WEIGHTED DISCRETE LEAST-SQUARES POLYNOMIAL APPROXIMATION USING RANDOMIZED QUADRATURES*

TAO ZHOU[†], AKIL NARAYAN[‡], AND DONGBIN XIU[§]

Abstract. We discuss the problem of polynomial approximation of multivariate functions using discrete least squares collocation. The problem stems from uncertainty quantification (UQ), where the independent variables of the functions are random variables with specified probability measure. We propose to construct the least squares approximation on points randomly and uniformly sampled from tensor product Gaussian quadrature points. We analyze the stability properties of this method and prove that the method is asymptotically stable, provided that the number of points scales *linearly* (up to a logarithmic factor) with the cardinality of the polynomial space. Specific results in both bounded and unbounded domains are obtained, along with a convergence result for Chebyshev measure. Numerical examples are provided to verify the theoretical results.

Key words. Least squares method, orthogonal polynomials, generalized polynomial chaos, uncertainty quantification

1. Introduction. We consider in this paper the classical problem of approximating a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $d \geq 1$, via the least squares method using orthogonal polynomials. The work is largely motivated by the field of parametric uncertainty quantification (UQ), where approximating functions of high dimensional ($d \gg 1$) random parameters is one of the salient challenges due to the fast growth of simulation cost (the “curse of dimensionality”).

One of the most widely used techniques in UQ is generalized polynomial chaos (gPC) [15, 40]. In gPC, the function $f(x)$, with $x \in \mathbb{R}^d$, is approximated via d -variate orthogonal polynomials, whose orthogonality is defined by the probability measure of the input variable x . This becomes one of the most effective procedures for UQ, and many numerical techniques aimed at constructing gPC approximations have been developed. After the earlier development of the intrusive type stochastic Galerkin methods, most of the recent efforts concerning gPC methods concentrate on non-intrusive-type stochastic collocation methods, especially after the establishment of high-order stochastic collocation methods [39]. In the stochastic collocation framework, one seeks to construct a gPC type orthogonal polynomial approximation via point-evaluations of f . Popular methods for achieving this include sparse grids approximation [13, 19, 3, 27, 1, 12, 5, 11], pseudo orthogonal projection [30, 37], interpolation [24], and ℓ_1 -minimization sparse approximation [10, 42, 43], to name a few. For a review of the gPC methodology, see [38].

Another collocation-type method that has received less attention in UQ is the method of least squares (LSQ). The least squares method has a long history and is widely used in many disciplines. In UQ-related work, LSQ has been adopted as a natural technique for the response surface method (RSM) and has been used in many

*The work of T. Zhou was supported by the National Natural Science Foundation of China Award Nos. 91130003 and 11201461. The work of D. Xiu was partially supported by AFOSR FA95501410022, DOE DESC0011615, and NSF 1418771. The work of A. Narayan was partially supported by NSF award 1318427.

[†]Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, the Chinese Academy of Sciences, Beijing, China. Email: tzhou@lsec.cc.ac.cn.

[‡]Department of Mathematics and Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112. Email: akil@sci.utah.edu

[§]Department of Mathematics and Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112. Email: dongbin.xiu@utah.edu.

practical engineering problems. Compared to the previously mentioned stochastic collocation methods, the LSQ is not as popular, largely because it requires one to oversample the function f . That is, if N is the cardinality of a linear space from which the least squares approximation is sought, and M be the number of samples, then one must have $M > N$ in order for LSQ to be a proper regularization problem. Regularization is important if one wishes to ameliorate perturbations in the data or spurious oscillations in reconstruction. Typically, a rule-of-thumb in practice is to use a linear oversampling rate, i.e., $M \approx cN$, where $c = 1.5 \sim 2$. This demands more resources if the samples of f are acquired either by expensive simulations or costly experiments. For large and complex problems, it is therefore difficult to construct a high-order least squares approximation, especially in high dimensions $d \gg 1$ when N becomes very large. Almost all existing engineering applications of the least squares method utilize low-order polynomials, e.g., linear, quadratic or cubic polynomials. Most research efforts have been devoted to the design of optimal sample locations for better approximations, e.g., [17]. To this end, aside from random sampling with Monte Carlo points, deterministic samples that can fill up the space systematically are investigated extensively, e.g., quasi Monte Carlo points, orthogonal arrays, etc. This has been the focus of the field of design of experiments (DOE), cf. [2, 4, 6, 16, 20, 29, 36], and the references therein.

Recently, there has been a surge of renewed interest in analyzing the method of least squares from a mathematical perspective. Based on Monte Carlo points (random samples), the stability and accuracy of the least squares method was studied in [9, 23, 22, 17]. It was shown that for Legendre approximations the widely used linear over-sampling rate $M \approx cN$ is asymptotically unstable for high-order polynomial reconstructions. To obtain an asymptotically stable approximation, a quadratic over-sampling rate is required, i.e. $M \propto N^2$. We emphasize that this is an asymptotic result. It does not contradict the fact that the least squares method can deliver good approximations with linear oversampling rates, since most practical applications utilize low and moderate order polynomials. Studies have also been performed on the use of deterministically-generated points, e.g., low discrepancy quasi-Monte Carlo points [8, 14, 21], where over-sampling rate of at least quadratic is required for asymptotic stability ([21]). In [41, 44], a kind of deterministic sampling strategy using Weil's theorem was proposed, yielding again a quadratic oversampling rate for stable approximation. Most of the existing work has considered bounded domains (e.g., bounded hypercubes) and the optimal estimate is usually obtained by sampling asymptotically according to the Chebyshev measure. Recent results in [34] address the least squares problem in unbounded domains where approximations are formed via Hermite and Laguerre functions.

In this paper, we present a study on the least squares method using orthogonal polynomials in a general setting; our procedure is applicable for both bounded and unbounded domains. In particular, we propose to randomly select the sample points from the tensor product point set of the Gaussian quadrature points, which are determined by the orthogonal polynomial basis in the approximation. Our method subsequently uses the Gaussian quadrature weights to precondition the least squares linear system. (A similar strategy was investigated in the context of compressed sensing in [33].) We establish stability conditions and convergence estimates for this method. Specifically, we prove that LSQ using the new sampling strategy is stable under an *almost* linear (up to a logarithmic factor) oversampling rate, i.e., it is stable subject to $\frac{M}{\log M} \sim \mu N$, where $\mu \geq 2$. More detailed estimates are presented for bounded domains with be-

ta measures, as well as for unbounded domains with Gaussian measure. Numerical examples are presented to verify the theoretical findings; the results suggest that our method is no worse than the standard LSQ using Monte Carlo points, and usually performs much better.

2. Problem setup. We consider a d -variate function $f : I_X \rightarrow \mathbb{R}$, where $I_X \subseteq \mathbb{R}^d$, $d \geq 1$, is the support. We make the common assumption that the domain I_X is a tensor-product domain. With $x = (x^1, \dots, x^d)$ the coordinate, we have $x^i \in I_{X^i}$ and $I_X = \times_{i=1}^d I_{X^i}$. We also equip each variable x^i with a (probability) measure $\rho^i : I_{X^i} \rightarrow \mathbb{R}^+$ and let $\rho(x) = \prod_{i=1}^d \rho^i(x^i)$.

Let $\Theta_M = \{x_1, \dots, x_M\} \subset I_X$ be a set of points at which the function values of f are available. Let $f_i = f(x_i)$, $i = 1, \dots, M$, be these values. We are concerned with the approximation of the function f based on $\{x_i, f_i\}_{i=1}^M$. Let V be a linear space from which the approximation is sought and let $N = \dim V$. We consider applying the method of least squares when $M > N$ (oversampling). This is achieved by finding a $v \in V$ such that the sum-of-squares error is minimized, that is,

$$v = \operatorname{argmin}_{p \in V} \frac{1}{M} \sum_{i=1}^M (p(x_i) - f_i)^2. \quad (2.1)$$

2.1. Polynomial spaces. In this paper, we are primarily concerned with the use of orthogonal polynomials, that is, the linear space V is a polynomial space. Although other type of functions can be employed in the least squares framework, polynomial approximation remains one of the most widely used ones. If Λ is an index set for multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$, then $P_\Lambda = \operatorname{span}\{x^\alpha : \alpha \in \Lambda\}$. We denote $|\alpha|_\infty = \max_i |\alpha_i|$ and $|\alpha|_1 = \sum_i |\alpha_i|$. The polynomial spaces typically considered in practice include

- Tensor product (TP) space of degree n , where

$$\mathbb{P}_n^{TP} = \operatorname{span}\{x^\alpha : |\alpha|_\infty \leq n\}. \quad (2.2)$$

The cardinality of this space is $\dim(\mathbb{P}_n^{TP}) = (n+1)^d$.

- Total degree (TD) space of degree n , where

$$\mathbb{P}_n^{TD} = \operatorname{span}\{x^\alpha : |\alpha|_1 \leq n\}. \quad (2.3)$$

The cardinality of this space is $\dim(\mathbb{P}_n^{TD}) = \binom{n+d}{n}$.

- Hyperbolic cross (HC) space of degree n , where

$$\mathbb{P}_n^{HC} = \operatorname{span}\{x^\alpha : \prod_i (\alpha_i + 1) \leq n + 1\}. \quad (2.4)$$

The cardinality of this space is difficult to specify precisely. An estimate from [22] states that

$$\dim(\mathbb{P}_n^{HC}) \leq \lfloor (n+1)(1 + \log(n+1))^{d-1} \rfloor.$$

Note that these spaces are uniquely determined by the set of multi-indices Λ . For example, for the widely used total-degree space $\mathbb{P}_n^{TD}$, the index set is $\Lambda_n^{TD} = \{\alpha : |\alpha|_1 \leq n\}$. For the spaces $\mathbb{P}_n^{TP}$, $\mathbb{P}_n^{TD}$, and $\mathbb{P}_n^{HC}$, we will use the notation Λ^{TP} , Λ^{TD} , and Λ^{HC} for their respective multi-index sets. Hereafter we will frequently use $\mathbb{P}_\Lambda$ to denote a general multi-variate polynomial space, whose index set is Λ .

2.2. Generalized polynomial chaos. The basic idea behind generalized polynomial chaos (gPC) is to use orthogonal polynomials in terms of random variables to approximate functions of those random variables. Consider, for $i = 1, \dots, d$, each component $x^i \in I_{X^i}$ with the probability density ρ^i , we define the univariate gPC basis via the following orthogonality

$$\mathbb{E} [\varphi_m^i(x^i) \varphi_\ell^i(x^i)] = \int_{I_{X^i}} \varphi_m^i(s) \varphi_\ell^i(s) \rho^i(s) ds = \delta_{m,\ell}, \quad m, \ell \geq 0, \quad (2.5)$$

where $\delta_{m,\ell}$ is the Kronecker delta function. The probability measure ρ^i then determines the type of orthogonal polynomials. For example, Gaussian distribution corresponds to Hermite polynomials, uniform distribution pairs with Legendre polynomials, etc. For a detailed account of the correspondence, see [40].

For the multivariate case $d > 1$, the gPC basis polynomials are defined as a product of the univariate gPC polynomials in each direction, i.e.,

$$\psi_\alpha(x) = \prod_{i=1}^d \varphi_{\alpha_i}^i(x^i). \quad (2.6)$$

Note that the polynomials in each direction need not be of the same univariate polynomial family. We then require the gPC polynomials to reside in a properly defined polynomial space $\mathbb{P}_{\Lambda_n}$, where Λ_n is an index set with “degree” n . This can correspond to the tensor product space (2.2), total degree space (2.3), hyperbolic cross space (2.4), or any other properly defined space.

We remark that for the multi-index set Λ_n , it is always possible to define an ordering scheme such that the multi-indices can be ordered via a single index. That is, we have

$$\{\psi_\alpha(x)\}_{\alpha \in \Lambda_n} \Leftrightarrow \{\psi_j(x)\}_{j=1}^{\dim(\mathbb{P}_{\Lambda_n})}. \quad (2.7)$$

Hereafter we will frequently interchange between the two representations, unless confusion arises otherwise.

2.3. Weighted least squares approximation. We now present the basic formulation for (weighted) least squares approximation using orthogonal polynomials. Let $(x_i, w_i)_{i=1}^M$ be a set of M points in I_X with the corresponding weights satisfying $w_i > 0$ and $\sum_{i=1}^M w_i = 1$, and $f_i = f(x_i)$ be the function values. Let $\mathbb{P}_{\Lambda_n}$ be a chosen polynomial space and $N = \dim(\mathbb{P}_{\Lambda_n})$. We now seek a polynomial $p_n \in \mathbb{P}_{\Lambda_n}$ such that

$$p_n = \operatorname{argmin}_{p \in \mathbb{P}_{\Lambda_n}} \sum_{i=1}^M w_i (p(x_i) - f_i)^2. \quad (2.8)$$

The least squares solution is well-defined when $M \geq N$, assuming the matrix $(\psi_j(x_i))_{i,j}$ has full rank. Let $\{\psi_\alpha(x), \alpha \in \Lambda_n\}$ be the orthogonal polynomials spanning $\mathbb{P}_{\Lambda_n}$, we can write

$$p_n = \sum_{\alpha \in \Lambda_n} c_\alpha \psi_\alpha = \sum_{i=1}^N c_i \psi_i(x), \quad (2.9)$$

where we have used the linear ordering of the multi-index set. Let $\mathbf{c} = (c_1, \dots, c_N)^\top$ be the coefficient vector, $\mathbf{f} = (f_1, \dots, f_M)^\top$ the data vector,

$$\mathbf{D} = (d_{ij})_{1 \leq i \leq M, 1 \leq j \leq N} \in \mathbb{R}^{M \times N} \quad d_{ij} = \psi_j(x_i), \quad (2.10)$$

be the Vandermonde-like matrix, often referred to as the design matrix, and

$$\mathbf{W} = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_M}). \quad (2.11)$$

Then, the least squares solution can be solved by identifying coefficients $\mathbf{c}$ that satisfy

$$\widehat{\mathbf{D}}^\top \widehat{\mathbf{D}} \mathbf{c} = \widehat{\mathbf{D}}^\top \widehat{\mathbf{f}}, \quad (2.12)$$

where

$$\widehat{\mathbf{D}} = \mathbf{W}\mathbf{D}, \quad \widehat{\mathbf{f}} = \mathbf{W}\mathbf{f}. \quad (2.13)$$

3. Method of random Gauss quadrature. We now present the method of random Gauss quadrature for least squares approximation. We will primarily focus on the use of the tensor product space $\mathbb{P}_n^{TP}$ (2.2), the total degree space $\mathbb{P}_n^{TD}$ (2.3), and the hyperbolic cross space $\mathbb{P}_n^{HC}$ (2.4), with $n \geq 1$ denoting the “degree” of the space. We use Λ_n to refer to the index set of one of these choices.

3.1. Random sampling of tensor Gauss points. Consider, for each variable x^i , $i = 1, \dots, d$, the gPC orthogonal polynomial family satisfying (2.5). Let $\varphi_{n+1}^i(x^i)$ be the $(n+1)$ -degree polynomial. It is well known that it has $n+1$ real and distinct zeros, i.e.,

$$\varphi_{n+1}^i(z_j^i) = 0, \quad j = 1, \dots, n+1.$$

Furthermore, the Gauss quadrature rule can be constructed on the zeros in such a way that

$$\sum_{j=1}^{n+1} w_j^i f(z_j^i) \approx \int_{X^i} f(x^i) \rho^i(x^i) dx^i,$$

with equality for any $f \in \mathbb{P}_{2n+1}$. Here w_j^i , $j = 1, \dots, n+1$, are the univariate Gauss quadrature weights associated with dimension i . For the orthonormal polynomial family φ_n^i (2.5), the degree- n Gauss quadrature weights can be computed explicitly as

$$w_k^i = \frac{1}{n+1} \lambda_n^i(z_k^i), \quad k = 1, \dots, n+1, \quad (3.1)$$

where λ_n^i is the (normalized) Christoffel function associated with dimension i :

$$\lambda_n^i(x^i) = \frac{n+1}{\sum_{j=0}^n (\varphi_j^i(x^i))^2}. \quad (3.2)$$

We proceed to tensorize the univariate Gauss quadrature rules. Let

$$\Theta_{n+1}^i = \{z_1^i, \dots, z_{n+1}^i\} \subset I_{X^i}, \quad i = 1, \dots, d, \quad (3.3)$$

be the one dimensional Gauss point set. We then take the tensor product to construct a d -dimensional point set

$$\Theta_n = \Theta_{n+1}^1 \otimes \dots \otimes \Theta_{n+1}^d. \quad (3.4)$$

Obviously, $\#\Theta_n = (n+1)^d$. As before, an ordering scheme can be employed to order the points via a single index, i.e., for each $j = 1, \dots, \#\Theta_n$,

$$z_j = (z_{\alpha_1}^1, \dots, z_{\alpha_d}^d), \quad j \longleftrightarrow \alpha \in \mathbb{N}_0^d,$$

it corresponds to a unique combination of $(\alpha_1, \dots, \alpha_d)$ with $|\alpha|_\infty \leq n$. Each point has the scalar weight

$$w_j = w_{\alpha_1}^1 \times \dots \times w_{\alpha_d}^d, \quad j = 1, \dots, \#\Theta_n,$$

with the same $j \leftrightarrow \alpha$ correspondence. This tensorized Gauss quadrature in the d -dimensional space I_X is exact for integrals of any polynomials in $\mathbb{P}_{2n+1}^{TP}$. We note that using (3.1) and (3.2), the d -dimensional Gauss quadrature weights are given by

$$w_j = \prod_{i=1}^d \frac{1}{\sum_{k=0}^n [\varphi_k^i(z_{\alpha_i}^i)]^2} = \frac{1}{\#\Theta_n} \prod_{i=1}^d \lambda_n^i(z_{\alpha_i}^i), \quad (3.5)$$

where $\#\Theta_n = (n+1)^d$.

Let $N = \dim(\mathbb{P}_{\Lambda_n})$ be the cardinality of the polynomial space and $M > N$ be the number of points used in the weighted least squares approximation. We now choose a number of points M randomly from the tensor Gauss points Θ_n under the uniform discrete probability distribution. (We implicitly assume throughout this paper that $M \leq (n+1)^d$.) Note that this can effectively be accomplished by choosing M indices randomly from the index set $\{1, \dots, \#\Theta_n\}$. Subsequently, we obtain a set of points and their corresponding weights, i.e.,

$$\Theta_M = \{x_{j_1}, \dots, x_{j_M}\}, \quad \{w_{j_1}, \dots, w_{j_M}\}.$$

We then proceed to solve the weighted least squares problem (2.8), or equivalently, (2.12).

By letting $K = \#\Theta_n = (n+1)^d$ be the total number of tensorized Gauss points, we define

$$\mathbf{D}_{\Theta_n} = (g_{j\ell})_{1 \leq j \leq K, 1 \leq \ell \leq N}, \quad g_{j\ell} = \psi_\ell(z_j), \quad (3.6)$$

to be the design matrix for the entire set of the tensor Gauss points. Let

$$\mathbf{W}_{\Theta_n} = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_K}), \quad \widehat{\mathbf{D}}_{\Theta_n} = \mathbf{W}_{\Theta_n} \mathbf{D}_{\Theta_n}. \quad (3.7)$$

Since the tensor-product Gauss quadrature rule is exact for all elements of $\mathbb{P}_{2n+1}^{TP}$, then $\mathbf{D}_{\Theta_n}$ is an orthogonal matrix if $\Lambda_n \subset \Lambda_n^{TP}$. Our sub-sampled design matrix $\widehat{\mathbf{D}}$ (2.13) is obtained by uniformly sampling, with probability $1/K$, M rows from the matrix $\widehat{\mathbf{D}}_{\Theta_n}$.

3.2. Stability analysis. We now discuss the stability of the least squares problem using random sub-sampling of tensor Gauss points. We employ an approach similar to the one used in [9] — we analyze how much the matrix $\widehat{\mathbf{D}}$ (2.10) deviates from an orthogonal matrix in probability. To this end, the following fact plays an important role.

THEOREM 3.1 ([35]). *Let $\mathbf{A}$ be an $M \times N$ matrix ($M > N$) whose rows $\mathbf{A}_i$ are independent isotropic random vectors in $\mathbb{R}^N$. Let τ be a number such that $\|\mathbf{A}_i\|_2 \leq \sqrt{\tau}$ almost surely for all i . Then for every $t > 0$, one has*

$$\sqrt{M} - t\sqrt{\tau} \leq \sigma_{\min}(\mathbf{A}) \leq \sigma_{\max}(\mathbf{A}) \leq \sqrt{M} + t\sqrt{\tau} \quad (3.8)$$

with probability at least $\gamma = 1 - 2N \cdot \exp(-ct^2)$, where $c > 0$ is an absolute constant.

A vector $\mathbf{X} \in \mathbb{R}^N$ is an *isotropic random vector* if

$$\mathbb{E}[\mathbf{X} \otimes \mathbf{X}] = \mathbf{I}, \quad (3.9)$$

where $\mathbb{E}[\cdot]$ is the expectation operator and $\mathbf{I}$ is the $N \times N$ identity matrix. Note also that (3.8) is equivalent to

$$\left\| \frac{1}{M} \mathbf{A}^\top \mathbf{A} - \mathbf{I} \right\| \leq \max(\theta, \theta^2), \quad \theta = t \sqrt{\frac{\tau}{M}}. \quad (3.10)$$

We will use this inequality to analyze the stability properties for the proposed least squares method. To avoid cluttering notation, we will set θ such that the right-hand-side of (3.10) becomes $1/2$.

We now state the following stability result.

THEOREM 3.2. *Consider the $(M \times N)$ design matrix $\widehat{\mathbf{D}}$ (2.13), obtained by uniformly sampling, with probability $1/K$, M rows out of the $(K \times N)$ matrix $\widehat{\mathbf{D}}_\Theta$ in (3.7). Then,*

$$\left\| \frac{K}{M} \widehat{\mathbf{D}}^\top \widehat{\mathbf{D}} - \mathbf{I} \right\| \leq \frac{1}{2} \quad (3.11)$$

with probability at least $\gamma = 1 - 2M^{1-\mu}$, for any $\mu \geq 2$, provided that

$$\frac{M}{\log M} > \frac{4\mu}{c} C_w C_b N, \quad (3.12)$$

where c is the absolute constant from Theorem 3.1, and

$$C_w = K \max_{1 \leq k \leq K} |w_k| = \max_{1 \leq k \leq K} \prod_{i=1}^d \lambda_n^i(z_k^i), \quad C_b = \max_{\alpha \in \Lambda_n, x \in \Theta_n} \psi_\alpha^2(x). \quad (3.13)$$

Proof. Let $\mathbf{X}_k$, $k = 1, \dots, K$, be the row vectors of $\widehat{\mathbf{D}}_\Theta$. Note that in order to form $\widehat{\mathbf{D}}$ we randomly choose M rows out of $\widehat{\mathbf{D}}_\Theta$ with equal probability $1/K$. Thus, row m of $\widehat{\mathbf{D}}$ is $\mathbf{X}_{k(m)}$, where $(k(m))_{m=1}^M$ are M iid random variables uniformly distributed on the discrete set $1, \dots, K$. By renormalizing these rows as $\widehat{\mathbf{X}}_{k(m)} = \sqrt{K} \mathbf{X}_{k(m)}$, then

$$\left\| \widehat{\mathbf{X}}_{k(m)} \right\|_2^2 = K w_{k(m)} \sum_{\alpha \in \Lambda_n} \psi_\alpha^2(z_{k(m)}) = \left(\prod_{i=1}^d \lambda_n^i(z_{k(m)}^i) \right) \left(\sum_{\alpha \in \Lambda_n} \psi_\alpha^2(z_{k(m)}) \right)$$

We define C_w as in (3.13) and note that with $N = \#\Lambda_n$,

$$\sum_{\alpha \in \Lambda_n} \psi_\alpha^2(z_{k(m)}) \leq N \max_{\alpha \in \Lambda_n} \psi_\alpha^2(z_{k(m)}).$$

Thus, with C_b defined in (3.13) we have $\|\widehat{\mathbf{X}}_{k(m)}\|_2 \leq \sqrt{C_w C_b N}$ with probability 1.

The scaled row $\widehat{\mathbf{X}}_{k(m)}$ is isotropic in the sense of (3.9). That is, consider the (n, ℓ) entry of this $(N \times N)$ matrix $\mathbb{E}[\widehat{\mathbf{X}}_{k(m)} \otimes \widehat{\mathbf{X}}_{k(m)}]$,

$$\begin{aligned} \mathbb{E} \left[\widehat{\mathbf{X}}_{k(m)}^\top \widehat{\mathbf{X}}_{k(m)} \right]_{n,\ell} &= K \mathbb{E} [w_{k(m)} \psi_n(z_{k(m)}) \psi_\ell(z_{k(m)})] = \sum_{k=1}^K w_k \psi_n(z_k) \psi_\ell(z_k) \\ &= \int \psi_n(x) \psi_\ell(x) \rho(x) dx = \delta_{n,\ell}, \end{aligned} \quad (3.14)$$

where the penultimate equality uses the degree- $(2N + 1)$ exactness of the Gauss quadrature rule. Having established that each row $\widehat{\mathbf{X}}_{k(m)}$ of $\widehat{\mathbf{D}}$ is independent and isotropic, we invoke Theorem 3.1 (or (3.10)), to conclude that with probability $1 - 2N \exp(-ct^2)$, we have

$$\left\| \frac{K}{M} \widehat{\mathbf{D}}^\top \widehat{\mathbf{D}} - \mathbf{I} \right\| \leq \max(\theta, \theta^2), \quad \theta = t \sqrt{\frac{C_w C_b N}{M}}. \quad (3.15)$$

The probability $1 - 2N \exp(-ct^2)$ is only non-trivial if $t \geq \sqrt{\frac{\log 2N}{c}}$. Making the stronger assumption that $t > \sqrt{\frac{\mu \log M}{c}}$, we have

$$\theta^2 \frac{M}{C_w C_b N} = t^2 > \frac{\mu}{c} \log M$$

The desired follows by letting $\theta = \frac{1}{2}$, completing the proof. $\square$

The constants C_w and C_b are the major quantities that we need to consider, and we give specific instances of these in the following sections. The quantity C_w involves a factor $K = (n + 1)^d$. For large d , it may appear to be too large to make Theorem 3.2 useful. However, the weights scale as $1/(n + 1)^d = 1/K$, which “balances” the factors. Additionally, the constant C_w is a product of univariate (normalized) Christoffel functions, about which much precise information regarding bounds is known.

3.2.1. Chebyshev measure and Chebyshev polynomials. Let us consider the Chebyshev measure,

$$\rho(x) = \prod_{j=1}^d \frac{1}{\pi \sqrt{1 - (x^j)^2}}, \quad x \in [-1, 1]^d.$$

The underlying polynomials are tensor products of one-dimensional orthonormal Chebyshev polynomials. The univariate polynomials $\varphi_n^j(x^j)$ satisfy the bound $|\varphi_n^j| \leq \sqrt{2}$, uniformly in j and n . Then we have

$$C_b = \max_{\alpha, k} \psi_\alpha^2(x_k) \leq 2^d. \quad (3.16)$$

See, for example, [31]. Moreover, the quadrature weights w_k satisfy $w_k \equiv (\frac{1}{n+1})^d$. Consequently,

$$C_w = K \max_k |w_k| \equiv 1. \quad (3.17)$$

Theorem 3.2 then implies that, for the Chebyshev measure,

$$\left\| \frac{K}{M} \widehat{\mathbf{D}}^\top \widehat{\mathbf{D}} - \mathbf{I} \right\| \leq \frac{1}{2} \quad \Leftrightarrow \quad \left\| \frac{1}{M} \mathbf{D}^\top \mathbf{D} - \mathbf{I} \right\| \leq \frac{1}{2} \quad (3.18)$$

with probability at least $1 - 2M^{1-\mu}$ provided that

$$\frac{M}{\log M} \gtrsim 2^d \mu N \quad (3.19)$$

Thus, for the Chebyshev measure the weighted least squares method using random samples from the tensor Gauss point is asymptotically stable for a linear oversampling rate, with a logarithmic factor.

3.2.2. Uniform measure and Legendre polynomials. Let now consider the uniform measure $\rho(x) = 1/2^d$, $x \in [-1, 1]^d$. The underlying polynomials are the tensor products of one-dimensional orthonormal Legendre polynomials. In this case, we have the following estimate ([28])

$$C_w = K \max_{\alpha} |w_{\alpha}| \sim (\pi/2)^d. \quad (3.20)$$

The univariate orthonormal Legendre polynomials satisfy $\max_{x \in [-1, 1]} \varphi_n^j(x) = \sqrt{2n+1}$ so that for the tensor product polynomial space

$$C_b = \max_{\alpha \in \Lambda_n} \psi_{\alpha}^2 \leq \max_{\alpha \in \Lambda_n} \prod_{j=1}^d (2\alpha_j + 1) \leq \max_{\alpha \in \Lambda} \left(\frac{\sum_{j=1}^d (2\alpha_j + 1)}{d} \right)^d \leq \left(1 + \frac{2}{d} A \right)^d, \quad (3.21)$$

where the second inequality uses the arithmetic-geometric mean inequality and $A \triangleq \max_{\alpha \in \Lambda_n} |\alpha|$. Such a constant C_{Λ} grows exponentially when $A > d$, and we are forced to require

$$\frac{M}{\log M} \gtrsim \left[\pi \left(\frac{A}{d} + \frac{1}{2} \right) \right]^d \quad (3.22)$$

for stability. This induces a large and fast growing number of points if the maximum degree of approximation A is large.

The stability estimate can be significantly improved if we consider the case of $A < d$. This is a practical situation, where the dimensionality d of the problem is so large that one can not afford to use high degree polynomial approximations. (Similar situation was also considered in [42], for compressed sensing approximation.) In this case, the quantity $\prod_{j=1}^d (2\alpha_j + 1)$ in (3.21) is maximized when n of the α_j 's are one and the rest are zero. We then obtain the following estimate

$$C_b = \max_{\alpha} \psi_{\alpha}^2 \leq \left(1 + \frac{2A}{d} \right)^d \leq 3^n. \quad (3.23)$$

And stability then requires

$$\frac{M}{\log M} \gtrsim \mu \left(\frac{3\pi}{2} \right)^d N. \quad (3.24)$$

Similar to the Chebyshev measure case, the Legendre least squares measure using the random samples from the tensor Gauss points is asymptotically stable for the linear over-sampling rate, with a logarithmic factor.

Furthermore, if one adopts the hyperbolic cross (HC) space (2.4), then

$$C_b^{HC} = \max_{\alpha \in \Lambda} \psi_{\alpha}^2 \leq \max_{\alpha \in \Lambda_n} \prod_{j=1}^d (2\alpha_j + 1) \leq 2(n+1). \quad (3.25)$$

Consequently, stability in HC space requires

$$\frac{M^{HC}}{\log M^{HC}} \gtrsim \mu (\pi/2)^d (n+1) N \quad (3.26)$$

We remark that the constant $(\pi/2)^d$ is introduced by a rough estimate for the quantity C_w .

REMARK 3.3. *Note that in general, the Chebyshev measure introduces a better constant against the uniform measure. One can then consider a weighted least-squares approach for the uniform measure, namely, to use a Chebyshev pre-conditioner to reduce the constant. Such an idea was introduced in [44] in the least-squares projection. One can use similar arguments as above to specify the constants for general types of measures (e.g. the beta measure with Jacobi polynomials, which is investigated in the following section.)*

We also remark that orthogonal polynomials in a compressed sensing framework are closely related to the maximum quantities used in the above discussion, particularly $C_b C_w$, and one can refer to [33, 18] for similar discussion in the compressed sensing framework.

3.2.3. Jacobi polynomials. The probability density

$$\rho(x) = C(\alpha, \beta)^d \prod_{j=1}^d (1 - x^j)^\alpha (1 + x^j)^\beta, \quad \alpha, \beta \geq 0$$

is the tensor-product Beta density function on $[-1, 1]^d$. The normalization coefficient is

$$C(\alpha, \beta) = \frac{\Gamma(\alpha + \beta + 2)}{\Gamma(\alpha + 1)\Gamma(\beta + 1)2^{\alpha+\beta+1}}.$$

The associated polynomial family ψ_α is the tensor-product Jacobi polynomials associated to parameters α and β . (One may consider different values of α and β in each dimension, which changes the following analysis in straightforward ways.)

For this polynomial family, it is known that the univariate Christoffel function satisfies

$$\lambda_n^j(x^j) \lesssim \pi \rho^j(x^j) (1 - (x^j)^2)^{1/2}.$$

See, e.g., [25]. The maximum of $(1 - x)^{\alpha+1/2}(1 + x)^{\beta+1/2}$ is attained at $x = \frac{\beta-\alpha}{\alpha+\beta+1}$, and so

$$\begin{aligned} C_w &\leq \max_{x \in [-1, 1]^d} \prod_{j=1}^d \lambda_n(x^j) \lesssim \left[C(\alpha, \beta) \frac{(2\alpha + 1)^\alpha (2\beta + 1)^\beta}{(\alpha + \beta + 1)^{\alpha+\beta+1}} \right]^d \\ &= \left[\frac{\Gamma(\alpha + \beta + 2) (\alpha + \frac{1}{2})^{\alpha+1/2} (\beta + \frac{1}{2})^{\beta+1/2}}{2\Gamma(\alpha + 1)\Gamma(\beta + 1)(\alpha + \beta + 1)^{\alpha+\beta+1}} \right]^d \end{aligned}$$

For $\alpha, \beta \leq 10$, the quantity under the brackets can be bounded by 2, so that

$$C_w \leq 2^d, \quad 0 \leq \alpha, \beta \leq 10.$$

Thus, C_w is well-behaved. However, we see growth of C_b similar to that which was observed for Legendre polynomials. We have

$$\left(\varphi_k^j \right)^2 \lesssim \frac{2e}{\pi (1 - x)^{\alpha+1/2} (1 + x)^{\beta+1/2}}.$$

Now since C_b is computed as the maximum over Gauss quadrature points, we will use the fact that $-1 + \frac{2\beta^2}{(2n+\alpha+\beta+3)^2} \leq x_k^j \leq 1 - \frac{2\alpha^2}{(2n+\alpha+\beta+1)^2}$ for the size- $(n+1)$ univariate quadrature rule; see [26, 32]. Defining $\gamma \triangleq \max\{\alpha, \beta\}$, we conclude for $\gamma > 0$:

$$\left(\varphi_k^j\right)^2 \lesssim \frac{e}{\gamma^2 \pi} (2n + \alpha + \beta + 1)^{2\gamma+1}$$

Then, of course, we require an unpleasant dependence of M on N :

$$\frac{M}{\log M} \gtrsim \mu \left[\frac{2e}{\gamma^2 \pi} (2n + \alpha + \beta + 1)^{2\gamma+1} \right]^d N,$$

Similar to the arguments in the previous section, for cases when the dimension of the polynomial space far exceeds the maximum degree, then the estimate for the required number of samples can be improved.

3.2.4. Gaussian measure and Hermite polynomials. Let us now consider the Gaussian measure on the unbounded real line, i.e., $\rho(x) = (\frac{1}{\sqrt{2\pi}})^d \exp\left(-\sum_{j=1}^d \frac{(x^j)^2}{2}\right)$, $x \in \mathbb{R}^d$. The corresponding orthogonal polynomials are the orthonormal Hermite polynomials $H_\alpha(x)$. However, it is shown in [34] that it is almost impossible to control the stability of the design matrix if the Hermite polynomials are used. This can also be observed by investigating the constant C_b , which becomes unbounded.

To circumvent to difficulty, the corresponding Hermite functions $\hat{H}_\alpha(x)$ should be used, instead of the Hermite polynomials. (We note that this is a standard practice for approximations in unbounded domains.) We then have the following estimates ([31])

$$C_b = \max_{\alpha} \hat{H}_{\alpha}^2 \leq \pi^{\frac{d}{2}}. \quad (3.27)$$

For the corresponding Gauss quadrature rule (for the Hermite *functions*, not polynomials), we have asymptotically $w_{\alpha} \sim (\frac{\pi}{\sqrt{2(n+1)}})^d$ (see also in [31], pp.258), such that

$$C_w = K \max_{\alpha} |w_{\alpha}| \leq \left(\frac{\pi}{2}\right)^{\frac{d}{2}}. \quad (3.28)$$

Then, the stability of the least squares method requires

$$\frac{M}{\log M} \gtrsim \left(\frac{\pi}{\sqrt{2}}\right)^d N \quad (3.29)$$

We remark that the estimate (3.28) is an asymptotic result. In practice, one always uses a finite order n . This introduces a constant that depends on the dimension d , i.e., $C_w \sim C(d)$, and it may be large for small degree n at large dimension d .

3.3. Convergence analysis. We now discuss the convergence property of the proposed least squares method using random samples of tensor Gauss points. For simplicity, we shall focus our attention on the Chebyshev measure, nevertheless, the framework easily applies to general measures. We also remark that the convergence proof in the following is standard, and one could find similar arguments in [7].

For the Chebyshev measure, by (3.18) we already have that

$$\Pr \left\{ \left\| \frac{1}{M} \mathbf{D}^{\top} \mathbf{D} - \mathbf{I} \right\| > \frac{1}{2} \right\} \leq 1 - 2M^{1-\mu} \quad (3.30)$$

provided that

$$\frac{M}{\log M} \gtrsim 2^d \mu N. \quad (3.31)$$

Note that equation (3.30) is equivalent to

$$\left| \|v\|_M^2 - \|v\|^2 \right| \leq \frac{1}{2} \|v\|^2, \quad \forall v \in \mathbb{P}_n^{TD}, \quad (3.32)$$

where $\|\cdot\|_M$ and $\|\cdot\|$ are the discrete norm and the continuous norm respectively, i.e.,

$$\|v\|_M^2 := \frac{1}{M} \sum_{j=1}^M v(y_j)^2, \quad \|v\| = \int_{[-1,1]^d} \rho(y) v^2(y) dy, \quad \forall v \in \mathbb{P}_n^{TD}, \quad (3.33)$$

where the y_j are the subsampled Gauss quadrature points. Now let f be the function to be approximated and f_n^{TD} be the discrete least-squares solution in $\mathbb{P}_n^{TD}$, and let

$$\epsilon_{\Lambda_n^{TD}}(f) := \inf_{v \in \mathbb{P}_n^{TD}} \|f - v\|_{L^\infty}. \quad (3.34)$$

Then, we shall show that with high probability, the least-squares projection error is bounded by the best approximation error. More precisely, we have the following convergence theorem

THEOREM 3.4. *Under the requirement $\frac{M}{\log M} \gtrsim \mu 2^d N$, one has*

$$\Pr \left\{ \|f - f_n^{TD}\| > (1 + \sqrt{2}) \epsilon_{\Lambda_n^{TD}}(f) \right\} \leq 1 - 2M^{1-\mu}. \quad (3.35)$$

Proof. We will show that, under any event satisfying (3.30), we have $\|f - f_n^{TD}\| < (1 + \sqrt{2}) \epsilon_{\Lambda_n^{TD}}$. Let $v \in \mathbb{P}_n^{TD}$ denote the minimizer from (3.34). For the least-squares solution $f_n^{TD} \in \mathbb{P}_n^{TD}$ obtained from a system that satisfies (3.30), we have

$$\|f - f_n^{TD}\| \leq \|f - v\| + \|v - f_n^{TD}\| \leq \|f - v\| + \sqrt{2} \|v - f_n^{TD}\|_M. \quad (3.36)$$

where we have used (3.32). Note that we have

$$\|f - v\|_M^2 = \|f - f_n^{TD}\|_M^2 + \|v - f_n^{TD}\|_M^2, \quad (3.37)$$

so that

$$\|f - f_n^{TD}\| \leq \|f - v\| + \sqrt{2} \|f - v\|_M \leq (1 + \sqrt{2}) \|f - v\|_{L^\infty}. \quad (3.38)$$

This completes the proof. $\square$

4. Numerical examples. We now provide numerical examples to demonstrate the properties of the least-squares method using randomized Gauss quadrature. We first investigate how the condition number of the design matrix $\text{cond}(\hat{\mathbf{D}}) = \sigma_{\max}(\hat{\mathbf{D}})/\sigma_{\min}(\hat{\mathbf{D}})$ is affected by the number of samples. This quantity directly reflects the stability of the method. Due to the random nature of the method, all reported results are averaged over 200 independent tests. In all the figures, we denote *Gaussian* as the results obtained by randomized Gauss quadrature, whereas *Random* stands for the results obtained by the standard least squares method using random (Monte Carlo) points. We remark that our estimate of $\frac{M}{\log M} \gtrsim \mu N$ actually implies that $M \gtrsim \mu N \log N$, and thus, we shall investigate both the linear dependence $M \sim cN$ and the *log-linear* dependence $M \sim cN \log N$ in our numerical tests.

4.1. Chebyshev measure and Chebyshev polynomial. We first show the numerical results for 10-dimensional ($d = 10$) hyperbolic-cross polynomial space $\mathbb{P}^{HC}$. The underlying measure is the Chebyshev measure and the basis functions are the corresponding Chebyshev polynomials. In Figures 4.1 and 4.2 the condition numbers obtained by the two sampling points (“Gaussian” – our subsampling strategy, and “Random” – Monte Carlo sampling from ρ) are shown for a linear oversampling of $M \sim 2N$, and *log-linear* oversampling rates of $M \sim 0.8N \log N$, and $M \sim 0.2N \log^2 N$, respectively. We notice that the performance of the two kinds of points is similar. Also, the two *log-linear* oversampling rates in Figure 4.2 in fact lead to decay of the condition number with increasing order of the basis polynomials. This suggests that the method is asymptotically stable. On the other hand, the linear oversampling, Figure 4.1 leads to a growth of the condition number with respect to the polynomial order, which suggests the method is asymptotically unstable.

FIG. 4.1. Condition number with respect to the polynomial degree in the 10-dimensional hyperbolic spaces (Chebyshev measure) with linear oversampling of $M = 3N$.

FIG. 4.2. Condition number with respect to the polynomial degree in the 10-dimensional hyperbolic spaces (Chebyshev measure) with an oversampling of $M = 1.5N \log N$ (left) and an oversampling rate of $M = 0.2N \log^2 N$ (right).

4.2. Uniform measure and Legendre polynomial. In Figure 4.3, we report the numerical results for the uniform measure with Legendre polynomial expansions in a 6-dimensional hyperbolic cross polynomial space. The condition number with Monte Carlo samples drawn from the uniform measure are also provided for comparison.

Again, a linear dependence (Figure 4.3, left) admits a growth of the condition number with respect to the polynomial order n , for both the random and the randomly chosen Gaussian points. While the *log-linear* rule (Figure 4.3, right) admits decay properties of the condition number against the polynomial order n , for the randomly chosen Gaussian points. The points randomly sampled from the Gauss quadrature points have a noticeably better performance for the uniform measure.

FIG. 4.3. Condition number with respect to the polynomial degree k in the 6-dimensional hyperbolic spaces (Uniform measure). Left: $M = 3N$. Right: $M = 2N \log N$.

4.3. Effect of dimensionality. In Figure 4.4, we examine the effect of the dimension d on the condition number with a log-linear oversampling rule. This can be understood by examining the constants C_w and C_b in our theoretical estimates. In Figure 4.4 (left), we report the numerical condition number for Chebyshev polynomial approximation in hyperbolic cross spaces against the polynomial degree n , for different dimension d ($d = 4, 6, 8$) with the same sample count scaling $M = N \log^2 N$. We notice that the dimension d does not affect the decay property of the condition number, and it seems that the higher the dimension d , the better conditioned the design matrix. Such an observation coincides with those in [22], pointing out that our estimate is still not optimal due to a large constant that depends on d . Similar results are presented in the right plot for the uniform measure, with $d = 3, 5, 7$.

FIG. 4.4. Condition number against polynomial degree k for different dimension d , with $M = N \log^2 N$. Left: the Chebyshev measure. Right: the uniform measure.

4.4. Convergence. In Figure 4.5, we report the convergence rate of the least-squares projection with randomly chosen Gaussian points, in the two dimensional total

degree space. The least-squares with Monte Carlo points are also reported for comparison. The underlying function to be approximated is $f(x) = \exp\left(-\sum_{i=1}^d c_i(x^i)^2\right)$. In the left-hand plot, we consider the Chebyshev measure, and note that the linear sample scaling rule results in the best approximation rate for random points ($\circ$) and Gaussian points ($*$), as well as for log-linear scaling ($\diamond$). However, the random Gaussian points work slightly better, and we remark that for the linear rule, instability may manifest when a large degree n is considered. In the right plot, we report the convergence for the uniform measure (for which the corresponding polynomials are Legendre polynomials). Again, the Gaussian points admit the best convergence pattern among the choices, both using linear rule ($*$) and the log-linear rule ($\diamond$). In contrast, the convergence rate for the random points ($\circ$ and $\triangleleft$) deteriorates when a large degree is used.

FIG. 4.5. Convergence rate against polynomial degree k for T_k^d with $d = 2$. Left: the Chebyshev measure. Right: the uniform measure.

4.5. Gaussian measure in unbounded domain. In Figure 4.6, we report the numerical results for the Gaussian measure with Hermite polynomial/function approximations. In the left picture, we plot the condition number against the polynomial order in the 4d hyperbolic cross space. Again, randomly chosen Gaussian points perform better than a Monte Carlo sampling method. In the right-hand plot, we test how the dimension d affects the condition number with a log-linear sample size scaling. Unlike the bounded measures in Figure 4.4, it seems that a higher dimension indeed introduces a larger condition number when the Gaussian measure is considered.

In Figure 4.7, we plot the approximation error with the Hermite functions for the underlying function $f(Y) = \exp\left(-\sum_{i=1}^d c_i x_i\right)$. Linear scaling of sample size results in very stable approach when the Gaussian points are used (blue) and the convergence is up to machine accuracy. In contrast, the convergence rate of the Monte Carlo points (red) deteriorates when large polynomial degree is used.

5. Conclusion. We discuss multivariate discrete least-squares projection onto polynomial spaces with a randomly chosen subgrid of the corresponding tensorized grid of Gaussian points. The framework provides a non-intrusive way to construct the generalized polynomial chaos expansions. We prove the stability and an optimal convergence estimate, provided the number of points scales *linearly* (up to a logarithmic factor) with the dimension of the polynomial space. The stability results include both the bounded measures such as the uniform and the Chebyshev measure, and

FIG. 4.6. Left: condition number against polynomial degree k for the 4-dimensional hyperbolic cross space (Gaussian measure). Right: Condition number against polynomial degree k for different dimension d , with $M = 1.5N \log N$.

FIG. 4.7. Left: Convergence rate against polynomial degree k for $\mathbb{P}_n^{TD}$ with $d = 2$. Left: the Chebyshev measure. Right: the uniform measure.

the unbounded measures which include the Gaussian measure. Our numerical results indicate that this method performs either markedly better (Hermite polynomials, Legendre polynomials) or no worse (Chebyshev polynomials) than standard Monte Carlo least-squares algorithms for discrete least-squares polynomial approximation.

REFERENCES

- [1] N. Agarwal and N.R. Aluru. A domain adaptive stochastic collocation approach for analysis of mems under uncertainties. *J. Comput. Phys.*, 228(20):7662–7688, 2009.
- [2] A.C. Atkinson, A.N. Donev, and R.D. Tobias. *Optimum Experimental Designs, with SAS*. Oxford Univ. Press, 2007.
- [3] I. Babuska, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.
- [4] R.A. Bailey. *Design of Comparative Experiments*. Cambridge Univ. Press, 2008.
- [5] M. Bieri. A sparse composite collocation finite element method for elliptic spDEs. Technical Report 2009-08, Seminar for Applied Mathematics, ETHZ, 2009.
- [6] G.E. Box, J.S. Hunter, and W.G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley, 2nd edition, 2005.
- [7] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone. Discrete least squares poly-

- nomial approximation with random evaluations-application to parametric and stochastic elliptic pdes. *EPFL, MATHICSE Technical Report*, 35/2013.
- [8] S.K. Choi, R. V Grandhi, R.A. Canfield, and C.L. Pettit. Polynomial chaos expansion with Latin hypercube sampling for estimating response variability. *AIAA J.*, (45):1191–1198, 2004.
 - [9] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.
 - [10] A. Doostan and H. Owhadi. A non-adaptive sparse approximation for PDEs with stochastic inputs. *J. Comput. Phys.*, 230(8):3015–3034, 2011.
 - [11] M. Eldred. Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, volume AIAA-2009-2249, 2009.
 - [12] J. Foo, X. Wan, and G.E. Karniadakis. The multi-element probabilistic collocation method (ME-PCM): error analysis and applications. *J. Comput. Phys.*, 227(22):9572–9595, 2008.
 - [13] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation methods for stochastic natural convection problems. *J. Comput. Phys.*, 225(1):652–685, 2007.
 - [14] Z. Gao and T. Zhou. On the choice of design points for least square polynomial approximations with application to uncertainty quantification. *Commun. Comput. Phys.*, 16:365–381, 2014.
 - [15] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., 1991.
 - [16] P. Goos and B. Jones. *Optimal Design of Experiments: A Case Study Approach*. Wiley, 2011.
 - [17] J. Hampton and A. Doostan. Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression. *Computer Methods in Applied Mechanics and Engineering*, 290:73–97, June 2015. arXiv:1410.1931 [math.ST].
 - [18] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363–386, January 2015. arXiv:1408.4157 [math.PR].
 - [19] X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *J. Comput. Phys.*, 228:3084–3113, 2009.
 - [20] R.L. Mason, R.F. Gunst, and J.L. Hess. *Statistical design and analysis of experiments with applications to engineering and science*. 1989.
 - [21] G. Migliorati and F. Nobile. Analysis of discrete least squares on multivariate polynomial spaces with evaluations in low-discrepancy point sets. *preprint*, 2014.
 - [22] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Approximation of quantities of interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces. *SIAM Journal on Scientific Computing*, 35(3):A1440–A1460, May 2013.
 - [23] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of the discrete L^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14(3):419–456, 2014.
 - [24] A. Narayan and D. Xiu. Stochastic collocation methods on unstructured grids in high dimensions via interpolation. *SIAM J. Sci. Comput.*, 34(3):A1729–A1752, 2012.
 - [25] P. Nevai. Géza freud, orthogonal polynomials and christoffel functions. a case study. *Journal of Approximation Theory*, 48(1):3–167, September 1986.
 - [26] P. Nevai, T. Erdélyi, and A. Magnus. Generalized jacobi weights, christoffel functions, and jacobi polynomials. *SIAM Journal on Mathematical Analysis*, 25(2):602–614, March 1994.
 - [27] F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.
 - [28] K. Petras. An asymptotic expansion for the weights of Gaussian quadrature formulae. *Acta Math. Hungar.*, 70(1-2):89–100, 1996.
 - [29] F. Pukelsheim. *Optimal Design of Experiments*. SIAM, 2006.
 - [30] M. Reagan, H. Najm, R. Ghanem, and O. Knio. Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. *Combustion & Flame*, 132:545–555, 2003.
 - [31] J. Shen, L.-L. Wang, and T. Tang. *Spectral Methods: Algorithms, Analysis and Applications*. Springer, 2011.
 - [32] G. Szegő. *Orthogonal Polynomials*. American Mathematical Society, Providence, RI, 1975.
 - [33] G. Tang and G. Iaccarino. Subsampled Gauss quadrature nodes for estimating polynomial chaos expansions. *SIAM/ASA J. Uncertainty Quantification*, 2(1):423–443, 2014.
 - [34] T. Tang and T. Zhou. On discrete least square projection in unbounded domain with random evaluations and its application to parametric uncertainty quantification. *SIAM J. Sci. Comput.*, 36:A2272–A2295, 2014.

- [35] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. <http://arxiv.org/abs/1011.3027>, 2011.
- [36] C.F. Wu and M.S. Hamada. *Experiments: planning, analysis, and optimization*. John Wiley & Sons, 2009.
- [37] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Comm. Comput. Phys.*, 2(2):293–309, 2007.
- [38] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, 2010.
- [39] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.
- [40] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, January 2002.
- [41] Z. Xu and T. Zhou. On sparse interpolation and the design of deterministic interpolation points. *SIAM J. Sci. Comput.*, 36:A1752–A1769, 2014.
- [42] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1 -minimization. *Int. J. Uncertainty Quantification*, 2(3):279–293, 2012.
- [43] X. Yang and G.E. Karniadakis. Reweighted l_1 minimization method for stochastic elliptic differential equations. *J. Comput. Phys.*, 248:87–108, 2013.
- [44] T. Zhou, A. Narayan, and Z. Xu. Multivariate discrete least-squares approximations with a new type of collocation grid. *SIAM J. Sci. Comput.*, 36:A2401–A2422., 2014.