Let X be the random variable uniformly distributed in Ω. Then the expectation of a random variable $f(X)$ is

$$\mathbb{E}[f] = \frac{1}{|\Omega|} \int_\Omega f(x) dx.$$

Suppose that we have N random samples $\{(X_j, f(X_j))\}_{j=1}^N$ at hand. We form an empirical approximation $A_N[f]$ of $\mathbb{E}[f]$ by

$$A_N[f] := \frac{1}{N} \sum_{j=1}^N f(X_j).$$

In the statistics literature, one can find many proofs showing that $A_N[f]$ is an unbiased estimator of $\mathbb{E}[f]$. Moreover, the following mean squared error ($\mathbb{MSE}$) estimate holds true [15]:

$$\mathbb{E}\left(\int_\Omega f(x)dx - |\Omega| A_N[f]\right)^2 = \frac{\mathbb{VAR}(f(X))}{N}, \tag{2.4}$$

in which $\mathbb{VAR}(f(X))$ denotes the variance of the random variable $f(X)$. There are two key features of the above Monte Carlo method. First, the $\mathbb{MSE}$ estimate $\mathcal{O}(N^{-1})$ is independent of dimensions. Second, the method is simple and robust since it only involves computing the average of N samples. This motivates us to derive comparable error estimates for and analyze the asymptotic normality of our quasi-interpolation scheme as shown in (1.2). The latter has a positive impact on the robustness of the scheme against noise. We caution that the presence of the parameters $h > 0$ and x adds two layers of difficulty. These error estimates and the analysis will be carried out in section 3.

3. Monte Carlo quasi-interpolation scheme. Let $f \in W_p^l(\mathbb{R}^d)$ $(l > d/2)$. Given function values $\{f(X_j)\}_{j=1}^N$ at scattered sites $\{X_j\}_{j=1}^N$, we assemble the approximant as given in (1.2). The novelty here is that we view the scattered sites $\{X_j\}_{j=1}^N$ as N independent copies of the uniformly distributed random variable X in Ω, which is reminiscent of the classical Monte Carlo method for estimating integrals reviewed in subsection 2.3. Set

$$\mathbb{MSE}(f,x) = \mathbb{E}\left(f(x) - Q_{\Psi,h}f(x)\right)^2, \quad x \in V.$$

Then $\mathbb{MSE}(f)$ is a continuous function on V. Our first result in this section concerns M-MSE estimate. Denote

$$\mathcal{C}_{h,\Omega}(f)(x) := \int_\Omega \Psi_h(x-y)f(y)dy.$$

LEMMA 3.1. *Suppose that Ψ satisfies the moment condition of order k. Let V be a bounded domain satisfying $V \subset \Omega$, and let* $\operatorname{dist}(\partial\Omega, \partial V) \geq c$ *for a fixed* $0 < c < 1$. *Here* $\operatorname{dist}(\partial\Omega, \partial V)$ *denotes the Hausdorff distance between $\partial\Omega$ and ∂V. Then, for any $f \in W_p^l(\mathbb{R}^d)$, $1 \leq p \leq \infty$, $l > d/2$, there exists a constant C that is independent of h such that*

$$\|\mathcal{C}_{h,\Omega}(f) - f\|_{p,V} \leq C\ h^m, \quad m = \min\{k, l\}. \tag{3.1}$$

Proof. We first use Minkowski's inequality to write

$$\|f - \mathcal{C}_{h,\Omega}(f)\|_{p,V} \leq \|f - \mathcal{C}_h(f)\|_{p,V} + \|\mathcal{C}_h(f) - \mathcal{C}_{h,\Omega}(f)\|_{p,V}.$$

Here we remind the reader that $\mathcal{C}_h$ is defined for L^p functions on $\mathbb{R}^d$ and $\mathcal{C}_{h,\Omega}$ for those functions when restricted on Ω.

Next, we show that both terms on the right-hand side of the above inequality are of the order $\mathcal{O}(h^m)$. Lemma 2.1 implies that there is a constant C_1 independent of h such that

$$\|f - \mathcal{C}_h(f)\|_{p,V} \leq \|f - \mathcal{C}_h(f)\|_p \leq C_1\ h^m.$$

Fix an $x \in V$. Since $\operatorname{dist}(\partial\Omega, \partial V) \geq c$, we have $\mathbb{R}^d \setminus \Omega \subset D(x,c) := \{t : |x-t| \geq c\}$. We apply Chebyshev's inequality to get

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d \setminus \Omega} f(t)\Psi_h(x-t)dt \right| \\
&\quad \leq h^{-d} \int_{D(x,c)} |f(t)\Psi((x-t)/h)|dt \\
&\quad \leq c^{-m}\ h^{-d} \int_{D(x,c)} |x-t|^m |f(t)\Psi((x-t)/h)|dt.
\end{aligned}$$

The substitution $y = (x-t)/h$ in the integral above leads to

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d \setminus \Omega} f(t)\Psi_h(x-t)dt \right| \\
&\quad \leq \frac{h^m\ \|f\|_\infty}{c^m} \int_{D(0,\ c/h)} |y|^m\ |\Psi(y)|dy \\
&\quad \leq \frac{h^m\ \|f\|_\infty}{c^m} \int_{\mathbb{R}^d} |y|^m\ |\Psi(y)|dy.
\end{aligned}$$

The last integral in the above inequality is finite because Ψ satisfies the moment condition of order k ($k \geq m$). This implies that

$$\left| \int_{\mathbb{R}^d \setminus \Omega} f(t)\Psi_h(x-t)dt \right| \leq C_2 h^m,$$

where C_2 is a constant independent of h and $x \in V$. Hence, we have

$$\|\mathcal{C}_h(f) - \mathcal{C}_{h,\Omega}(f)\|_{\infty,V} \leq C_2 h^m$$

and therefore that

$$\|\mathcal{C}_h(f) - \mathcal{C}_{h,\Omega}(f)\|_{p,V} \leq C_2 |V|^{1/p} h^m.$$

Thus, the lemma holds with $C = C_1 + C_2|V|^{1/p}$. ∎

In what follows, we will denote C, unless otherwise specified, as universal constant whose value may change from line to line.

THEOREM 3.1. *Suppose that Ψ satisfies the moment condition of order $l > d/2$. Let V be a bounded domain satisfying $V \subset \Omega$, and let $\operatorname{dist}(\partial\Omega, \partial V) \geq c$ for a fixed $0 < c < 1$. Then there exists a constant C independent of h such that*

$$\sup_{x \in V} \mathbb{MSE}(f,x) \leq C\ (h^{2l} + h^{-d}N^{-1}), \quad f \in W^l_\infty(\mathbb{R}^d). \tag{3.2}$$

In particular, setting $h = N^{-1/(2l+d)}$, we obtain the following estimate:

$$\sup_{x \in V} \mathbb{MSE}(f,x) \leq C\ N^{-2l/(2l+d)}, \quad f \in W^l_\infty(\mathbb{R}^d).$$

Proof. Fix an $x \in V$. We write

$$\begin{aligned}\mathbb{MSE}(f,x) &= \mathbb{E}\left(f(x) - \int_\Omega f(t)\Psi_h(x-t)dt + \int_\Omega f(t)\Psi_h(x-t)dt - Q_{\Psi,h}f(x)\right)^2 \\ &= \left(f(x) - \int_\Omega f(t)\Psi_h(x-t)dt\right)^2 + \mathbb{E}\left(\int_\Omega f(t)\Psi_h(x-t)dt - Q_{\Psi,h}f(x)\right)^2. \end{aligned} \tag{3.3}$$

We will estimate separately the two terms on the right-hand side of the above inequality. For the first term, we make use of Lemma 3.1 (for the case $p = \infty$) to derive

$$\sup_{x\in V}\left|f(x) - \int_\Omega f(t)\Psi_h(x-t)dt\right|^2 \le C\ h^{2l}. \tag{3.4}$$

To estimate the second term on the right-hand side of (3.3), we first bound the variance of the random variable

$$Y_x := \frac{1}{|\Omega|}\Psi_h(x-X)f(X)$$

uniformly distributed in Ω as follows:

$$\mathbb{VAR}(Y_x) \le \|f\|_\infty^2 \int_{\mathbb{R}^d}\Psi_h^2(y)dy = h^{-d}\ \|f\|_\infty^2 \int_{\mathbb{R}^d}\Psi^2(y)dy.$$

We then resort to (2.4) to obtain

$$\mathbb{E}\left(\int_\Omega f(t)\Psi_h(x-t)dt - Q_{\Psi,h}f(x)\right)^2 \le C\ h^{-d}\ N^{-1}. \tag{3.5}$$

Combining the above two inequalities (3.4) and (3.5) yields the inequality

$$\mathbb{MSE}(f,x) \le C(h^{2l} + h^{-d}N^{-1}), \quad x \in V.$$

This estimate does not depend on $x \in V$. Hence, the desired result of the theorem follows. ∎

Remark 3.1. We make two timely comments about the optimal value of the scale parameter.

1. Inequality (3.2) shows that a smaller h yields a more favorable error estimate in the approximation of f by a convolution, while the opposite is true in the approximation of the convolution by $Q_{\Psi,h}f$. Minimizing the error estimate in (3.2) with respect to h gives the optimal choice $h = N^{-1/(2l+d)}$ for the scale parameter.
2. Intuitively, the error estimate in (3.2) is the expected pointwise convergence rate of $Q_{\Psi,h}f(x)$ to $f(x)$, while the scale parameter $h = N^{-1/(2l+d)}$.

In many applications (e.g., uncertainty quantification [3, 23, 50, 59]), the mean L^p-convergence rate is preferred. We will establish the mean L^p-convergence rate for our quasi-interpolants for p in the range of $1 \le p \le 2$ in the following theorem.

THEOREM 3.2. *Suppose that Ψ satisfies the moment condition of order l. Set $h = N^{-1/(2l+d)}$. Then there exists a constant C independent of h such that*

$$\mathbb{E}\left(||Q_{\Psi,h}f - f||^p_{L^p(V)}\right) \le C\ N^{-pl/(2l+d)}, \quad f \in W_p^l(\mathbb{R}^d), \quad 1 \le p \le 2.$$

Proof. For p in the range of $1 \leq p \leq 2$, and for $x \in V$, we first write

$$f(x) - Q_{\Psi,h}f(x) = (f(x) - \mathcal{C}_{h,\Omega}(f)(x)) + (\mathcal{C}_{h,\Omega}(f)(x) - Q_{\Psi,h}f(x)) .$$

We then make use of Jensen's inequality on the right-hand side of the above inequality to derive

$$\mathbb{E}\left(||Q_{\Psi,h}f - f||^p_{L^p(V)}\right) \leq 2^{p-1}\left(\|\mathcal{C}_{h,\Omega}(f) - f\|^p_{L^p(V)} + \mathbb{E}\,\|\mathcal{C}_{h,\Omega}(f) - Q_{\Psi,h}f\|^p_{L^p(V)}\right).$$

By Lemma 3.1, we can bound the first term on the right-hand side of the above inequality as follows:

$$\|\mathcal{C}_{h,\Omega}(f) - f\|^p_{L^p(V)} \leq Ch^{pl}. \tag{3.6}$$

To bound the second term, we first use Fubini's theorem and then Hölder's inequality to obtain

$$\begin{aligned}
&\mathbb{E}\,\|\mathcal{C}_{h,\Omega}(f) - Q_{\Psi,h}f\|^p_{L^p(V)} \\
&\quad = \int_V \left[\mathbb{E}\,|\mathcal{C}_{h,\Omega}(f)(x) - (Q_{\Psi,h}f)(x)|^p\right] dx \\
&\quad \leq |V|^{\frac{2-p}{2}} \left(\int_V \left[\mathbb{E}\,|\mathcal{C}_{h,\Omega}(f)(x) - (Q_{\Psi,h}f)(x)|^2\right] dx\right)^{p/2}.
\end{aligned}$$

Similarly to inequality (3.5) in the proof of Theorem 3.1,[1] we have

$$\mathbb{E}\,|\mathcal{C}_{h,\Omega}(f)(x) - (Q_{\Psi,h}f)(x)|^2 \leq Ch^{-d}N^{-1}, \quad x \in V.$$

It then follows that

$$\mathbb{E}\,\|\mathcal{C}_{h,\Omega}(f) - Q_{\Psi,h}f\|^p_{L^p(V)} \leq C|V|\left(h^{-d}N^{-1}\right)^{p/2}. \tag{3.7}$$

Combining inequalities (3.6) and (3.7), we obtain

$$\mathbb{E}\left(||Q_{\Psi,h}f - f||^p_{L^p(V)}\right) \leq C\left[h^{pl} + \left(h^{-d}N^{-1}\right)^{p/2}\right].$$

Set $h = N^{-1/(2l+d)}$ in the above inequality to get the desired result of Theorem 3.2. ∎

Remark 3.2. Maz'ya and Schmidt constructed quasi-interpolants for multivariate scattered data within the paradigm of approximate approximations [47, 48, 49]. But their quasi-interpolants generally do not converge to target functions. Corollary 3.2 shows that our quasi-interpolation scheme gives an optimal convergence rate in the context of nonparametric kernel regression in statistics (see Theorem 2 in [7]). In addition, the quasi-interpolants here are easy to construct and require weaker conditions on kernels than the ones frequently used in kernel regression. These have made our scheme more efficient and less stringent in terms of applicability and implementation.

[1]In establishing inequality (3.5), we need the fact that $\|f\|_\infty < \infty$, which follows from the following special form of the Sobolev embedding theorem:

$$\|f\|_\infty \leq C\|f\|_{W^l_p(\mathbb{R}^d)}, \quad l > d/2, \quad 1 \leq p \leq \infty,$$

where C is a constant independent of f.

Remark 3.3. For simplicity, the Monte Carlo technique here is employed to uniformly distributed sampling sites. It is worth mentioning that the same idea works for any probabilistic integration technique.

The second half of this section is devoted to the derivations of the asymptotic normality and confidence interval tests for our quasi-interpolants. For these purposes, we will enforce the following additional decay condition on the function Ψ which we use as the kernel for the quasi-interpolation scheme: There exist two positive constants C and δ such that

$$|\Psi(x)| \le \frac{C}{(1+|x|)^{d+\delta}}, \quad x \in \mathbb{R}^d. \tag{3.8}$$

THEOREM 3.3. *Suppose that Ψ satisfies the moment condition of order l as well as the decay condition in* (3.8). *For a fixed $x \in V$, let $Y_N(x)$ be the random variable defined by*

$$Y_N(x) = (Q_{\Psi,h}f(x) - \mathbb{E}[Q_{\Psi,h}f(x)])/\sqrt{\mathbb{VAR}(Q_{\Psi,h}f(x))}, \quad h = N^{-1/(2l+d)}.$$

Assuming that $\mathbb{VAR}(Q_{\Psi,h}f(x)) \neq 0$ for each fixed $x \in V$. Then, as $N \to \infty$, the sequence of random variables $Y_N(x)$ converges in distribution to the standard normal distribution.

Proof. Fix an $x \in V$. We first write $Y_N(x)$ as a certain linear combination of N independent copies of the random variable $\tilde{\xi}(x)$ defined by

$$\tilde{\xi}(x) = |\Omega| f(X)\Psi_h(x-X) - \mathbb{E}[Q_{\Psi,h}f(x)],$$

in which X is the random variable uniformly distributed in Ω. We see immediately that $\mathbb{E}[\tilde{\xi}(x)] = 0$ and that

$$\begin{aligned} \mathbb{VAR}(\tilde{\xi}(x)) &= \mathbb{VAR}(|\Omega| f(X)\Psi_h(x-X)) \\ &= N\mathbb{VAR}\left(\frac{|\Omega|}{N}\sum_{j=1}^{N} f(X_j)\Psi_h(x-X_j)\right) \\ &= N\mathbb{VAR}(Q_{\Psi,h}f(x)). \end{aligned}$$

Proceeding with some simple calculations while using the assumption $\mathbb{VAR}(Q_{\Psi,h}f(x)) > 0$, we obtain

$$\begin{aligned} Y_N(x) &= \frac{1}{\sqrt{\mathbb{VAR}(Q_{\Psi,h}f(x))}}\left(\frac{|\Omega|}{N}\sum_{j=1}^{N} f(X_j)\Psi_h(x-X_j) - \mathbb{E}[Q_{\Psi,h}f(x)]\right) \\ &= \frac{1}{N}\sum_{j=1}^{N} \frac{|\Omega| f(X_j)\Psi_h(x-X_j) - \mathbb{E}[Q_{\Psi,h}f(x)]}{\sqrt{\mathbb{VAR}(Q_{\Psi,h}f(x))}} \\ &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} \frac{\tilde{\xi}_j(x)}{\sqrt{\mathbb{VAR}(\tilde{\xi}_j(x))}}. \end{aligned} \tag{3.9}$$

Here $\{\tilde{\xi}_j(x)\}_{j=1}^N$ are N independent copies of the random variable $\tilde{\xi}(x)$. Unfortunately, the random variable $\tilde{\xi}(x)$ depends on h, and therefore on N, which has prevented us from applying the central limit theorem directly. To overcome this obstacle, we consider the characteristic function $\varphi_{Y_N(x)}$ (see [52]) of $Y_N(x)$,

$$\varphi_{Y_N(x)}(t) := \mathbb{E}\left(\exp\left(itY_N(x)\right)\right).$$

For each fixed $x \in V$, we derive from (3.9) that

$$\begin{aligned}\varphi_{Y_N(x)}(t) &= \mathbb{E}\left(\exp\left(it\sum_{j=1}^{N}\frac{\tilde{\xi}_j(x)}{\sqrt{N}\sqrt{\mathbb{VAR}(\tilde{\xi}_j(x))}}\right)\right)\\ &= \varphi^N_{\tilde{\xi}_1(x)}\left(\frac{t}{\sqrt{N}\sqrt{\mathbb{VAR}(\tilde{\xi}_1(x))}}\right).\end{aligned}$$

Using Taylor's expansion of the exponential function in the above equation, we have

$$\varphi_{Y_N(x)}(t) = \left(1 - \frac{t^2\mathbb{E}(\tilde{\xi}_1^2(x))}{2N\mathbb{VAR}(\tilde{\xi}_1(x))} + o\left(\frac{t^2}{N\mathbb{VAR}(\tilde{\xi}_1(x))}\right)\right)^N.$$

Letting $N \to \infty$, we get

$$\begin{aligned}\lim_{N\to\infty}\varphi_{Y_N(x)}(t) &= \lim_{N\to\infty}\left(1 - \frac{t^2\mathbb{E}(\tilde{\xi}_1^2(x))}{2N\mathbb{VAR}(\tilde{\xi}_1(x))} + o\left(\frac{t^2}{N\mathbb{VAR}(\tilde{\xi}_1(x))}\right)\right)^N\\ &= \exp\left(-\frac{t^2}{2\mathbb{VAR}(\tilde{\xi}_1(x))}\mathbb{E}(\tilde{\xi}_1^2(x))\right)\\ &= \exp\left(-\frac{t^2}{2}\right).\end{aligned}$$

This is the desired result of the theorem. □

Remark 3.4. It is worth noting that the assumption $\mathbb{VAR}(Q_{\Psi,h}f(x)) \neq 0$ for each fixed $x \in V$ is rather innocuous. In fact, if $\mathbb{VAR}(Q_{\Psi,h}f(x)) = 0$ for some fixed $x \in V$, then re-checking the situation in which Cauchy-Schwarz inequality becomes equality shows that the product $\Psi_h(x-y)\, f(y)$ is a constant for all $y \in \Omega$. However, the decay condition (3.8) ensures that as $h \downarrow 0$,

$$\Psi_h(0) \to \infty \quad \text{and} \quad \Psi_h(a) \to 0 \quad \text{for any } a \neq 0,$$

which has eliminated all possibilities in which $\mathbb{VAR}(Q_{\Psi,h}f(x)) = 0$ for some fixed $x \in V$ except for the case when $f \in W_p^r(\mathbb{R}^d)$ $(r > d/2)$ is identically zero.

The asymptotic normality of $Y_N(x)$ allows us to construct pointwise local asymptotic confidence intervals [51] for our quasi-interpolant $Q_{\Psi,h}f(x)$. Let $\alpha < 1$ be a small positive number (which is called the confidence level in statistics [40]). Set

$$f_\alpha^L(x) = \mathbb{E}\left(Q_{\Psi,h}f(x)\right) - t_{1-0.5\alpha}\sqrt{\mathbb{VAR}(Q_{\Psi,h}f(x))}$$

and

$$f_\alpha^U(x) = \mathbb{E}\left(Q_{\Psi,h}f(x)\right) + t_{1-0.5\alpha}\sqrt{\mathbb{VAR}(Q_{\Psi,h}f(x))},$$

where $t_{1-0.5\alpha}$ is the $(1-0.5\alpha)$th quantile of the standard normal distribution [40]. Then, based on Theorem 3.3, $[f_\alpha^L,\ f_\alpha^U(x)]$ is the asymptotic $(1-\alpha)$ confidence interval of $Q_{\Psi,h}f(x)$ at each x [51]. That is, for each fixed $x \in V$, we have

$$\lim_{N\to\infty}\mathbb{P}\left\{f_\alpha^L(x) \leq Q_{\Psi,h}f(x) \leq f_\alpha^U(x)\right\} = 1 - \alpha.$$

In the remaining part of this section, we investigate the robustness of our quasi-interpolation scheme in the presence of noise. Suppose that sample data are given in the form

$$\{(X_k, f(X_k) + \varepsilon_k)\}_{k=1}^N,$$

where $\{\varepsilon_j\}_{k=1}^N$ are independently and identically distributed Gaussian noise with zero mean and finite variance σ^2 that are independent of x. Suppose that under the influence of the noise, we have constructed a new quasi-interpolant $Q^*_{\Psi,h}f$,

$$Q^*_{\Psi,h}f(x) = \frac{|\Omega|}{N}\sum_{k=1}^N (f(X_k) + \varepsilon_k)\Psi_h(x - X_k). \tag{3.10}$$

We have the following result.

Proposition 3.1. *Suppose that Ψ satisfies the moment condition of order l. Let $Q^*_{\Psi,h}f(x)$ be defined as in (3.10). For a fixed $x \in V$, let $\mathbb{MSE}^*(f,x) = \mathbb{E}[Q^*_{\Psi,h}f(x) - f(x)]^2$ be the mean integrated squared error between $Q^*_{\Psi,h}f(x)$ and $f(x)$. Then, there exists a constant C independent of h and σ^2, such that*

$$\sup_{x\in V} \mathbb{MSE}^*(f,x) \le C\ \left(h^{2l} + \frac{\|f\|_\infty^2 + \sigma^2|\Omega|}{Nh^d}\right). \tag{3.11}$$

Proof. We have

$$\begin{aligned}
\mathbb{E}(Q^*_{\Psi,h}f(x) - f(x))^2 &= \mathbb{E}\left(\frac{|\Omega|}{N}\sum_{k=1}^N (f(X_k) + \varepsilon_k)\Psi_h(x - X_k) - f(x)\right)^2 \\
&\le 2\,\underbrace{\mathbb{E}\left(\frac{|\Omega|}{N}\sum_{k=1}^N f(X_k)\Psi_h(x - X_k) - f(x)\right)^2}_{P_1} \\
&\quad + 2\,\underbrace{\mathbb{E}\left(\frac{|\Omega|}{N}\sum_{k=1}^N \varepsilon_k\Psi_h(x - X_k)\right)^2}_{P_2}.
\end{aligned}$$

By Theorem 3.1, there is a constant C independent of h, such that

$$P_1 \le \frac{2\|f\|_\infty^2}{Nh^d} + C\ h^{2l}.$$

It remains to estimate P_2. Note that

$$\begin{aligned}
P_2 &= \frac{2|\Omega|^2}{N^2}\sum_{k=1}^N \mathbb{E}(\varepsilon_k^2\Psi_h^2(x - X_k)) \\
&= \frac{2\sigma^2|\Omega|^2}{N}\mathbb{E}(\Psi_h^2(x - X_k)) \\
&\le \frac{2\sigma^2|\Omega|}{Nh^d}\|\Psi\|_{L^2(\mathbb{R}^d)}^2.
\end{aligned}$$

The desired result then follows. ∎

Remark 3.5. Based on inequality (3.11), we conclude the following:

1. If $\sigma^2 \le \|f\|_\infty^2/|\Omega|$, then we set $h = N^{-1/(2l+d)}$. Consequently, inequality (3.11) becomes
$$\sup_{x\in V} \mathbb{MSE}^*(f,x) \le C\ N^{-2l/(2l+d)}.$$
2. If $\sigma^2 \ge \|f\|_\infty^2/|\Omega|$, then we set $h = (\sigma^2/N)^{1/(2l+d)}$, which morphs inequality (3.11) into
$$\sup_{x\in V} \mathbb{MSE}^*(f,x) \le C\ (\sigma^2/N)^{2l/(2l+d)}.$$

In both cases above, C is a constant independent of N. This may be considered as a recipe for selecting optimal values for the scale parameter in the presence of noise. Generally speaking, a larger scale parameter value (corresponding to a smoother quasi-interpolant) shall be selected for data with a higher noise level. The error estimates in both cases demonstrate that the quasi-interpolation scheme is robust.

Asymptotic normality and confidence interval test results also hold true for the mean $L_p(1 \le p \le 2)$ error estimates. We state these results in the following corollaries without giving proofs.

COROLLARY 3.1. *Suppose that Ψ satisfies the moment condition of order l. Then with the optimal value of the scale parameter set*
$$h = \max\left(N^{-1/(2l+d)}, (\sigma^2/N)^{1/(2l+d)}\right),$$
we have the inequality
$$\mathbb{E}(\|Q^*_{\Psi,h}f - f\|_p^p) \le C\ \max\left(N^{-pl/(2l+d)}, (\sigma^2/N)^{2l/(2l+d)}\right),$$
where C is a constant independent of N.

COROLLARY 3.2. *For a fixed $x \in V$, let $\mathbb{VAR}(Q^*_{\Psi,h}f(x))$ be the variance of $Q^*_{\Psi,h}f(x)$. Let*
$$Y^*_N(x) = (Q^*_{\Psi,h}f(x) - \mathbb{E}[Q^*_{\Psi,h}f(x)])/\sqrt{\mathbb{VAR}(Q^*_{\Psi,h}f(x))},$$
assuming that $\mathbb{VAR}(Q^*_{\Psi,h}f(x)) > 0$. Suppose that the optimal value of the scale parameter h is set as in Corollary 3.1*. Then, as $N \to \infty$, the sequence of random variables $Y^*_N(x)$ converges in distribution to the standard normal distribution.*

The asymptotic normality result of $Q^*_{\Psi,h}f$ allows us to identify its asymptotic $(1-\alpha)$ confidence interval as $[f^{*,L}_\alpha(x), f^{*,U}_\alpha(x)]$. Here,
$$f^{*,L}_\alpha(x) = \mathbb{E}\left(Q^*_{\Psi,h}f(x)\right) - t_{1-0.5\alpha}\sqrt{\mathbb{VAR}(Q^*_{\Psi,h}f(x))}$$
and
$$f^{*,U}_\alpha(x) = \mathbb{E}\left(Q^*_{\Psi,h}f(x)\right) + t_{1-0.5\alpha}\sqrt{\mathbb{VAR}(Q^*_{\Psi,h}f(x))}.$$

4. Numerical simulations. This section is devoted to discussing three numerical simulations in details to substantiate the theoretical results established in the previous section.

Simulation I. We employ our quasi-interpolation scheme to approximate the following test functions, respectively:

$$f_{\text{test}} := \begin{cases} \dfrac{1}{1+x^2} & \text{in } \mathbb{R}, \\ \text{Franke's function [32]} & \text{in } \mathbb{R}^2, \\ \sin(2\pi x)\cos(2\pi y)\sin(2\pi \mathrm{z}) & \text{in } \mathbb{R}^3. \end{cases} \tag{4.1}$$

TABLE 1
A priori and a posteriori convergence rates of our quasi-interpolation with compactly supported kernels.

d/k	2	4	6
1	(0.40, 0.43, 0.46)	(0.44, 0.46, 0.49)	(0.46, 0.47, 0.50)
2	(0.33, 0.43, 0.42)	(0.4, 0.47, 0.47)	(0.43, 0.46, 0.49)
3	(0.29, 0.35, 0.39)	(0.36, 0.43, 0.47)	(0.40, 0.46, 0.49)

TABLE 2
A priori and a posteriori convergence rates of our quasi-interpolation with Gaussian kernels.

d/k	2	4	6
1	(0.40, 0.44, 0.46)	(0.44, 0.49, 0.48)	(0.46, 0.48, 0.50)
2	(0.33, 0.41, 0.42)	(0.4, 0.43, 0.49)	(0.43, 0.51, 0.50)
3	(0.29, 0.43, 0.48)	(0.36, 0.52, 0.55)	(0.40, 0.55, 0.61)

TABLE 3
A priori and a posteriori convergence rates of our quasi-interpolation with Gaussian kernels.

2	4	6
(0.133, 0.158, 0.156)	(0.211, 0.215, 0.238)	(0.261, 0.270, 0.280)

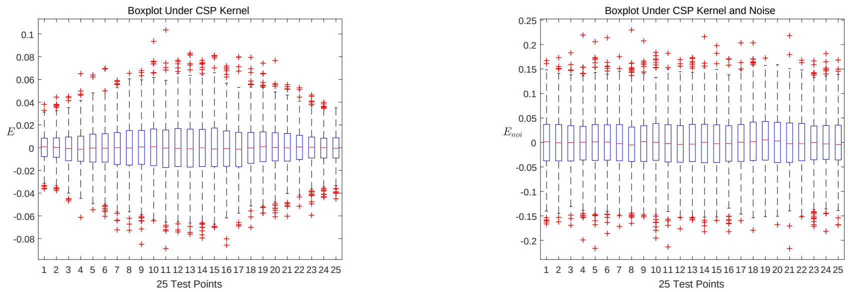
We take the unit ball as our sampling domain, that is, $\Omega = \{x : |x| \leq 1\}$, and the approximation domain is chosen to be $V = \{x : |x| \leq 0.92\}$. Theoretically, we can choose our approximation domain as $\{x : |x| \leq 1 - c\}$ for any fixed $0 < c < 1$. However, in a realistic computational environment, c needs to be no less than h. Since the smallest h value we choose is 0.08, we set the value of c to be 0.08. In what follows, we report the results of several rounds of simulations. In each round, we select a different value of N (whose values will be specifically given below) and randomly generate a $1000 \times N$ array $\{X_j^{(i)}\}_{i=1,j=1}^{1000,N}$ of points in Ω, with which we build 2×1000 quasi-interpolants: $Q_{\Psi,h}^{(i)} f$ and $Q_{\Psi,h}^{*(i)} f$ (1000 each). The latter carries a centered isotropic Gaussian noise with variance $\sigma^2 = 4$. To simulate the corresponding M-MSE of the quasi-interpolant $Q_{\Psi,h} f$, we design the following empirical M-MSE:

$$\text{M-MSE}_{\text{emp}}(Q_{\Psi,h} f) = \max_{1 \leq k \leq M} \frac{1}{1000} \sum_{i=1}^{1000} \left(Q_{\Psi,h}^{(i)} f(t_k) - f(t_k) \right)^2,$$

$$\text{M-MSE}_{\text{emp}}(Q_{\Psi,h}^* f) = \max_{1 \leq k \leq M} \frac{1}{1000} \sum_{i=1}^{1000} \left(Q_{\Psi,h}^{*(i)} f(t_k) - f(t_k) \right)^2.$$

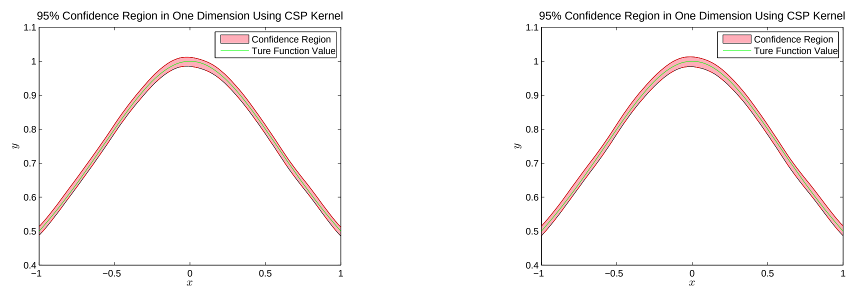
Here $M = 500$, and the point set $\{t_k\}_{k=1}^M$ is randomly generated in the prediction domain V. A posteriori convergence rates of our quasi-interpolants in terms of $\text{M-MSE}_{\text{emp}}(Q_{\Psi,h}^* f)$ and $\text{M-MSE}_{\text{emp}}(Q_{\Psi,h}^* f)$ using the designated kernels satisfying moment conditions of various orders in different dimensions are provided. A posteriori convergence rates given in Tables 1–3 are obtained for $N = 2^i$, $(i = 6, 7, 8, \ldots, 15)$.

In the first numerical experiment, we approximate the test functions using as kernel function the compactly supported radial kernels $\Phi_{d,k}$ constructed in [35] with $z_0 = 1, z_1 = 2, z_2 = 3$, in which d denotes the dimension $(d = 1, 2, 3)$ and k the order of moment conditions $(k = 2, 4, 6)$ for the kernels. A posteriori and a priori convergence rates of the quasi-interpolants with kernel functions $\Phi_{d,k}$ above are provided in Table 1. Each entry of the table is given in the format of an ordered triplet, indicating



(a) Box plots of $E(x)$ at 25 test points (b) Box plots of $E_{noi}(x)$ at 25 test points

FIG. 1. *Box plots of the errors of our quasi-interpolation with a compactly supported kernel.*



(a) 95% confidence interval of $Q_{\Psi,h}f(x)$ at 100 test points (b) 95% confidence interval of $Q^*_{\Psi,h}f(x)$ at 100 test points

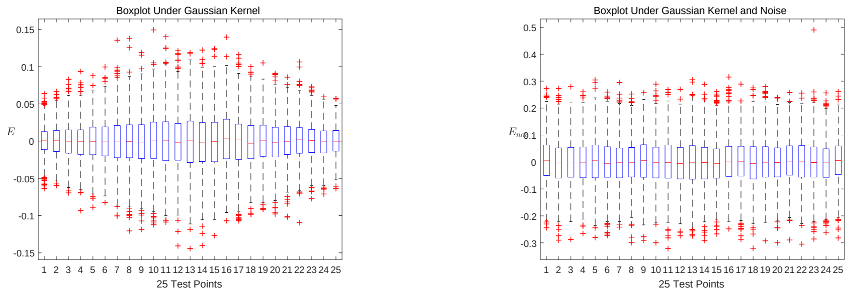
FIG. 2. 95% *confidence intervals of our quasi-interpolation with a compactly supported kernel.*

from left to right the a priori convergence rate and the a posteriori convergence rate of our quasi-interpolation for the data with and without the same Gaussian noise as above. We see from the table that the a posteriori convergence rates of our quasi-interpolants are better than the corresponding a priori convergence rates with and without the presence of noise. This indicates that our quasi-interpolation scheme is efficient and stable for practical applications.

In the second numerical experiment, we demonstrate the asymptotic normality of $Q_{\Psi,h}f(x)$ and $Q^*_{\Psi,h}f(x)$ by providing box plots of their approximation errors $E(x) = Q_{\Psi,h}f(x) - f(x)$ and $E_{noi}(x) = Q^*_{\Psi,h}f(x) - f(x)$, respectively, at 25 test points in Figure 1. In these box plots, we employ the compactly supported kernel above, satisfying the moment condition of order four in approximation of Frank's function [32].

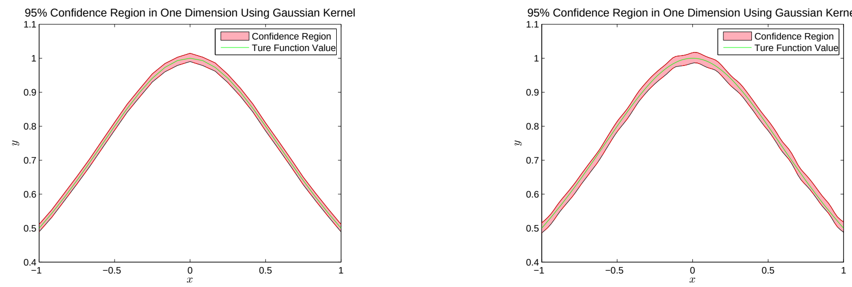
In Figure 1, subfigure (a) shows the box plots (mentioned above) for $E(x)$, and subfigure (b) shows those for $E_{noi}(x)$. These box plots show symmetry with respect to medians, which demonstrates the asymptotic normality of the quasi-interpolants $Q_{\Psi,h}f(x)$ and $Q^*_{\Psi,h}f(x)$ [39].

In the last numerical experiment of Simulation I, we use the asymptotic normality property to test 95% confidence intervals of the quasi-interpolants $Q_{\Psi,h}f(x)$ and $Q^*_{\Psi,h}f(x)$ at 100 points, shown in Figure 2. Here we choose $f(x) = 1/(1 + x^2)$ as the test function. Our numerical experiments show that the approximant $f(x)$ lies in the 95% confidence intervals of the quasi-interpolants for sampling data with or without



(a) Box plots of $E(x)$ at 25 test points (b) Box plots of $E_{noi}(x)$ at 25 test points

FIG. 3. *Box plots of the errors of our quasi-interpolation with a Gaussian kernel.*



(a) 95% confidence intervals of $Q_{\Psi,h}f(x)$ at 100 test points (b) 95% confidence intervals of $Q^*_{\Psi,h}f(x)$ at 100 test points

FIG. 4. 95% *confidence intervals of our quasi-interpolation with a Gaussian kernel.*

noise, which shows the efficiency and stability of our quasi-interpolation scheme.

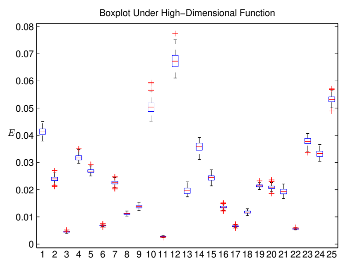
Simulation II. This is essentially a do-over of all numerical experiments carried out in Simulation I with the new kernel function $\Phi_{d,k}$ that is a linear combination of the Gaussian kernel explicitly constructed in [35] with $z_0 = 1, z_1 = 2, z_2 = 3$, $c = 0.1$. Here we remind the reader that we use c^2 in place of the σ^2 as in [35] because we denote σ^2 as the variance of noise in the current article. We provide the simulation results in Figures 3 and 4.

Simulation III. In the last simulation, we employ our quasi-interpolation scheme with Gaussian kernels satisfying moment conditions of respective orders two, four, and six [35] to approximate the following function defined on the unit cube of 11-dimensional Euclidean space:

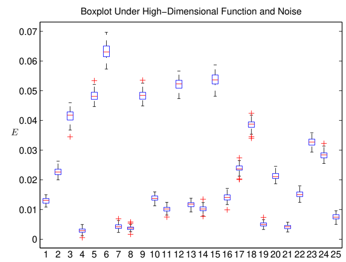
$$f(x) = \prod_{j=1}^{11} \left(\sin\left[\frac{\pi}{2} \left(x_j + \frac{j}{11} \right) \right] \right)^{5/j}, \quad x \in [0,1]^{11}. \tag{4.2}$$

Here we have written a point $x \in [0,1]^{11}$ in the 11-tuplet format: $x = (x_1, \dots, x_{11})$. A priori and a posteriori convergence rates of our quasi-interpolants are provided in Table 3.

Moreover, to demonstrate the asymptotic normality property of our quasi-interpolants $Q_{\Psi,h}f$ and $Q^*_{\Psi,h}f$, we provide corresponding box plots for the quasi-interpolants employing a Gaussian kernel satisfying a moment condition of order four, shown in Figure 5.



(a) Box plots of $E(x)$ at 25 test points



(b) Box plots of $E_{noi}(x)$ at 25 test points

FIG. 5. *Box plots of the errors of our quasi-interpolation with a Gaussian kernel satisfying a moment condition of order four.*

5. Conclusions and discussion. In this paper, we propose and study a new quasi-interpolation scheme for multivariate scattered data from a stochastic perspective, which features synergy between convolution operators and statistical integration using the Monte Carlo method. We establish results concerning optimal choice of the scaling parameter, maximal mean integrated square error analysis, L_p $(1 \leq p \leq 2)$ mean convergence rate, and asymptotic normality, as well as stability of the quasi-interpolation scheme. Numerical simulations using the quasi-interpolation scheme to approximate several benchmark functions are demonstrated. Both theoretical analysis and numerical results show that the scheme is simple, efficient, practical, and stable for real applications. Our approach will open doors to broader and deeper interdisciplinary research among diverse areas, such as approximation theory, machine learning, and data mining. Finally, we point out that constructing numerical algorithms from the stochastic perspective is a central theme in the new, surging research discipline of probabilistic numerics [10, 38].

Acknowledgments. We thank two anonymous referees for their constructive suggestions that have enhanced the mathematical structure and the exposition of the article.

REFERENCES

[1] R. A. Adams and J. J. Fournier, *Sobolev Spaces*, Elsevier/Academic Press, 2003.
[2] A. Aldroubi and K. Gröchenig, *Nonuniform sampling and reconstruction in shift-invariant spaces*, SIAM Rev., 43 (2001), pp. 585–620, https://doi.org/10.1137/S0036144501386986.
[3] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034, https://doi.org/10.1137/050645142.
[4] G. Backus and F. Gilbert, *The resolving power of gross earth data*, Geophys. J. R. Astr. Soc., 16 (1968), pp. 169–205.
[5] R. Beatson and W. Light, *Quasi-interpolation in the absence of polynomial reproduction*, in Numerical Methods of Approximation Theory, Vol. 9, D. Braess and L. L. Schumaker, eds., Birkhäuser, 1992, pp. 21–39.
[6] R. Beatson and M. Powell, *Univariate multiquadric approximation: Quasi-interpolation to scattered data*, Constr. Approx., 8 (1992), pp. 275–288.
[7] M. Belkin, A. Rakhlin, and A. B Tsybakov, *Does Data Interpolation Contradict Statistical Optimality?*, preprint, https://arxiv.org/abs/1806.09471v1, 2018.
[8] C. de Boor and G. Fix, *Spline approximation by quasi-interpolants*, J. Approx. Theory, 8 (1973), pp. 19–45.
[9] C. de Boor, *Quasi-interpolants and approximation power of multivariate splines*, in Com-

putation of Curves and Surfaces, M. Gasca and C. A. Micchelli, eds., Kluwer Academic Publishers, 1990, pp. 313–345.
[10] F. X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic, *Probabilistic integration: A role in statistician computation?*, Statist. Sci., 34 (2019), pp. 1–22.
[11] M. Buhmann, *Convergence of univariate quasi-interpolation using multiquadrics*, IMA J. Numer. Anal., 8 (1988), pp. 365–383.
[12] M.Buhmann, *On quasi-interpolation with radial basis functions*, J. Approx. Theory, 72 (1993), pp. 103–130.
[13] M. Buhmann, N. Dyn, and D. Levin, *On Quasi-interpolation by radial basis functions with scattered centres*, Constr. Approx., 11 (1995), pp. 239–254.
[14] M.Buhmann and F. Dai, *Pointwise approximation with quasi-interpolation by radial basis functions*, J. Approx. Theory, 192 (2015), pp. 156–192.
[15] R. Caflisch, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numer., 7 (1998), pp. 1–49.
[16] C. K. Chui and H. Diamond, *A natural formulation of quasi-interpolation by multivariate splines*, Proc. Amer. Math. Soc., 99 (1987), pp. 643–646.
[17] C. K. Chui and M. J. Lai, *A multivariate analog of Marsden's identity and a quasi-interpolation scheme*, Constr. Approx., 3 (1987), pp. 111–122.
[18] C. K. Chui and H. Diamond, *A characterization of multivariate quasi-interpolation formulas and its applications*, Numer. Math., 57 (1990), pp. 105–121.
[19] E. W. Cheney, W. Light, and Y. Xu, *On kernels and approximation orders*, in Approximation Theory (Memphis, TN, 1991), Lecture Notes in Pure Appl. Math. 138, G. A. Anastassiou, ed., Dekker, 1992, pp. 227–242.
[20] E. W. Cheney and J. J. Lei, *Quasi-interpolation on irregular points*, in Approximation and Computation: A Festschrift in Honor of Walter Gautschi (West Lafayette, IN, 1993), Internat. Ser. Numer. Math. 119, R. V. M. Zahar, ed., Birkhäuser Boston, 1994, pp. 121–135.
[21] F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bull. Amer. Math. Soc., 39 (2001), pp. 1–49.
[22] F. Cucker and D. X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monogr. Appl. Comput. Math., Cambridge University Press, 2007.
[23] I. Cialenco, G. Fasshauer, and Q. Ye, *Approximation of stochastic partial differential equations by a kernel-based collocation method*, Inter. J. Comput. Math., 89 (2012), pp. 2543–2561.
[24] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, 2007.
[25] N. Dyn, I. R. H. Jackson, D. Levin, and A. Ron, *On multivariate approximation by integer translates of a basis function*, Israel J. Math., 78 (1992), pp. 95–130.
[26] N. Dyn and A. Ron, *Radial basis function approximation: From gridded centers to scattered centers*, Proc. London Math. Soc., 71 (1995), pp. 76–108.
[27] M. Evans and T. Swartz, *Approximating Integrals via Monte Carlo and Deterministic Methods*, Oxford University Press, 2000.
[28] G. Fasshauer, *Toward approximate moving least squares approximation with irregularly spaced centers*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1231–1243.
[29] G. Fasshauer and J. Zhang, *Scattered data approximation of noisy data via iterated moving least squares*, in Proceedings of Curve and Surface Fitting: Avignon 2006, T. Lyche, J. L. Merrien, and L. L. Schumaker, eds., Nashboro Press, 2007, pp. 150–159.
[30] G. Fasshauer and J. Zhang, *On choosing "optimal" shape parameters for RBF approximation*, Numer. Algorithms, 45 (2007), pp. 345–368.
[31] G. Fasshauer and J. Zhang, *Iterated approximate moving least squares approximation*, in Advances in Meshfree Techniques, Comput. Methods Appl. Sci. 5, V.M.A. Leitão, C.J.S. Alves, and C. Armando Duarte, eds., Springer, 2007, pp. 221–239.
[32] R. Franke, *A Critical Comparison of Some Methods for Interpolation of Scattered Data*, Technical report NPS-53-79-003, Naval Postgraduate School, Monterey, CA, 1979.
[33] W. W. Gao and Z. M. Wu, *Quasi-interpolation for linear functional data*, J. Comput. Appl. Math., 236 (2012), pp. 3256–3264.
[34] W. W. Gao and Z. M. Wu, *Constructing radial kernels with higher-order moment conditions*, Adv. Comput. Math., 43 (2017), pp. 1355–1375.
[35] W. W. Gao and X. Zhou, *Multiscale radial kernels with high-order generalized Strang-Fix conditions*, Numer. Algorithms, 2019, https://doi.org/10.1007/s11075-019-00820-9.
[36] C. Graham and D. Talay, *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*, Springer-Verlag, 2013.
[37] P. Grohs, *Quasi-interpolation in Riemannian manifolds*, IMA J. Numer. Anal., 33 (2013), pp. 849–874.
[38] P. Hennig, M. A. Osborne, and M. Girolami, *Probabilistic numerics and uncertainty in*

computations, Proc. A, 471 (2015), 20150142.

[39] K. Vipin, P. N. Tan, and M. Steinbach, *Introduction to Data Mining*, Pearson, 2006.

[40] J. Lamperti, *Probability*, Benjamin, 1981.

[41] F. Lanzara, V. Maz'ya, and G. Schmidt, *Approximate approximations from scattered data*, J. Approx. Theory, 145 (2007), pp. 141–170.

[42] F. Lanzara, V. Maz'ya, and G. Schmidt, *Approximation of solutions to multidimensional parabolic equations by approximate approximations*, Appl. Comput. Harmon. Anal., 36 (2014), pp. 167–182.

[43] F. Lanzara, V. Maz'ya, and G. Schmidt, *Fast cubature of volume potentials over rectangular domains by approximate approximations*, Appl. Comput. Harmon. Anal., 41 (2016), pp. 749–767.

[44] J. Lei, R.-Q. Jia, and E. W. Cheney, *Approximation from shift-invariant spaces by integral operators*, SIAM J. Math. Anal., 28 (1997), pp. 481–498, https://doi.org/10.1137/S0036141095279869.

[45] B. Z. Li, B. L. He, and D. X. Zhou, *Approximation on variable exponent spaces by linear integral operators*, J. Approx. Theory, 223 (2017), pp. 29–51.

[46] W. Light and E. W. Cheney, *Quasi-interpolation with translates of a function having noncompact support*, Constr. Approx., 8 (1992), pp. 35–48.

[47] V. Maz'ya and G. Schmidt, *On approximate approximations using Gaussian kernels*, IMA J. Numer. Anal., 16 (1996), pp. 13–29.

[48] V. Maz'ya and G. Schmidt, *Construction of basis functions for high order approximate approximations*, in Mathematical Aspects of Boundary Elements Methods, Chapman and Hall/CRC Res. Notes Math. 414, 1998, pp. 191–202.

[49] V. Maz'ya and G. Schmidt, *On quasi-interpolation with non-uniformly distributed centers on domains and manifolds*, J. Approx. Theory, 110 (2001), pp. 125–145.

[50] J. Mohring, R. Milk, A. Ngo, O. Klein, O. Iliev, M. Ohlberger, and P. Bastian, *Uncertainty quantification for porous media flow using multilevel Monte Carlo*, in International Conference on Large-Scale Scientific Computing, Lecture Notes in Comput. Sci. 9374, I. Lirkov, S. Margenov, and J. Waśniewski, eds., Springer, 2015, pp. 145–152.

[51] H. Müller, U. Stadtmüller, and T. Schmitt, *Bandwidth choice and confidence intervals for derivatives of noisy data*, Biometrika, 74 (1987), pp. 743–749.

[52] B. Patrick, *Probability and Measure*, John Wiley & Sons, 1995.

[53] T. Ramming and H. Wendland, *A kernel-based discretisation method for first order partial differential equations*, Math. Comp., 87 (2018), pp. 1757–1781.

[54] H. Speleers and C. Manni, *Effortless quasi-interpolation in hierarchical spaces*, Numer. Math., 132 (2016), pp. 155–184.

[55] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[56] Z. M. Wu and R. Schaback, *Shape preserving properties and convergence of univariate multiquadric quasi-interpolation*, Acta Math. Appl. Sinica (Engl. Ser.), 10 (1994), pp. 441–446.

[57] Z. M. Wu and J. P. Liu, *Generalized Strang-Fix condition for scattered data quasi-interpolation*, Adv. Comput. Math., 23 (2005), pp. 201–214.

[58] Z. M. Wu, X. P. Sun, and L. M. Ma, *Sampling scattered data with Bernstein polynomials: Stochastic and deterministic error estimates*, Adv. Comput. Math., 38 (2013), pp. 187–205.

[59] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, 2010.

[60] J. Yoon, *Approximation in* $L_p(\mathbb{R}^d)$ *from a space spanned by the scattered shifts of a radial basis function*, Constr. Approx., 17 (2001), pp. 227–247.