SIAM J. NUMER. ANAL.
Vol. 58, No. 4, pp. 2059–2078

OPTIMALITY AND REGULARIZATION PROPERTIES OF QUASI-INTERPOLATION: DETERMINISTIC AND STOCHASTIC APPROACHES*

WENWU GAO†, GREGORY E. FASSHAUER‡, XINGPING SUN§, AND XUAN ZHOU¶

Abstract. Probabilistic numerics aims to study numerical algorithms from a stochastic perspective. This field has recently evolved into a surging interdisciplinary research area (between numerical approximation and probability theory) attracting much attention from the data science community at large. Motivated by this development, we incorporate a stochastic viewpoint into our study of multivariate quasi-interpolation for irregularly spaced data, a subject traditionally investigated in the realm of deterministic function approximation. We first construct quasi-interpolants directly from irregularly spaced data and show their optimality in terms of a certain quadratic functional on a weighted Hilbert space. We then derive the approximation order of our quasi-interpolants via a two-step procedure. In the first step, we approximate a target function by scaled integral operators (with an error term referred to as bias). In the second step, we discretize the underlying convolution integral at the irregularly spaced data sites (with an error term called variance). The final approximation order is obtained as an optimal trade-off between bias and variance. We also show that the scale parameter of the integral operators governs the regularization effect of the quasi-interpolation scheme, and find an optimal parameter value range to fine-tune the subtle balance between bias and variance under some additional assumptions on the distribution of the data sites. It is worth noting that evaluation of integrals is not needed in the implementation of our quasi-interpolation scheme, and that our quasi-interpolants are easy to construct. Numerical simulation results, including approximating the classical bore hole test function in eight dimensional space, provide evidence that our quasi-interpolation scheme is robust and capable of providing accurate generalizations.

Key words. bias-variance decomposition, optimality, quasi-interpolation, (quasi-)Monte Carlo method, regularization method

AMS subject classifications. 41A05, 41A63, 41A65, 65D05, 65D10, 65D15

DOI. 10.1137/19M1266496

1. Introduction.

Let Ω be a bounded domain in $\mathbb{R}^d$ with Lipschitz boundary. Let $f \in W_p^l(\mathbb{R}^d)$, $1 \le p \le \infty$, $l > d/2$. Here l is a natural number, and the Sobolev space $W_p^l(\mathbb{R}^d)$ will be defined in subsection 2.1. Suppose that a data set $\{(x_j, f(x_j))\}_{j=1}^N$ has been acquired in which $\Xi_N := \{x_j\}_{j=1}^N$ is a discrete subset of Ω.

*Received by the editors June 5, 2019; accepted for publication (in revised form) May 4, 2020; published electronically July 6, 2020.
https://doi.org/10.1137/19M1266496
Funding: The work of the authors was supported by the National Natural Science Foundation of China grants 11871074, 11501006, 61672032, the NSFC Key Project grants 11631015, 91330201, the SGST grant 12DZ 2272800, the Joint Research Fund by the National Natural Science Foundation of China and Research Grants Council of Hong Kong grant 11461161006, the Fund of Introducing Leaders of Science and Technology of Anhui University grant J10117700057, the 4th Project of Cultivating Backbone of Young Teachers of Anhui University grant J01005138, and the Anhui Provincial Science and Technology Major Project grant 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 Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO 80401 (wenwugao528@163.com).
‡Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO 80401 (fasshauer@mines.edu).
§Corresponding author. Department of Mathematics, Missouri State University, Springfield, MO 65897 (XSun@MissouriState.edu).
¶School of Mathematical Sciences, Fudan University, Shanghai, People's Republic of China (15110180025@fudan.edu.cn).

In the current article, we study the approximation of f on Ω by $Q_{\psi,h}(f)(x)$ defined by

$$Q_{\psi,h}(f)(x) := \sum_{j=1}^{N} f(x_j)\psi_{h,N}(x - x_j), \quad h > 0, \quad x \in \Omega.$$

Here the function $\psi \in W_p^l(\mathbb{R}^d)$ $(1 \le p \le \infty,\ l > d/2)$ is prescribed, and $\psi_{h,N}(x) := N^{-1}\, h^{-d}\, |\Omega|\, \psi(x/h)$. In the literature, the approximation scheme considered herein is often referred to as quasi-interpolation based on Ξ_N; see [9, 48]. If the set Ξ_N is given in a deterministic fashion (e.g., a low discrepancy set), then $Q_{\psi,h}$ is called a quasi-interpolation operator; if Ξ_N is generated by a set of random variables $\{X_j\}_{j=1}^{M}$ $(M \le N)$ distributed in Ω, then $Q_{\psi,h}$ is called a *stochastic* quasi-interpolation operator; see [50].

Literature abounds in the study of quasi-interpolation based on gridded data; see [7, 15] and the references therein. We particularly mention here that Schoenberg [40] proposed a simple quasi-interpolation scheme in which all the basis functions are just translations and dilations of a given kernel. The action of the translation group on the grid lends itself to an array of powerful mathematical machineries (e.g., Fourier transform) in this area, culminating in the so called "Strang–Fix conditions," characterizing completely the approximation order of the underlying quasi-interpolation scheme [32]. In comparison to quasi-interpolation on a grid, the body of literature on quasi-interpolation on irregularly spaced data pales in both size and richness of results which we take as an opportunity to review in the following. Beatson and Powell [3] constructed three univariate multiquadric quasi-interpolation schemes defined on a bounded interval based on irregularly spaced data. Wu and Schaback [49] improved Beatson and Powell's results by introducing a new multiquadric quasi-interpolation scheme. They also investigated convergence rates and shape-preserving properties of these multiquadric quasi-interpolation schemes. Ling's multilevel scheme for multiquadric quasi-interpolation [35] also ensures shape preservation. Wu, Sun, and Ma [50] extended the classical Bernstein polynomial approximation to include irregularly spaced data sites, and derived both deterministic and stochastic error estimates. By introducing a multiquadric trigonometric kernel, Gao and Wu [20] extended multiquadric quasi-interpolation to univariate periodic data with irregularly spaced centers. Their results include (both deterministic and stochastic) approximation orders of all derivatives (up to a certain order) of a target function. They also studied stability of the algorithms that implement the quasi-interpolation schemes; see [21] and [23]. Backus and Gilbert [2] proposed a general quasi-interpolation framework for linear functionals, including discrete function values at irregularly spaced data sites as a special case. Later, Bos and Salkauskas [5] showed that the Backus–Gilbert approach also works in quasi-interpolation schemes associated with moving least squares [31, 34]. However, the Backus–Gilbert approach entails solving a minimization problem at each prediction point, which has somewhat hindered its implementation. Buhmann, Dyn, and Levin [8] constructed quasi-interpolants for multivariate scattered data approximation via radial basis functions with quasi-uniform centers. Dyn and Ron [16] introduced the "two-step construction" to deal with the mathematical difficulties stemming from quasi-interpolation at scattered sites. More precisely, they built a quasi-interpolant based on gridded data as an intermediate step to obtain a quasi-interpolant based on scattered centers. This approach was further generalized by Yoon [51]. Wu and Liu [48] constructed quasi-interpolation for multivariate scattered data by viewing quasi-interpolation as a two-step procedure: first convolution and then discretization of the

convolution integrals. But the discretization utilizes the hypervolumes of Dirichlet cells of all the data sites, which amounts to solving a large scale moving least squares problem.

Recently, Gao et al. [19] took a stochastic viewpoint to study quasi-interpolation for irregularly spaced data. Their effort was in part motivated by the works on probabilistic numerics [26] and on stochastic Bernstein approximation [50]. Gao et al. first approximated a Sobolev space function by a scaled integral operator, and then used the Monte Carlo interaction to discretize the convolution integral. The procedure decomposes the approximation error into two parts: the convolution error and the discretization error. The former can be derived using moment conditions [10] for the convolution kernel [9, 32], and the latter relies on the theory of the Monte Carlo integration method. In the process, introducing the convolution operators is solely for the purpose of facilitating error analysis. Implementation of the quasi-interpolation scheme does not require any numerical handling of the involved convolution integral.

In the current article, we set up a theoretical framework to study the optimality and regularization properties of the underlying quasi-interpolation scheme. We also investigate conditions on kernels, sampling centers, and weights for the quasi-interpolation scheme so that a desirable rate of convergence is achieved. Our starting point is to construct a broad class of quasi-interpolants $Q_h^* f$ in a rational formulation (i.e., (3.1)) that include almost all of those discussed in the aforementioned references as well as those in the research field of kernel regression. Moreover, the rational formulation allows us to prove these quasi-interpolants' optimality property from either the kernel regression perspective [36, 46] or the weighted approximation perspective [12, 39, 45]. To obtain a desirable convergence rate for our quasi-interpolants, we make additional assumptions on the distribution of these data sites by viewing them either as independent copies of a random variable distributed in the given domain, or as point sets with low L^q-discrepancy (cf. inequality (2.7)).

We study the regularization property of the quasi-interpolation scheme as a two-step approximation procedure. We first approximate a target function by a scaled convolution integral operator, and then make use of a probabilistic integration method to approximate the convolution integral. Such a two-step procedure naturally leads to a bias-variance decomposition of the mean squared error of our quasi-interpolants, which in essence treats the regularization property of quasi-interpolants as a delicate balance between its bias and variance. In principle, this is akin to the general statistical concept of regularization proposed by Bickel and Li [4]. On the one hand, the bias and variance tradeoff viewpoint of regularization identifies the scale parameter in the scaled kernel as a regularization parameter controlling a balance between bias (approximation accuracy) and variance (prediction ability) of quasi-interpolation. On the other hand, this two-step procedure assists us to efficiently select weights and sampling centers of quasi-interpolation by utilizing optimal quadrature rules that have been extensively discussed in the literature; see, e.g., [6, 13].

The current article is structured as follows. In section 2, we introduce notations and definitions, and review some important results which we use to establish ours in the current paper. These include *convolution*, *moment conditions*, and *probabilistic integration*. In section 3, we state our main quasi-interpolation schemes and study their optimality, regularization, and approximation properties. Section 4 is devoted to an exposition of numerical simulation results, including that of approximating a bore hole function in eight dimensional space, from which readers witness the simplicity, efficiency, and robustness of our quasi-interpolation scheme.

2. Preliminaries and logistics.

2.1. Convolution integral operators. Our target functions are from Sobolev spaces of integer orders [1] to comply with the requirement of the moment condition (Definition 2.1). Let $1 \le p \le \infty$, and 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

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

Let $l > d/2$ be a positive integer and $W_p^l(\mathbb{R}^d)$ 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 $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index, $|\alpha| = \sum_{j=1}^d \alpha_j$, and $D^\alpha f$ denote the distributional derivatives of f. We adopt the Sobolev norms

$$\|f\|_{W_p^l(\mathbb{R}^d)} = \sum_{0 \le |\alpha| \le l} \|D^\alpha f\|_p$$

and the Sobolev seminorm

$$|f|_{l,p} = \sum_{|\alpha| = l} \|D^\alpha f\|_p$$

for any function f in the Banach space $W_p^l(\mathbb{R}^d)$.

Let $\psi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ and $\psi_h = h^{-d}\psi(x/h)$ for $h > 0$. Given a function $f \in W_p^l(\mathbb{R}^d)$, we first define a convolution operator $\mathcal{C}_h$ as

$$\mathcal{C}_h : f \mapsto \psi_h * f$$

with

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

It is well known that $\mathcal{C}_h(f)$ is an approximate identity on $L^p(\mathbb{R}^d)$. What is more interesting to us is the following. If ψ satisfies the moment condition (to be discussed in the next subsection), then an exact approximation order of f by $\mathcal{C}_h(f)$ can be given in terms of h and the Sobolev seminorm $|f|_{l,p}$.

2.2. Moment conditions.

Definition 2.1 (moment condition). *Let k be a positive integer. We say that a function ψ satisfies the moment condition of order k if*

$$\int_{\mathbb{R}^d} |x|^k |\psi(x)| dx < \infty \quad \text{and} \tag{2.1}$$

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

Here $\delta_{\alpha,0}$ is a Kronecker delta symbol, i.e.,

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

Moment conditions have been reformulated in several equivalent forms. Cheney and coauthors [9, 10, 11, 32] showed that the moment condition of order k is equivalent to the fact that the convolution operator reproduces polynomials up to degree $k-1$. Wu and Liu [48] characterized moment conditions in the frequency domain as generalized Strang–Fix conditions.

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) $\mathcal{F}$ and $\mathcal{F}^{-1}$ for a function $f \in L^1(\mathbb{R}^d)$ is given by

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \xi \cdot x} dx, \quad \mathcal{F}^{-1}(f)(x) = \int_{\mathbb{R}^d} f(\xi) e^{2\pi i \xi \cdot x} d\xi.$$

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

$$f(x) = \mathcal{F}^{-1}(\mathcal{F}(f)(\xi))(x) = \int_{\mathbb{R}^d} \mathcal{F}(f)(\xi) e^{2\pi i \xi \cdot x} d\xi.$$

In this case, both f and $\mathcal{F}(f)$ are continuous on $\mathbb{R}^d$. If ψ satisfies the moment condition of order k, then for every α with $|\alpha| \le k$, $D^\alpha \mathcal{F}(\psi)$ is continuous on $\mathbb{R}^d$, and (2.2) implies that

$$D^\alpha \mathcal{F}(\psi)(0) = \delta_{\alpha,0}, \ 0 \le |\alpha| \le k-1. \tag{2.3}$$

If ψ satisfies the decay condition as shown in (2.1), then (2.3) also implies the moment condition of order k as shown in (2.2). As such, (2.3) is referred to as the generalized Strang–Fix condition in Wu and Liu [48]. Equation (2.3) offers a convenient way to verify if a given ψ satisfies the moment conditions (2.2). If $\psi \in L^1(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} \psi(x) dx = 1,$$

then ψ satisfies the moment condition of order $k = 1$. Making use of (2.3), some authors have applied various techniques to construct functions that satisfy higher-order moment conditions [22, 48].

The following result was proved in [32, Theorem 2.1].

LEMMA 2.2. *Assume that ψ satisfies the moment condition of order k. Let l be a positive integer. Then the following error estimate*

$$\|\mathcal{C}_h(f) - f\|_p \le C_{m,d} |f|_{m,p} h^m, \quad f \in W_p^l(\mathbb{R}^d),$$

holds true. Here $m = \min\{k, l\}$ and $C_{m,d}$ is a positive constant that is independent of h.

2.3. Discretization of convolution integrals. In this subsection, we will introduce two methods utilized in section 3 to discretize the convolution integral in Lemma 2.2. Let Ω be a bounded domain with Lipschitz boundary and $\mathcal{B}$ the Borel sigma algebra on Ω. We declare here once and for all that $(\Omega, \mathcal{B}, \mu)$ is a probability space and X is a random variable following the law μ. Let $f : \Omega \to \mathbb{R}$ be a measurable function. Write

$$\mathbb{E} f(X) := \int_\Omega f(x) d\mu(x).$$

Let N be a positive integer, and let $\{X_i\}_{i=1}^N$ be N independent copies of X. The first discretization method we use is the well-known Monte Carlo integration formula given by

$$MC[f(X)] := \frac{|\Omega|}{N}\sum_{j=1}^{N} f(X_j),$$

in which $|\Omega|$ denotes the Lebesgue measure of Ω.

Our second method is to discretize convolution integrals by using pertinent function values on a low-discrepancy set, which belongs to a large family of quadrature rules often referred to as "Quasi-Monte Carlo methods" (QMC) [14].

To elaborate, we follow [13] to introduce some notations.

Let u be a subset of $(1{:}d) := \{1, 2, \ldots, d\}$. Let x_u be the set of components $x^{(j)}$ of $x = (x^{(1)}, x^{(2)}, \ldots, x^{(d)}) \in \mathbb{R}^d$ with $j \in u$, and $x_{u^c,1}$ the point in $\mathbb{R}^d$ derived from x by replacing with 1 the components $x^{(j)}$ of x for $j \in (1{:}d) \setminus u$ (the complement of u in $(1{:}d)$). A clarification of the two symbols x_u and $x_{u^c,1}$ is in order. We treat the former merely as a symbol. For instance, in what follows, we use $\partial^{|u|} f/\partial x_u$ to denote the mixed first partial derivative of f with respect to the components contained in x_u. The latter is an element of the vector space $\mathbb{R}^d$, and hence obeys all the rules of operations in the vector space $\mathbb{R}^d$. Let $\mathbb{1}_{[0,x]}$ be the indicator function of the d-dimensional rectangle $[0, x]^d := [0, x^{(1)}] \times \cdots \times [0, x^{(d)}]$, i.e., $\mathbb{1}_{[0,x]^d}(y) = 1$, if $y \in [0, x]^d$; $\mathbb{1}_{[0,x]^d}(y) = 0$, otherwise. Let $\mathbb{P}$ be the set of sampling centers, i.e., $\mathbb{P} = \{t_j\}_{j=1}^N$. Let $\Delta_{\mathbb{P}}(x)$ be the local discrepancy function [13] of $\mathbb{P}$ in d dimensions, i.e.,

$$\Delta_{\mathbb{P}}(x) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{[0,x]^d}(t_j) - \prod_{i=1}^{d} x^{(i)}.$$

Then the L^q-discrepancy $\|\Delta_{\mathbb{P}}\|_{q,q'}$ of $\mathbb{P}$ is defined via

$$\|\Delta_{\mathbb{P}}\|_{q,q'} := \left(\sum_{u\subseteq(1:d)} \left(\int_{[0,1]^{|u|}} \left|\Delta_{\mathbb{P}}(x_{u^c,1})\right|^q dx_u\right)^{q'/q}\right)^{1/q'}, \quad 1 \le q, q' \le \infty. \tag{2.4}$$

We follow the stipulations from [13] in defining the variation $\|f\|_{p,p'}$ of f in the sense of Hardy and Krause [38] by

$$\|f\|_{p,p'} := \left(\sum_{u\subseteq(1:d)} \left(\int_{[0,1]^{|u|}} \left|\frac{\partial^{|u|}}{\partial x_u} f(x_{u^c,1})\right|^p dx_u\right)^{p'/p}\right)^{1/p'}, \quad 1 \le p, p' \le \infty. \tag{2.5}$$

The generalized Koksma–Hlawka inequality bounds the error of a QMC quadrature rule as follows (see [13, 38]):

$$\left|\int_{I^d} f(x)dx - QMC[f]\right| \le \|f\|_{p,p'}\|\Delta_{\mathbb{P}}\|_{q,q'},\ 1 \le p, p', q, q' \le \infty,\ \tfrac{1}{p}+\tfrac{1}{q} = 1,\ \tfrac{1}{p'}+\tfrac{1}{q'} = 1. \tag{2.6}$$

A key feature of the inequality (2.6) is that, for a given integrand f with $\|f\|_{p,p'} < \infty$, the error bound of the QMC method depends only on the L^q-discrepancy of sampling centers [27, 28, 29, 37, 42]. Thus, the crucial point of a QMC method is how to choose sampling centers with an optimal low L^q-discrepancy, which is compatible with the

framework of the worst-case error estimate as shown by Kuo, Schwab, and Sloan [29]. In most QMC computations, one adopts low-discrepancy sampling centers satisfying

$$\|\Delta_{\mathbb{P}}\|_{q,q'} \leq B(d)\frac{(\ln N)^d}{N} \tag{2.7}$$

with $B(d)$ being some positive constant depending only on d. Classical examples of low-discrepancy sampling centers include Halton sequences [25], Sobol′ sequences [42], digital nets [38], lattice rules [38, 41], etc. In particular, Halton sequences [25] satisfy the above inequality with $q,q' = \infty$. QMC methods have been widely used in many practical applications [14]. For example, Lemieux [33] provided applications of QMC in finance, Heinrich and Keller discussed applications of QMC in computer graphics, Kuo, Schwab, and Sloan [30, 29] applied QMC to constructing finite elements for solving stochastic partial differential equations.

3. The main results. The data sites based on which we build our quasi-interpolants can be either acquired in deterministic ways or generated by independent random variables. For clarity, we summarize our important notations for the data sites and their conventional use in the following remark.

Remark 3.1. The notion of data sites satisfying some prescribed distribution (such as uniformly random for Monte Carlo quadrature or low-discrepancy points for QMC) is needed only for our derivation of regularization parameters in section 3.2 which optimally balance bias and variance, and the error estimates we derive in that context. The optimality results of section 3.1 as well as the general discussion of our regularization property hold for arbitrarily scattered data. Therefore, to maintain notational clarity, we use $\{X_i\}_{i=1}^N$ to denote N independent copies of a random variable X, while we use $\{t_i\}_{i=1}^N$ to denote not only low-discrepancy points for QMC quadrature rules, but also a set of N scattered points. Readers can easily tell the meaning from a specific context.

Let ψ be a kernel satisfying the moment condition of order k. Let $r_h(x) = \sum_{j=1}^N \omega_j \psi_h(x - t_j)$. We are interested in the following quasi-interpolant built in a rational form:

$$Q_h^* f(x) := \sum_{i=1}^N Y_i \frac{\omega_i \psi_h(x - t_i)}{r_h(x)}. \tag{3.1}$$

Obviously, the feasibility of this quasi-interpolation scheme requires that the denominator $r_h(x)$ be nonzero.

3.1. Optimality. Suppose $\psi(x) > 0$ for all $x \in \Omega$. Let $L^2(\Omega, \psi_h)$ denote the Hilbert space with inner product

$$\langle f, g\rangle_{\psi_h} := \int_\Omega f(x)\ g(x) r_h(x) dx, \quad f, g \in L^2(\Omega, \psi_h).$$

For fixed N and h, the norm induced by the inner product $\langle\cdot,\cdot\rangle_{\psi_h}$ on $L^2(\Omega, \psi_h)$ is equivalent to that on the usual $L^2(\Omega)$. Our quasi-interpolation scheme essentially provides an optimal algorithm in which the two limit processes $N \to \infty$ and $h \to 0^+$ are coordinated so that the unit ball of $L^2(\Omega, \psi_h)$ approximates well that of $L^2(\Omega, \mu)$ for a certain class of measures μ. Here $L^2(\Omega, \mu)$ denotes the Hilbert space with inner product

$$\langle f, g\rangle_\mu := \int_\Omega f(x)\ g(x) d\mu(x), \quad f, g \in L^2(\Omega, \mu).$$

For the given set $\{t_i, Y_i\}_{i=1}^N$ of scattered data, we define the following quadratic loss functional $I_{N,h}$ on $L^2(\Omega, \psi_h)$:

$$I_{N,h}(f) := \int_\Omega \sum_{i=1}^N (Y_i - f(x))^2 \omega_i \psi_h(x - t_i) dx.$$

It is clear from the above definition that the larger the N value and smaller the h value, the more the functional $I_{N,h}$ weighs the error between Y_i and $f(x)$ for those x that are near t_i.

Let V_{N,ψ_h} denote the space $\text{span}\{\psi_h(\cdot - t_i)\}_{i=1}^N$, and V^*_{N,ψ_h} the space span $\{\psi_h(\cdot - t_i)/r_h(\cdot)\}_{i=1}^N$. The following theorem shows that the quasi-interpolant as shown in (3.1) is the unique minimizer of the functional $I_{N,h}(f)$ restricted to V^*_{N,ψ_h}.

THEOREM 3.2. *Let $\{t_i, Y_i\}_{i=1}^N$ be a given set of scattered data. Suppose that ψ is positive and that the N translates of ψ_h, $\psi_h(\cdot - t_i)$, $1 \le i \le N$, are linearly independent over $\mathbb{R}$ as functions defined on Ω. Then, the quasi-interpolant $Q_h^* f$ defined in* (3.1) *with positive weights $\{\omega_i\}_{i=1}^N$ is the unique minimizer of the functional $I_{N,h}$ over V^*_{N,ψ_h}. In particular, the following equation holds true:*

$$I_{N,h}(Q_h^* f) = \min_{g \in V^*_{N,\psi_h}} I_{N,h}(g).$$

Proof. Write a $g \in V^*_{N,\psi_h}$ as

$$g(x) = \sum_{i=1}^N c_i \frac{\psi_h(x - t_i)}{r_h(x)}.$$

Restricting $I_{N,h}$ on V^*_{N,ψ_h}, we may consider $I_{N,h}$ as a quadratic function from $\mathbb{R}^N$ to $\mathbb{R}_+$ given in the form

$$I_{N,h}[(c_1, c_2, \ldots, c_N)] = \int_\Omega \sum_{i=1}^N \left(Y_i - \sum_{j=1}^N c_j \frac{\psi_h(x - t_j)}{r_h(x)} \right)^2 \omega_i \psi_h(x - t_i) dx$$

in which the quadratic term $c_i\ c_j$ has coefficient

$$A_{ij} := \left\langle \frac{\psi_h(x - t_i)}{r_h(x)}, \frac{\psi_h(x - t_j)}{r_h(x)} \right\rangle_{\psi_h}.$$

Thus, for any fixed x, the $N \times N$ coefficient matrix $A := (A_{ij})$ of the corresponding quadratic form is a Gram matrix. Since the N functions $\psi_h(\cdot - t_i)$, $1 \le i \le N$, are linearly independent over $\mathbb{R}$ as functions defined on Ω, the Gram matrix A is positive definite. It then follows that the function $I_{N,h}$ has a unique critical point, which is also the point where $I_{N,h}$ reaches its absolute minimum. To find this point, we take partial derivatives of $I_{N,h}$ with respect to each $c_k, 1 \le k \le N$, to get

$$\frac{\partial I_{N,h}}{\partial c_k} = -2 \int_\Omega \sum_{i=1}^N \left(Y_i - \sum_{j=1}^N c_j \frac{\psi_h(x - t_j)}{r_h(x)} \right) \frac{\psi_h(x - t_k)}{r_h(x)} \omega_i \psi_h(x - t_i) dx, \quad 1 \le k \le N.$$

Setting these partial derivatives equal to zero and employing the definition of $r_h(x)$, the equations simplify to

$$\int_\Omega \frac{\psi_h(x - t_k)}{r_h(x)} \sum_{i=1}^N (\omega_i Y_i - c_i) \psi_h(x - t_i) dx = 0, \quad 1 \le k \le N.$$

Hence, the unique critical point is given by $c_i = \omega_i\ Y_i$, which is the desired result. ∎

Similarly, we can show that, at any given prediction point x, the quasi-interpolant $Q_h^* f(x)$ is the unique pointwise minimizer of the following weighted quadratic functional:

$$I_{N,h}^*[f](x) = \sum_i (Y_i - f(x))^2 \omega_i \psi_h(x - t_i), \quad f \in L^2(\Omega, \psi_h),$$

from the space V_{N,ψ_h}^*. This is another indication that the quasi-interpolation scheme is simple, robust in terms of availability, stability, and generalization ability.

3.2. Regularization. Regularization is a commonly used technique for data mining [24, 47]. It aims to find a physically meaningful and stable approximation by coupling some prior information (e.g., smoothness, sparsity, convexity) together with sampling data. Using regularization, we can often recover some of the prior information lost during the sampling process by a balancing act on bias and variance (or between approximation accuracy and generalization ability) [43].

Smoothness is a natural and arguably the most popular prior information in the regularization literature. It can be imposed both explicitly and implicitly [43]. In the celebrated Tikhonov's regularization scheme [44], smoothness was imposed explicitly by adding a roughness penalty term to a goodness-of-fit measure. The final approximation is obtained by solving a penalized optimization problem. The balance between approximation accuracy and smoothness is controlled by the coefficient (regularization parameter) of the added term. We point out here that penalization is not the only form of regularization. In kernel regression [36], smoothness is imposed implicitly via a bandwidth parameter in the regression kernel. The final regressor is obtained by minimizing the mean square error, which balances a trade-off between bias and variance. In addition, approximation of a function with a truncated basis, iterated approximation, boosting, subsampling, etc., can also be viewed as regularization [43]. Bickel and Li [4] provided a general statistical concept of regularization by viewing regularization as a two-step approximation process. They first construct an approximation sequence with a parameter controlling the bias. They then select a value for the parameter so that the squared-bias term and the variance term are of the same order of magnitude, and the mean squared error is consequently minimized. This general framework includes many regularization approaches in the current literature for data fitting. For example, the commonly used Tikhonov regularization can be viewed as a procedure in which one first constructs an approximant

$$f_\lambda^*(x) := (\phi_1(x), \phi_2(x), \ldots, \phi_N(x))(A + \lambda I)^{-1}(f(x_1), f(x_2), \ldots, f(x_N))^T$$

and then chooses an appropriate parameter λ to balance the bias and variance of f_λ^*. Here $\{\phi_j\}_{j=1}^N$ is a basis for the prescribed approximating function space, and I is an identity matrix and $A = (\phi_j(x_i))$. One can take a similar view on the iterated quasi-interpolation proposed by Fasshauer and Zhang [18] in which an approximation sequence

$$f_M^*(x) := (\phi_1(x), \phi_2(x), \ldots, \phi_N(x)) \sum_{i=0}^{M} (I - A)^i (f(x_1), f(x_2), \ldots, f(x_N))^T$$

is constructed first and then an appropriate nonnegative integer M is selected to achieve the balance between the bias and the variance. In this subsection, we show that our quasi-interpolants in the rational form have regularization properties [43] comparable to those described in [4], in which the scale parameter of the scaled kernel of the quasi-interpolation scheme plays the role of the regularization parameter. For

this purpose, we consider the quasi-interpolant $Q_h^* f$ in the context of a general form of kernel regression, and derive the mean squared error of our quasi-interpolation scheme. However, instead of working within the framework of kernel regression to get the error estimate, which requires more stringent conditions on both the sampling data and the kernel, we follow the two-step approximation procedure advocated by Gao et al. [19] in which they first approximate a target function by a convolution operator, and then discretize the underlying convolution integral to obtain the quasi-interpolants.

To this end, we define the following alternative form of a quasi-interpolant:

$$Q_h f(x) = \sum_j Y_j \omega_j \psi_h(x - t_j). \tag{3.2}$$

We note that $Q_h f$ differs from $Q_h^* f$ by the factor $r_h^{-1}(x)$. For $1 \le p \le \infty$, let $\|f\|_{p,\Omega} := \|f \cdot \mathbb{1}_\Omega\|_p$. We have

$$\begin{aligned} \|Q_h^* f - f\|_{p,\Omega} &\le \|Q_h^* f - Q_h f\|_{p,\Omega} + \|Q_h f - f\|_{p,\Omega} \\ &\le \|Q_h^* f\|_{p,\Omega} \cdot \|r_h - 1\|_{p,\Omega} + \|Q_h f - f\|_{p,\Omega}. \end{aligned} \tag{3.3}$$

If $r_h(x) = \sum_j \omega_j \psi_h(x - t_j) \equiv 1,\ x \in \Omega$, then $Q_h^* f = Q_h f$, and the inequalities above become equality. Suppose that ψ is positive and that $f \in W_p^l(\mathbb{R}^d)\,(l > d/2)$. Then for any $1 \le p \le \infty$, we have

$$\|Q_h^* f\|_{p,\Omega} \le |\Omega|^{1/p}\ \|Q_h^* f\|_{\infty,\Omega} \le \frac{|\Omega|^{1/p}\ \|f\|_{\infty,\Omega}}{r_h(x)}\ \sum_{j=1}^N \omega_j \psi_h(x - t_j) = |\Omega|^{1/p}\ \|f\|_{\infty,\Omega}. \tag{3.4}$$

Suppose now that we have established an approximation order in the form

$$\|Q_h f - f\|_{p,\Omega} = \mathcal{O}(h^\alpha), \quad f \in W_p^l(\mathbb{R}^d) \quad (l > d/2),$$

where α is a positive constant. Then it follows from inequality (3.3) that

$$\|Q_h^* f - f\|_{p,\Omega} = \mathcal{O}(h^\alpha), \quad f \in W_p^l(\mathbb{R}^d) \quad (l > d/2).$$

This allows us to dominate the approximation error of the quasi-interpolant $Q_h^* f$ by that of $Q_h f$, which is advantageous to our analytical setup. At each given prediction point x, we view the sampling centers $\{X_i\}_{i=1}^N$ as N independent copies of the random variable X uniformly distributed on Ω. It follows that the expectation of the random variable (the quasi-interpolant) $Q_h f(x)$ is

$$\mathbb{E}\left[Q_h f(x)\right] = \mathcal{C}_{h,\Omega}(f)(x) := \int_\Omega \psi_h(x - y) f(y) dy.$$

LEMMA 3.3 (bound of the bias). *Suppose that ψ satisfies the moment condition of order k. Let V be a bounded domain satisfying $V \subset \Omega$ and $\operatorname{dist}(\partial\Omega, \partial V) \ge c$ for a fixed $0 < c < 1$. Here $\operatorname{dist}(\partial\Omega, \partial V)$ denotes the Hausdorff distance between $\partial\Omega$ and ∂V. Then, for any $f \in W_p^l(\mathbb{R}^d),\ 1 \le p \le \infty,\ l > d/2$, there exists a constant C that is independent of h such that*

$$\|\mathcal{C}_{h,\Omega}(f) - f\|_{p,V} \le C\ h^m, \quad m = \min\{k, l\}. \tag{3.5}$$

Proof. We first use the Minkowski inequality to write

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

Here we remind readers 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 2.2 implies that there is a constant C_1 independent of h such that

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

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

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

We carry out the substitution $y = (x-t)/h$ in the integral above to continue:

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

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

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

where C_2 is a constant independent of h and $x \in V$ and, furthermore, that

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

and therefore that

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

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

The result of Lemma 3.3 shows that in order to build a quasi-interpolant of the form $Q_h f$ it is sufficient to discretize the underlying convolution integral $\mathcal{C}_h(f)(x)$ on the bounded domain Ω rather than the whole space $\mathbb{R}^d$. A noticeable blemish is that the approximation error is only obtained on V, a slightly smaller region than the domain Ω. In the following, we will focus on discretizing $\int_\Omega f(t)\psi_h(x-t)dt$.

Method 1: Monte Carlo quadrature rule. Let $\{X_i\}_{i=1}^N$ be N independent copies of the random variable X uniformly distributed on Ω. For a fixed $x \in V$, we discretize $\int_\Omega f(t)\psi_h(x-t)dt$ via

$$MC[f](x) := \frac{|\Omega|}{N} \sum_{j=1}^{N} f(X_j)\psi_h(x - X_j),$$

which yields the following result. To make the statement of our result less ponderous, we will write C as a universal constant whose value may vary from line to line.

THEOREM 3.4 (bound of the mean squared error). *Suppose that ψ satisfies the moment condition of order k and that $f \in W_2^l(\mathbb{R}^d)$ with $l > d/2$. Let $\{X_i\}_{i=1}^N$ be N independent copies of the random variable X uniformly distributed on Ω, and let V be as described in Lemma* 3.3. *Then there exists a positive constant C independent of h and N, such that the following inequality holds true:*

$$\mathbb{E}\,\|Q_h f - f\|_{2,V}^2 \le C\left(h^{2m} + \frac{1}{Nh^d}\right), \quad m = \min\{l, k\}. \tag{3.6}$$

If we, in particular, select $h = (1/N)^{1/(2m+d)}$, then inequality (3.6) *yields the following estimate:*

$$\mathbb{E}\|Q_h f - f\|_{2,V}^2 \le C\ N^{-\frac{2m}{2m+d}}. \tag{3.7}$$

Lemma 3.3 and Theorem 3.4 have appeared in [19]. We include a proof of Lemma 3.3 here for assurance of application, as the result will be used in the remainder of this section.

Remark 3.5. As stated in terms of the bias-variance error decomposition, h^{2m} is the bias error term, and $\frac{1}{Nh^d}$ the variance error term. The choice of the scale parameter value $h = (1/N)^{1/(2m+d)}$ is an optimal trade-off between the bias error term and the variance error term.

Remark 3.6. Although Theorem 3.4 deals with the uniformly distributed random variable X, we can use a similar argument to show that the same result of Theorem 3.4 is true if we replace X by a random variable Z with law μ that is absolutely continuous with respect to the Lebesgue measure on Ω, and concentrated on V.

Method 2: QMC quadrature rule. The discussion herein follows closely the narrative contained in the second half of subsection 2.3. Here we assume that $\Omega = I^d := [0,1]^d$ is the d-dimensional unit cube. (In Lemma 3.3 below, we will use the notation $I_c^d := [c, 1-c]^d$ for a fixed $0 < c < 1/2$.) Accordingly, we discretize $\int_{I^d} f(t)\psi_h(x-t)dt$ using

$$QMC[f](x) := N^{-1}\sum_{j=1}^{N} f(t_j)\psi_h(x - t_j).$$

Here $\{t_j\}_{j=1}^N$ is a prescribed set of sampling centers whose L^q-discrepancy satisfies inequality (2.7).

PROPOSITION 3.7 (bound of the variance using the QMC quadrature rule). *Suppose that $\psi \in W_\infty^l(\mathbb{R}^d)$ and $f \in W_p^l(\mathbb{R}^d)$, $l \ge d$. Assume that $\{t_j\}_{j=1}^N$ is a point set of sampling centers whose L^q-discrepancy satisfies inequality* (2.7), *$1/p + 1/q = 1$. Then there exists a positive constant C independent of h and N, such that*

$$\sup_{x\in\mathbb{R}^d}\left|\int_{I^d} f(t)\psi_h(x-t)dt - Q_h f(x)\right| \le C\ \frac{\ln^d N}{N}\ h^{-(q+1)d/q}. \tag{3.8}$$

Proposition 3.7 follows outright from inequalities (2.6), (2.7), and the results in Lemma 3.8 (below) which gives an upper bound for the total variation of $f\psi_h$ in the sense of Hardy and Krause, which we will simply refer to as "the variation of $f\psi_h$"; see (2.5).

LEMMA 3.8. *Suppose that $\psi \in W_\infty^l(\mathbb{R}^d)$, and $f \in W_p^l(\mathbb{R}^d)$, $l \geq d$. Then there exists a constant $C > 0$ independent of h, so that the variation of $f\psi_h$ satisfies the following inequality:*

$$\sup_{x\in\mathbb{R}^d} \|f(\cdot)\psi_h(x-\cdot)\|_{p,p'} \leq C\ h^{(1-2p)d/p}, \quad 1 \leq p, p' \leq \infty.$$

Proof. Fix an $x \in \mathbb{R}^d$ and a pair of $1 \leq p, p' \leq \infty$. We use the definition of variation (2.4) to write

$$\|f(\cdot)\psi_h(x-\cdot)\|_{p,p'} = \left(\sum_{u\subseteq(1:d)} \left(\int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|}}{\partial t_u} \left(f(t_{u^c,1})\psi_h(x - t_{u^c,1}) \right) \right|^p dt_u \right)^{p'/p} \right)^{1/p'}.$$

Applying the Leibniz product rule for differentiation, Jensen's inequality, and the substitution: $z = (x - t_{u^c,1})/h$, we derive

$$\begin{aligned}
&\|f(\cdot)\psi_h(x-\cdot)\|_{p,p'} \\
&\quad := \left(\sum_{u\subseteq(1:d)} \left(\int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|}}{\partial t_u} \left(f(t_{u^c,1})\psi_h(x - t_{u^c,1}) \right) \right|^p dt_u \right)^{\frac{p'}{p}} \right)^{\frac{1}{p'}} \\
&\quad = \left(\sum_{u\subseteq(1:d)} \left(\int_{[0,1]^{|u|}} \left| \sum_{v\subseteq u} \binom{|u|}{|v|} \frac{\partial^{|v|}}{\partial t_v} f(t_{u^c,1}) \frac{\partial^{|u|-|v|}}{\partial t_{u\setminus v}} \psi_h(x - t_{u^c,1}) \right|^p \right. \right. \\
&\qquad\qquad \leq 2^{\frac{p-1}{p}}\ h^{-d} \left(\sum_{u\subseteq(1:d)} h^{|u|(1-p)/p} \sum_{v\subseteq u} h^{|v|} \binom{|u|}{|v|} \right. \\
&\qquad\qquad\qquad \left. \left. \times \left(\int_{\mathbb{R}^{|u|}} \left| f^{(|v|)}(x - hz)\psi^{(|u|-|v|)}(z) \right|^p dz_u \right)^{\frac{p'}{p}} \right)^{\frac{1}{p'}} \\
&\qquad\qquad \leq 2^{\frac{p-1}{p}}\ h^{(1-2p)d/p} \sum_{u\subseteq(1:d)} \sum_{v\subseteq u} \binom{|u|}{|v|} \\
&\qquad\qquad \times \left(\int_{\mathbb{R}^{|u|}} \left| f^{(|v|)}(z)\psi^{(|u|-|v|)}(x - hz) \right|^p dz_u \right)^{\frac{p'}{p}} \Bigg)^{\frac{1}{p'}}.
\end{aligned}$$

Since $\psi \in W_\infty^l(\mathbb{R}^d)$, $f \in W_p^l(\mathbb{R}^d)$, $l \geq d$, the integral

$$\left(\int_{\mathbb{R}^{|u|}} \left| f^{(|v|)}(z)\psi^{(|u|-|v|)}(x - hz) \right|^p dz_u \right)^{1/p}$$

is bounded above for any $v \subseteq u \subseteq (1:d)$. The above estimate is independent of $x \in \mathbb{R}^d$, and the desired result follows. ∎

Lemma 3.3 and Proposition 3.7 imply the following theorem.

THEOREM 3.9. *Suppose that ψ satisfies the moment condition of order k and that $\psi \in W_\infty^l(\mathbb{R}^d)$ and $f \in W_p^l(\mathbb{R}^d)$ $(l \geq d)$. Assume that $\{t_j\}_{j=1}^N$ is a point set of sampling centers whose L^q-discrepancy satisfies inequality (2.7), $1/p + 1/q = 1$. Then for any r in the range $1 \leq r \leq \infty$, there exists a constant C independent of h and N, such that*

$$\|Q_h f - f\|_{r,I_c^d} \leq C \left(h^m + \frac{\ln^d N}{N} h^{-(q+1)d/q} \right), \quad m = \min\{l, k\}. \tag{3.9}$$

If we, in particular, select $h = (\ln^d N/N)^{q/(mq+(q+1)d)}$, then we have

$$\|Q_h f - f\|_{r,I_c^d} \leq C \left(\frac{\ln^d N}{N} \right)^{mq/(mq+(q+1)d)}. \tag{3.10}$$

Remark 3.10. In Theorem 3.9, the two parameters r and q can vary independently in the range $[1, \infty]$. Now suppose that $1/p + 1/q = 1$. The optimal upper bound for the Hardy–Krause variation in Lemma 3.8 is given by $p = 1$. The optimal error estimate provided by Theorem 3.9 is given by $q = \infty$, which yields

$$\|Q_h f - f\|_{r,I_c^d} \leq C \left(\frac{\ln^d N}{N} \right)^{m/(m+d)}, \quad f \in W_1^l(\mathbb{R}^d), \quad 1 \leq r \leq \infty, \quad l \geq d.$$

In bounding the variance error term in Lemma 3.3, we have chosen a QMC quadrature rule based on a point set with L^q-discrepancy satisfying inequality (2.7). To accelerate the convergence rate, one can adopt recently developed lattice rules or digital nets to discretize the underlying convolution integral. For more details on QMC quadrature rules, we refer readers to the recent review paper [13].

4. Numerical simulations. This section provides several numerical simulation results to support the theoretical analysis in the previous sections. Our simulations pertain to examples discussed in [17] (i.e., sections 16.3.1 and 16.3.2) to substantiate the validity of our quasi-interpolation scheme. We choose the unit cube $[0,\ 1]^d$ ($d = 7$ in Example 1 and $d = 8$ in Example 2) as the domain for our simulations. We compute function values at N random sampling centers over $[0, 1]^d$ for the Monte Carlo (MC) method or N Halton points for the QMC method, and use them as sampling data to construct quasi-interpolants using (i) the standard Gaussian kernel

$$\psi_2(x) = \left(\frac{1}{\sqrt{2\pi}} \right)^d e^{-x^2/2},$$

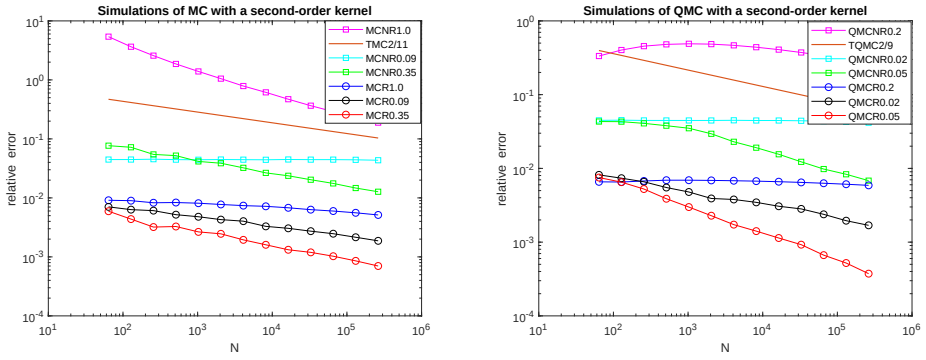
and (ii) a modified Gaussian kernel

$$\psi_4(x) = \frac{9\psi_2(x) - 3^d \psi_2(x/2)}{8}.$$

We note that ψ_2 satisfies the moment condition of order 2 and ψ_4 of order 4.

Example 1: Approximation of the seven-dimension piston function. The seven-dimension piston function takes the composite form via a chain of nonlinear functions [17]

$$C(x) = 2\pi \sqrt{\frac{W}{m + S^2 \frac{P_0 V_0}{T_0} \frac{T_a}{V^2}}}.$$



(a) (b)

FIG. 1. *Numerical simulations of $Q_h^* f$ and $Q_h f$ using MC and QMC with a standard Gaussian kernel under different choices of C_1 and C_2.*

Here the function

$$V = \frac{S}{2m}\left(\sqrt{A^2 + 4m\frac{P_0 V_0}{T_0}T_a} - A\right), \quad A = P_0 S + 19.62M - \frac{mV_0}{S},$$

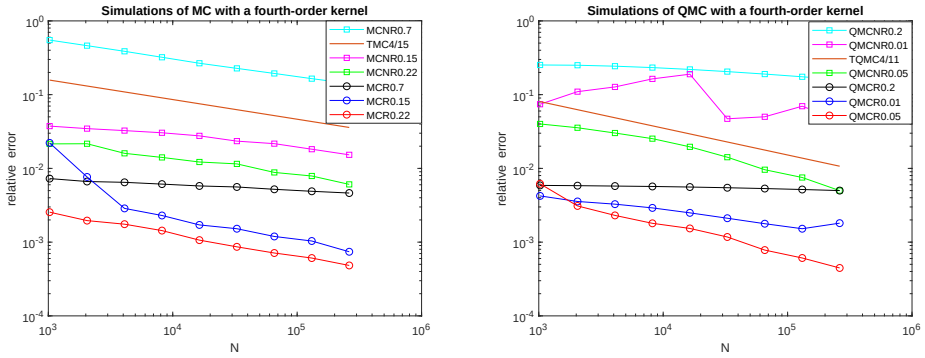
models the circular motion of a piston within a cylinder and consequently the piston function C is the time it takes to complete one cycle, in seconds. The input variable $x = (W, S, V_0, m, P_0, T_a, T_0)$ has the following usual input ranges:

$W \in [30,\ 60]$,	piston weight (kg),
$S \in [0.005,\ 0.020]$,	piston surface area (m^2),
$V_0 \in [0.002,\ 0.010]$,	initial gas volume (m^3),
$m \in [1000,\ 5000]$,	spring coefficient (N/m),
$P_0 \in [90000,\ 110000]$,	atmospheric pressure (N/m^2),
$T_a \in [290,\ 296]$,	atmospheric pressure (K),
$T_0 \in [340,\ 360]$,	filling gas temperature (K).

To compare $Q_h^* f$ with $Q_h f$, we employ both of them to approximate the target function using, respectively, the MC method and the QMC method. We choose $h = C_1(1/N)^{1/(2m+d)}$ for the MC method and $h = C_2(\ln^d N/N)^{1/(m+d)}$ for the QMC method. We exercise caution to choose the values of the positive constants C_1 and C_2 for both ψ_2 (satisfying the second-order moment condition) and ψ_4 (satisfying the fourth-order moment condition). Note that Halton points satisfy inequality (2.7) with $q = \infty$ [28]. These results are shown in Figures 1 and 2.

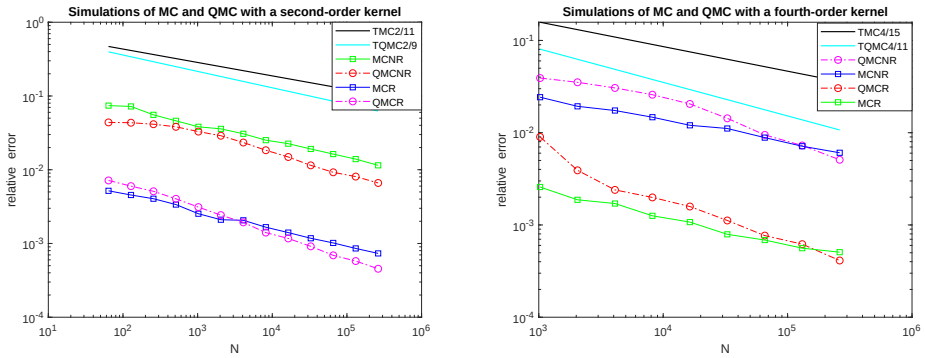
In the figures, TMC ($TQMC$) denotes the line whose slope is the theoretical approximation order of quasi-interpolation with the MC (QMC) method. $MCRconst$ ($MCNRconst$) denotes relative root mean squared errors of $Q_h^* f$ ($Q_h f$) with the MC (QMC) method under different values of C_1, while $QMCRconst$ ($QMCNRconst$) denotes corresponding ones of $Q_h^* f$ ($Q_h f$) with the QMC method and different values of C_2. Others are similar. Relative root mean squared errors are computed at 500 random prediction points over the domain $[0.1,\ 0.9]^d$, under $N = 2^i,\ i = 6, 7, \ldots, 18$, for Figure 1 and $N = 2^i,\ i = 10, 11, \ldots, 18$, for Figure 2.

These figures demonstrate clearly that the constants C_1 and C_2 do influence approximation accuracy. While there is no theoretical guidance, we settle upon optimal



(a) (b)

FIG. 2. *Numerical simulations of $Q_h^* f$ and $Q_h f$ with a fourth-order kernel ψ_4 under different choices of C_1 and C_2.*



(a) (b)

FIG. 3. *Numerical simulations of quasi-interpolation $Q_h^* f$ and $Q_h f$ with the MC method and the QMC method for the piston function.*

choices of C_1 and C_2 by trial and error. To compute the a posteriori convergence order, we first compute empirical convergence orders using approximation errors of each two consecutive N's. Then, we compute the final a posteriori convergence order based on these empirical ones using the least squares method. Corresponding numerical results are provided in Figure 3.

In Figure 3(a), the regularization parameter h is chosen as $h = 0.35N^{-1/(2m+d)}$ for the MC method and $h = 0.05((\ln N)^d/N)^{1/(m+d)}$ for the QMC method with $m = 2$, $d = 7$, and N ranging from 2^6 to 2^{18}. Relative root mean squared errors at 500 random test points under N are provided in this subfigure. The slopes of the black and cyan lines are theoretical convergence orders 2/11 for MC and 2/9 for QMC based on Theorem 3.4. The final a posteriori approximation order of $Q_h^* f$ with the MC method is 0.2462, while that of $Q_h f$ with the MC method is 0.2153. The rates of convergence shown here seem to be better than what Lemma 3.3 has determined which is 2/11. Besides, the final a posteriori approximation order of $Q_h^* f$ with the QMC method is 0.3431, while that of $Q_h f$ with the QMC method is 0.2498. These two orders are all higher than the theoretically derived order 2/9.

In Figure 3(b), the regularization parameter h is chosen as $h = 0.22N^{-1/(2m+d)}$ for the MC method and $h = 0.05((\ln N)^d/N)^{1/(m+d)}$ for the QMC method with $m = 4$, $d = 7$, and N ranging from 2^{10} to 2^{18}. In this subfigure, the slope of the black line is the theoretical convergence order 4/15 for MC, while that of the cyan line is the theoretical convergence order 4/11 for QMC. The final a posteriori approximation order of $Q_h^* f$ with the MC method is 0.2944, while that of $Q_h f$ with the MC method is 0.2676. These two orders are both higher than the theoretically derived order 4/15. Besides, the final a posteriori approximation order of $Q_h^* f$ with the QMC method is 0.4407, while the one of $Q_h f$ with the QMC method is 0.3744. These two orders are larger than the theoretically derived order 4/11.

Remark 4.1. Based on these simulation results as shown in the above three figures, we have the following observations:

- The constants C_1 and C_2 influence numerical results of both $Q_h^* f$ and $Q_h f$.
- Both a posteriori approximation orders of $Q_h^* f$ and $Q_h f$ are higher than the theoretically derived orders, which validifies the efficiency of the underlying quasi-interpolants.
- Quasi-interpolation $Q_h^* f$ usually performs better than $Q_h f$ in terms of approximation errors, a posteriori approximation orders, and sensitivity to choices of C_1 and C_2.
- Both quasi-interpolants $Q_h^* f$ and $Q_h f$ made with kernel ψ_4 (satisfying the fourth-order moment condition) perform exceptionally well when employed to approximate the piston function.

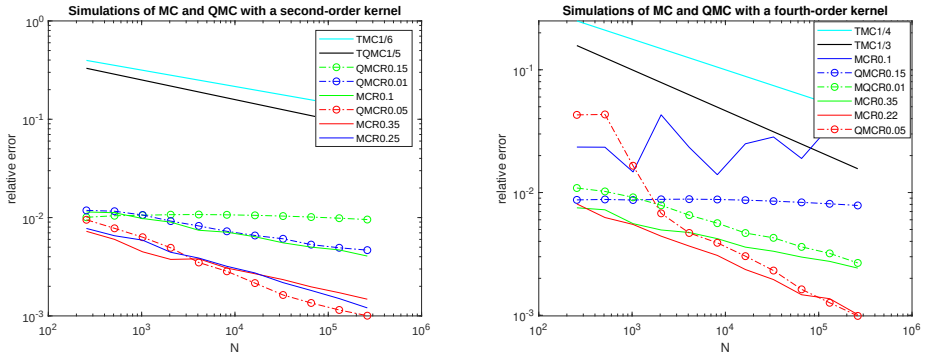
Example 2: Approximation of the eight-dimension borehole function. We adopt the notations in [17] and define the eight-dimension borehole function in the form

$$f(x) = \frac{2\pi T_u(H_u - H_\ell)}{\ln\left(\frac{r}{r_w}\right)\left(1 + \frac{2LT_u}{\ln\left(\frac{r}{r_w}\right)r_w^2 K_w} + \frac{T_u}{T_\ell}\right)}.$$

This function models the flow of water through a borehole in an impermeable layer of rock separating an upper from a lower aquifer. It has been commonly used as a function for testing numerical algorithms in computer experiments due to its simplicity and quick evaluation. The input variable $x = (r_w, r, T_u, H_u, T_\ell, H_\ell, L, K_w) \in \mathbb{R}^8$ has the following usual input ranges:

$r_w \in [0.05, 0.15]$,	radius of borehole (m),
$r \in [100, 50000]$,	radius of influence (m),
$T_u \in [63070, 115600]$,	transmissivity of upper aquifer (m^2/yr),
$H_u \in [990, 1110]$,	potentiometric head of upper aquifer (m),
$T_\ell \in [63.1, 116]$,	transmissivity of lower aquifer (m^2/yr),
$H_\ell \in [700, 820]$,	potentiometric head of lower aquifer (m),
$L \in [1120, 1680]$,	length of borehole (m),
$K_w \in [9855, 12045]$,	hydraulic conductivity of borehole (m/yr).

The output is the water flow rate, in m^3/yr. For this particular target function, initial simulation results (not reported here) show that $Q_h^* f$ performs better than $Q_h f$. As such, we herein only provide the numerical simulation results of $Q_h^* f$ when employed to approximate the borehole function; see Figure 4.



(a) (b)

FIG. 4. *Numerical simulations of quasi-interpolation $Q_h^* f$ and $Q_h f$ with the MC method and the QMC method for the borehole function.*

We choose $h = C_1(1/N)^{1/(2m+d)}$ for the MC method and $h = C_2(\ln^d N/N)^{1/(m+d)}$ for the QMC method, and exercise caution in selecting different values for the constants C_1 and C_2. TMC ($TQMC$) denotes the line whose slope is the theoretical approximation order of quasi-interpolation with the MC (QMC) method. $MCRconst$ ($QMCRconst$) denotes relative root mean squared errors of $Q_h^* f$ with the MC (QMC) method and different values of C_1 (C_2). Relative root mean squared errors are computed at 500 random prediction points over the domain $[0.1,\ 0.9]^d$ under $N = 2^i,\ i = 8, 9, \ldots, 18$.

The results of the figure demonstrate in a similar fashion that the values of constants C_1, C_2 influence the numerical results. Again, we choose the optimal constant values by trial and error. In Figure 4(a), the regularization parameter h is set to be $h = 0.25N^{-1/(2m+d)}$ for the MC method and $h = 0.05((\ln N)^d/N)^{1/(m+d)}$ for the QMC method with $m = 2$ and $d = 8$. The final a posteriori approximation order of $Q_h^* f$ with the MC (QMC) method is 0.2642 (0.3332), which is higher than the theoretically derived order 1/6 (1/5). Figure 4(b) shows numerical simulation results of quasi-interpolants built with the kernel ψ_4 (satisfying the fourth-order moment condition). The regularization parameter h is chosen as $h = 0.22N^{-1/(2m+d)}$ for the MC method and $h = 0.05((\ln N)^d/N)^{1/(m+d)}$ for the QMC method with $m = 4$, $d = 8$. The final a posteriori approximation order of $Q_h^* f$ with the MC (QMC) method is 0.2933 (0.5382), which is higher than the theoretically derived order 1/4 (1/3).

Acknowledgments. We express our gratitude to Luis Tenorio for valuable discussions that help us improve the presentation of our paper. We are also indebted to two anonymous reviewers who have pointed out a number of oversights and inaccuracies in the initial draft of the manuscript, the corrections of which have enhanced the mathematical structure and the exposition of the paper. The first author thanks the Chinese Scholarship Council for sponsoring the visit to Colorado School of Mines.

REFERENCES

[1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, Pure Appl. Math. (Amst.) 140, 2nd ed., Academic Press, Amsterdam, 2003.

[2] G. BACKUS AND F. GILBERT, *The resolving power of gross earth data*, Geophys. J. R. Astronomic. Soc., 16 (1968), pp. 169–205.

[3] R. BEATSON AND M. POWELL, *Univariate multiquadric approximation: Quasi-interpolation to scattered data*, Constr. Approx., 8 (1992), pp. 275–288.

[4] J. P. Bickel and B. Li, *Regularization in statistics*, Test, 15 (2006), pp. 271–344.
[5] L. P. Bos and K. Salkauskas, *Moving least-squares are Backus-Gilbert optimal*, J. Approx. Theory, 59 (1989), pp. 267–275.
[6] F. X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic, *Probabilistic integration: A role in statistical computation?*, Statist. Sci., 34 (2019), pp. 1–22.
[7] M. Buhmann, *Convergence of univariate quasi-interpolation using multiquadrics*, IMA J. Numer. Anal., 8 (1988), pp. 365–383.
[8] M. Buhmann, N. Dyn, and D. Levin, *On quasi-interpolation by radial basis functions with scattered centres*, Constr. Approx., 11 (1995), pp. 239–254.
[9] E. W. Cheney and J. J. Lei, *Quasi-interpolation on irregular points*, in Approximation and Computation: A Festschrift in Honor of Walter Gautschi, R. V. M. Zahar, ed., Internat. Ser. Numer. Math. 119, Birkhäuser, Boston, 1994, pp. 121–135.
[10] E. W. Cheney, W. Light, and Y. Xu, *On kernels and approximation orders*, in Approximation Theory (Memphis, TN), Lect. Notes in Pure Appl. Math. 138, Dekker, New York, 1992, pp. 227–242.
[11] E. W. Cheney and W. A. Light, *A Course in Approximation Theory*, Brooks/Cole, Pacific Grove, CA, 2009.
[12] A. Cohen and G. Migliorati, *Optimal weighted least squares methods*, SMIA J. Comput. Math., 3 (2017), pp. 181–203.
[13] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer., 22 (2013), pp. 133–288.
[14] J. Dick and F. Pillichshammer, *Digital Nets and Sequences*, Cambridge University Press, Cambridge, 2010.
[15] 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.
[16] N. Dyn and A. Ron, *Radial basis function approximation: From gridded centers to scattered centers*, Proc. London Math. Soc. (3), 71 (1995), pp. 76–108.
[17] G. E. Fasshauer and M. J. McCourt, *Kernel-based Approximation Methods using* Matlab, Interdiscip. Math. Sci. 19, World Scientific, Hackensack, NJ, 2015.
[18] G. E. Fasshauer and J. G. Zhang, *Iterated approximate moving least squares approximation*, in Advances in Meshfree Techniques, Springer, Dordrecht, The Netherlands, 2007, pp. 221–240.
[19] W. W. Gao, X. P. Sun, Z. M. Wu, and X. Zhou, *Multivariate Monte Carlo Approximation Based on Scattered Data*, manuscript.
[20] W. W. Gao and Z. M. Wu, *A quasi-interpolation scheme for periodic data based on multiquadric trigonometric B-splines*, J. Comput. Appl. Math., 271 (2014), pp. 20–30.
[21] W. W. Gao and Z. M. Wu, *Approximation orders and shape preserving properties of the multiquadric trigonometric B-spline quasi-interpolant*, Comput. Math. Appl., 69 (2015), pp. 696–707.
[22] W. W. Gao and Z. M. Wu, *Constructing radial kernels with higher-order moment conditions*, Adv. Comput. Math., 43 (2017), pp. 1355–1375.
[23] W. W. Gao and R. Zhang, *Multiquadric trigonometric spline quasi-interpolation for numerical differentiation of noisy data: A stochastic perspective*, Numer. Algorithms, 77 (2018), pp. 243–259.
[24] G. Girosi, M. Jones, and T. Poggio, *Regularization theory and neural networks architectures*, Neural Comput., 7 (1995), pp. 219–269.
[25] J. H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*, Numer. Math., 2 (1960), pp. 84–90.
[26] P. Hennig, M. A. Osborne, and M. Girolami, *Probabilistic numerics and uncertainty in computations*, Proc. A, 47 (2015), 20150142.
[27] F. J. Hickernell, *Quadrature error bounds with applications to lattice rules*, SIAM J. Numer. Anal., 33 (1996), pp. 1995–2016.
[28] F. J. Hickernell, *A generalized discrepancy and quadrature error bound*, Math. Comput., 67 (1998), pp. 299–322.
[29] F. Y. Kuo, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo methods for high dimensional integration: The standard weighted-space setting and beyond*, ANZIAM J., 53 (2011), pp. 1–37.
[30] F. Y. Kuo, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 3351–3374.
[31] P. Lancaster and K. Salkauskas, *Surfaces generated by moving least squares methods*, Math. Comput., 37 (1981), pp. 141–158.

[32] 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.
[33] C. Lemieux, *Monte Carlo and quasi-Monte Carlo sampling*, Springer Ser. Statist., Springer, New York, 2009.
[34] D. Levin, *The approximation power of moving least-squares*, Math. Comp., 67 (1998), pp. 1517–1531.
[35] L. Ling, *A univariate quasi-multiquadric interpolation with better smoothness*, Comput. Math. Appl., 48 (2004), pp. 897–912.
[36] H. Müller, U. Stadtmüller, and T. Schmitt, *Bandwidth choice and confidence intervals for derivatives of noisy data*, Biometrika, 74 (1987), pp. 743–749.
[37] H. Niederreiter, *Quasi-Monte Carlo methods and pseudo-random numbers*, Bull. Amer. Math. Soc., 84 (1978), pp. 957–1041.
[38] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conf. Ser. Appl. Math. 63, SIAM, Philadelphia, 1992.
[39] D. Ruppert and W. P. Wand, *Multivariate locally weighted least squares regression*, Ann. Statist., 22 (1994), pp. 1346–1370.
[40] I. J. Schoenberg, *Spline functions and the problem of graduation*, Proc. Natl. Acad. Sci. (USA), 52 (1964), pp. 947–950.
[41] I. H. Sloan and S. Joe, *Lattice Methods for Multiple Integration*, Oxford University Press, Oxford, 1994.
[42] I. M. Sobol′, *On the distribution of points in a cube and the approximate evaluation of integrals*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 86–112.
[43] L. Tenorio, *Statistical regularization of inverse problems*, SIAM Rev., 43 (2001), pp. 347–366.
[44] A. N. Tikhonov, *On the stability of inverse problems*, C. R. Dokl. Acad. Sci. URSS (N. S.), 39 (1943), pp. 176–179.
[45] V. Totik, *Weighted Approximation with Varying Weight*, Springer, Berlin, 1994.
[46] A. B. Tsybakov, *Introduction to Nonparamteric Estimation*, Springer Ser. Statist., Springer, New York, 2009.
[47] K. Vipin, P. N. Tan, and M. Steinbach, *Introduction to Data Mining*, Pearson, Boston, 2006.
[48] Z. M. Wu and J. P. Liu, *Generalized Strang–Fix condition for scattered data quasi-interpolation*, Adv. Comput. Math., 23 (2005), pp. 201–214.
[49] Z. M. Wu and R. Schaback, *Shape preserving properties and convergence of univariate multiquadric quasi-interpolation*, Acta Math. Appl. Sin., 10 (1994), pp. 441–446.
[50] 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.
[51] J. Yoon, *Approximation in $L_p(R^d)$ from a space spanned by the scattered shifts of a radial basis function*, Constr. Approx., 17 (2001), pp. 227–247.