Commun. Comput. Phys.
doi: 10.4208/cicp.020215.070515a

Vol. **18**, No. 1, pp. 1-36
July 2015

REVIEW ARTICLE

Stochastic Collocation on Unstructured Multivariate Meshes

Akil Narayan[1] and Tao Zhou[2,*]

[1] *Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112, USA.*
[2] *LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, the Chinese Academy of Sciences, Beijing, China.*

Received 2 February 2015; Accepted 7 May 2015

Abstract. Collocation has become a standard tool for approximation of parameterized systems in the uncertainty quantification (UQ) community. Techniques for least-squares regularization, compressive sampling recovery, and interpolatory reconstruction are becoming standard tools used in a variety of applications. Selection of a collocation mesh is frequently a challenge, but methods that construct geometrically *unstructured* collocation meshes have shown great potential due to attractive theoretical properties and direct, simple generation and implementation. We investigate properties of these meshes, presenting stability and accuracy results that can be used as guides for generating stochastic collocation grids in multiple dimensions.

AMS subject classifications: 6502, 65C20

Key words: Stochastic collocation, unstructured methes, least-squares, compressive sampling, least interpolation.

Contents

1 Introduction 2
2 Generalized polynomial chaos 3
3 Regression 8
4 Compressive sampling 17
5 Interpolation 24
6 Summary 31

*Corresponding author. *Email addresses:* akil@sci.utah.edu (A. Narayan), tzhou@lsec.cc.ac.cn (T. Zhou)

 ©2015 Global-Science Press

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

1 Introduction

The field of uncertainty quantification has enjoyed much attention in recent years as theoreticians and practitioners have tackled problems in the diverse areas of stochastic analysis, exascale applied computing, high-dimensional approximation, and Bayesian learning. The advent of high-performance computing has led to an increasing demand for efficiency and accuracy in predictive capabilities in computational models.

One of the persistent problems in Uncertainty Quantification (UQ) focuses on parameterized approximation to differential systems: Let u be a state variable for a system that is the solution to a physical model

$$\mathcal{L}(u;t,x,\omega)=0, \tag{1.1}$$

Above $x\in\mathbb{R}^p$ for $p=1,2,3$ is a spatial variable, $t\in\mathbb{R}$ is a temporal variable, and $\omega\in\Omega$ is a probabilistic event that encodes randomness on a complete probability space $(\Omega,\mathcal{F},\mathcal{P})$. We assume that the model (1.1) defines a map $\omega\mapsto u(t,x,\omega)$ with $u:\Omega\to B$ that is well-posed almost surely for some appropriate space B of (x,t)-dependent functions.

The operator $\mathcal{L}$ may represent any mathematical model of interest; examples that are popular in modern applied communities are elliptic partial differential equations, systems of time-dependent differential equations, parametric inverse problems, and data-driven optimization; e.g., [1, 17, 37, 57, 62, 69, 71]. The system defined by the operator $\mathcal{L}$ may include boundary value constraints, initial value prescriptions, physical domain variability, or any combination of these [81, 88, 89, 93].

The sought system response $u(t,x,\omega)$ is random/stochastic, given by the solution to (1.1). The stochastic dependence in (1.1) given by the event ω is frequently approximated by a d-dimensional random variable $Z(\omega)$. In some cases this parameterization of randomness is straightforward: e.g., in a Bayesian framework when ignorance about the true value of a vector of parameters is modeled by treating this parameter set as a random vector in (1.1). In contrast, it is common in models for an infinite-dimensional random field to contribute to the stochasticity, and in these cases parameterization is frequently accomplished by some finite-dimensional truncation procedure, e.g., via the Karhunen-Loeve expansion [44, 90], and this reduces the stochastic dependence in (1.1) to dependence on a random vector Z. In either case, a modeler usually wants to take $d\triangleq\dim Z$ as large as possible to encode more of the random variability in the model.

Under an assumption of model validity, the larger the stochastic truncation dimension d, the more accurate the resulting approximation. (Even when model validity is suspect, one can devise metamodeling procedures to capture model form error [50].) Therefore, it is mathematically desirable to take d as large as possible. We rewrite (1.1) to emphasize dependence on the d-dimensional random variable Z:

$$\mathcal{L}(u;t,x,Z)=0. \tag{1.2}$$

In this article we are ultimately interested in approximating $u(x,t,Z)$ or some functional

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

of it, and concentrate on the task of approximating u as a function of Z. This is the standard *modus operandi* for non-intrusive methods.

A major challenge for modern uncertainty quantification is the curse of dimensionality. Coined by Richard Bellman [3], this refers to the exponentially-increasing computational cost of resolving variability with respect to an increasing number of parameters. The trade-off that one frequently makes is that a large d induces an accurate stochastic truncation, but results in a computationally challenging problem since u depends on a d-dimensional parameter.

When Z is high-dimensional, model reduction techniques such as proper orthogonal decomposition methods [21, 36] or reduced basis methods [67, 70] are useful. In many situations these methods are powerful in their own right, robustly addressing problems in the scientific community; however, many implementations of these approaches are intrusive, meaning that significant rewrite of large legacy codebases is required. The focus of this paper is directed towards a different approach: non-intrusive response construction using multivariate polynomial collocation. "Non-intrusive" effectively means that existing black-box tools can be used in their current form. In particular we will focus on weighted methods, which are of concern for *stochastic collocation* methods. Stochastic collocation entails polynomial approximation in parameter (Z) space using either interpolation or regularized collocation approximation. These approaches have become extremely popular [20, 89, 91] for their efficiency and effectiveness. In many situations of interest, polynomial approximations converge to the true response exponentially with respect to the polynomial degree.

In stochastic collocation, a polynomial surrogate that predicts variability in parameter space is constructed from point-evaluations of the model response (1.2) at an ensemble of fixed parameter values $Z_n \in D$; we will call this ensemble of parameter values a *grid* or *mesh*, or a collection of *nodes*. While much work exists on geometrically structured meshes (e.g. tensor-product lattices or sparse grids), we will focus on *unstructured* meshes, which we believe is fertile ground deserving of much attention. Our use of terms 'structured' versus 'unstructured' refers to visual appearance of a lattice or geometric regularity of the mesh distribution in multivariate space. Obviously use of such a term is a subjective matter, and our goal is not to taxonomically classify collocation methods as structured or unstructured. Instead, the goal of this paper is to highlight some recent collocation strategies that distribute collocation nodes in an apparently unstructured manner; many of these recently developed methods produce approximation meshes that have attractive theoretical and computational properties. We remark that there are many alternative approaches in the UQ community, and one can refer to [80,89] for more detail discussions. However, in this current project, we shall focus only on the collocation approaches.

2 Generalized polynomial chaos

The generalized Polynomial Chaos method (gPC) [92] is essentially the strategy of approximating the Z-dependence of $u(x,t,Z)$ from (1.2) by a Z-polynomial. Let Z be a

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

random variable with density function $\rho(z)$, so that $P[Z\in A]=\int_A \rho(z)\mathrm{d}z$ for any Borel set A contained in the domain of Z, $A\subset D\subset\mathbb{R}^d$.

Hereafter, we will subsume t-dependence into the variable x and write $u(x,Z)\triangleq u(x,t,Z)$. A gPC method proceeds by making the ansatz that the variability of u in the random variable Z is described by a polynomial:

$$u_N(x,Z)=\sum_{n=1}^{N}\widehat{c}_n(x)\phi_n(Z), \tag{2.1}$$

for a prescribed multivariate polynomial basis ϕ_n and unknown coefficient functions $\widehat{c}_n(x)$. The gPC approach is to choose the basis ϕ_n to be the family of L^2_ρ-orthogonal polynomials,

$$\mathbb{E}\phi_n(Z)\phi_m(Z)=\int_{\mathbb{R}^d}\phi_n(z)\phi_m(z)\rho(z)\mathrm{d}z=\delta_{n,m}$$

with $\delta_{n,m}$ the Kronecker delta function. We also make the assumption that these polynomials are complete in the corresponding ρ-weighted L^2 space; this is satisfied if, for example, ρ is continuous and decays at least as fast as $\exp(-|z|)$ as $|z|\to\infty$. More intricate conditions can be found in [41].

In many cases of practical interest, it is natural to assume that the components of Z are mutually independent. In this case, the multivariate functions ϕ_n decompose into products of univariate functions. If the components of Z are independent, then $D=\prod_{j=1}^d I_j$ for univariate intervals $I_j\subset\mathbb{R}$ and $\rho(z)=\prod_{j=1}^d\rho_j\left(z^{(j)}\right)$, with $z=\left(z^{(1)},z^{(2)},\cdots,z^{(d)}\right)$ the components of z. Then the multivariate orthogonal polynomials are products of univariate orthogonal polynomials:

$$\phi_\alpha(z)=\prod_{j=1}^{d}\phi_{\alpha_j}^{(j)}\left(z^{(j)}\right),\quad \int_{I_j}\phi_n^{(j)}(z)\phi_m^{(j)}(z)\rho_j(z)\mathrm{d}z=\delta_{n,m}.$$

We have introduced multi-index notation: $\alpha=(\alpha_1,\cdots,\alpha_d)\in\mathbb{N}_0^d$ is a multi-index, $z^\alpha=\prod_{j=1}^d\left(z^{(j)}\right)^{\alpha_j}$, and $|\alpha|=\sum_{j=1}^d\alpha_j$. Depending on the situation, we will alternate between integer and multi-index notation for the polynomial ansatz:

$$u_N(x,Z)=\sum_{n=1}^{N}\widehat{c}_n(x)\phi_n(Z)=\sum_{\alpha\in\Lambda}\widehat{c}_\alpha(x)\phi_\alpha(Z),$$

where $\Lambda\subset\mathbb{N}_0^d$ is an index set with cardinality $|\Lambda|=N$. Any convenient mapping between $1,\cdots,N$ and Λ may be used. The choice of Λ defines the polynomial approximation space.

Some classical polynomial spaces to which u_N may belong are the tensor-product,

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

total degree, and hyperbolic cross spaces, respectively:

$$\Lambda_{d,k}^P = \left\{ \alpha \in \mathbb{N}_0^d \,|\, \max_j \alpha_j \leq k \right\}, \qquad P_k^d = \text{span}\left\{ z^\alpha \,|\, \alpha \in \Lambda_{d,k}^P \right\}, \tag{2.2a}$$

$$\Lambda_{d,k}^T = \left\{ \alpha \in \mathbb{N}_0^d \,|\, |\alpha| \leq k \right\}, \qquad T_k^d = \text{span}\left\{ z^\alpha \,|\, \alpha \in \Lambda_{d,k}^T \right\}, \tag{2.2b}$$

$$\Lambda_{d,k}^H = \left\{ \alpha \in \mathbb{N}_0^d \,|\, \prod_j \alpha_j + 1 \leq k+1 \right\}, \qquad H_k^d = \text{span}\left\{ z^\alpha \,|\, \alpha \in \Lambda_{d,k}^H \right\}. \tag{2.2c}$$

We have chosen particular definitions of these spaces above, but there are generalizations. E.g., dimensional anisotropy can be used to 'bias' the index set toward more important dimensions, or a different ℓ^p norm ($0 < p < 1$) can be placed on index space to tailor the hyperbolic cross spaces. The dimensions of T_k^d and P_k^d are

$$t_k^d \triangleq \dim T_k^d = \binom{d+k}{k}, \quad p_k^d \triangleq \dim P_k^d = (k+1)^d. \tag{2.3}$$

The dimension of H_k^d has the following upper bound [61]:

$$h_k^d \doteq \dim H_k^d \leq \lfloor (k+1)(1+\log(k+1))^{d-1} \rfloor.$$

For index sets Λ that do not fall into the categories defined by (2.2), we will use the notation $P(\Lambda)$ to denote the corresponding polynomial space.

For the index sets (2.2), we immediately see the curse of dimensionality: the dimensions of T_k^d and P_k^d increase exponentially with d, although t_k^d is smaller than p_k^d. The indices in the sets $\Lambda_{d,k}^P$, $\Lambda_{d,k}^T$, and $\Lambda_{d,k}^H$ are graphically plotted in Fig. 1 for $d=2$ and polynomial degree $k=25$.

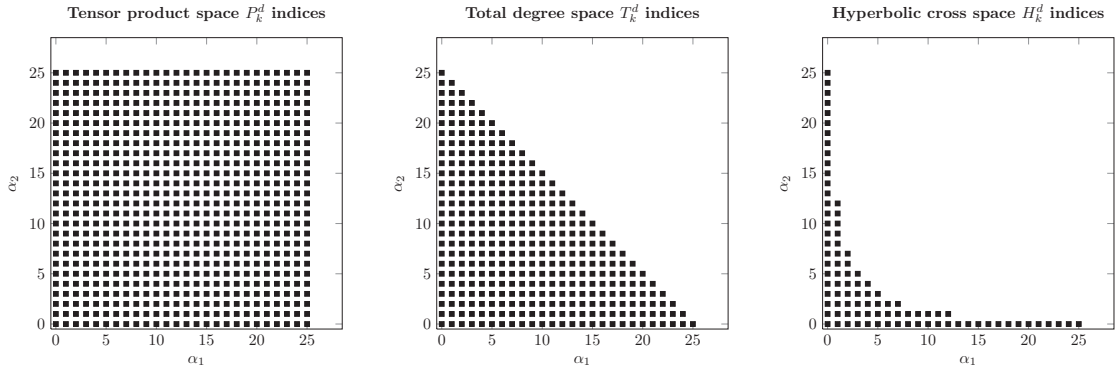


Figure 1: Indices associated with the tensor product set $\Lambda_{d,k}^P$ (left), the total degree set $\Lambda_{d,k}^T$ (center), and the hyperbolic cross set $\Lambda_{d,k}^H$ (right). All sets have dimension $d=2$ and degree $k=25$.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

This highlights a challenge with gPC in high dimensions: the number of degrees of freedom required to resolve highly oscillatory structure grows exponentially with dimension. (Indeed, this is a challenge for any non-adapted multivariate approximation scheme.) In the next section we narrow our focus to collocation schemes.

2.1 Stochastic collocation

The determination of $\widehat{c}_n(x)$ in (2.1) is the main difficulty, and one way to proceed is to ask that at some predetermined realizations $\{z_m\}_{m=1}^M$ of Z, the ansatz match the actual response:

$$u_N(x,z_m) \approx u(x,z_m), \quad m=1,\cdots,M, \tag{2.4}$$

where the approximation $\approx$ and the relation between N and M are discussed later in this section. (We will allow the ensemble of realizations to be randomly generated in some cases, but we will continue to use lowercase notation z_m to denote specific samples of the random variable Z.) This is a stochastic collocation approach and is *non-intrusive*: we need to compute $u(x,z_m)$, but this is accomplished by simply setting $Z=z_m$ in (1.2) and solving M realizations of the model equation. Therefore, existing deterministic solvers can be utilized. This ability to reuse existing solvers is one of the major strengths of non-intrusive (in particular, collocation) strategies.

In contrast, *intrusive* methods generally require a nontrivial mathematical reformulation of the model (1.2), and usually necessitate novel algorithm development. One popular intrusive method for gPC is the stochastic Galerkin approach, where u_N is specified by imposing that some probabilistic moments of the model equation residual vanish. Because these moment equations are coupled and are a novel formulation compared to the original type of equation, existing deterministic solvers of (1.2) cannot be used. The advantage of intrusive methods compared to non-intrusive methods is that one can usually make more formal mathematical statements about convergence of u_N with an intrusive formulation. For details of intrusive methods, see [44]. Although intrusive methods are advantageous in many situations, in this paper we only consider non-intrusive collocation approaches.

The focus of this paper discusses how to enforce the collocation conditions (2.4). We investigate three situations:

- $M>N$: when there are more constraints than degrees of freedom, we employ regression to attain a solution;
- $M<N$: with more degrees of freedom than constraints, we may seek sparse solutions and appeal to the theory of compressive sampling;
- $M=N$: with an equal number of linear constraints and degrees of freedom, we may enforce interpolation.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

Examples of sampling strategies that we consider as 'structured' are tensor-product constructions and sparse grid constructions. The former is quickly seen as infeasible for large dimensionality d. If we have an m-point one-dimensional grid (such as a Clenshaw-Curtis grid), then an isotropic tensorization has $M=m^d$ samples; this dependence on d is usually not computationally acceptable.

Sparse grids are unions of anistropically-tensorized grids, and have proven very effective [2, 19, 43, 66] at approximating high-dimensional problems. However, the sparse grids adherence to rigid and predictable layouts have the potential weakness of missing important features that do not line up with cardinal directions or coordinates in multivariate space, especially earlier in the computation when adaptive methods are seeded with isotropic grids. We believe that unstructured sample designs have the potential to mitigate these shortcomings.

2.2 Multivariate collocation

We introduce notation that is used throughout this article. In particular, we reserve the notation N and M to denote the number of terms in the expansion (2.1) and the number of collocation points in the ensemble (2.4), respectively. We will also use k to denote the polynomial degree of an index set Λ when this index set corresponds to one of the choices (2.2).

The main goal is to compute the coefficient functions $\widehat{c}_n$ from (2.1), and this can be decomposed into smaller linear algebra problems. In practice the realization $u(x,z_m)$ (as a solution to (1.2) with $Z=z_m$) is usually computed via a spatial discretization as an R-dimensional vector $\mathbf{u}(z_m)$. The type of spatial discretization to which $\mathbf{u}(z_m)$ corresponds typically does not influence Z-approximation strategies if non-intrusive methods are considered.

Let $\mathbf{A}$ be an $M\times N$ matrix with entries $(\mathbf{A})_{m,n}=\phi_n(x_m)$ with ϕ_n the orthogonal gPC basis. Then the conditions (2.4) can be written as

$$(\mathbf{A}\otimes\mathbf{I})(\mathbf{C})\approx\mathbf{U},$$

where $\mathbf{C}$ is an $RN\times 1$ vector with entries $C_{r,n}=\widehat{c}_{n,r}$ ordered lexicographically in (r,n), $\mathbf{U}$ is an $RM\times 1$ vector with entries $U_{r,m}=(\mathbf{u}(z_m))_r$ ordered lexicographically, and $\mathbf{I}$ is the $R\times R$ identity matrix. Then clearly the above system can be rewritten as the following decoupled series of systems:

$$\mathbf{A}\mathbf{c}_r\approx\mathbf{u}_r,\quad r=1,\cdots,R, \tag{2.5}$$

where $\mathbf{c}_r$ has entries $(\mathbf{c}_r)_n=\widehat{c}_{n,r}$, and $\mathbf{u}_r$ has entries $(\mathbf{u}_r)_m=(\mathbf{u}(x_m))_r$.

Thus the stochastic collocation solution to (1.2) under the polynomial ansatz (2.1) with conditions (2.4) is given by the solution to (2.5). Therefore, as is common in the stochastic collocation community, we focus entirely on solving the model problem

$$\mathbf{A}\mathbf{c}\approx\mathbf{u}, \tag{2.6}$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

for the coefficient vector $\mathbf{c}$ and given data $\mathbf{u}$ obtained from non-intrusive queries of the mode problem (1.2).

The three major approximations for (2.6) that we consider are:

1. Regression/regularization: we enforce (2.6) in the least-squares sense over the nodal array;

2. Interpolation: we enforce exact equality in (2.6);

3. Compressive sensing: under the assumption that the exact solution coefficients $c_{k,n}$ are sparse in n, we attempt to recover a sparse solution $\mathbf{c}$ to (2.6).

Our focus in this paper is on presentation on types of 'unstructured', high-order collocation approximation methods. We will introduce the above approximation methods and present some recently-developed unstructured mesh designs that complement each method. Our discussion revolves on the following considerations:

1. For a given mesh, how does the stability of the problem scale with the approximation order N and the sample size M?

2. For a given mesh, how is the accuracy of the reconstruction affected by the approximation order N and the sample size M?

3. Can we find or generate a mesh for which the approximation problem has 'nice' stability or accuracy properties?

These concerns undergird most of our future discussion.

3 Regression

Least-squares regression is one of the most classical approaches to collocation approximation, with vast literature for recovery with noisy data [7, 8, 46]. In UQ applications, the uncertainty of the input parameters can be treated as random variables Z. It is therefore reasonable to assume that all the stochasticity/uncertainty in the model is described by Z. Thus, we shall focus on the least-squares regression with noise-free data $\mathbf{u}$. Such UQ-focused least-squares regression (sometimes called point collocation [47]) has been widely explored [30, 55, 61, 74, 101].

For the least-squares approach we consider here, the gPC coefficients $\widehat{c}_n$ are estimated by minimizing the mean square error of the response approximation, i.e., one finds the least-squares solution u_N that satisfies

$$u_N := \underset{p\in P(\Lambda)}{\operatorname{argmin}} \frac{1}{M}\sum_{m=1}^{M}\left(p(z_m)-u(z_m)\right)^2, \tag{3.1}$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

where the index set Λ may be any general index set, e.g. P_k^d or T_k^d. For convenience we also introduce the discrete inner product on parameter space

$$\langle u,v\rangle_M = \frac{1}{M}\sum_{m=1}^{M} u(z_m)v(z_m) \tag{3.2}$$

and the corresponding discrete norm $\|u\|_M = \langle u,u\rangle_M^{1/2}$.

The formulation (3.1) is equivalent to requiring that the model problem (2.6) is satisfied in the following algebraic sense

$$\mathbf{c} = \underset{\mathbf{v}\in\mathbb{R}^N}{\operatorname{argmin}} ||\mathbf{A}\mathbf{v}-\mathbf{u}||_2. \tag{3.3}$$

Alternatively, the solution to the least-squares problem (3.3) can also be computed by solving an $N\times N$ system (the "normal equations"):

$$\hat{\mathbf{A}}\mathbf{c} = \hat{\mathbf{u}} \tag{3.4}$$

with

$$\hat{\mathbf{A}} := \mathbf{A}^\top\mathbf{A} = \left(M\langle\phi_\alpha,\phi_\beta\rangle_M\right)_{\alpha,\beta\in\Lambda}, \tag{3.5a}$$

$$\hat{\mathbf{u}} := \mathbf{A}^\top\mathbf{u} = \left(M\langle u,\phi_\alpha\rangle\right)_{\alpha\in\Lambda}. \tag{3.5b}$$

We will describe three kinds of such unstructured collocation grids, for which the corresponding theoretical analysis has been addressed recently and is under active development.

3.1 Monte Carlo sampling

Monte Carlo (MC) sampling is a natural choice for least-squares regression. For example, one generates independent and identically-distributed (iid) samples from a random variable with density ρ (recall that Z is a random variable with density ρ), and these samples form the nodal array $\{z_m\}$. This choice of sampling is certainly justifiable: It is straightforward to establish that the discrete formulation (3.3) converges to the L^2_ρ-optimal continuous formulation as $M\to\infty$:

$$\lim_{M\to\infty}\ \underset{p\in P(\Lambda)}{\operatorname{argmin}}\frac{1}{M}\sum_{m=1}^{M}\left(p(z_m)-u(z_m)\right)^2 = \underset{p\in P(\Lambda)}{\operatorname{argmin}}\mathbb{E}\left(p(Z)-u(Z)\right)^2.$$

It is thus not surprising that this least-squares formulation with random samples is popular [9,47,55,74]. In practical computations, the number of design samples M drawn from the input distribution scales linearly with the dimension of the approximating polynomial space N; taking $M\simeq cN$ with c between 2 and 3 is a common choice. Thus we desire

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

stability and convergence results under the assumption that the sample set size M scales *linearly* with the approximation space dimension N. Such results are not yet definitively available, but below we discuss what has been accomplished in this direction.

Perhaps the simplest example is MC with Z a uniformly-distributed random variable on a compact interval: Consider ρ the uniform measure on $[-1,1]$. The analysis in [30] shows that if one generates $M \sim N^2$ iid MC design samples drawn from ρ, then the spectral properties of the least-squares design matrix are controlled, implying stability for recovery with regression.

Theorem 3.1. [30] *Let ρ be the uniform measure on $[-1,1]$, and $\Lambda = \Lambda_{1,N-1}^T$. For any $r > 0$, assume that $\frac{M}{\log M} \geq CrN^2$ for a universal constant C. Then*

$$\Pr\left[\left|\left|\left|\hat{\mathbf{A}} - \mathbf{I}\right|\right|\right| \geq \frac{1}{2}\right] \leq 2M^{-r}.$$

This stability result can also be used to prove near-best approximation properties. These results are extended in [27] to multidimensional polynomial spaces with arbitrary lower index sets[†]. I.e., let ρ be the uniform measure on $[-1;1]^d$, and assume quadratic dependence $M \sim N^2$ as in the univariate case. If Λ in (3.5) is a lower set, then similar quadratic scaling $M \propto N^2$ ensures stability and near best approximation of the method independent of the dimension d.

If instead one considers sampling from the Chebyshev measure

$$\rho_c(z) = \prod_{j=1}^{d} \left(1 - \left(z^{(j)}\right)^2\right)^{-1/2}, \quad z \in [-1,1]^d \tag{3.6}$$

with the associated Chebyshev polynomial basis ϕ_α, then the quadratic dependence can be weakened to $M \sim N^{\frac{\ln 3}{\ln 2}}$, where $\frac{\ln 3}{\ln 2} \approx 1.58$. Such a technique was applied to a class of elliptic PDE models with stochastic coefficients, and an exponential convergence rate in expectation was established [27].

The analysis for unbounded state spaces D is less straightforward; for these unbounded domains, some of the tools used to establish the results above cannot directly be applied. Nevertheless, for the univariate exponential density function $\rho(z) = \exp(-z^2)$, the authors in [79] use a mapping technique in conjunction with weighted polynomials to establish stability. With this choice of ρ, the orthogonal family of Hermite polynomials $\phi_n(z) = H_n(z)$ would be the gPC basis choice. Consider approximation not with polynomials, but instead with the Hermite *functions*:

$$\psi_n(z) = \exp(-z^2/2)\phi_n(z) = \exp(-z^2/2)H_n(z).$$

[†]A set Λ is a *lower set* in this paper if for any $\alpha \in \Lambda$, all indices below it also lie in Λ: i.e., $\{\beta \in \mathbb{N}_0^d \,|\, \beta \leq \alpha\} \subset \Lambda$ holds for any $\alpha \in \Lambda$. Here, the ordering $\leq$ is the partial ordering on $\mathbb{N}_0^d$.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

The collocation samples z_m are *not* generated with respect to the density ρ. Instead, one first generates samples ξ_m of a uniform random variable on the interval $[-1,1]$ and subsequently maps these to the real line via the transformation

$$z(\xi)=\frac{L}{2}\log\frac{1-\xi}{1+\xi},\quad \xi\in(-1,1),$$

where the scalar L is a free parameter. Thus, the least-squares design matrix $\hat{\mathbf{A}}$ from (3.4) is now given by

$$\hat{\mathbf{A}}=\left(M\langle\psi_i,\psi_j\rangle_M\right)_{i,j=1,\cdots,N},\quad \langle u,v\rangle_M=\frac{1}{M}\sum_{m=1}^{M}u(z(\xi_m))v(z(\xi_m)). \tag{3.7}$$

The authors in [79] prove stability of the formulation (3.7), requiring only linear scaling of M with respect to N (modulo logarithmic factors).

Theorem 3.2. [79] *For any $r>0$, assume $\frac{M}{\log M}\gtrsim rN$ and $L\gtrsim\sqrt{N}$. Then the least-squares design matrix from* (3.7) *is stable in the sense*

$$\Pr\left[|||\hat{\mathbf{A}}-\mathbf{I}|||\geq\frac{5}{8}\right]\leq 2M^{-r},$$

The constant C is universal.

Because the hermite functions ψ_m are weighted versions of the Hermite polynomials H_m, one can consider the above a statement about stability of a weighted least-squares approximation problem.

In this subsection, Monte Carlo grids are themselves randomly generated, so most statements about stability and accuracy of these least-squares formulations are probabilistic in nature, e.g., convergence with high probability or convergence of the solution expectation. In the next subsection we consider one type of deterministically-generated mesh.

3.2 The Weil points

In certain applications, a judicious, deterministic choice of samples may provide several advantages over randomly-generated samples. In [101], the authors present a novel constructive analysis for the deterministically-generated *Weil points*. Suppose that $\mathcal{M}$ is a prime number; [101] proposes the following sample set:

$$\mathcal{W}_{\mathcal{M}}:=\left\{z_{j+1}=\cos(y_{j+1}):y_j=2\pi\left(j,j^2,\cdots,j^d\right)/\mathcal{M},\ j=0,\cdots,\lfloor\mathcal{M}/2-1\rfloor\right\}, \tag{3.8}$$

where $\lfloor\mathcal{M}/2\rfloor$ gives the integer part of $\mathcal{M}/2$. Note that the number of the points in the above grid is $M=\lfloor\mathcal{M}/2\rfloor$. In fact, with z_j as in (3.8), one can show that the points $\{z_j\}_0^{M-1}$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

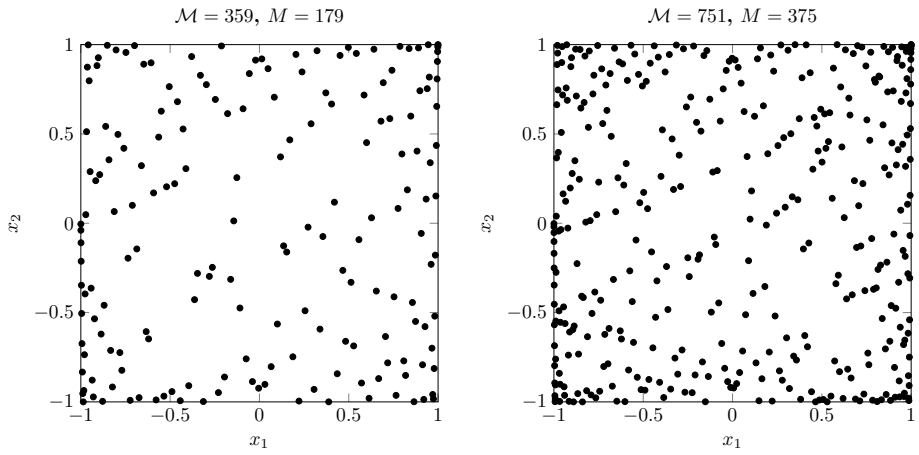


Figure 2: Weil grids in two dimensions ($d=2$). Left: prime number seed $\mathcal{M}=359$ resulting in $M=179$ points. Right: prime number seed $\mathcal{M}=751$ resulting in $M=375$ points.

coincide with the set of points $\{z_j\}_M^{\mathcal{M}}$, and so this latter half of the set is discarded. Examples of two-dimensional Weil grids are shown in Fig. 2 with $(\mathcal{M},M)=(359,179)$ and $(\mathcal{M},M)=(751,375)$.

The above collocation grid $\mathcal{W}_M$, designed originally for approximation when ρ is the Chebyshev measure, is motivated by the following formula of André Weil (hence the eponymous title "Weil points"):

Theorem 3.3 (Weil's formula [87]). *Let $\mathcal{M}$ be a prime number. Suppose $f(w)=m_1w+m_2w^2+\cdots+m_dw^d$ and there is a $j, 1\le j\le d$, such that $\mathcal{M}\nmid m_j$, then*

$$\left|\sum_{j=0}^{\mathcal{M}-1} e^{\frac{2\pi i f(j)}{\mathcal{M}}}\right| \le (d-1)\sqrt{\mathcal{M}}. \tag{3.9}$$

This formula plays a central role in deriving several properties about the Weil points, including least-squares stability results. Using Weil's formula, [101] shows that the Weil points distribute asymptotically according to the Chebyshev measure:

Theorem 3.4. [101] *Let $\mathcal{W}_{M_K}=\{z_{1,K},\cdots,z_{M_K,K}\}$ be the deterministic sampling set* (3.8) *generated by the K'th prime number. We define the empirical measure of the $\mathcal{W}_{M_K}$:*

$$\nu_K := \frac{1}{M_K}\sum_{j=1}^{M_K}\delta_{z_{j,K}},$$

where δ_z is the Dirac measure centered at z, and let ν_c be the Chebyshev measure with density $\mathrm{d}\nu_c(z)=\rho_c(z)$ from (3.6). *Then $\nu_K\to\nu_c$ in the weak-$*$ topology as $K\to\infty$.*

Having established what kind of measure the Weil points (3.8) sample according to, we turn to stability. Assume that ρ is the Chebyshev density; namely, we use the tensorized Chebyshev polynomials as the gPC polynomial basis. By utilizing the Weil points

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

in the least-squares framework, one can obtain estimates for the components of the design matrix $\hat{\mathbf{A}}$ by using Weil's formula Theorem 3.3. These estimates, in conjunction with Gerschgorin's Theorem, result in the following stability result:

Theorem 3.5. [101] *Suppose that* $\mathbf{I}$ *is the size-*N *identity matrix,* $\hat{\mathbf{A}}$ *is the design matrix associated with the Chebyshev gPC basis, and the Weil points are generated with the prime number seed* $\mathcal{M}$*, yielding* M *points. If* $M \geq C(d)\cdot N^2$*, then the following stability result holds*

$$\left|\left|\left| \frac{2^{d+1}}{M}\hat{\mathbf{A}} - \mathbf{I} \right|\right|\right| \leq \frac{1}{2},$$

where $|||\cdot|||$ *is the spectral norm.*

Therefore the corresponding least-squares problem admits a unique solution, provided that $M \geq C(d)\cdot N^2$. Stability results for least-squares problems are one ingredient for deriving convergence results. For example, the above stability result yields the following bound on least-squares error:

Theorem 3.6. [101] *Let* $f \in L^2_\rho$ *be a multivariate function, and let* p_N^* *be the* L^∞*-best approximating polynomial from* $P(\Lambda)$*, and let* $P_M^N f$ *be the least-squares* $P(\Lambda)$*-solution with the Weil points. If the prime number seed satisfies* $\mathcal{M} \geq C(d)\cdot N^2$*, then*

$$\|f - P_M^N f\|_{L^2_\rho} \leq C\|f - p_N^*\|_{L^\infty}.$$

The above theorem implies the L^2_ρ error for the least-squares solution is comparable to the L^∞ optimal approximation. However, we have assumed the quadratic dependence of the number of design points on the degree of freedom of the approximation space. This is stronger than the dependence $\mathcal{M} \sim N^{\frac{\ln 3}{\ln 2}}$ for the Monte Carlo sampling from the Chebyshev measure as discussed in Section 3.1. Nevertheless the construction here, being deterministic, does not suffer from probabilistic qualifiers on convergence results. (E.g., convergence "with high probability".) Moreover, we shall show with numerical examples that least-squares regression with Weil points performs comparably to Monte Carlo sampling.

For general (non-Chebyshev) types of orthogonal polynomial approximations on compact domains, [101] also proposed the following weighted least squares approach

$$u_{N,w} = P_{M,w}^N u = \underset{p\in P(\Lambda)}{\operatorname{argmin}} \frac{1}{M}\sum_{m=1}^{M} w_m \left(u(z_m) - p(z_m)\right)^2, \tag{3.10}$$

with positive weights w_m defined by

$$w_m = \frac{\rho(z_m)}{\rho_c(z_m)} = \pi^d \rho(z_m) \prod_{q=1}^{d} \left(1 - \left(z_m^{(q)}\right)^2\right)^{1/2},$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

where $z_m^{(q)}$ is the qth component of the grid point z_m. The basis set ϕ_α is the ρ-gPC basis, orthonormal under the weight ρ, and ρ_c is the Chebyshev density (3.6). The choice of the weights w_m above ensures that the induced change of measure makes the Weil-sampled w_m-weighted norm equivalent to the sought ρ-weighted norm.

An example of this will be illustrative: suppose we let ρ be the uniform (probability) density $\rho(z)\equiv 2^{-d}$ on $[-1,1]^d$. Then we have

$$w_m = \frac{\rho(z_m)}{\rho_c(z_m)} = (\pi/2)^d \prod_{q=1}^{d} \left(1-\left(z_m^{(q)}\right)^2\right)^{1/2}. \tag{3.11}$$

Since w_m is applied to the quadratic form (3.10), we are effectively preconditioning $f(z_m)$ with $\sqrt{w_m}$. Thus, if the ϕ_j are tensor-product Legendre polynomials (orthonormal under the uniform density), then we are preconditioning our expansion as

$$\sum_{j=1}^{N} \widehat{c}_j \phi_j(z) \longrightarrow \sum_{j=1}^{N} \widehat{c}_j \sqrt{w_j}\phi_j(z) = \sum_{j=1}^{N} \widehat{c}_j \left(\prod_{q=1}^{d}\left(1-(z^{(q)})^2\right)^{1/4} \phi_j(z)\right).$$

This type of preconditioning is known to produce well-conditioned design matrices in the context of ℓ^1 minimization for Legendre approximations [73]. Of course if $\rho \propto \rho_c$, then we obtain constant weights. Therefore, the weights proposal (3.11) reduces to well-known preconditioning techniques for some special cases.

3.3 Structured random points

Although a structured grid, such as the tensor-product grid, will frequently be impractical in computations for $d \gg 1$, the grid itself may be used as a candidate set on which to extract a subset of points that is useful for approximation. One way to do this is to randomly sample a subset grid from a high-cardinality structured candidate set, thus producing a subset grid that is essentially unstructured.

For example, assume that we try to find an approximation in the hyperbolic cross space H_k^d with $N=h_k^d$. We assume that the functions $\phi_\alpha(x)$ that span H_k^d are tensor-product Chebyshev polynomials. Such an expansion can also be viewed as an expansion in the tensor product space P_k^d, i.e.

$$u(z) = \sum_{\alpha\in\Lambda_{d,k}^P} \widehat{c}_\alpha \phi_\alpha(z), \quad \widehat{c}_\alpha = 0 \quad \text{if} \quad \alpha \in \Lambda_{d,k}^P \backslash \Lambda_{d,k}^H. \tag{3.12}$$

Now let $\{y_m\}_{m=1}^{p_k^d}$ be the tensorized grid of the one dimensional Chebyshev Gauss quadrature points, where $p_k^d=(k+1)^d$. With such a tensor grid $\{y_m\}$, one can exactly recover any polynomial in P_k^d by the generalized discrete Fourier transform. The normal equations design matrix is the identity matrix in this case. Because $h_k^d \ll t_k^d$ when d is large, we have

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

the freedom to choose a subset of points with cardinality M from the full tensor grid, with M satisfying

$$h_k^d < M < p_k^d. \tag{3.13}$$

We will select these points z_m randomly with the equal probability law on the candidate set $\{y_m\}$. This idea is proposed in [100], and it is shown that, using the Chebyshev density ρ and associated gPC basis, the design matrix is stable with probability at least $1-2M^{1-\mu}$, for any $\mu \geq 2$, provided that

$$\frac{M}{\log M} > C\mu N. \tag{3.14}$$

The framework in [100] applies to approximations with densities ρ more general than Chebyshev measures. In particular, it includes measures on bounded domains such as the uniform measure, and on unbounded measures such as the Gaussian measure.

3.4 Numerical tests

We now provide some numerical examples to illustrate the stability and the convergence properties of the least-squares approach with the design points described above. Many more related tests are available in [61, 100, 101]. We first investigate how the number of collocation points affects the condition number, $\mathrm{cond}(\hat{\mathbf{A}}) = \frac{\sigma_{\max}(\hat{\mathbf{A}})}{\sigma_{\min}(\hat{\mathbf{A}})}$, where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values, respectively. These results are shown in Fig. 3 for the 4-dimensional hyperbolic space H_k^4. The orthogonal polynomial measure ρ is the Chebyshev measure and the basis elements are tensor-product Chebyshev polynomials. We test the three design sampling methods described in the previous sections: "MC" corresponds to the Monte Carlo design of Section 3.1, "Weil points" corresponds to the Weil points design of Section 3.2, and "Gaussian" refers to random subset sampling from a tensor-product Gauss quadrature grid from Section 3.3.

Because the design matrices with the structured points and the random points are random matrices, we repeat each of these tests 200 times and report the mean condition number. We also plotted the error bars for the two kinds of random samples corresponding to

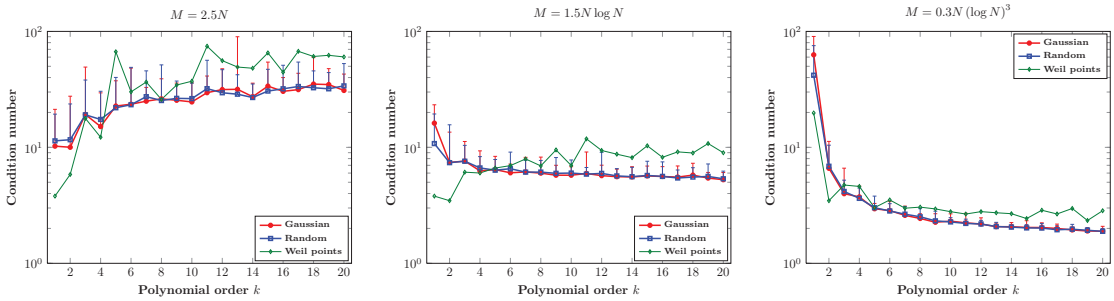


Figure 3: Condition number with respect to the polynomial degree k in the 4-dimensional hyperbolic spaces. Left: $M = \frac{5}{2}N$. Middle: $M = \frac{3}{2}N\log N$. Right: $M = 0.3N\log^3 N$.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

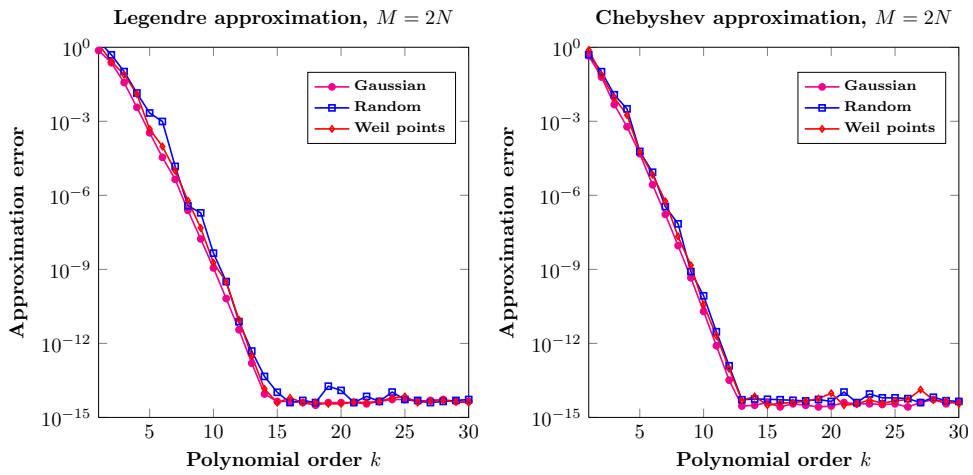


Figure 4: Approximation error in the two dimensional total degree space. Left: Chebyshev approximation. Right: Legendre approximation.

one standard deviation above the mean. The left plot of Fig. 3 shows results for linear scaling of M with respect to N, i.e. $M \sim \frac{5}{2}N$, and the middle plot shows log-linear dependence $M \sim \frac{3}{2}N\log N$. The right plot shows and even stronger dependence, $M \sim 0.3N\log^3 N$, We note that the performance of all three methods is similar, in the sense that the log-linear dependence (middle and right plots) admits stable condition numbers with respect to the polynomial order k. In contrast, the linear rule (left plot) admits modest growth of the condition number with respect to the polynomial order k. We also observe that the Weil points perform slightly worse compared to the other two types; however, we emphasize that the Weil sampling method and the corresponding analysis is deterministic. Not shown is quadratic dependence, $M \sim N^2$, because the Fig. 3 shows that log-linear dependence is enough to guarantee stability. Thus, there seems to be room for improving the quadratic dependence estimate in [101].

Finally, we test the convergence rate of the least-squares approaches in $d = 2$ dimensions. The target function is smooth, chosen as $u(z) = \exp^{-\sum_{i=1}^{d} a_i z_i}$, where the parameters $\{a_i\}$ are generated randomly. We measure the error in the L^∞ norm, computed on a set of 2000 points that are independent samples (i.e., different from the design points) from a uniform distribution on $[-1,1]^2$. The errors with respect to the polynomial order k from the two-dimensional total degree space T_k^d are shown in Fig. 4. In the left-hand pane, the underlying measure is the Chebyshev measure and the basis functions ϕ_α are Chebyshev polynomials. In the right-hand plot, the underlying measure is the uniform measure and we use the Legendre polynomials as basis elements. In this framework, we will use the preconditioned version of least-squares (3.11) with Chebyshev-like design points (i.e., random Chebyshev points, the Weil points, and points randomly selected from Chebyshev Gauss points). In all the plots, we have used the linear rule $M = 2N$ which is dependence that is more feasible in practical computations.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

The results given in Fig. 4 illustrate that the linear rule display the exponential convergence with respect to k, for both the least-squares and its preconditioned version. The convergence stagnates at machine precision, which is expected. Such a result also points out a gap between the M/N dependence necessary to achieve optimality in current theory, and the condition that in practice yields an optimal convergence rate.

Remark 3.1. We note that Quasi Monte Carlo points are deterministically-generated and can also be used in the least-squares framework, and one can find such investigations in [28, 42]. We also remark that the random parameters here are assumed to be supported in $[-1,1]^d$. One may of course encounter problems with unbounded parameters, e.g., problems with Gaussian/Gamma parameters. Recent work in [63, 79] investigates such situations. In particular, the recent work [63] provides a novel Christoffel function weighted least squares approach, with a equilibrium measure sampling strategy, and they show numerically that the linear dependence seems enough to obtain the stability for both the bounded and unbounded problems.

4 Compressive sampling

Compressive sampling (CS), or compressive sensing, considers recovery the of a sparse representations from limited data, and in applications this is usually considered when there is insufficient information about the target function. In the framework of this paper, this occurs when the number of samples M is less than the cardinality of the polynomial space for the approximation N. CS is an emerging and maturing area of research in signal processing that aims at recovering sparse signals accurately from a small number of their random projections (see e.g. [23, 24, 26, 29, 39] and references therein). A sparse signal is simply a signal that has only few significant coefficients when linearly expanded into a non-adapted basis, in our case the $\{\phi_\alpha\}$. Thus, when the sample count M is smaller than the approximation dimension N then one can use to CS approaches to construct a polynomial approximation, under the assumption that the underlying target function is sparse in ϕ_α. The typical approach in CS is to minimize a norm of the polynomial, with the constraint of matching the data. For sparsity, one seeks to minimize the ℓ_0 norm $\|\mathbf{c}\|_0$ of the coefficient vector, and under certain conditions the minimizing coefficient vector is identical to that which minimizes the ℓ_1 norm $\|\mathbf{c}\|_1$. Solving the latter minimization problem is preferred in practice because such problems are convex optimization problems.

The success of the CS lies in the assumption that in practice many target functions are sparse: what appear to be a signal with many features may contain only a small number of notable terms when transformed to the frequency domain. This is indeed the case in many UQ problems. For instance, solutions to linear elliptic PDEs with high-dimensional random coefficients admit sparse representations with respect to the gPC basis under some mild conditions [6, 31, 83].

For stochastic collocation methods in the CS framework, we are interested in the case that the number M of solution samples is much smaller than the number N of unknown

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

coefficients. One then seeks to a solution $\mathbf{c}$ with the minimum number of non-zero terms. This can be formulated as the optimization problem

$$\mathbf{P}_0\colon \operatorname{argmin}||\mathbf{c}||_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{c}=\mathbf{u}, \tag{4.1}$$

where $||\mathbf{c}||_0 : \#\{\alpha : \widehat{c}_\alpha \neq 0\}$ is the ℓ^0 norm on vectors and should be interpreted as the number of non-zero components of $\mathbf{c}$. In general, the global minimum solution of $\mathbf{P}_0$ is not unique and is NP-hard to compute. Fortunately, under *restricted isometry* conditions on the design matrix, the computed ℓ_1 minimizer approximates the ℓ_0 minimizer very well, even with noisy measurements [25,26]. The ℓ_1 minimization problem in our context is

$$\mathbf{P}_1\colon \operatorname{argmin}||\mathbf{c}||_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{c}=\mathbf{u}. \tag{4.2}$$

(Above, $\|\cdot\|_1$ is the ℓ^1 norm on vectors.) The advantage of the $\mathbf{P}_1$ formulation is that it is a convex problem, and so computational solvers for convex problems may be leveraged [97,99].

In practice, since the approximation basis is truncated to a finite number of functions, the approximation $\mathbf{u}$ may not be exact. Further, the measurement vector $\mathbf{u}$ may be corrupted by noise. These factors lead to the relaxation of the equality in (4.2) and reformulate the problem under the Basis Pursuit Denoising form:

$$\mathbf{P}_{0,\varepsilon}\colon \operatorname{argmin}||\mathbf{c}||_0 \quad \text{subject to} \quad ||\mathbf{A}\mathbf{c}-\mathbf{u}||_2 \leq \varepsilon, \tag{4.3a}$$

$$\mathbf{P}_{1,\varepsilon}\colon \operatorname{argmin}||\mathbf{c}||_1 \quad \text{subject to} \quad ||\mathbf{A}\mathbf{c}-\mathbf{u}||_2 \leq \varepsilon. \tag{4.3b}$$

In what follows, we will focus on the traditional ℓ_1-minimization form 4.2. Consider the ℓ^p error in the best s-term approximation of a coefficient vector $\mathbf{c}\in\mathbb{R}^N$

$$\sigma_{s,p}(\mathbf{c}) = \inf_{||\mathbf{y}||_0\leq s} ||\mathbf{y}-\mathbf{c}||_p. \tag{4.4}$$

Clearly, $\sigma_{s,p}(\mathbf{c})=0$ if $\mathbf{c}$ is s-sparse, i.e., $||\mathbf{c}||_0 \leq s$. Let $\mathbf{A}$ be an $M\times N$ matrix. Define the restricted isometry constant (RIC) $\delta_s < 1$ to be the smallest positive number such that the inequality

$$(1-\delta_s)||\mathbf{c}||_2^2 \leq ||\mathbf{A}\mathbf{c}||_2^2 \leq (1+\delta_s)||\mathbf{c}||_2^2 \tag{4.5}$$

holds for all $\mathbf{c}\in\mathbb{R}^N$ satisfying $\|\mathbf{c}\|_0 \leq s$. If the above holds, the matrix $\mathbf{A}$ is said to satisfy the s-restricted isometry property (RIP) with restricted isometry constant δ_s. Now, we are ready to state the following sparse recovery for RIP matrices.

Theorem 4.1. [72] *Let $\mathbf{A}\in\mathbb{R}^{M\times N}$ be a matrix with RIC such that $\delta_s \leq 0.307$. For a given $\tilde{\mathbf{c}}\in\mathbb{R}^N$, let $\mathbf{c}^\#$ be the solution of the ℓ_1-minimization*

$$\operatorname{argmin}||\mathbf{c}||_1 \quad \textit{subject to} \quad \mathbf{A}\mathbf{c}=\mathbf{A}\tilde{\mathbf{c}}. \tag{4.6}$$

Then the reconstruction error satisfies

$$||\mathbf{c}^\#-\tilde{\mathbf{c}}||_2 \leq C\frac{\sigma_{s,1}(\tilde{c})}{\sqrt{s}} \tag{4.7}$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

for some constant $C>0$ that depends only on δ_s. In particular, if $\tilde{\mathbf{c}}$ is s-sparse then reconstruction is exact, i.e., $\mathbf{c}^{\#}=\tilde{\mathbf{c}}$.

The theorem above indicates that, as with regression, the choice of nodal array is of great importance: A 'good' nodal array will lead to a design matrix $\mathbf{A}$ with an acceptable RIC so that Theorem 4.1 may be invoked for convergence.

4.1 Monte Carlo sampling

As in the least-squares framework, the MC sampling method is still promising in the CS framework. Indeed, much of the pioneering CS work employed MC sampling [23, 24, 26, 39]. However, using such an idea for UQ applications is relatively new. Some of the first work in this area was investigated in [40, 58], where the authors applied CS ideas to stochastic collocation and obtained some key properties, such as the probability under which the sparse random response function can be recovered.

The authors in [72, 73] investigated the recovery of expansions that are sparse in a univariate Legendre polynomial basis. Their analysis relies strongly on RIP results from bounded orthonormal polynomial systems, which we summarize in the following:

Theorem 4.2. [72] *Let $\{\phi_n\}$ be a bounded orthonormal system, namely,*

$$\sup_n ||\phi_n||_\infty = \sup_n \sup_z |\phi_n(z)| \le L \tag{4.8}$$

for some constant $L\ge 1$. Let $\mathbf{A}\in\mathbb{R}^{M\times N}$ be the interpolation matrix with entries $\left\{a_{n,m}=\phi_{n-1}(z_m)\right\}_{1\le n\le N,1\le m\le M}$, and let $\mathbf{W}$ be the $M\times M$ diagonal matrix with entries $w_{m,m}=(\pi/2)^{1/2}(1-z_m^2)^{1/4}$, where the points $\{z_m\}_{1\le m\le M}$ are iid samples drawn from the one-dimensional Chebyshev measure (3.6)*. Assume that there is a $\delta>0$ such that*

$$M > C\delta^{-2}L^2 s\log^3(s)\log(N). \tag{4.9}$$

Then with probability at least $1-N^{-\gamma\log^3(s)}$, the RIC δ_s of $\frac{1}{\sqrt{M}}\mathbf{WA}$ satisfies $\delta_s\le\delta$. Here the $C,\gamma>0$ are universal constants.

The authors in [73] use the above use to provide the following recoverability result for one dimensional sparse Legendre polynomial expansions:

Theorem 4.3. [73] *Let z_m be iid samples from the Chebyshev measure, and let $\mathbf{A}\in\mathbb{R}^{M\times N}$ be the corresponding Legendre polynomial ϕ_n design matrix with entries*

$$\left\{a_{m,n}=\phi_{n-1}(z_m)\right\}_{1\le n\le N,1\le m\le M},$$

and let $\mathbf{W}$ be a diagonal matrix with entries $w_{m,m}=(\pi/2)^{-1/2}(1-z_m^2)^{1/4}$. Assume that

$$M > Cs\log^3(s)\log(N). \tag{4.10}$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

Let $\tilde{\mathbf{c}} \in \mathbb{R}^N$ be any coefficient vector, and consider the solution $\mathbf{c}^\#$ to the following ℓ^1 optimization problem:

$$\mathbf{c}^\# = \operatorname{argmin} ||\mathbf{c}||_1 \quad \textit{subject to} \quad \mathbf{WAc} = \mathbf{WA}\tilde{\mathbf{c}}. \tag{4.11}$$

Then with high probability the solution $\mathbf{c}^\#$ is within a factor of the best s-term approximation error. I.e.,

$$Pr\left[\left\| \mathbf{c}^\# - \tilde{\mathbf{c}} \right\|_2 \leq \frac{C\sigma_{s,1}(\tilde{\mathbf{c}})}{\sqrt{s}} \right] \geq 1 - N^{-\gamma \log^3(s)}.$$

Above, C and γ are universal constants.

Note that in the above theorem, the random samples are drawn from the Chebyshev measure, namely, the CS method above is a Legendre-preconditioned ℓ_1-minimization with Chebyshev samples. The preconditioned/weighted Legendre polynomials have a uniform bound [78] when the weight is $(1-z^2)^{1/4}$, and this is the main reason to define the weight matrix $\mathbf{W}$ to obtain Theorem 4.2 or Theorem 4.3.

The above result is extended in [96] to high-dimensional problems, for both the original ℓ_1-minimization and the preconditioned ℓ_1-minimization. We summarize these results with the following theorem:

Theorem 4.4. [96] *Let $\{\phi_n\}_{n=0}^{N-1}$ be the Legendre polynomial bases of the total degree space T_k^d, and let $u(z) = \sum_{n=0}^{N-1} \widehat{c}_n \phi_n$ be an arbitrary polynomial with coefficient vector $\tilde{\mathbf{c}}$. For some nodal array $\{z_m\}_{1\leq m\leq M}$, let $\mathbf{A}$ and $\mathbf{W}$ be the corresponding $M\times N$ design matrix and $M\times M$ diagonal preconditioner/weight matrix, respectively, with entries*

$$a_{m,n} = \phi_{n-1}(z_m), \quad w_{m,m} = \left(\frac{\pi}{2}\right)^{-d/2} \prod_{j=1}^{d} \left(1 - \left(z_m^{(j)}\right)^2\right)^{1/4}.$$

1. *Let $\{z_m\}_{1\leq m\leq M}$ be i.i.d. random samples drawn from the uniform measure, and assume that*

$$M > 3^k s \log^3(s) \log(N).$$

Then with high probability, the solution $\mathbf{c}^\#$ to the direct ℓ^1 minimization problem (4.6) *is within a factor of the best s-term error:*

$$Pr\left[\left\| \mathbf{c}^\# - \tilde{\mathbf{c}} \right\|_2 \leq \frac{C\sigma_s(\tilde{\mathbf{c}})_1}{\sqrt{s}} \right] \geq 1 - N^{-\gamma \log^3(s)}.$$

2. *Let $\{z_m\}_{1\leq m\leq M}$ be i.i.d. random samples drawn from the Chebyshev measure, and assume that*

$$M > 2^d s \log^3(s) \log(N).$$

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

Then with high probability, the solution $\mathbf{c}^{\#}$ *to the preconditioned/weighted* ℓ^1 *minimization problem* (4.11) *is within a factor of the best s-term error:*

$$Pr\left[\left\|\mathbf{c}^{\#}-\tilde{\mathbf{c}}\right\|_2 \leq \frac{C\sigma_s(\tilde{\mathbf{c}})_1}{\sqrt{s}}\right] \geq 1-N^{-\gamma\log^3(s)}.$$

For both of the above cases, the constants C *and* γ *are universal.*

The above results implies that in very high dimensional spaces $(d>k)$, the Chebyshev preconditioned ℓ^1-minimization may be less efficient than the direct ℓ^1-minimization, because it may require more sample points when $2^d>3^k$. Of course when $d=1$ only the $k=0$ trivial case satisfies this inequality, so in one dimension the preconditioned case is more effective [73].

We remark that some modified CS methods have also been investigated in the random sampling framework, one can refer to [68,98] for the weighted (re-weighted) approaches.

4.2 Deterministic sampling

Although random sampling methods have been widely used in the CS framework, a judicious, deterministic choice of points may provide several advantages over randomly-generated points. Deterministic CS sampling and recoverability is a well-studied field for recovery of sparse trigonometric polynomials [16,38,48,49,94].

However, such a framework has not been widely investigated for UQ applications, especially for recoverability of general types of sparse polynomial expansions. In [95], the authors use the Weil points (3.8) to recover sparse Chebyshev polynomials. They use the Weil exponential sum formula, Theorem 3.3, to control the incoherence parameter of the design matrix, which in turn can be used to ascertain RIC information. More precisely, we have

Theorem 4.5. [95] *Let*

$$u=\sum_{\alpha\in\Lambda} c_\alpha\phi_\alpha$$

be an arbitrary Chebyshev polynomial expansion in $P(\Lambda)$ *with coefficient vector* $\boldsymbol{c}$*. Suppose that* $\mathcal{M}\geq C(\Lambda)s^2$ *is a prime number, and assume that* $\boldsymbol{c}^{\#}$ *is given by the* ℓ_1*-minimization problem* (4.6) *with the design matrix* $\mathbf{A}$ *being the evaluations of the Chebyshev bases on the Weil points generated by prime number seed* $\mathcal{M}$*. Then*

$$\|\mathbf{c}^{\#}-\mathbf{c}\|_2 \lesssim \frac{\sigma_{s,1}(\tilde{\mathbf{c}})}{\sqrt{s}}. \tag{4.12}$$

Thus, ℓ^1 minimization can recover s-sparse Chebyshev polynomials, provided that the number of the Weil samples scales quadratically with the sparsity s, i.e., $M\sim C(\Lambda)s^2$. The results apply to any high dimensional polynomial spaces with downward closed

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

multi-index sets, such as the T_k^d, P_k^d and H_k^d. However, different spaces result in different constants $C(\Lambda)$, and the estimates for $C(\Lambda)$ in [95] are not optimal. Although the Weil estimates obtained are not optimal in the sense that they require that the number of sample points M scale quadratically on the sparsity s. Nevertheless, we will show via numerical tests that the Weil points have a similar recoverability properties when compared with MC samples.

The results indicating that the Weil points perform comparably to MC sampling (see Section 4.3) are not necessarily surprising: We have shown in Theorem 3.4 that the Weil points distribute asymptotically according to the Chebyshev measure. Thus, it might be expected that the Weil points produce similar performance as MC Chebyshev samples. Moreover, [95] also proposed the preconditioned ℓ^1-minimization to handle general sparse polynomials with the Weil points. However, such a framework has the similar drawback as we discussed in the last section: In very high dimensional spaces $(d>k)$, the preconditioned ℓ^1-minimization may be less efficient than direct ℓ^1-minimization.

Remark 4.1. Similar to Section 3.3, one may use what we refer to as structured random points (samples randomly chosen from a tensor-product Gauss quadrature grid, or some other candidate set) to recover sparse polynomials. The idea is similar to Section 3.3, so we will not discuss it in detail; However, we will test the numerical performance of this method in the next section.

4.3 Examples

This section parallels Section 3.4. We will compare the performance of MC samples, Weil points, and random subsampling from a structured tensor-product Gauss quadrature grid. We are interested in the recovery performance with different kinds of points via preconditioned/weighted ℓ_1 minimization. We assume the target (exact) function has a polynomial form, i.e. $u(z)=\sum_{\alpha\in\Lambda}\widehat{c}_\alpha\phi_\alpha(z)$ with $\|\widehat{c}\|_0=s$, and attempt to recover this vector. For a given sparsity level s, we fix s coefficients of the polynomial while keeping the rest of the coefficients zero. The values of the s non-zero coefficients are drawn as iid samples from a standard normal distribution. We repeat the experiment 100 times for each fixed sparsity s and calculate the success rate on these 100 runs. ("Success" here means that $\|\hat{\mathbf{c}}-\mathbf{c}^\#\|_{\ell^\infty}\leq 10^{-4}$.)

As in Section 3.4, we use the terms "MC", "Weil", and "Gaussian" to refer to Monte Carlo sampling (Sections 3.1 and 4.1), Weil points sampling (Sections 3.2 and 4.2), and subset sampling from a tensor-product Gauss quadrature grid (Section 3.3), respectively.

To solve the ℓ_1 minimization, we employ the available tool SPGL1 from [85] that was implemented in Matlab. We will conduct two groups of tests, the first using a Chebyshev polynomial expansion, and the second using a Legendre polynomial expansion using the preconditioning strategy of Theorems 4.3 and 4.4.

The first test is a low dimensional test, with the index set Λ corresponding to the two dimensional total degree space T_k^2. In the left-hand plot of Fig. 5, we show the recovery

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

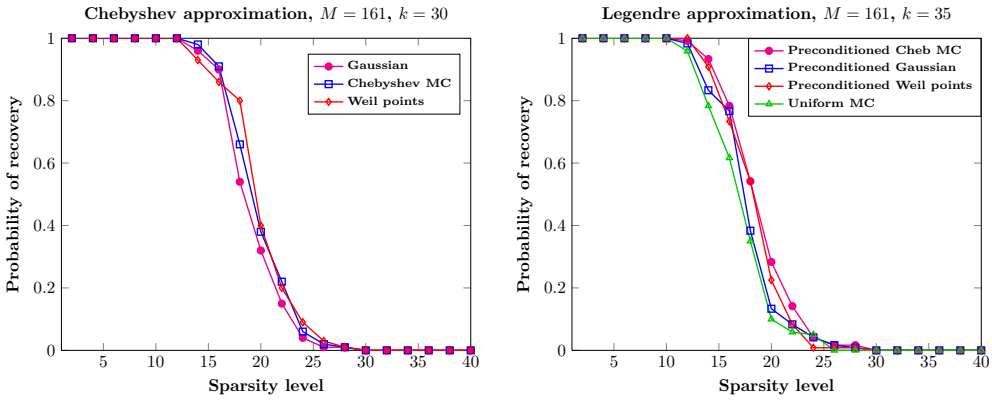


Figure 5: Recovery probability with respect to sparsity s in the two-dimensional total degree space T_k^2. Left: recovery of sparse Chebyshev polynomial representations. Right: recovery of sparse Legendre polynomial representations.

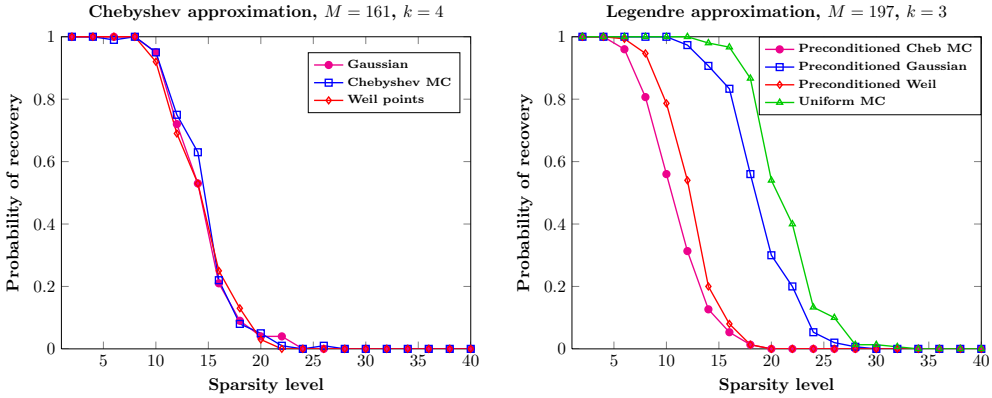


Figure 6: Recovery probability with respect to sparsity s in the 15-dimensional total degree space. Left: recovery of sparse Chebyshev polynomial representations. Right: recovery of sparse Legendre polynomial representations.

rate for sparse Chebyshev polynomials (with k=20,d=2, and M=74) as a function of sparsity level s. We see that the three kinds of design points have very similar performance. In the right-hand plot, we show the recovery results for sparse Legendre polynomials (with k=35,d=2 and M=66) using the preconditioned formulation (4.11). We have also tested *direct* ("unpreconditioned") recovery with MC samples drawn from the *uniform* measure. For this low dimensional test, the preconditioned results are slightly better compared to direct ℓ_1-minimization.

The second example we consider is the 15-dimensional total degree space T_k^{15}. In Fig. 6, the left-hand plot shows the recovery results for sparse Chebyshev polynomials (with $k=4,d=15$), with respect to the sparsity level s and a fixed number of design points ($M=81$). Again, we see that the three kinds of design points have very similar performance. The right-hand plot provides recovery results for the sparse Legendre polynomials (with k=4,d=15, and M=97). For this higher dimensional test, the direct

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

ℓ_1-minimization with MC uniform points performs better than the preconditioned versions (with Weil points, the Chebyshev MC points, and the Gaussian points). This is a direct manifestation of the analysis provided by Theorem 4.4. Finally, we remark that within the preconditioned tests shown in Fig. 6, the Gaussian points have a noticeably performance than the other two.

5 Interpolation

The last type of approximation we consider is interpolation: given a general array $\mathcal{Z}_M = \{z_1,\cdots,z_M\}$ of unique nodes and data $u_m = u(z_m)$, we wish to construct a polynomial p satisfying

$$p(z) = \sum_{\alpha\in\Lambda} c_\alpha \phi_\alpha(z), \quad p(z_m) = u_m, \quad m = 1,\cdots,M, \tag{5.1}$$

for some index set Λ with $|\Lambda| = M = N$. In the context of the model problem (2.6), we have the $N\times N$ system with $\mathbf{A}$ as defined in previous sections:

$$\mathbf{Ac} = \mathbf{u}, \quad (A)_{n,\alpha} = \phi_\alpha(z_n). \tag{5.2}$$

Usually we want p to be as "simple" as possible, and this translates into the prescription of Λ, e.g., requiring that p have smallest total degree. One of the challenges for a naïve approach to interpolation is that one may not have existence and uniqueness: the classical Mairhuber-Curtis theorem [32, 56] ensures that no matter which Λ is chosen, there will exist a set of points on which interpolation is not unisolvent for $P(\Lambda)$.

The above is in contrast to the univariate interpolation problem where, with N nodes, $\deg p$ is known to be exactly $N-1$ regardless of (distinct) nodal choice, and the resulting polynomial is unique given data. In the multivariate case, the space from which the interpolant is drawn cannot be *a priori* specified if we allow the grid z_m to be freely defined; the approximation space must be identified once the nodes z_n and/or the data u_n is provided.

In Section 5.1 we present one way that uniquely chooses a smallest-degree interpolatory polynomial given data *locations* z_m, which depends on the gPC weight function ρ that determines the orthogonal basis ϕ. The construction is *independent* of the data $\mathbf{u}$. This polynomial is the Least Orthogonal Interpolant and allows us to make smallest-degree polynomial interpolation well-defined in multiple dimensions. This is followed in Section 5.2 by a discussion of Lebesgue constants for accuracy and stability. Finally, Sections 5.3 and 5.4 describes conditions under which an unstructured nodal array may have well-behaved Lebesgue constant.

5.1 Least orthogonal interpolation

Our task in this section is to define and present an operator that prescribes ρ-dependent, smallest-degree, unisolvent polynomial interpolation in the multivariate setting. As pre-

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

viously discussed, this is *not* the only choice of interpolation operator, but it has the advantages of being (1) ρ-dependent, (2) automatically constructible simply with evaluations of the basis ϕ, and (3) is computed via an extension of standard linear algebraic matrix factorizations.

The Least Orthogonal Interpolant (LOI) is a generalization of the 'least interpolation' construction [33–35]. The LOI uses information about the nodal array z_m along with the density ρ to construct this interpolant [65]. If ρ is the density function corresponding to a multivariate standard normal random variable, then the LOI coincides with the traditional least interpolant.

Given a multivariate nodal distribution z_m, define L^2_ρ expansions of Dirac distributions centered at the z_m:

$$\delta_{z_m}(\cdot) = \sum_{|\alpha|=0}^{\infty} \phi_\alpha(z_m)\phi_\alpha(\cdot). \tag{5.3}$$

Since δ_{z_m} is a representation of the reproducing kernel, one can directly verify that the formal sum on the right-hand side is the L^2_ρ-representor for point-evaluation: $u(z_m) = \langle u, \delta_{z_m}\rangle_\rho$ for any continuous u in L^2_ρ. For any polynomial p, one can define the degree-k L^2_ρ projection operators:

$$\mathcal{P}_k p = \sum_{|\alpha|\le k} \langle p, \phi_\alpha\rangle_\rho \phi_\alpha.$$

Note that these are effectively the action of a series truncation of δ_{z_m} on p. Finally, we introduce the 'least-ρ' operation:

$$p_{\downarrow,\rho} = P_{\widehat{k}} p, \quad \widehat{k} = \min\{k \in \mathbb{N} \,|\, P_k \neq 0\}.$$

So $p_{\downarrow,\rho}$ is effectively the first nonzero "Taylor series" contribution from an expansion of the form (5.3). Because the projection operators P_k depend on ρ, the polynomial $p_{\downarrow,\rho}$ likewise depends on ρ. Given a collection of M nodes $\mathcal{Z} = \{z_1, \cdots, z_M\}$, the polynomial images of δ_{z_m} under $(\cdot)_{\downarrow,\rho}$ form a polynomial space:

$$\Pi_{\mathcal{Z}} = \operatorname{span}\left\{ g_{\downarrow,\rho} \,|\, g \in \operatorname{span}\{\delta_{z_1}, \cdots, \delta_{z_M}\}\right\}.$$

The polynomial space $\Pi_{\mathcal{Z}}$ is the least orthogonal polynomial space for interpolation.

Theorem 5.1. [65] *The space $\Pi_{\mathcal{Z}}$ has dimension M, and for any continuous u there is a unique $p \in \Pi_{\mathcal{Z}}$ such that $u_m = p(z_m)$ for $m = 1, \cdots, M$. In particular, there exist Lagrange functions $\ell_m(z) \in \Pi_{\mathcal{Z}}$ such that the interpolant $p \in \Pi_{\mathcal{Z}}$ can expressed as*

$$p(z) = \sum_{m=1}^{M} u_m \ell_m(z), \quad \ell_m(z_n) = \delta_{m,n}, \tag{5.4}$$

with $\delta_{m,n}$ the Kronecker delta.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

The interpolant p is the Least Orthogonal interpolant. It depends on the weight function ρ. When ρ is a the Gaussian density function for a standard normal random variable, $\Pi_{\mathcal{Z}}$ coincides with the traditional 'least interpolant' polynomial space of [34]. While the definition of the space $\Pi_{\mathcal{Z}}$ is fairly abstract, the construction is accomplished in familiar ways: assuming we know $k = \deg\Pi_{\mathcal{Z}}$, let the index set $\Lambda = \mathcal{I}_{d,k}^T$. Form the rectangular design matrix $\mathbf{A}$ from (2.5). The space $\Pi_{\mathcal{Z}}$ can be computed with a combination of LU and QR operations on this matrix, and the operation count is asymptotically similar to standard interpolatory matrix factorization schemes. Using these operations one obtains the following factorization for the input rectangular matrix $\mathbf{A}$

$$\mathbf{PA} = \mathbf{LUH}. \tag{5.5}$$

For a size-M interpolation problem, the matrices $\mathbf{L}$ and $\mathbf{U}$ are standard $M \times M$ lower- and upper-triangular matrices corresponding to an LU factorization, and $\mathbf{P}$ is the associated $M \times M$ row permutation matrix. The rectangular matrix $\mathbf{H}$ contains entries that identify the space Π_Z.

Once data $\mathbf{u}$ is given, the coefficients $\mathbf{c}$ from (5.1) that define the interpolant are computed as

$$\mathbf{c} = \widetilde{\mathbf{H}}^T \mathbf{U}^{-1} \mathbf{L}^{-1} \mathbf{P} \mathbf{u}.$$

The new matrix $\widetilde{\mathbf{H}}$ is a rectangular matrix with orthonormal rows that can be obtained from $\mathbf{H}$ in a trivial manner. In order to compute $\mathbf{c}$, the operation count is asymptotically identical to any standard LU factorization for square matrix inversion. We refer to [65] for the details.

5.2 Accuracy and the Lebesgue constant

The existence of Lagrange interpolants from Theorem 5.1 allow us to extend univariate notions of stability and accuracy to multivariate settings. With a grid $\mathcal{Z}$ and the associated Lagrange polynomials ℓ_m defined in (5.4), we can define a $\mathcal{Z}$-dependent, ρ-weighted Lebesgue constant:

$$\Lambda_\rho(Z) = \max_{z\in\Omega} \rho(z) \sum_{m=1}^{M} \frac{|\ell_m(z)|}{\rho(z_m)}. \tag{5.6}$$

In the unweighted case $\rho \equiv 1$, this reverts to the more familiar formula for the Lebesgue constant. The introduction of the weight ρ in the above expression is the proper way to generalize the Lebesgue constant to the weighted interpolation problem [54]. Note that one need not use the LOI space $\Pi_{\mathcal{Z}}$ in the above definition of Λ_ρ.

On compact domains, the presence of a weight ρ bounded away from 0 and ∞ does not significantly affect the interpolation problem in the context of (5.6), as in such a case one can declare the weighted norm equivalent to the unweighted one with appropriate

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

bounding constants. However, on unbounded domains it becomes a necessary consideration. In particular we will consider exponential weights of the form $\rho(z)=\exp(-|z|^\nu)$ for $\nu\geq 1$ on $D=\mathbb{R}^d$.

The error in an interpolation procedure can be understood in terms of the Lebesgue constant, which is the operator norm for weighted interpolation. Let $\mathcal{I}_N$ be any interpolation operator; the LOI construction from the previous section is one such choice. We assume that $\mathcal{I}_N$ maps continuous functions to some polynomial space P_N, i.e., $\mathcal{I}_N: C_\rho(D)\to P_N$. The classical Lebesgue Lemma ties the operator norm Λ_ρ of $\mathcal{I}_N$ to the interpolant error:

$$\sup_{z\in D}\rho(z)|u(z)-\mathcal{I}_N u(z)|\leq\left(1+\Lambda_\rho\right)E_{\rho,N}(u). \tag{5.7}$$

The term $E_{\rho,N}(u)$ is the best (smallest) possible error in the ρ-weighted supremum norm when approximating u by any element in P_N. The Lebesgue Lemma is one way of "separating" the total interpolation error into two parts: first the inherent error $E_{\rho,N}$ depending on the data u and the choice of approximation space, and second the instability Λ_ρ due to the interpolation procedure. Thus, we are ultimately interested in constructing nodal arrays $\mathcal{Z}$ with small Lebesgue constant. In the multivariate setting, this is still an open problem.

Values for acceptable Lebesgue constants in one dimension are well-understood: on bounded intervals $D\subset\mathbb{R}$ it is well-known that if a nodal set $\mathcal{Z}$ is uniformly distributed on the interval, then Λ grows exponentially as the cardinality of $\mathcal{A}$ is increased (see [84] and references therein), and thus the interpolation problem is unstable. On the other hand, for certain special configurations of $\mathcal{A}$ that cluster nodes towards the boundary like the arcsine density $\rho_c=(1-z^2)^{-1/2}$, the Lebesgue constant exhibits logarithmic growth in the mesh cardinality, yielding a stable interpolation scheme [18, 45]. Results on the one-dimensional unbounded real line are likewise well-established [59, 60, 77].

In our multivariate setting, we will generally be concerned with whether or not an array of nodes exhibits exponentially-growing Lebesgue constant. In fact, since $E_{\rho,N}(u)$ usually decreases exponentially with N for u an analytic function, then *subexponential* growth of Λ_ρ can be used in Lebesgue's Lemma (5.7) to conclude convergence of the interpolation scheme.

While concrete results about Λ_ρ are available for the univariate case, these results are currently lacking for the multivariate case. What is known are necessary conditions that a grid should satisfy in order to have a "good" Lebesgue constant: The following sections introduce these conditions, which stem from pluripotential theory. All these results do not directly appeal to the geometry of the grid, and so they apply to any unstructured grids.

5.3 Unweighted polynomial approximation

There is a limited understanding of how one constructs robust interpolation grids in high dimensions. For canonical domains, many approaches yield effective sampling meshes

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

(e.g., the 'Padua points' in two dimensions [13,22]). Tensor-product domains yield good results when the dimension is small ($d \leq 3$). Mapping techniques can produce good nodes for curvilinear domains and blending techniques work well for triangular and tetrahedral domains [86]. Approaches that fit more within the unstructured framework are directly optimized Fekete nodes [4,82], approximate Fekete points [15,76], and discrete Leja sequences [14]. The interested reader may consult these references to find discussions of how to construct multivariate interpolation grids; here we will only discuss theoretical ways of characterizing a 'good' interpolation grid.

Consider the case where D is a compact domain in $\mathbb{R}^d$ and ρ is the uniform weight function. Thus, this reduces to unweighted interpolation where the weight appearing in (5.6) may be set to unity. We are interested in characterizing limiting behaviors of nodal sets $\mathcal{A}$ as $N = |\mathcal{Z}| \to \infty$. Let $\mathcal{Z}_N$ denote a nodal set of size N. We do not require monotonicity, i.e. we do not enforce $\mathcal{Z}_N \subset \mathcal{Z}_{N+1}$ for any N. We recall that t_k^d is the dimension of the total-degree polynomial space from (2.3). The following discussion requires that we restrict our attention to nodal sets with total-degree cardinality $N = t_k^d$ for some degree k. We are concerned with the asymptotic behavior of selecting a 'good' nodal set Z_N; i.e., we are concerned with the behavior of stability metrics of $\mathcal{Z}_N$ as $N \to \infty$.

The following notation will be needed: $\mathcal{Z}_N = \{z_{1,N}, \cdots, z_{N,N}\}$ is a set of N nodes in D. We define the following constant as the sum the polynomial degrees of any basis for T_k^d:

$$s_k^d \triangleq \sum_{j=1}^{k} j t_j^{d-1}, \quad t_j^0 \triangleq 1. \tag{5.8}$$

In this section, we will fix $d \geq 1$, and relate the cardinality N of each grid to the polynomial degree k:

$$N_k \triangleq t_k^d. \tag{5.9}$$

Given a degree k, a set of nodes $\mathcal{Z}_{N_k}^*$ is called a set of *Fekete* nodes for the space T_k^d if it maximizes the design matrix determinant over all possible configurations:

$$\mathcal{Z}_{N_k}^* \triangleq \underset{\mathcal{Z}=(y_1,\cdots,y_{N_k}) \subset D}{\operatorname{argmax}} |\det \mathbf{A}(\mathcal{Z})|.$$

Above, $\mathbf{A}$ is the (square) design matrix with the *monomial* basis $z^\alpha \in T_k^d$ on the nodes y_n. The behavior of the maximal determinant achieved by $\mathcal{Z}_{N_k}^*$ defines the transfinite diameter:

$$\delta(D) \triangleq \lim_{k \to \infty} \left|\det \mathbf{A}\left(\mathcal{Z}_{N_k}^*\right)\right|^{1/s_k^d}.$$

For compact $D \subset \mathbb{R}^d$, this limit exists. An array of nodes $\{\mathcal{Z}_N\}_N$ whose determinant (raised to the $1/s_k^d$) limits to δ is called *asymptotically* Fekete.

One final concept we will need is that of equilibrium measures. For a compact $D \subset \mathbb{R}^d$, we introduce the pluripotential equilibrium measure μ_D of the set D [51,75]. This

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

measure is the multidimensional analogue of the univariate extremal logarithmic energy measure. On a (tensor-product) interval, the measure μ_D is the (tensor-product) arcsine measure.

The Lebesgue constant, the concept of Fekte nodes, and the equilibrium measure are intimately connected. As described in [10], the following theorem considers three ways in which one can characterize the behavior of the array of nodes $\mathcal{Z}_N$ for $N = t_k^d$:

Theorem 5.2. [5, 10, 12] *Consider the three properties an array Z_{N_k} may satisfy:*

1. *Subexponential Lebesgue constant growth:* $\lim_{k\to\infty}(\Lambda(\mathcal{Z}_{N_k}))^{1/k} = 1$.
2. *Asymptotically Fekete:* $\lim_{k\to\infty}|\det\mathbf{A}(\mathcal{Z}_{N_k})|^{1/s_k^d} = \delta(D)$.
3. *Distribution according to the equilibrium measure:*

$$\lim_{k\to\infty}\frac{1}{N_k}\sum_{n=1}^{N_k}\delta_{z_{n,N_k}} = \mu_D.$$

Then $1 \Rightarrow 2$ *and* $2 \Rightarrow 3$*; and the reverse implications are false.*

The proof that $1 \Rightarrow 2$ is given in [12] for one dimension and extends to multiple dimensions as shown in [10]. That $2 \Rightarrow 3$ is proven for one dimension in [10] and in [5] for the multivariate case. The explanation in [5] is relatively abstract, and an accompanying informal discussion can be found in [52, 53].

These properties give insight into how one should sample an interpolation grid in multiple dimensions: In order to have a stable interpolation operator for interpolation in T_k^d as measured by the Lebesgue constant, one *must* asymptotically sample according to the measure μ_D. I.e., sampling according to this measure is a necessary (but not sufficient) condition to obtain a well-bounded Lebesgue constant.

In the absence of constructive ways to achieve provably well-behaved Lebesgue constants in high dimensions on arbitrary compact domains, one usually devises methods that produce asymptotically Fekete nodes, or nodes that distribute according to μ_D. While such methods are still under active development, promising preliminary results are given by constructing "approximate" Fekete points [76] or "discrete" Leja sequences [14]. A slightly different route may be taken by optimizing a so-called Fejer-constraint that is related to the cardinal Lagrange polynomials of the Least Orthogonal Interpolant [65]. We also note that the d-dimensional Weil points on a hypercube domain D distribute according to μ_D; this is statement of Theorem 3.4; whether or not these points are asymptotically Fekete is unknown, although use of the Weil points for interpolation is limited by the restriction on their cardinality.

5.4 Weighted polynomial approximation

In this section we consider *weighted* approximation on the unbounded domain $D = \mathbb{R}^d$ with exponential weights. We seek to present an extension of Theorem 5.2 to the weighted

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

case; in order to do so we will need *contraction* factors to account for the unbounded nature of D.

On $D=\mathbb{R}^d$, we consider weights of the form $\rho(z)=\exp\left(\|z\|^r\right)$ for $r\geq 1$, where $\|z\|$ is the Euclidean 2-norm of $z\in\mathbb{R}^d$. This family of weights includes the ubiquitous Gaussian density for $r=2$. It is convenient and common to consider the log-weight, $Q(z)\triangleq -\log\rho = \|z\|^r$. As before, we work on approximation for the space T_k^d, define s_k^d as in (5.8), and restrict attention to cardinalities N_k in (5.9). Given a set of nodes $\mathcal{Z}=\{z_1,\cdots z_N\}$, then for any $c>0$ we use the notation $c\mathcal{Z}=\{cz_1,\cdots,cz_N\}$.

The weighted Lebesgue constant Λ_ρ is as in (5.6). The notions of Fekete arrays and equilibrium measures carry over to the weighted case with some modifications: an array is a weighted Fekete array for T_k^d if it maximizes a weighted determinant:

$$\mathcal{Z}_{N_k,Q}^* \triangleq \underset{\mathcal{Z}=(y_1,\cdots,y_{N_k})\subset D}{\operatorname{argmax}} |\det\mathbf{A}(\mathcal{Z})| \prod_{m=1}^{N_k}\rho^k(y_j),$$

where again the design matrix $\mathbf{A}$ has monomial entries: $(A)_{n,\alpha}=z_n^\alpha$. As with the unweighted case, the limit of the maximal determinants exists [53];

$$\delta_\rho(D)\triangleq \lim_{k\to\infty}\left|\det\mathbf{A}\left(\mathcal{Z}_{N_k}^*\right)\prod_{m=1}^{N_k}\rho^k(y_j)\right|^{1/s_k^d},$$

and any array of points whose weighted determinant behaves like $\delta_\rho(D)$ is called *asymptotically* weighted Fekete. For our domain $D=\mathbb{R}^d$ we will make use of the *ρ-weighted* equilibrium measure $\mu_{D,Q}$ from weighted pluripotential theory [11,51,75]. We mention that the support $\operatorname{supp}\mu_{D,Q}$ of the equilibrium measure is a compact set (even though $D=\mathbb{R}^d$ is unbounded).

For a degree k, the weighted situation requires a contraction factor $k^{-1/r}$, where r is the exponent in the log-weight $Q=\|z\|^r$. Interpolation grids with subexponentially growing weighted Lebesgue constant are *not* asymptotically weighted Fekete arrays, but contracted versions of them are.

Theorem 5.3. [5,65] *Consider the four properties an array $\mathcal{Z}_{N_k}$ may satisfy:*

1. *Subexponential weighted Lebesgue constant growth:* $\lim_{k\to\infty}\left(\Lambda_\rho(\mathcal{Z}_{N_k})\right)^{1/k}=1$.
2. *Contracted asymptotically weighted Fekete:*

$$\lim_{k\to\infty}\left|\det\mathbf{A}\left(k^{-1/r}\mathcal{Z}_{N_k}\right)\prod_{n=1}^{N_k}\rho^k(z_n)\right|^{1/s_k^d}=\delta_\rho(D).$$

2a. *Uniform contracted boundedness: There is compact set $S\supset\operatorname{supp}\mu_{\Omega,Q}$ such that $k^{-1/r}\mathcal{Z}_{N_k}\subset S$ for all $k\in\mathbb{N}$.*

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

3. *Distribution according to the weighted equilibrium measure:*

$$\lim_{k\to\infty}\frac{1}{N_k}\sum_{n=1}^{N_k}\delta_{k^{-1/r}z_{n,N_k}}=\mu_{D,Q}.$$

Then $1\Rightarrow 2$ *and* $(2+2a)\Rightarrow 3$*; and the reverse implications are false*‡.

The weighted case on the whole space $\mathbb{R}^d$ here reveals an interesting twist: We do not directly want to sample with respect to $\mu_{D,Q}$; among the justifying reasons is that the measure has compact support, which will be of limited use in approximation with polynomials on an unbounded domain. However, if we sample points so that grids that are progressively contracted by $k^{-1/r}$ distribute according to $\mu_{D,Q}$, then this is a necessary condition for controlled Lebesgue constant growth.

To our knowledge it is unknown whether multidimensional *weighted approximate/discrete* Fekete/Leja arrays distribute according to the weighted pluripotential equilibrium measure. However, we anticipate that such a result is true, and if so would provide a powerful computational approach for generating optimal unstructured meshes. The one-dimensional affirmative answer to this question from [64] gives hope to this possibility.

6 Summary

The ability to construct polynomial approximations to functions of high-dimensional inputs is of great interest and importance in modern uncertainty quantification settings. We have explored recent advances in non-intrusive stochastic collocation methods for doing this using a selection of geometrically unstructured meshes. With respect to N degrees of freedom in the polynomial expansion, and M data points, we have presented results for least-squares regularization for overdetermined systems ($M>N$), interpolatory reconstruction for square systems ($M=N$), and compressive sampling/sparse reconstruction for underdetermined systems ($M<N$).

The results in this field as presented here are far from optimal and settled. We expect a great deal of upcoming and future work on the theory of approximation on unstructured meshes will make such an approach one of the preferred methods for collocative approaches in high-dimensional approximation.

Acknowledgments

T. Zhou is supported by the National Natural Science Foundation of China (Award Nos. 91130003 and 11201461).

‡The uniform contracted boundedness condition 2a is a technicality. Although we suspect that this condition is ultimately unnecessary, a direct proof that $2\Rightarrow 3$ seems to not yet be available.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

References

[1] I. Babuska, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Review*, 52(2):317–355, January 2010.

[2] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12(4):273–288, March 2000.

[3] R. E. Bellman. *Adaptive control processes: a guided tour*. Princeton University Press, 1961.

[4] E. Bendito, A. Carmona, A.M. Encinas, and J.M. Gesto. Estimation of Fekete points. *Journal of Computational Physics*, 225(2):2354–2376, August 2007.

[5] R. Berman, S. Boucksom, and D. Nyström. Fekete points and convergence towards equilibrium measures on complex manifolds. *Acta Mathematica*, 207(1):1–27, 2011.

[6] M. Bieri and C. Schwab. Sparse high order FEM for elliptic sPDEs. *Computer Methods in Applied Mechanics and Engineering*, 198(13-14):1149–1170, March 2009.

[7] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Universal algorithms for learning theory. ii. piecewise polynomial functions. *Constr. Approx.*, 2(26):127–152, 2007.

[8] P. Binev, A. Cohen, W. Dahmen, R. A. Devore, and V. N. Temlyakov. Universal algorithms for learning theory part I: Piecewise constant functions. *Journal of Machine Learning Research*, 6:1297–1321, 2005.

[9] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, March 2011.

[10] T. Bloom, L. Bos, C. Christensen, and N. Levenberg. Polynomial interpolation of holomorphic functions in $\mathbb{C}$ and $\mathbb{C}^n$. *Rocky Mountain Journal of Mathematics*, 22(2):441–470, June 1992.

[11] T. Bloom and N. Levenberg. Weighted pluripotential theory in c^n. *American Journal of Mathematics*, 125(1):57–103, February 2003.

[12] L. Bos. *Near optimal location of points for Lagrange interpolation in several variables.* PhD thesis, University of Toronto, 1981.

[13] L. Bos, M. Caliari, S. De Marchi, M. Vianello, and Y. Xu. Bivariate lagrange interpolation at the padua points: The generating curve approach. *Journal of Approximation Theory*, 143(1):15–25, November 2006.

[14] L. Bos, S. De Marchi, A. Sommariva, and M. Vianello. Computing multivariate fekete and leja points by numerical linear algebra. *SIAM Journal on Numerical Analysis*, 48(5):1984, 2010.

[15] L. P. Bos and N. Levenberg. On the calculation of approximate fekete points: the univariate case. *Electronic Transactions on Numerical Analysis*, 30:377–397, 2008.

[16] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Mathematical Journal*, 159(1):145–185, July 2011.

[17] J. Breidt, T. Butler, and D. Estep. A measure-theoretic computational method for inverse sensitivity problems i: Method and analysis. *SIAM Journal on Numerical Analysis*, 49(5):1836–1859, 2011.

[18] L. Brutman. On the lebesgue function for polynomial interpolation. *SIAM Journal on Numerical Analysis*, 15(4):694, 1978.

[19] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13(-1):147–269, 2004.

[20] J. Burkardt and M. Eldred. Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. In *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition*. American Institute of Aeronautics and Astronautics, 2009.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

[21] J. Burkardt, M. Gunzburger, and H.-C. Lee. POD and CVT-based reduced-order moeling of navier-stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 196(1-3):337–355, December 2006.
[22] M. Caliari, S. De Marchi, and M. Vianello. Bivariate polynomial interpolation on the square at new nodal sets. *Applied Mathematics and Computation*, 165(2):261–274, June 2005.
[23] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 8(59):1207–1223, 2006.
[24] E. Candès and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *IEEE Trans. Inform. Theory*, 12(51):4203–4215, 2005.
[25] E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, December 2005.
[26] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006.
[27] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone. Discrete least squares polynomial approximation with random evaluations-application to parametric and stochastic elliptic pdes. *EPFL, MATHICSE Technical Report*, 35/2013.
[28] 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 Journal*, 42(6):1191–1198, 2004.
[29] A. Cohen, W. Dahmen, and R. A. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.
[30] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, 13(5):819–834, 2013.
[31] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best n-term galerkin approximations for a class of elliptic spdes. *Found. Comp. Math.*, 6(10):615–646, 2010.
[32] P. Curtis. n-parameter families and best approximation. *Pacific Journal of Mathematics*, 9(4):1013–1027, December 1959.
[33] C. de Boor. Gauss elimination by segments and multivariate polynomial interpolation. In *Proceedings of the conference on Approximation and computation: a fetschrift in honor of Walter Gautschi: a fetschrift in honor of Walter Gautschi*, pages 11–22, West Lafayette, Indiana, United States, 1994. Birkhauser Boston Inc.
[34] C. de Boor and A. Ron. Computational aspects of polynomial interpolation in several variables. *Mathematics of Computation*, 58(198):705–727, April 1992.
[35] C. de Boor and A. Ron. The least solution for the polynomial interpolation problem. *Mathematische Zeitschrift*, 210(1):347–378, December 1992.
[36] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag. Low-dimensional mdoels for complex geometry flows: Application to grooved channels and circular cylinders. *Physics of Fluids A: Fluid Dynamics (1989-1993)*, 3(10):2337–2354, October 1991.
[37] B. J. Debusschere, H. N. Najm, A. Matta, O. M. Knio, R. G. Ghanem, and O. P. Le Maitre. Protein labeling reactions in electrochemical microchannel flow: Numerical simulation and uncertainty propagation. *Physics of Fluids*, 15:2238, 2003.
[38] R. A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23(4-6):918–925, August 2007.
[39] D.L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 4(52):1289–1306, 2006.
[40] A. Doostan and H. Owhadi. A non-adapted sparse approximation of pdes with stochastic inputs. *J. Comput. Phys.*, 8(230):3015–3034, 2011.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

[41] O.G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(02):317–339, 2012.
[42] Z. Gao and T. Zhou. Choice of nodal sets for least square polynomial chaos method with application to uncertainty quantification. *Communications in Computational Physics*, 16:365–381, 2014.
[43] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18(3):209–232, January 1998.
[44] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., 1991.
[45] R. GÃijnttner. Evaluation of lebesgue constants. *SIAM Journal on Numerical Analysis*, 17(4):512–520, August 1980.
[46] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics, Springer-Verlag, Berlin, 2002.
[47] S. Hosder, R. W. Walters, and M. Balch. Point-collocation nonintrusive polynomial chaos method for stochastic computational fluid dynamics. *AIAA Journal*, 48(12):2721–2730, December 2010.
[48] M. A. Iwen. Simple deterministically constructible rip matrices with sublinear fourier sampling requirements. *43rd Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD*, 2009.
[49] M. A. Iwen. Combinatorial sublinear-time fourier algorithms. *Foundations of Computational Mathematics*, 3(10):303–338, 2010.
[50] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, January 2001.
[51] M. Klimeck. *Pluripotential Theory*. Oxford University Press, Oxford, 1991.
[52] N. Levenberg. Weighted pluripotential theory results of berman-boucksom. *arXiv:1010.4035*, October 2010.
[53] N. Levenberg. Ten lectures on weighted pluripotential theory. *Dolomites Research Notes on Approximation*, 5:1–59, 2012.
[54] D. Lubinsky. A survey of weighted polynomial approximation with exponential weights. *Surveys in Approximation Theory*, 3:1–105, 2007.
[55] 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*, Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics, May 2009.
[56] J.C. Mairhuber. On haar's theorem concerning chebychev approximation problems having unique solutions. *Proceedings of the American Mathematical Society*, 7(4):609, August 1956.
[57] Y. Marzouk and D. Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 6(4):826–847, October 2009.
[58] L. Mathelin and K. A. Gallivan. A compressed sensing approach for partial differential equations with random input data. *J. Comput. Phys.*, 12(4):919–954, 2012.
[59] D. M. Matjila. Bounds for lebesgue functions for freud weights. *Journal of Approximation Theory*, 79(3):385–406, December 1994.
[60] D. M. Matjila. Bounds for the weighted lebesgue functions for freud weights on a larger interval. *Journal of Computational and Applied Mathematics*, 65(1-3):293–298, December 1995.
[61] G. Migliorati, F. Nobile, E. Schwerin, and R. Tempone. Analysis of the discrete l^2 projection

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

on polynomial spaces with random evaluations. *Foundations of Computational Mathematics, doi:10.1007/s10208-013-9186-4*, 2014.

[62] H. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual review of fluid mechanics*, 41(1):35–52, 2009.

[63] A. Narayan, J. Jackman, and T. Zhou. A Christoffel function wighted least squares algrithm for collocation approximations. *ArXiv: 1412.4305*, 2014.

[64] A. Narayan and J. Jakeman. Adaptive leja sparse grid constructions for stochastic collocation and high-dimensional approximation. *SIAM Journal on Scientific Computing*, 36(6):A2952–A2983, 2014.

[65] A. Narayan and D. Xiu. Stochastic collocation methods on unstructured grids in high dimensions via interpolation. *SIAM Journal on Scientific Computing*, 34(3):A1729–A1752, June 2012.

[66] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, January 2008.

[67] A. T. Patera and G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. MIT, version 1.0 edition, 2007.

[68] J. Penga, J. Hampton, and A. Doostan. A weighted ℓ^1-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267(0):92–111, 2014.

[69] G. PoÃńtte, B. DesprÃl's, and D. Lucor. Uncertainty quantification for systems of conservation laws. *Journal of Computational Physics*, 228(7):2443–2467, April 2009.

[70] C. Prudhomme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70, 2002.

[71] R. Pulch and D. Xiu. Generalised polynomial chaos for a class of linear conservation laws. *Journal of Scientific Computing*, 51(2):293–312, May 2012.

[72] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical Foundations and Numerical Methods for Sparse Recovery, Fornasier, M. (Ed.) Berlin,New York (DE GRUYTER)*, pages 1–92, 2010.

[73] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ^1-minimization. *Journal of Approximation Theory*, 164(5):517–533, May 2012.

[74] S. Hosder, R. Walters, and M. Balch. Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables. In *48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Structures, Structural Dynamics, and Materials and Co-located Conferences. American Institute of Aeronautics and Astronautics, April 2007.

[75] E. Saff and V. Totik. *Logarithmic Potentials with External Fields*. Springer, Berlin, 1997.

[76] A. Sommariva and M. Vianello. Computing approximate fekete points by QR factorizations of vandermonde matrices. *Computers & Mathematics with Applications*, 57(8):1324–1336, April 2009.

[77] J Szabados. Weighted lagrange and hermite-fejér interpolation on the real line. *Journal of Inequalities and Applications*, 1997(2):481267, 1997.

[78] G. Szegö. *Orthogonal Polynomials*. American Mathematical Society, Providence, RI, 1975.

[79] T. Tang and T. Zhou. On discrete least square projection in unbounded domain with random evaluations and its application to parametric uncertainty quantification. *SIAM Journal on Scientific Computing*, 36(5):A2272–A2295, 2014.

[80] T. Tang and T. Zhou. Recent developments in high order numerical methods for uncer-

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press

tainty quantification (in Chinese). *Sciences in China: Mathematics*, 45(6), 2015.
[81] D. M. Tartakovsky and D. Xiu. Stochastic analysis of transport in tubes with rough walls. *Journal of Computational Physics*, 217(1):248–259, September 2006.
[82] M. A. Taylor, B. A. Wingate, and R. E. Vincent. An algorithm for computing fekete points in the triangle. *SIAM Journal on Numerical Analysis*, 38(5):1707–1720, 2000.
[83] R.A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 2(27):232–261, 2007.
[84] L. N Trefethen and J. A. C Weideman. Two results on polynomial interpolation in equally spaced points. *Journal of Approximation Theory*, 65(3):247–260, June 1991.
[85] E. van den Berg and M. Friedlander. Spgl1: A solver for large-scale sparse reconstruction. *http://www.cs.ubc.ca/labs/scl/spgl1*, 2007.
[86] T. Warburton. An explicit construction of interpolation nodes on the simplex. *Journal of Engineering Mathematics*, 56(3):247–262, November 2006.
[87] A. Weil. On some exponential sums. *Proceedings of the National Academy of Sciences of the United States of America*, 34(5):204–207, May 1948. PMID: 16578290 PMCID: PMC1079093.
[88] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys*, pages 293–309, 2007.
[89] D. Xiu. Fast numerical methods for stochastic computations: A review. *Communications in Computational Physics*, 5(2-4):242–272, 2009.
[90] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, July 2010.
[91] D. Xiu and Jan S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, January 2005.
[92] 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.
[93] D. Xiu and G. E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137–167, May 2003.
[94] Z. Xu. Deterministic sampling of sparse trigonometric polynomials. *Journal of Complexity*, 27(2):133–140, April 2011.
[95] Z. Xu and T. Zhou. On sparse interpolation and the design of deterministic interpolation points. *SIAM Journal on Scientific Computing*, 36(4):A1752–A1769, 2014.
[96] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ^1-minimization. *International Journal for Uncertainty Quantification*, 2(3):279–293, 2012.
[97] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ^1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, January 2011.
[98] X. Yang and G. E. Karniadakis. Reweighted ℓ_1 minimization method for stochastic elliptic differential equations. *Journal of Computational Physics*, 248:87–108, September 2013.
[99] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ^1-minimization with applications to compressed sensing. *SIAM J. Imaging Sciences*, 1(1):143–168, 2008.
[100] T. Zhou, A. Narayan, and D. Xiu. Weighted discrete least-squares polynomial approximation using randomized quadratures. *submitted*, 2014.
[101] T. Zhou, A. Narayan, and Z. Xu. Multivariate discrete least-squares approximations with a new type of collocation grid. *SIAM Journal on Scientific Computing*, 36(5):A2401–A2422, January 2014.

https://doi.org/10.4208/cicp.020215.070515a Published online by Cambridge University Press