

STOCHASTIC COLLOCATION METHODS VIA L_1 MINIMIZATION USING RANDOMIZED QUADRATURES

LING GUO, AKIL NARAYAN, TAO ZHOU, AND YUHANG CHEN

ABSTRACT. In this work, we discuss the problem of approximating a multivariate function by polynomials via ℓ_1 minimization method, using a random chosen sub-grid of the corresponding tensor grid of Gaussian points. The independent variables of the function are assumed to be random variables, and thus, the framework provides a non-intrusive way to construct the generalized polynomial chaos expansions, stemming from the motivating application of uncertainty quantification. We provide theoretical analysis on the validity of the approach. The framework includes both the bounded measures such as the uniform and the Chebyshev measure, and the unbounded measures which include the Gaussian measure. Several numerical examples are given to confirm the theoretical results.

1. INTRODUCTION

Stochastic computation has received intensive attention in recent years, due to the pressing need to conduct uncertainty quantification (UQ) in practical computing. One of the most widely used techniques in UQ is generalized polynomial chaos (gPC), see e.g. [19, 45, 46]. In gPC, the stochastic function $f(x)$, with $x \in R^d$, is approximated via d - variate orthogonal polynomials, whose orthogonality is defined by the probability measure of the input variable \mathbf{x} . This becomes one of the most effective procedures for UQ, and many numerical techniques on how to construct gPC approximations have been developed. For practical computing, the gPC stochastic collocation algorithm is highly popular because it allows one to repetitively use existing deterministic simulation codes and to render the construction of gPC approximation a post-processing step. In the stochastic collocation framework, one seeks to construct a gPC type orthogonal polynomial approximation via point-evaluations of f . Popular methods for achieving this include sparse grids approximation [2, 1, 3, 14, 16, 17, 26, 33], pseudo orthogonal projection [37], interpolation [30], and least squares approach [40, 51, 9, 29] to name a few. For a review of the gPC methodology, see [31].

The challenge is in high-dimensional spaces, where the number of collocation nodes grows fast. Since each node represents a full-scale deterministic simulation, the total number of nodes one can afford is often limited, especially for large-scale problems. This represents a significant difficulty in constructing a gPC-type approximation using the existing approaches: it is often not possible to construct a good polynomial approximation using a very limited number of simulations in a large dimensional random space.

A more recent development in signal analysis is compressive sensing, also known as compressed sampling. Compressive sensing (CS) deals with the situation when there is insufficient information about the target function. This occurs when the number of samples is less than the cardinality of the

Ling Guo. Department of Mathematics, Shanghai Normal University, Shanghai, China. L. Guo was partially supported by NSFC (11101287), Shanghai Leading Academic Discipline Project (No. S30405) and ECInstitutes of Shanghai Municipal Education Commission (E03004).

Akil Narayan. Mathematics Department and Scientific Computing and Imaging Institute, University of Utah, University of Utah, Salt Lake City, UT 84112. A. Narayan was partially supported by AFOSR FA9550-15-1-0467 and DARPA N660011524053.

Tao Zhou. LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, Beijing, China. T. Zhou work was supported the National Natural Science Foundation of China (Award Nos. 91530118 and 11571351).

Yuhang Chen. Mathematics Department and Scientific Computing and Imaging Institute, University of Utah, University of Utah, Salt Lake City, UT 84112.

polynomial space for the approximation. CS then seeks to construct a polynomial approximation by minimizing the norm of the polynomial, typically its ℓ^1 norm or ℓ^0 norm (count the nonzero components). Following the seminal work of [7, 8, 6, 11], the theory of CS has generated an enormous amount of interest in many disciplines and resulted in many newer theoretical results and practical implementations. The success of the CS methods lies in the assumption that in practice many target functions (signals) are sparse, in the sense that what appear to be rough signals in the time/space domain may contain only a small number of notable. The works of [4, 42, 13, 10] show that under weak assumptions, solutions to the stochastic elliptic PDEs exhibit sparsity in the properly chosen basis.

In a recent work [13], the idea of CS has been extended to stochastic collocation and resulted in a highly flexible method. With CS, one can employ arbitrary nodal sets with an arbitrary number of nodes. This can be very helpful in practical computations. In [13], some key properties, such as the probability under which the sparse random response function can be recovered, are studied. Then Yan, Guo and Xiu extends this work in [48], which focus on the recoverability of stochastic solutions in high-dimensional random spaces with random sampling. This is relevant because in UQ simulations the dimensionality is often determined by the number of random parameters and can be very large. The work in [20] concerned with convergence analysis and sampling strategies to recover a sparse stochastic function in both Hermite and Legendre PC expansions from ℓ_1 -minimization problem. 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. In [47], the authors use Weil points to recover sparse Chebyshev polynomials. Tang and Zhou proposed a structured sampling method on quadrature points to recover sparse Legendre polynomials in [39].

In this work, we investigate the compressive sensing approach for stochastic collocation method, using a random chosen sub-grid from the corresponding tensor grid of Gaussian points. We will provide theoretical analysis on the convergence of such an approach. The main contribution of this work is to show that the method proposed in [39] is actually applicable for most random variables of interest, including measures on bounded domains (e.g., the uniform and Beta distributions) and unbounded domains (e.g., normal distributions). We also provide with several numerical examples to confirm the theoretical results.

Our main result is a sample count criterion for sparse recovery for general polynomial spaces with general measures. The precise statement is given by Theorem 4.1 and can be summarized as follows: suppose we wish to recover a multivariate polynomial expansion of $f(\mathbf{X})$ from a finite-dimensional subspace of dimension N whose maximum polynomial degree in any dimension is less than n . To do this, we use M collocation samples and perform an ℓ^1 minimization procedure. The M samples are chosen randomly (with the uniform probability law) from a tensor-product Gaussian quadrature grid with n points in each dimension. This method produces an approximation comparable to the best s -term expansion of f with high probability if

$$M \gtrsim L(n)s$$

where

- $L(n) \leq C^d$ if \mathbf{X} has independent components, with each component having a Beta distribution.
- $L(n) \leq (Cn)^{2d/3}$ if \mathbf{X} is normally-distributed.
- $L(n) \leq (Cn)^{2d/3}$ if \mathbf{X} has independent components, with each components having a one-sided exponential distribution.

In all cases the constant C is independent of the degree n and the dimension d , and depends only on the (one-dimensional) marginal distributions of \mathbf{X} . In the first case, the C^d dependence was established in [39] assuming the special case of \mathbf{X} a uniform random variable on a hypercube. In the latter two cases where \mathbf{X} has unbounded state space, the $n^{2d/3}$ dependence seems unpleasant, but this rate is essentially sharp if one uses our strategy for analysis and insists on sub-sampling from a tensor-product Gauss quadrature grid. Despite this dependence, we note that our analysis

Symbol(s)	
d, i	Dimension d , index $1 \leq i \leq d$.
\mathbf{n}, n_i	Multivariate tensor-product quadrature rule size \mathbf{n} , with n_i points in dimension i
\mathbf{X}, X^i	Random vector \mathbf{X} , components X^i
Γ, Γ^i	State space of \mathbf{X} and X^i , respectively
ρ, ρ^i	Joint density of \mathbf{X} , marginal density of X^i , respectively
$\phi_{\mathbf{k}}, \varphi_{k_i}^i$	Multivariate gPC basis element of degree $ \mathbf{k} $, dimension- i univariate gPC basis element of index k_i , respectively. Each element has polynomial degree $ \mathbf{k} $, and k_i , respectively.
$\Lambda, \Lambda_{\mathbf{n}}^P$	General multi-index set, multi-index set $\{\mathbf{k} \in \mathbb{N}_0^d \mid \mathbf{k} \leq \mathbf{n}\}$, respectively.

Table 1. Notation used throughout this article.

is quite general, extending to any random vector \mathbf{X} whose independent components have Beta or Exponential (one- or two-sided) distributions with essentially any shape parameters. Finally, we would like to mention the work by Li and Zhang [25], where a interpolation scheme (with $M = N$) on a sub-set of Gaussian quadrature points is proposed, and the sub-set therein is chosen according to the corresponding value of the joint probability density function of the random input. We expect that the idea in [25] may be useful also for the compressed sensing approach, and this will be investigated in our future studies.

The rest of the paper is organized as follows. In Section 2, we introduce the gPC approximation, set up the ℓ_1 -minimization problem and definitions and theorems used in the compressed sensing approach. Section 3 gives a short view of Gaussian quadrature, introduces the discrete transform as a discretization of orthogonal polynomial system, and gives the theorems for recovering polynomial chaos when sampling from the discretized system. Several numerical tests are provided in Section 4, and we finally give some conclusions in Section 5.

2. THE SETUP

Let $\mathbf{X} = (X^1, \dots, X^d)^\top$ be a random vector with d mutually independent components; each X^i takes values in $\Gamma^i \subset \mathbb{R}$. Since the variables $\{X^i\}_{i=1}^d$ are mutually independent, they have marginal probability density functions ρ^i associated with random variable X^i that completely characterize the distribution of \mathbf{X} . Define $\Gamma := \otimes_{i=1}^d \Gamma^i \subset \mathbb{R}^d$, and let $\rho(\mathbf{x}) = \prod_{i=1}^d \rho^i(x^i) : \Gamma \rightarrow \mathbb{R}^+$ denote the joint probability density function (PDF) of \mathbf{X} .

In a simple stochastic collocation setup, we consider a d -variate function $f : \Gamma \rightarrow \mathbb{R}$, and wish to recover information about this function from a finite (ideally small) set of function evaluations. Let $\theta_M = \{\mathbf{x}_1, \dots, \mathbf{x}_M\} \in \Gamma$ be a set of points at which the function values of f are available, and denote these values as $f_m = f(\mathbf{x}_m), m = 1, \dots, M$. We are concerned with the approximation of the function f based on $\{\mathbf{x}_m, f_m\}_{m=1}^M$. Let V be a linear space from which the approximation is sought and let $N = \dim V$. This paper is concerned with the special case of the above setup where V is a polynomial subspace, and the $N = \dim V$ coefficients defining the approximation are underdetermined from $M < N$ samples.

Much of the rest of this section is tasked with introducing our requisite notation. A summary of much of this notation is given in Table 1.

2.1. Generalized polynomial chaos. We are primarily concerned with the approximation f in a polynomial subspace V . In particular, we seek this approximation using the Generalized polynomial chaos (gPC) framework. The basic idea of gPC is to represent the function f as a polynomial of the random variables X^i ; the basis for approximation is typically taken as a set of orthogonal polynomials. For each marginal density ρ^i , we can define the univariate gPC basis elements φ_n^i , polynomials of degree n , via the orthogonality relation

$$(1) \quad \mathbb{E}[\varphi_n^i(X^i)\varphi_\ell^i(X^i)] = \int_{\Gamma^i} \varphi_n^i(s)\varphi_\ell^i(s)\rho^i(s)ds = \delta_{n,\ell}, \quad n, \ell \geq 0,$$

with $\delta_{n,\ell}$ the Kronecker delta function. Up to a multiplicative sign, this defines the polynomials φ_n^i uniquely; thus the probability measure ρ^i determines the type of orthogonal polynomial basis. For example, the Gaussian (normal) distribution yields the Hermite polynomials, the uniform distribution pairs with Legendre polynomials, etc. For a detailed account of the correspondence, see [46]. In this paper, each marginal density ρ^i may be associated with any of the classical orthogonal polynomial families: this includes Beta distributions (Legendre, Chebyshev, and Jacobi polynomials), normal distributions (Hermite polynomials), and exponential distributions (Laguerre polynomials).

For the multivariate ($d > 1$) case, we will use standard multi-index notation. For some $\mathbf{n} \in \mathbb{N}_0^d$, it has components denoted n_i , i.e., $\mathbf{n} = (n_1, \dots, n_d)$. Its modulus is its ℓ^1 norm, i.e., the sum of its components $|\mathbf{n}| = \sum_{i=1}^d n_i$. The factorial is the product of its componentwise factorials, $\mathbf{n}! = \prod_{i=1}^d n_i!$, where we take $0! = 1$. Given another d -dimensional index \mathbf{j} , we have

$$\mathbf{j} + \mathbf{n} = (j_1 + n_1, \dots, j_d + n_d), \quad \mathbf{n}^{\mathbf{j}} = \prod_{i=1}^d n_i^{j_i}.$$

Boldface explicit constants, e.g., $\mathbf{3}$ are multi-indices with the explicit value repeated:

$$\mathbf{3} = (3, \dots, 3) \in \mathbb{N}_0^d.$$

A partial ordering is defined on multi-indices:

$$\mathbf{j} < \mathbf{n} \iff j_i < n_i \quad \forall i = 1, \dots, d,$$

with a similar definition for \leq . We will occasionally use multi-index operations on d -dimensional vectors whose entries are non-integers; given the definitions above, the meaning of such notation should be clear.

One convenient representation for a multivariate gPC basis is as a product of the univariate gPC polynomials in each direction. For a multi-index $\mathbf{n} \in \mathbb{N}_0^d$, we have

$$(2) \quad \phi_{\mathbf{n}}(\mathbf{x}) = \prod_{i=1}^d \varphi_{n_i}^i(x^i).$$

The product functions $\phi_{\mathbf{n}}$ are L^2 orthogonal under the joint probability density function ρ for \mathbf{X} :

$$(3) \quad \mathbb{E}[\phi_{\mathbf{n}}(\mathbf{X})\phi_{\mathbf{j}}(\mathbf{X})] = \int_{\Gamma} \phi_{\mathbf{n}}(\mathbf{x})\phi_{\mathbf{j}}(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} = \delta_{\mathbf{n},\mathbf{j}},$$

where $\delta_{\mathbf{n},\mathbf{j}} = \prod_{i=1}^d \delta_{n_i,j_i}$.

A gPC expansion for f is given by

$$(4) \quad f(\mathbf{X}) = \sum_{\mathbf{j} \in \mathbb{N}_0^d} c_{\mathbf{j}}\phi_{\mathbf{j}}(\mathbf{X}),$$

where the coefficients $c_{\mathbf{j}}$ are the unknowns that must be computed from available knowledge of the function f . If $f(\mathbf{X})$ has finite variance, then this expansion is well-defined and convergent in the L^2 sense under fairly weak assumptions on the distribution of \mathbf{X} [15].

For computational purposes, the above gPC expansion must be truncated. One widely-used approach is to approximate f in a *total-degree* space $V = T_n$:

$$\Lambda_n^T = \left\{ \mathbf{k} \in \mathbb{N}_0^d \mid \sum_{i=1}^d k_i \leq n \right\}, \quad T_n = \text{span} \{ \phi_{\mathbf{k}} \mid \mathbf{k} \in \Lambda_n^T \}.$$

The dimension of T_n is

$$(5) \quad N = \#\Lambda_n^T \triangleq \dim T_n = \binom{d+n}{n}.$$

Another common index space is the *tensor-product* index space; for a maximum degree $n \in \mathbb{N}_0$,

$$\Lambda_n^P = \left\{ \mathbf{k} \in \mathbb{N}_0^d \mid \max_i k_i \leq n \right\}, \quad P_n = \text{span} \{ \phi_{\mathbf{k}} \mid \mathbf{k} \in \Lambda_n^P \}.$$

Note that the tensor-product space is larger than the total degree space, $\Lambda_n^T \subset \Lambda_n^P$, and that its cardinality is $\#\Lambda_n^P = (n+1)^d$. We can define anisotropic versions of the tensor-product space: Given a multi-index $\mathbf{n} \in \mathbb{N}_0^d$, we can allow polynomials up to degree n_i in dimension i ,

$$\Lambda_{\mathbf{n}}^P = \{\mathbf{k} \in \mathbb{N}_0^d \mid \mathbf{k} \leq \mathbf{n}\}, \quad P_{\mathbf{n}} = \text{span} \{\phi_{\mathbf{k}} \mid \mathbf{k} \in \Lambda_{\mathbf{n}}^P\}.$$

The total degree space T_n is a more practical space to use in computations compared to the tensor-product space P_n because the latter has extremely large dimension when d or n grows.

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

$$\{\phi_{\mathbf{k}}(\mathbf{x})\}_{\mathbf{k} \in \Lambda} \Leftrightarrow \{\phi_j(\mathbf{x})\}_{j=1}^N,$$

for some one-to-one correspondence between $j \in \{1, \dots, N\}$, with $N = \#\Lambda$. For any such Λ , V is the N -dimensional polynomial subspace formed from the span of $\phi_{\mathbf{k}}$ for $\mathbf{k} \in \Lambda$.

One of our main tasks is to estimate the projected function of f in the total degree space T_n , i.e.

$$(6) \quad f_n = \sum_{\mathbf{k} \in \Lambda_n^T} c_{\mathbf{k}} \phi_{\mathbf{k}} = \sum_{j=1}^N c_j \phi_j.$$

using interpolation conditions on some data $f(\mathbf{x}_m), m = 1, \dots, M$.

2.2. Compressive sensing approach. We now present the basic formulation for the stochastic collocation methods in the compressed sensing framework. In the framework of this paper, the compressive sensing approach can be described as follows. Given a set of M realizations $\{\mathbf{x}_i\}_{i=1}^M$, with corresponding outputs $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_M)]^\top$, we now seek a solution that satisfies

$$(7) \quad \Psi \mathbf{c} = \mathbf{f}$$

where $\mathbf{c} = (c_1, \dots, c_N)^\top$ is the coefficient vector of gPC (6), and

$$\Psi = (\Psi_{ij})_{1 \leq i \leq M, 1 \leq j \leq N} \in \mathbb{R}^{M \times N} \quad \Psi_{ij} = \phi_j(\mathbf{x}_i),$$

is the Vandermonde-like matrix, often referred to as the design matrix.

This problem is determined when $M = N$, overdetermined when $M > N$, and underdetermined when $M < N$. It is the underdetermined case that is considered here. This is often encountered in practice, especially in high dimensions with the total degree space T_K since $\dim T_K \sim K^d$ can be extremely large. In general when $d \gg 1$, the cardinality N of most standard polynomial spaces becomes extremely large, even when the order of the polynomials is moderate. On the other hand, in many practical applications the evaluation of the target function $f(\mathbf{x}_i)$ is expensive and one often has much smaller number of samples M than the number of gPC coefficients N , i.e. $M \ll N$. Thus problem (7) becomes ill-posed and we need some form of regularization to obtain a unique solution.

One efficient method uses an ℓ_1 -minimization algorithm which, under certain conditions, provides a means of identifying sparse coefficient vectors from a limited amount of data. A polynomial chaos expansion is defined as s -sparse when $\|\mathbf{c}\|_0 \leq s$, i.e the number of non-zero coefficients, does not exceed s . An ℓ_1 -minimization scheme attempts to find dominant gPC coefficients by solving the optimization problem

$$(8) \quad \text{argmin} \|\mathbf{c}\|_1 \quad \text{subject to} \quad \Psi \mathbf{c} = \mathbf{f},$$

where $\|\cdot\|_1$ is the ℓ_1 norm on vectors. The advantage of the above formulation (in contrast to optimizing over $\|\mathbf{c}\|_0$) is that it is a convex problem, and so computational solvers for convex problems may be leveraged (see [50, 49]). Under certain conditions, the solution to the ℓ_1 optimization problem also solves the ℓ_0 problem (see Theorem 2.1 below). This ℓ_1 minimization problem is often referred to as *Basis Pursuit*. Other types of minimization problems such as *Basis Pursuit Denoising* and *Least Absolute Shrinkage Operator* (LASSO) can be found in [12, 41] and references therein.

2.3. Recovery via ℓ_1 minimization. The ability of ℓ_1 -minimization method (8) to determine the dominant coefficients of the gPC expansion is determined by the properties of the measurement matrix Ψ and the sparsity of \mathbf{c} . We need some definitions to make this precise.

Definition 2.1. *The error of the best s -term approximation of a vector $\mathbf{c} \in R^N$ in the ℓ^p -norm is defined as*

$$(9) \quad \sigma_{s,p}(\mathbf{c}) = \inf_{\|\mathbf{y}\|_0 \leq s} \|\mathbf{y} - \mathbf{c}\|_p.$$

Clearly, $\sigma_{s,p}(\mathbf{c}) = 0$ if \mathbf{c} is s -sparse.

Definition 2.2 (restricted isometry constant [7, 8]). *Let \mathbf{D} 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*

$$(10) \quad (1 - \delta_s)\|\mathbf{c}\|_2^2 \leq \|\mathbf{D}\mathbf{c}\|_2^2 \leq (1 + \delta_s)\|\mathbf{c}\|_2^2$$

holds for all $\mathbf{c} \in R^N$ of sparsity at most s . Then, the matrix \mathbf{D} is said to satisfy the s -restricted isometry property (RIP) with restricted isometry constant δ_s .

Theorem 2.1 (Sparse recovery for RIP-matrices [8, 36, 5]). *Let $\mathbf{D} \in R^{M \times N}$ be a matrix with RIC satisfying $\delta_s < 0.307$. For any given $\tilde{\mathbf{c}} \in R^N$, let $\mathbf{c}^\#$ be the solution of the ℓ_1 -minimization*

$$(11) \quad \operatorname{argmin} \|\mathbf{c}\|_1 \quad \text{subject to} \quad \mathbf{D}\mathbf{c} = \mathbf{D}\tilde{\mathbf{c}}.$$

Then the reconstruction error satisfies

$$(12) \quad \|\mathbf{c}^\# - \tilde{\mathbf{c}}\|_2 \leq C \frac{\sigma_{s,1}(\tilde{\mathbf{c}})}{\sqrt{s}}$$

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., $\mathbf{c}^\# = \tilde{\mathbf{c}}$.

We consider a system $\{\psi_k(\mathbf{x})\}_k$ that is orthonormal with respect to a density $\nu(\mathbf{x})$ a *bounded orthonormal system* if it satisfies:

$$(13) \quad \max_{1 \leq k \leq N} \|\psi_k\|_\infty^2 = \max_{1 \leq k \leq N} \sup_{\mathbf{x} \in \operatorname{supp} \nu} |\psi_k(\mathbf{x})|^2 \leq L(N) < \infty.$$

If the bound L is independent of N , then we call it a *uniformly bounded orthonormal system*, satisfying

$$(14) \quad \sup_{1 \leq k} \|\psi_k\|_\infty^2 = \sup_{1 \leq k} \sup_{\mathbf{x} \in \operatorname{supp} \nu} |\psi_k(\mathbf{x})|^2 \leq L,$$

for some $L \geq 1$. For such systems, there is a precise undersampling rate for which the coefficient vector can be recovered with high probability when we solve (8).

Theorem 2.2 (RIP for bounded orthonormal systems [34, 36]). *Let $\mathbf{D} \in R^{M \times N}$ be the interpolation matrix with entries $\{d_{ij} = \psi_j(\mathbf{x}_i)\}$ from (8), where the points $\mathbf{x}_i, i = 1, \dots, M$, are i.i.d. random samples drawn from the orthogonalizing measure ν for the bounded orthonormal system $\{\psi_j\}$ satisfying (13). For some $\delta > 0$, assuming that*

$$(15) \quad M \geq C\delta^{-2}Ls \log^3(s) \log(N),$$

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

The Monte Carlo (MC) sampling method utilizing iid samples is very promising in the CS framework, and the result above is one of the foundational tools in this regard. The authors in [36] use the concept of bounded orthonormal systems to quantify recovery of sparse expansion coefficients in a univariate Legendre polynomial basis. They exploit the fact that weighting Legendre polynomials by the factor $(1-x^2)^{1/4}$ makes these polynomials a uniformly bounded system. The strategy is then to perform recovery with weighted/preconditioned Legendre polynomials, sampling from the appropriate biased measure that retains orthogonality; in this case this biased measure is the Chebyshev measure.

This idea of preconditioning systems is a theme in recent ℓ_1 recovery procedures. Let Ψ denote an $M \times N$ matrix containing evaluations of the N multivariate polynomials ϕ_n at collocation locations \mathbf{x}_m , with \mathbf{W} a diagonal $M \times M$ matrix:

$$(16a) \quad (\Psi)_{m,n} = \phi_n(\mathbf{x}_m), \quad (\mathbf{W})_{m,m} = w_m > 0,$$

where w_m are weights that will be specified. We can introduce an $M \times N$ weighted matrix \mathbf{D} that serves as the matrix used to recover sparse coefficients in the weighted setting, defined as

$$(16b) \quad \mathbf{D} = \sqrt{\mathbf{W}}\Psi.$$

For an unknown function $f(\mathbf{x})$, the $M \times 1$ vector \mathbf{f} with entries $(f)_m = f(\mathbf{x}_m)$ contains evaluations of f . The standard “unweighted” ℓ_1 optimization solves,

$$(17) \quad \operatorname{argmin} \|\mathbf{c}\|_1 \quad \text{subject to} \quad \Psi\mathbf{c} = \mathbf{f},$$

whereas a weighted version is given by

$$(18) \quad \operatorname{argmin} \|\mathbf{c}\|_1 \quad \text{subject to} \quad \mathbf{D}\mathbf{c} = \sqrt{\mathbf{W}}\mathbf{f}.$$

Note that we introduce \mathbf{W} as a matrix of positive weights but use only its square root in the formulations (16b) and (18). While cumbersome at present, this choice will be notationally convenient later.

We can now summarize some existing methods for sparse recovery of multivariate Legendre expansions. The main results for recovery in the formulations (17) and (18) attempt to show that the system matrices (Ψ and \mathbf{D} , respectively) satisfy the conditions of Theorem 2.2 and thus can invoke Theorem 2.1 to show convergence of the algorithm.

The authors in [48] extend the univariate results of [36] to high-dimensional problems, for both the original (i.e., unweighted) ℓ_1 -minimization and the preconditioned (i.e., weighted) ℓ_1 -minimization.

Theorem 2.3 (Recoverability with multivariate Legendre Polynomials [48]). *Let $\{\phi_j\}_{j=0}^{N-1}$ be the multivariate Legendre polynomial basis elements of the total degree space T_n^d , and let $f(\mathbf{x}) = \sum_{j=0}^{N-1} \tilde{c}_j \phi_j$ be an arbitrary polynomial with coefficient vector $\tilde{\mathbf{c}}$. For some nodal array $\{\mathbf{x}_i\}_{1 \leq i \leq M}$, let the weights w_m in (16) be defined by*

$$(19) \quad w_m = \left(\frac{2}{\pi}\right)^d \prod_{n=1}^d \left(1 - (\mathbf{x}_i^n)^2\right)^{1/2}.$$

1. Assume $d \geq n$. With $\{\mathbf{x}_i\}_{1 \leq i \leq M}$ i.i.d random samples drawn from the uniform measure on $[-1, 1]^d$, and if

$$M > 3^n s \log^3(s) \log(N),$$

then with high probability the solution \mathbf{c}^\sharp to the direct ℓ_1 minimization problem (17) is within a factor of the best s -term error:

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

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

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

Then with high probability, the solution \mathbf{c}^\sharp to the preconditioned/weighted ℓ_1 minimization problem (18) is within a factor of the best s -term error:

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

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

3. RANDOM SAMPLING OF GAUSSIAN QUADRATURE POINTS

We now present the method of random Gauss quadrature for sparse polynomial recovery via ℓ_1 minimization problem. The basic idea is to use the results from the previous section to conclude that subsampling a tensor-product Gaussian quadrature produces an accurate recovery procedure for CS. The two basic ingredients are (i) Tensor-product Gaussian quadrature can be used to define discrete measures under which polynomials are orthogonal, (ii) weighted polynomials under a Gaussian quadrature rule have quantifiable bounds.

3.1. Tensor grid of Gaussian points. Let $\varphi_n^i(x^i)$ be the degree- n orthonormal polynomial corresponding to the density ρ^i . It is well known that φ_n^i has n real and distinct zeros. I.e., there are n distinct nodes z_k^i , $k = 1, \dots, n$, such that

$$(20) \quad \varphi_n^i(z_k^i) = 0, \quad k = 1, \dots, n.$$

Furthermore, an interpolatory quadrature with weights w_k rule can be constructed on the zeros, satisfying

$$(21) \quad \sum_{k=1}^n w_k^i f(z_k^i) = \int_{X^i} f(x^i) \rho^i(x^i) dx^i,$$

for any polynomial f of degree $2n - 1$ or less; this is the Gaussian quadrature rule.* Here $w_k^i, k = 1, \dots, n$, are the univariate Gauss quadrature weights associated with dimension i . The n -point Gauss quadrature weights can be computed explicitly as

$$(22) \quad w_k^i = \lambda_n^i(z_k^i), \quad k = 1, \dots, n,$$

where λ_n^i is the 2-norm Christoffel function associated with dimension i :

$$(23) \quad \lambda_n^i(x^i) = \frac{1}{\sum_{k=0}^{n-1} (\varphi_k^i(x^i))^2}.$$

We define a discrete probability measure defined on the support of the n -point Gauss quadrature nodes. With δ_z the Dirac measure centered at z ,

$$(24) \quad \nu_n^i \triangleq \frac{1}{n} \sum_{k=1}^n \delta_{z_k^i}$$

The measure ν_n^i is the uniform empirical probability measure associated with the discrete set $\{z_1^i, \dots, z_n^i\}$.

We proceed to tensorize the univariate Gaussian quadrature rules. Let

$$(25) \quad \Theta_n^i = \{z_1^i, \dots, z_n^i\} \subset \Gamma^i, \quad i = 1, \dots, d$$

be the one dimensional n -point Gauss set associated to the i 'th dimension. We then take tensor products to construct a d -dimensional point set. Let n_i for $i = 1, \dots, d$ denote the Gauss quadrature rule size for dimension i . We collect these sizes into the multi-index $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{N}^d$. The tensor product set is then

$$(26) \quad \Theta_{\mathbf{n}} = \Theta_{n_1}^1 \otimes \dots \otimes \Theta_{n_d}^d$$

The cardinality of this set is $|\Theta_{\mathbf{n}}| = \prod_{i=1}^d n_i$. As before, an ordering scheme can be employed to order the points via a single index, i.e. for each $j = 1, \dots, |\Theta_{\mathbf{n}}|$,

$$z_j \leftrightarrow \mathbf{z}_{\mathbf{k}} = (z_{k_1}^1, \dots, z_{k_d}^d), \quad \mathbf{1} \leq \mathbf{k} \leq \mathbf{n}$$

*Note that z_k^i depends on the value of n , but we omit explicit notation indicating this dependence.

Each point has the scalar weight

$$(27) \quad w_{\mathbf{k}} = \prod_{i=1}^d w_{k_i}^i, \quad \mathbf{1} \leq \mathbf{k} \leq \mathbf{n}$$

This tensorized Gauss quadrature in the d -dimensional space Γ exactly integrates any polynomial in the tensor space $P_{2\mathbf{n}-1}$. We note that using (22) and (23), the d -dimensional Gauss quadrature weights for $\Theta_{\mathbf{n}}$ are given by

$$(28) \quad w_{\mathbf{k}} = \lambda_{\mathbf{n}}(z_{\mathbf{k}}) \triangleq \prod_{i=1}^d \lambda_{n_i}^i(z_{k_i}^i) = \prod_{i=1}^d \frac{1}{\sum_{k=0}^{n_i-1} [\varphi_k^i(z_{k_i}^i)]^2}$$

The uniform empirical probability measure on the set $\Theta_{\mathbf{n}}$ is given by

$$(29) \quad \nu_{\mathbf{n}} = \bigotimes_{i=1}^d \nu_{n_i}^i = \frac{1}{\prod_{i=1}^d n_i} \sum_{\mathbf{k} \leq \mathbf{n}} \delta_{z_{\mathbf{k}}} = \sum_{\mathbf{k} \leq \mathbf{n}} \prod_{i=1}^d \frac{1}{n_i} \delta_{z_{k_i}^i}$$

Note that iid sampling from $\nu_{\mathbf{n}}$ is equivalent to sampling uniformly from a tensor-product Gauss quadrature grid.

We will use the notation $\mathbb{E}_{\nu_{\mathbf{n}}}$ and $\mathbb{E}_{\nu_{n_i}^i}$ to denote expectations under the measures defined in (24) and (29), respectively. I.e.,

$$\mathbb{E}_{\nu_{n_i}^i} f(X^i) \triangleq \frac{1}{n_i} \sum_{k=1}^{n_i} f(z_k^i), \quad \mathbb{E}_{\nu_{\mathbf{n}}} f(\mathbf{X}) \triangleq \frac{1}{\prod_{i=1}^d n_i} \sum_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{n}} f(z_{\mathbf{k}})$$

3.2. Orthogonal matrices from Gaussian quadrature. The defining property of the univariate Gauss quadrature rule allows one to conclude that weighted Vandermonde-like matrices formed on the quadrature nodes are orthogonal matrices. Later, we will need the notion of an orthogonal rectangular matrix.

Definition 3.1. An $M \times N$ matrix \mathbf{D} with $M \geq N$ is orthogonal if $\mathbf{D}^T \mathbf{D} = \mathbf{I}_{N \times N}$.

With φ_k^i the degree- k polynomial from the univariate ρ^i -orthonormal family, consider the $n_i \times n_i$ Vandermonde-like matrix Ψ^i with entries

$$(30) \quad (\Psi^i)_{j,k} = \varphi_{k-1}^i(z_j^i), \quad j, k = 1, \dots, n_i$$

where z_j^i are the nodes of the n_i -point Gaussian quadrature rule introduced in (20). We also need a diagonal matrix Σ^i containing the quadrature weights:

$$(31) \quad (\Sigma^i)_{j,k} = w_j^i \delta_{j,k}, \quad j, k = 1, \dots, n_i,$$

with $\delta_{j,k}$ the Kronecker delta. The degree of exactness of each Gaussian quadrature rule implies

$$\delta_{j,k} = \int_{\Gamma^i} \varphi_{j-1}(s) \varphi_{k-1}(s) \rho^i(s) ds = \sum_{k=1}^{n_i} \varphi_{j-1}(z_k^i) \varphi_{k-1}(z_k^i) w_k^i = (\Psi^{iT} \Sigma^i \Psi^i)_{j,k}$$

for $j, k = 1, \dots, n_i$. The fact that the Kronecker delta is equal to the (j, k) element of the matrix $\Psi^{iT} \Sigma^i \Psi^i$ indicates that this matrix product is the identity matrix. So we have proven:

Lemma 3.1.A ([18]). The $n_i \times n_i$ matrix

$$(32) \quad \mathbf{D}^i = (\Sigma^i)^{1/2} \Psi^i,$$

defined by (30) and (31) is an orthogonal matrix.

A straightforward consequence of this is that the $(\prod_{i=1}^d n_i) \times (\prod_{i=1}^d n_i)$ matrix that is the tensor product of the d univariate matrices \mathbf{D}^i is also an orthogonal matrix:

Lemma 3.1.B ([39]). *The $(\prod_{i=1}^d n_i) \times (\prod_{i=1}^d n_i)$ matrix,*

$$(33) \quad \mathbf{D} = \bigotimes_{i=1}^d \mathbf{D}^i,$$

is an orthogonal matrix.

In [39], the matrix \mathbf{D} was called a “discrete orthogonal matrix” (DOM). Note that the same properties hold if \mathbf{D} has more rows than columns. (I.e., the marginal quadrature rule order dominates the marginal polynomial degree of the basis.)

Lemma 3.1.C. *Let $\mathbf{m}, \mathbf{n} \in \mathbb{N}^d$ satisfy $\mathbf{m} \geq \mathbf{n}$. For each i , define the $m_i \times n_i$ matrix Ψ^i with entries*

$$(\Psi^i)_{q,k} = \varphi_{k-1}(z_q^i), \quad k = 1, \dots, n_i, \quad q = 1, \dots, m_i.$$

With Σ^i the $m_i \times m_i$ diagonal weighting matrix given in (31), define the $m_i \times n_i$ matrix,

$$\mathbf{D}^i = (\Sigma^i)^{1/2} \Psi^i.$$

Then the $(\prod_{i=1}^d m_i) \times (\prod_{i=1}^d n_i)$ matrix

$$(34) \quad \mathbf{D} = \bigotimes_{i=1}^d \mathbf{D}^i$$

is an orthogonal matrix.

The most general version of this statement is that any matrix \mathbf{D} whose columns are a subset of the appropriate tensor-product space is orthogonal.

Lemma 3.1.D. *Let $\Lambda \in \mathbb{N}_0^d$ be a finite multi-index set, and let \mathbf{n} be any multi-index satisfying*

$$\Lambda \subseteq \Lambda_{\mathbf{n}-\mathbf{1}}^P.$$

With $N = |\Lambda|$, let $\mathbf{k}(1), \dots, \mathbf{k}(N)$ represent any enumeration of the elements of Λ . With $M = \prod_{i=1}^d n_i$, let $\mathbf{m}(1), \dots, \mathbf{m}(M)$ denote any enumeration of the elements in $\{\mathbf{m} \mid \mathbf{1} \leq \mathbf{m} \leq \mathbf{n} + \mathbf{1}\}$. Then the $M \times N$ matrix \mathbf{D} with entries

$$(D)_{j,\ell} = \sqrt{w_{\mathbf{m}(j)}} \phi_{\mathbf{k}(\ell)}(\mathbf{z}_{\mathbf{m}(j)}), \quad 1 \leq j \leq M, \quad 1 \leq \ell \leq N$$

is an orthogonal matrix.

That \mathbf{D} is an orthogonal matrix is interesting because it indicates that it is a well-conditioned matrix. If, in addition, the mass of the matrix is equidistributed across all its entries, then Theorem 2.2 implies that it has a small RIP constant. Thus, if equidistribution holds, then it might be possible to use its rows as a discrete candidate set to subsample for a compressive sampling strategy. This was explored in [39] when \mathbf{X} is a uniform random variable. Here we extend these results to more general cases. We delay the analysis and convergence results until Section 4.

3.3. Compressive sampling via Gaussian quadrature subsampling. A simple algorithm can now be presented for compressive sampling recovery of a random function $f(\mathbf{X})$: we can subsample M rows from the matrix \mathbf{D} in the previous section and use them to perform sparse recovery. A more detailed algorithm is as follows:

- (1) Given an index set Λ , find $\mathbf{n} \in \mathbb{N}^d$ such that $\Lambda \subseteq \Lambda_{\mathbf{n}-\mathbf{1}}^P$.
- (2) Generate the n_i -point Gauss quadrature rule $\{z_k^i, w_k^i\}_{k=1}^{n_i}$ for $i = 1, \dots, d$. The resulting full tensor-product rule is $\Theta_{\mathbf{n}} = \{(\mathbf{z}_{\mathbf{k}}, w_{\mathbf{k}})\}_{\mathbf{1} \leq \mathbf{k} \leq \mathbf{n}}$, but it need not be constructed explicitly.
- (3) According to the uniform probability law, randomly choose (and construct) M points $\{(\mathbf{y}_m, v_m)\}_{m=1}^M \subset \Theta_{\mathbf{n}}$.

(4) Generate an M -row matrix \mathbf{D} from these M points and index set Λ , having entries

$$(35) \quad (\mathbf{D})_{m,n} = \sqrt{v_m} \phi_{\mathbf{k}(n)}(\mathbf{y}_m),$$

where $\mathbf{k}(1), \dots, \mathbf{k}(N)$ is an enumeration of the elements in Λ . Also form the $M \times M$ diagonal matrix \mathbf{W} with entries $(\mathbf{W})_{m,m} = v_m$, and collect the M evaluations of the function $f(\mathbf{y}_m)$ into the vector \mathbf{f} .

(5) Solve (18) for the coefficients \mathbf{c} .

This procedure amounts to subsampling the product Gaussian quadrature rule in order to perform compressive sampling. One remaining question is how large M should be so that we can guarantee recovery. The authors in [39] show that if \mathbf{X} is a uniform random variable, then this strategy requires $M \gtrsim 3^d s$ samples to recover an approximately s -sparse vector. We provide a similar analysis for more general random variables by using analysis presented in the next section. Numerical results from the above algorithm are shown in Section 5 for various probability densities for \mathbf{X} .

4. ANALYSIS OF GAUSSIAN QUADRATURE SUBSAMPLING

Here we present analysis of the algorithm shown in Section 3.3. The essential question is how many samples M are required so that we can guarantee a faithful recovery of some function f . We proceed to show this by analyzing the entries of the weighted Vandermonde-like matrix \mathbf{D} . The entries of this matrix can be viewed as non-polynomial functions orthonormal under the discrete measure ν_n defined in (24). Using properties of orthogonal polynomials, we can determine the maximum magnitude of these functions, allowing us to use Theorem 2.2 to determine a sufficient number of samples M . Our main result is stated in Theorem 4.1.

4.1. Discrete orthonormal systems. The polynomials φ_k^i are orthonormal under the orthogonalizing measure ρ^i . In this section we show that specially weighted versions of φ_k^i are orthonormal under the discrete measures ν_n^i , defined in (24). We show similar statements for the multivariate polynomials $\phi_{\mathbf{k}}$ under the measure ν_n . These results stem from the accuracy of the Gauss quadrature rules introduced in the previous section: The four parts of Lemma 4.1 presented below are restatements of the four parts of Lemma 3.1. We show brief proofs for the first two parts.

Lemma 4.1.A. *Let $n \in \mathbb{N}$ be fixed. Then the n functions*

$$(36) \quad \{\psi_{k,n}^i(x^i)\}_{k=0}^{n-1} \triangleq \left\{ \sqrt{n\lambda_n^i(x^i)} \varphi_k^i(x^i) \right\}_{k=0}^{n-1}$$

are orthonormal under the probability measure ν_n^i defined in (24).

Proof. The result follows by direct calculation and use of the exactness of the Gauss quadrature rule, and is essentially equivalent to Lemma 3.1.A. For $j, k < n$ we have:

$$\begin{aligned} \mathbb{E}_{\nu_n^i} [\psi_{k,n}^i(X^i) \psi_{j,n}^i(X^i)] &\stackrel{(24)}{=} \frac{1}{n} \sum_{p=0}^{n-1} \psi_{j,n}^i(z_p^i) \psi_{k,n}^i(z_p^i) \\ &\stackrel{(36)}{=} \frac{1}{n} \sum_{p=0}^{n-1} n\lambda_n^i(z_p^i) \varphi_j^i(z_p^i) \varphi_k^i(z_p^i) \\ &\stackrel{(22)}{=} \sum_{p=0}^{n-1} w_p^i \varphi_j^i(z_p^i) \varphi_k^i(z_p^i) \\ &\stackrel{(21)}{=} \int_{\Gamma_i} \varphi_j^i(x^i) \varphi_k^i(x^i) \rho^i(x^i) dx^i \\ &\stackrel{(1)}{=} \delta_{j,k}. \end{aligned}$$

□

Note the factor of \sqrt{n} in (36), and the fact that $\psi_{k,n}^i$ depends on n . The result above generalizes in a straightforward manner to the multivariate case.

Lemma 4.1.B. *Let $\mathbf{n} \in \mathbb{N}^d$ be fixed. Then the $\prod_{i=1}^d n_i$ functions*

$$(37) \quad \{\psi_{\mathbf{k},\mathbf{n}}(z)\}_{\mathbf{k} \in \Lambda_{\mathbf{n}-1}^P} \triangleq \left\{ \prod_{i=1}^d \psi_{k_i, n_i}^i(z^i) \right\}_{\mathbf{k} \in \Lambda_{\mathbf{n}-1}^P}$$

are orthonormal under the probability measure $\nu_{\mathbf{n}}$ defined in (29).

Proof. We again use a direct calculation. For $\mathbf{0} \leq \mathbf{j}, \mathbf{k} < \mathbf{n}$ we have:

$$\begin{aligned} \mathbb{E}_{\nu_{\mathbf{n}}} [\psi_{\mathbf{j}}(\mathbf{X}) \psi_{\mathbf{k}}(\mathbf{X})] &= \mathbb{E}_{\nu_{n_1}^1 \otimes \dots \otimes \nu_{n_d}^d} \left[\prod_{i=1}^d \psi_{j_i}^i(Z^i) \psi_{k_i}^i(Z^i) \right] \\ &= \prod_{i=1}^d \mathbb{E}_{\nu_{n_i}^i} \left[\prod_{i=1}^d \psi_{j_i}^i(Z^i) \psi_{k_i}^i(Z^i) \right] \\ &\stackrel{\text{Lemma 4.1.A}}{=} \prod_{i=1}^d \delta_{j_i, k_i} = \delta_{\mathbf{j}, \mathbf{k}}, \end{aligned}$$

In the second equality above we have used the fact that a random variable associated to the measure $\nu_{\mathbf{n}}$ has independent components Z^i with marginal distributions given by $\nu_{n_i}^i$. \square

Of course, one can use a quadrature rule of higher-degree accuracy than required and still retain orthogonality of the resulting matrix.

Lemma 4.1.C. *Let $\mathbf{m}, \mathbf{n} \in \mathbb{N}^d$ satisfy $\mathbf{m} \geq \mathbf{n}$. The $\prod_{i=1}^d n_i$ functions*

$$\{\psi_{\mathbf{k},\mathbf{m}}(z)\}_{\mathbf{k} \in \Lambda_{\mathbf{n}-1}^P},$$

are orthonormal under the probability measure $\nu_{\mathbf{m}}$.

Finally, any non-tensor-product polynomial space of finite dimension can be encapsulated in a tensor-product space, and the quadrature rule associated to the tensor-product space exactly integrates elements from the original space.

Lemma 4.1.D. *Given a finite index set Λ , let $\mathbf{n} \in \mathbb{N}^d$ be such that $\Lambda \subseteq \Lambda_{\mathbf{n}-1}^P$. With $N = |\Lambda|$, then the N functions*

$$\{\psi_{\mathbf{k},\mathbf{n}}(z)\}_{\mathbf{k} \in \Lambda},$$

are orthonormal under the probability measure $\nu_{\mathbf{n}}$.

4.2. Bounded discrete orthonormal systems. We have established in the previous section that the collection of functions $\{\psi_{\mathbf{k},\mathbf{n}}\}_{\mathbf{k} \in \Lambda}$ are $\nu_{\mathbf{n}}$ -orthonormal for any $\Lambda \subseteq \Lambda_{\mathbf{n}-1}^P$. The recovery result in Theorem 2.2 then guarantees RIP properties associated to iid sampling strategies based on sup-norm bounds of these functions, to which we now turn.

Consider the scalar random variable X^i with $n \in \mathbb{N}$ fixed. The $\psi_{k,n}^i$ are ν_n^i -orthonormal, and so we must determine the bound

$$L_i(n) \triangleq \max_{0 \leq k \leq n-1} \sup_{x^i \in \text{supp } \nu_n^i} |\psi_{k,n}^i(x^i)|^2 = \max_{0 \leq k \leq n-1} \max_{1 \leq j \leq n} |\psi_{k,n}^i(z_j^i)|^2,$$

where z_j^i are the n -point Gaussian quadrature nodes. The $\psi_{k,n}$ functions are polynomials weighted by a Christoffel function. Much is known about the behavior of these functions, and can be used to estimate the bound L . The following lemmas establish the behavior of $L_i(n)$ for general distributions of X^i . The proofs of all the below Lemmas are in the Appendix.

When the polynomial family φ_k that defines the weighted functions $\psi_{k,n}$ corresponds to almost any Jacobi polynomial family, then the bounding constant $L_i(n)$ is *independent* of n , and in this case the $\psi_{k,n}$ are a uniformly bounded system for $k \leq n$.

Lemma 4.2.A. Let $X^i \sim B(\gamma + 1, \delta + 1)$ be a univariate Beta-distributed random variable with shape parameters $\gamma, \delta \geq -\frac{1}{2}$ on the domain $X^i = [-1, 1]$. Thus, the polynomials φ_k^i are Jacobi polynomials with parameters δ, γ . Then

$$L_i(n) \leq C$$

The constant C is uniform in n , $C = C(\gamma, \delta)$.

We show in the left-hand plot of Figure 1 the evolution of $L_i(n)$ as a function of the quadrature node size n , along with its dependence on symmetric parameters $\gamma = \delta$.

The strict n -independent bound obtained above for Jacobi polynomial families is, unfortunately, not true for polynomials orthogonal with respect to exponential weights on unbounded domains. Instead, the bound depends on the degree n . However, the dependence is relatively mild.

Lemma 4.2.B. Let X^i be a random variable whose density ρ^i is exponential on the real line:

$$\rho^i(x) \propto \exp(-|x|^\alpha), \quad x \in \mathbb{R}$$

for some $\alpha > \frac{3}{2}$. Then the weighted polynomials $\psi_{k,n}^i$ satisfy

$$L_i(n) \leq Cn^{2/3},$$

where the constant C is uniform in n , i.e., $C = C(\alpha)$.

Note that the above case covers $\alpha = 2$, corresponding to a normally-distributed X^i and a gPC basis of Hermite polynomials. The one-sided exponential bound is similar; we state it separately because its proof in the Appendix requires a different set of results.

Lemma 4.2.C. Let X^i be a random variable whose density ρ^i is exponential on the half real line:

$$\rho^i(x) \propto \exp(-|x|^\alpha), \quad x \geq 0$$

for some $\alpha > \frac{3}{4}$. Then the weighted polynomials $\psi_{k,n}^i$ satisfy

$$L_i(n) \leq Cn^{2/3},$$

where the constant C is uniform in n , i.e., $C = C(\alpha)$.

Remark 4.1. We expect the conclusion of Lemma 4.2.C to be valid for the more general weight $\rho^i \propto x^\mu \exp(-|x|^\alpha)$ for some $\mu \geq -\frac{1}{2}$. This would require some estimates on zeros of the associated orthogonal polynomials along with behavior of the associated Christoffel function. These estimates are essentially present in [23, 24]. Since this would necessitate a more technical analysis with dubious payoff, we do not pursue this here.

Of special note is the result of Lemma 4.2.C with $\alpha = 1$, corresponding to an exponential random variable X^i and a gPC basis of Laguerre polynomials.

The behavior of the bounds established above are shown in the right-hand plot of Figure 1 for Hermite polynomials (Lemma 4.2.B with $\alpha = 2$) and Laguerre polynomials (Lemma 4.2.C with $\alpha = 1$).

4.3. Multivariate expansions. This section contains our main results that utilize the univariate supremum bounds established in the previous section. Given $\mathbf{n} \in \mathbb{N}^d$, we define the product bound $L(\mathbf{n})$:

$$(38) \quad L(\mathbf{n}) = \prod_{i=1}^d L_i(n_i)$$

Our procedure runs into the familiar curse of dimensionality by requiring the sample count to dominate s times the product of all d of the L_i factors.

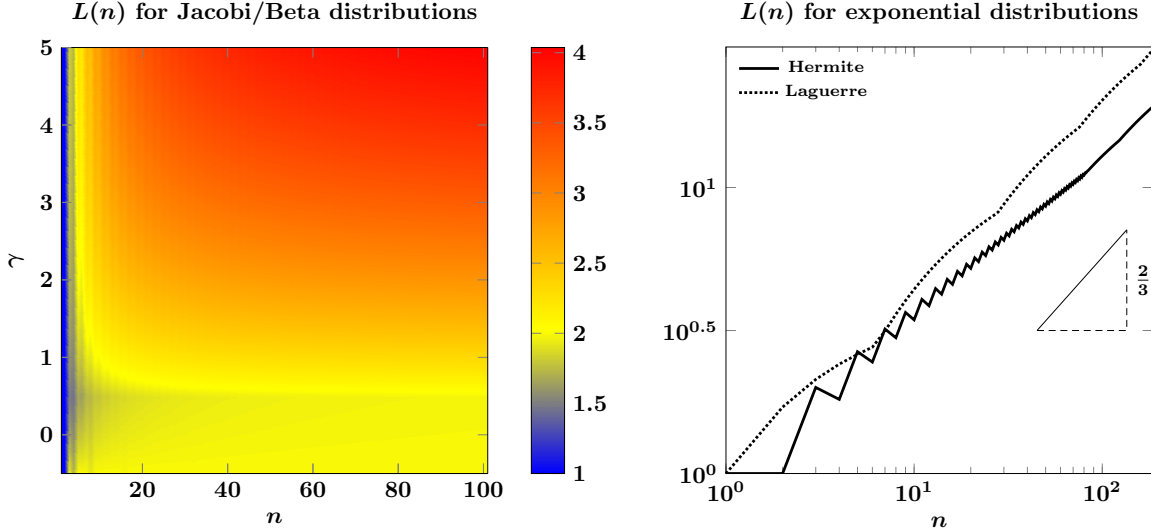


Figure 1. Left: Bound L for Jacobi polynomials with symmetric parameters $\gamma = \delta$. Right: Bound for two exponential-type densities: Hermite polynomials with $\rho = \exp(-x^2)$ on \mathbb{R} , and Laguerre polynomials with $\rho = \exp(-x)$ on $[0, \infty)$.

Theorem 4.1. *Let Λ be a finite index set with size N . Let \mathbf{n} be the smallest multi-index such that $\Lambda \subseteq \Lambda_{\mathbf{n}-1}^P$. (This defines \mathbf{n} uniquely.) Choose M samples randomly without replacement from the measure $\nu_{\mathbf{n}}$. (This is the algorithm presented in Section 3.3.) Assume M satisfies*

$$(39) \quad M \geq L(\mathbf{n}) C_1 s \log^3(s) \log(N),$$

where C_1 is a universal constant. Under these conditions, then for any $\mathbf{c} \in \mathbb{R}^N$, let \mathbf{c}^\sharp be the solution obtained by the weighted ℓ^1 optimization problem defined by (18). Then,

$$\Pr \left[\|\mathbf{c} - \mathbf{c}^\sharp\|_2 \leq \frac{C_2 \sigma_{s,1}(\mathbf{c})}{\sqrt{s}} \right] \leq 1 - N^{-\gamma \log^3(s)},$$

where C_2 and γ are universal constants. The individual factors $L_i(n_i)$ in (38) that comprise $L(\mathbf{n})$ in (39) are bounded by the three parts of Lemma 4.2 when the components of \mathbf{X} have the appropriate distributions.

The proof of the above Theorem, utilizing Lemmas 4.2, is given in Appendix A.4. The above theorem quantifies the size of M so that the subsampling algorithm in Section 3.3 converges with high probability. In particular, we frame this result slightly differently compared to the conclusion of Theorem 2.2: we make an explicit choice for the RIP constant δ so that recoverability is guaranteed (cf. Theorem 2.1).

Lemmas 4.2 cover many of the standard univariate distributions for X^i , in both the bounded and unbounded cases. Thus, the theorem above applies to very general cases of a random variable \mathbf{X} with independent components. We point out some special cases of our result:

- If ρ is uniform over $\Gamma = [-1, 1]^d$, so that the $\varphi_{\mathbf{k}}$ are tensor-product Legendre polynomials, then $L(\mathbf{n})$ satisfies

$$L(\mathbf{n}) \leq C^d,$$

with C the univariate bound in Lemma 4.2.A with $\gamma = \delta = 1$. We notice that the above constant depends exponentially on the dimension d . The authors in [39] show that this constant is essentially $C = 3$. Results from Figure 1 suggest that a sharper result would be $C = 2$, although this is not proven.

- If ρ corresponds to a standard Gaussian density function over \mathbb{R}^d , so that the $\varphi_{\mathbf{k}}$ are tensor-product Hermite polynomials, then each univariate $L_i(n_i)$ satisfies an $n_i^{2/3}$ bound, so that

$$(40a) \quad L(\mathbf{n}) \leq C^{2d/3} \mathbf{n}^{2/3}, \quad \mathbf{2/3} = (2/3, 2/3, \dots, 2/3) \in \mathbb{R}^d,$$

with C the constant given in Lemma 4.2.B. If the maximum polynomial degree $n_i - 1 \triangleq n - 1$ is uniform for all dimensions, then we have

$$(40b) \quad L(\mathbf{n}) \leq (Cn)^{2d/3},$$

This result unfortunately exhibits not only exponential dependence on the parametric d , but also algebraic dependence on the maximum polynomial degree n . Nevertheless, this bound on the supremum of the Gauss quadrature-weighted polynomials is sharp. However, the results in Figure 1 show that for polynomial degree $n - 1 = 9$, we empirically observe $Cn^{2/3} \lesssim 4$, so that for a degree-9 polynomial approximation, the requirement (39) states

$$(40c) \quad M \geq 4^d s \log^3(s) \log(N), \quad (n \leq 10)$$

We note that in high dimensions one is more likely to use low-degree approximations (small n) so that in high dimensions this requirement is comparable to the bounded-case sample count criterion.

- If ρ is given by $\rho(\mathbf{x}) \propto \exp(-\|\mathbf{z}\|_1)$ for $\mathbf{z} \in [0, \infty)^d$, then we also obtain the set of bounds (40) under the same conditions.

Finally, we remark that, once \mathbf{n} is identified as the multi-index identifying maximum the polynomial degree in each dimension, one may choose to subsample from the measure $\nu_{\mathbf{m}}$, where $\mathbf{m} \geq \mathbf{n}$. When \mathbf{X} is a bounded random variable, this would produce the same bound (39) with \mathbf{n} , even though one subsamples from $\nu_{\mathbf{m}}$. However, if \mathbf{X} has an exponential density (either one-sided or two-sided exponential), our analysis suggests that such a strategy would have a bound L that behaves like $L \sim \mathbf{m}^{2/3} > \mathbf{n}^{2/3}$. The penalty for this unbounded case makes sense: using Gauss quadrature rules with $\mathbf{m} > \mathbf{n}$ results in possible sampling of points that lie in regions where degree- \mathbf{n} polynomials weighted by $\lambda_{\mathbf{m}}$ are decaying quickly to 0, making recovery by sampling in these regions difficult. It is likely that one can improve the estimates in [21] to result in a tighter bound, but we still expect this bound to be greater than $\mathbf{n}^{2/3}$.

5. NUMERICAL EXAMPLES

We now provide with some numerical examples to test the theoretical findings and the convergence properties of the sparse recovery approach, with randomly chosen Gaussian grid. For the implementation of the ℓ_1 -minimization, we employ available tools such as Spectral Projected Gradient algorithm (SPGL1) from [44] that was implemented in the MATLAB package SPGL1 [43].

5.1. Polynomial function recovery. We first assume the target function has a polynomial form and choose a sparsity level s and then fix s coefficients of the polynomial while keeping the rest of the coefficients zero. The values of the s non-zero coefficients are drawn iid from a standard normal distribution. This procedure produces target coefficients that we seek to recover using the ℓ_1 -minimization algorithms.

In what follows, we use the following terms in figures to describe our recovery procedures:

- *Random* – sampling iid from the orthogonality measure ρ and solving the unweighted ℓ_1 optimization problem (17).
- *PreChebyshev* – sampling iid from the Chebyshev measure, i.e., the measure with probability density

$$v(\mathbf{x}) = \frac{1}{\pi^d \prod_{i=1}^d \sqrt{1 - x_i^2}},$$

and solving the preconditioned optimization problem (18) using the weights (19).

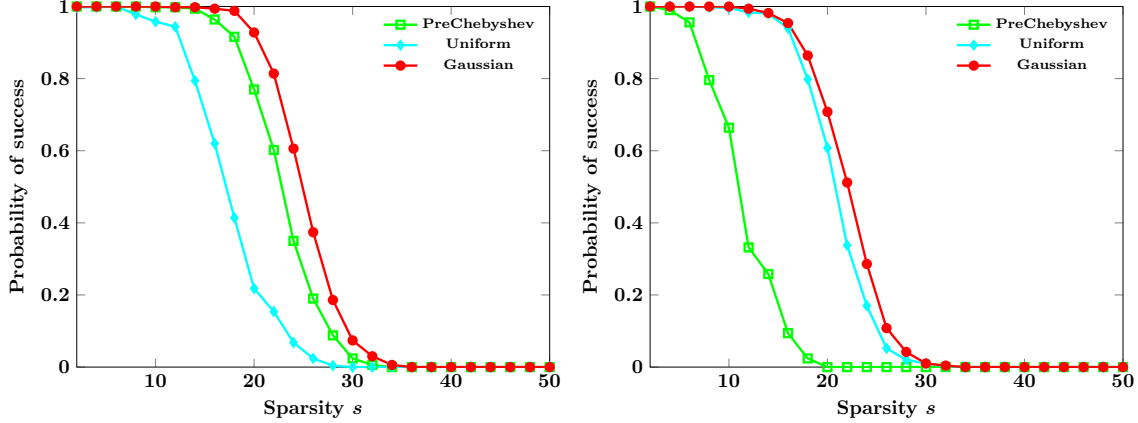


Figure 2. Recovery probability with respect to sparsity s of Legendre polynomials with fixed sample number $M = 85$. Left: $d = 2$, $\mathbf{n} = \mathbf{21}$ ($N = 231$). Right: $d = 10$, $\mathbf{n} = \mathbf{4}$ ($N = 286$).

- *Gaussian* – subsampling from a Gaussian quadrature grid using the Gaussian quadrature weights to solve the preconditioned problem (18). (This is the method proposed in this paper.)
- *Chebyshev* (resp. *Uniform*) – sampling iid from the Chebyshev (resp. uniform) measure and solving the unweighted optimization problem (17).

We recall our notation: n_i is the number of Gaussian quadrature points in dimension i , with $n_i - 1$ being the maximum polynomial degree in dimension i from the index set Λ . Thus, choosing \mathbf{n} and d defines the total degree space Λ . We define the size of Λ to be N , coinciding with the number of columns in the matrix \mathbf{D} that is input to the ℓ_1 optimization problem (18). The number of samples we use is M , and is the number of points subsampled from the tensor-product Gaussian quadrature grid, coinciding with the number of rows of the matrix \mathbf{D} .

5.1.1. *Uniform measure and Legendre polynomials.* The first test is the recovery of sparse Legendre polynomials, with the index set Λ corresponding to the two dimensional total degree space \mathbb{T}_{n-1}^2 and \mathbb{T}_{n-1}^{10} , respectively. We note that in this case our method coincides with the method in [39]. We examine the frequency of successful recoveries when the number of sample points is fixed at $M = 85$. This is accomplished by conducting 500 trials of the algorithms and counting the successful ones. A recovery is considered successful when the resulting coefficient vector \mathbf{c} satisfies $\|\mathbf{c} - \hat{\mathbf{c}}\| \leq 10^{-3}$. In the left-hand plot of Figure 2, we show the recovery rate for sparse Legendre polynomials (with $d = 2$, $n_1 = n_2 = 11$, and thus $N = 66$) as a function of sparsity level s . In the right-hand plot of Figure 2, we show the recovery rate for sparse Legendre polynomials (with $d = 10$, $\mathbf{n} = \mathbf{4}$, and $N = 286$) as a function of sparsity level s . We have also tested preconditioned recovery results with MC Chebyshev samples. In the cases we have tested, the Gaussian subsampling method works as well or better than the other methods, both for low and high dimensions.

5.1.2. *Chebyshev measure and Chebyshev polynomials.* The second test is the recovery of sparse Chebyshev polynomials, with the index set Λ corresponding to the two dimensional total degree space \mathbb{T}_{n-1}^2 and \mathbb{T}_{n-1}^{10} , respectively. We examine the probability of successful recoveries when the number of sample points is fixed at $M = 85$. In the left-hand plot of Figure 3, we show the recovery rate for sparse Chebyshev polynomials (with $d = 2$, $\mathbf{n} = \mathbf{11}$, and $N = 66$) as a function of sparsity level s . In this low-dimensional case, the results are similar to recovery of Legendre polynomials in Figure 2. In the right-hand plot of Figure 3, we show the recovery rate for sparse Chebyshev polynomials (with $d = 10$, $\mathbf{n} = \mathbf{4}$, and $N = 286$) as a function of sparsity level s . Again we see that the subsampling Gaussian quadrature case performs better than other methods, although the improvement in the high-dimensional cases is more minor.

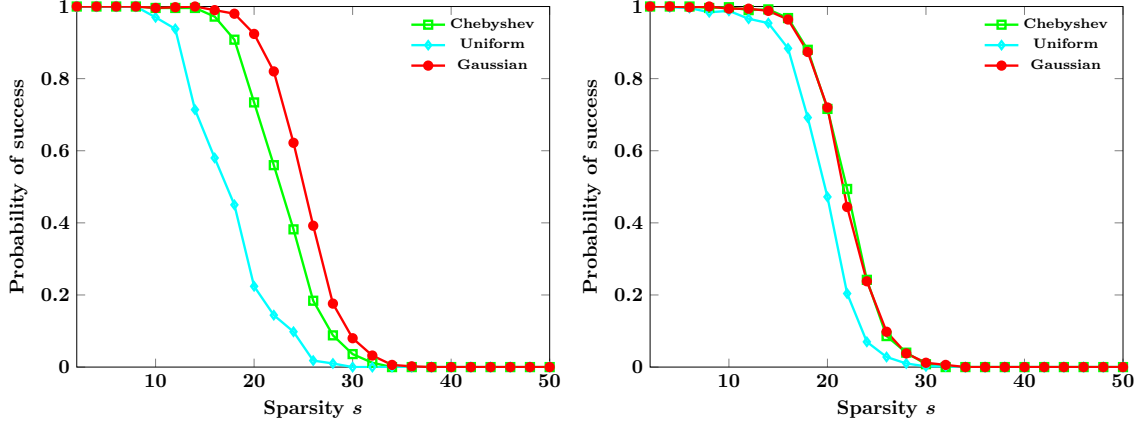


Figure 3. Recovery probability with respect to sparsity s of Chebyshev polynomials with fixed sample number $M = 85$. Left: $d = 2$, $\mathbf{n} = \mathbf{21}$ ($N = 231$). Right: $d = 10$, $\mathbf{n} = \mathbf{4}$ ($N = 286$).

5.1.3. *Gaussian measure with Hermite polynomials.* In Fig.4, we report the numerical results for the Gaussian measure with Hermite polynomials approximation. We examine the maximum coefficient error, $\|\mathbf{c} - \hat{\mathbf{c}}\|_\infty$, as we increase the number of sample points. In the left-hand plot of Figure 4, we show the convergence rate for sparse Hermite polynomials (with $d = 2$, $\mathbf{n} = \mathbf{21}$, and $N = 231$) as a function of number of sample points. In the right-hand plot of Figure 4, we show the convergence rate for sparse Hermite polynomials (with $d = 10$, $\mathbf{n} = \mathbf{4}$, and $N = 286$) as a function of number of sample points. The sparsity in both cases is $s = 5$. Although subsampling from a tensor-product grid works well in low dimensions, in the high-dimensional $d = 10$ case, we see that sampling according to the orthogonality measure produces better results. This is consistent with earlier observations [48].

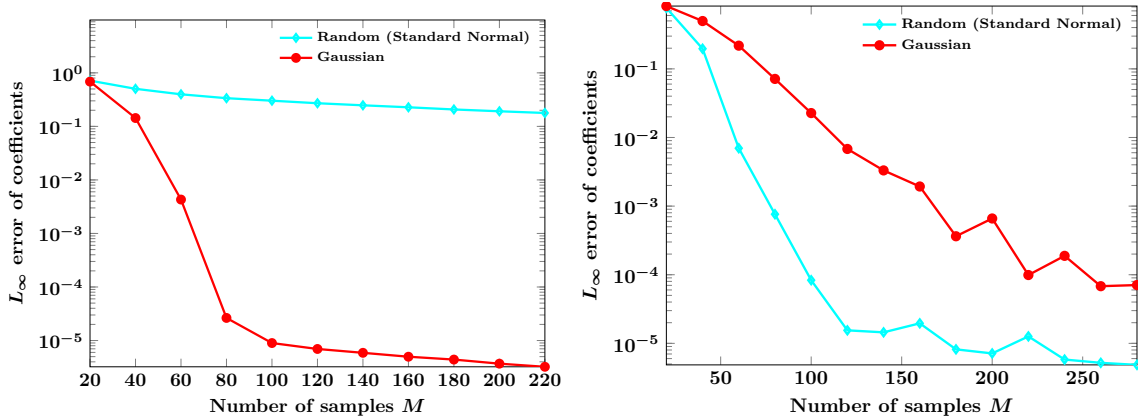


Figure 4. Recovery error with respect to number of sample points of Hermite polynomials. Left: $d = 2$, $\mathbf{n} = \mathbf{21}$, and $N = 231$. Right: $d = 10$, $\mathbf{n} = \mathbf{4}$, and $N = 286$. For both plots, the error shown is the average over 500 trials.

5.2. **Recovery of analytical functions.** In general, functions do not have a finite representation in the orthogonal polynomials, but instead have “approximately” sparse representations. Here, we consider a few functions of this form.

We report the numerical error with Legendre polynomials for the underlying high-degree monomial function $f(\mathbf{x}) = x_1^{10}x_2^{10}$ and high-dimensional Generalized Rosenbrock $f(\mathbf{x}) = \sum_{i=1}^{10} (1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2$. We attempt to recover a sparse representation of these functions in a Legendre polynomial basis. The left-hand plot of Figure 5 shows recovery of the monomial function (high degree, low dimension), and the right-hand plot shows recovery of the Rosenbrock function (low degree, high dimension). In both of these cases, subsampling from a Gaussian quadrature grid produces superior results when compared against standard alternatives.

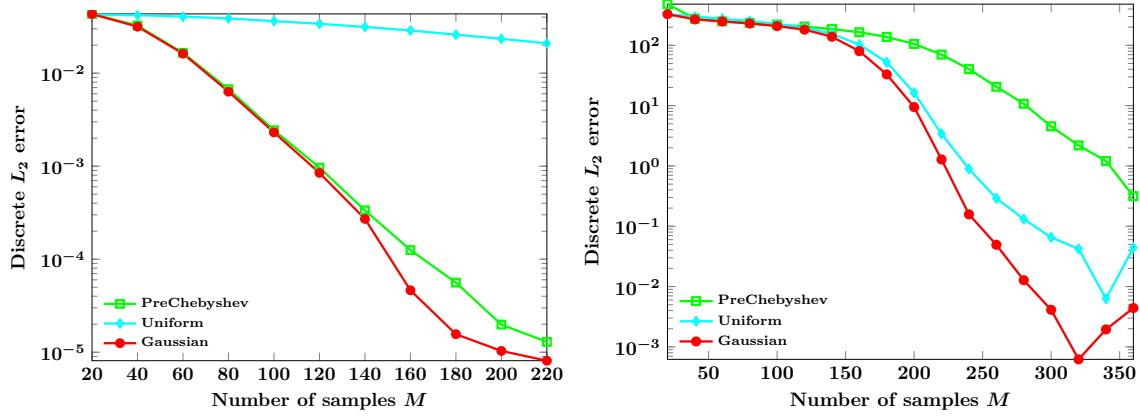


Figure 5. Convergence rate against number of sample points. Left: High-degree monomial function ($d = 2, n = 21$). Right: Generalized Rosenbrock ($d = 10, n = 5$). For both plots, the error is averaged over 500 trials.

We run a similar test using Hermite polynomials for two different functions in two dimensions: $f_1(\mathbf{x}) = 2^{-0.2x_1^2 - 0.2x_2^2}$ and $f_2(\mathbf{x}) = e^{-0.6x_1 - 0.6x_2}$. The recovery results are shown in Figure 6, where we see that in these low-dimensional cases, it is not always clear that subsampling from a Gaussian quadrature grid produces better results than sampling from the orthogonality measure. However, this does appear to be true if one can afford to take a larger sample count M .

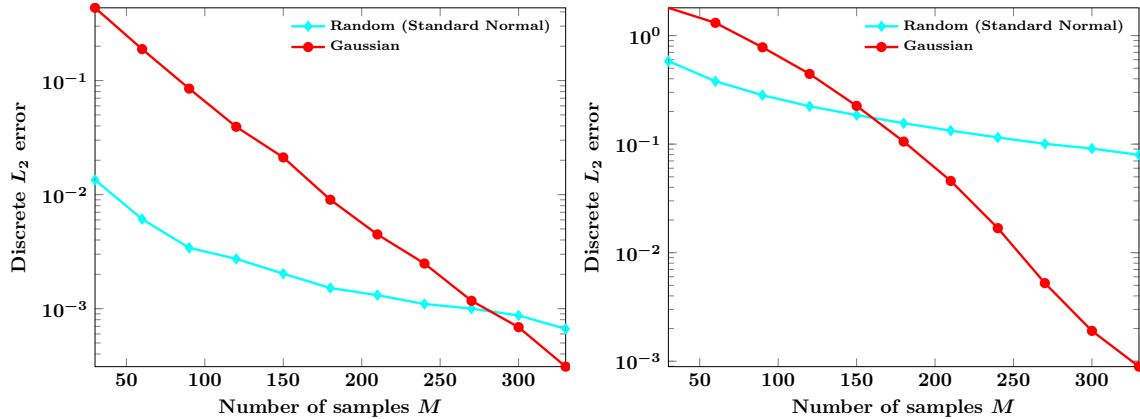


Figure 6. Convergence rate against number of sample points. Left: $f_1(\mathbf{x}) = 2^{-0.2x_1^2 - 0.2x_2^2}$ ($d = 2, n = 26$). Right: $f_2(\mathbf{x}) = e^{-0.6x_1 - 0.6x_2}$ ($d = 2, n = 25$). In both plots, the error is averaged over 500 trials.

5.3. A simple ODE with random inputs. We consider a simple random ODE problem with Gaussian random input:

$$(41) \quad \frac{du}{dt} = -k(X)u, \quad u(0) = 1,$$

where $k(X)$ is a function of a Gaussian random variable X ; thus the Hermite functions will be used as the approximation basis. To illustrate the idea, we set $k(X) = \beta X$. We approximate u , i.e., we recover the solution

$$u(t, X) \sim \sum_{n=0}^{N-1} c_n(t) \varphi_n(X),$$

by attempting to find a sparse representation of the coefficients c_n . Thus we are in fact approximating the function $\tilde{f} = e^{-\frac{y^2}{2}} u^2(t, y)$. We are interested in the second moment of the solution, i.e.

$$Q = \mathbb{E}u^2(t, X) \propto \int e^{-\frac{x^2}{2}} u^2(t, x) dx.$$

For each collocation points, one has to solve the ODE to get the information $u(t, y_i)$. The numerical convergence results for the quantity of interest Q is shown in Figure 7 with $\beta = -0.65$ and $t = 1$. In this case, the Gaussian quadrature subsampling strategy works very well.

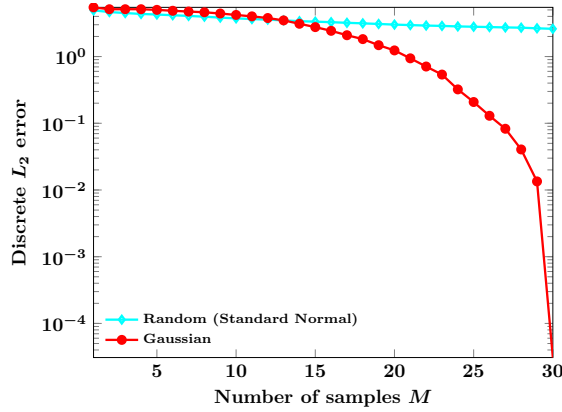


Figure 7. Convergence rate against number of sample points. $\beta = -0.65$, with $n = 30$, with error averaged over 500 trials.

5.4. PDE with random inputs. This section illustrates the computational performance of our algorithm for the following stochastic linear elliptic problems in two spatial dimensions.

$$(42) \quad \begin{cases} -\nabla \cdot (a(\mathbf{y}, \omega) \nabla u(\mathbf{y}, \omega)) = f(\mathbf{y}, \omega) & \text{in } \mathcal{D} \times \Omega, \\ u(\mathbf{y}, \omega) = 0 & \text{on } \partial \mathcal{D} \times \Omega, \end{cases}$$

with spatial domain $\mathcal{D} = [0, 1]^2$. For these numerical examples we take a deterministic load $f(\mathbf{y}, \omega) = \cos(y_1) \sin(y_2)$ and construct the random diffusion coefficient $a_N(\mathbf{y}, \omega)$ with one-dimensional spatial dependence as in [2, p. 2336]:

$$\log(a_N(\omega, \mathbf{y}) - 0.5) = 1 + X_1(\omega) \left(\frac{\sqrt{\pi}L}{2} \right)^{1/2} + \sum_{i=2}^3 \zeta_i f_i(\mathbf{y}) X_i(\omega),$$

where

$$\zeta_i := (\sqrt{\pi}L)^{1/2} \exp\left(\frac{-(\lfloor \frac{i}{2} \rfloor \pi L)^2}{8} \right), \quad \text{if } i > 1,$$

and

$$f_i(\mathbf{y}) := \begin{cases} \sin\left(\frac{\lfloor \frac{i}{2} \rfloor \pi y_1}{L_p}\right), & \text{if } i \text{ even,} \\ \cos\left(\frac{\lfloor \frac{i}{2} \rfloor \pi y_1}{L_p}\right), & \text{if } i \text{ odd.} \end{cases}$$

Here $\{X_i\}_{i=1}^3$ are mutually independent and are each uniformly distributed on the interval $[-1, 1]$. Thus a family of Legendre polynomials is used to approximation as a function of \mathbf{X} . For $y_1 \in [0, 1]$, let $L_c = 1/64$ be a desired physical correlation length for $a(\mathbf{y}, \omega)$. Then the parameter L_p and L are $L_p = \max\{1, 2L_c\}$ and $L = \frac{L_c}{L_p}$, respectively. The deterministic elliptic equations are solved by a standard finite element method and the spatial domain \mathcal{D} is partitioned into 648 triangles with 1369 unknowns.

As the exact solution is not available, we use a high level sparse grid collocation method to obtain the reference solution. In Fig. 8, we see plots of error in ℓ_2 norm of the mean and standard deviation between the reference and ℓ_1 -minimization for various sampling techniques as a function of the number of samples M .

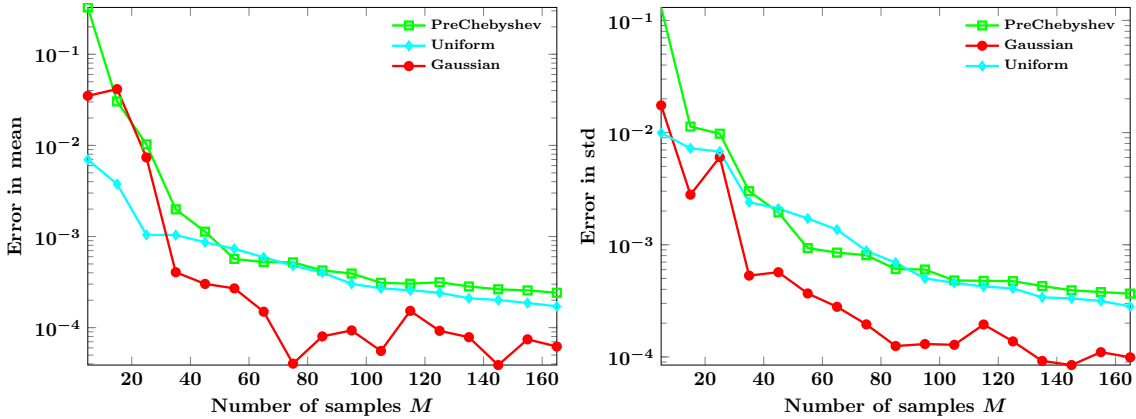


Figure 8. Error in ℓ_2 norm of the mean and standard deviation between the reference and ℓ_1 -minimization for the various sampling method as a function of the number of samples M . $d = 3$, $\mathbf{n} = 11$ ($N = 286$)

6. CONCLUSION

We investigate the problem of approximating a multivariate function via ℓ_1 minimization methods. Such strategies presume that the underlying function is sparse, or approximately sparse, in the approximating basis. The ability to exactly recover a sparse representation with a small number of samples has been the focus of much research in compressive sampling. We propose sampling by randomly subsampling a tensor-product grid of Gaussian quadrature points. This procedure was investigated in [39] for Legendre polynomials (corresponding to a uniformly distributed random variable). We have provided a nontrivial extension in both analysis and numerical results, covering general Beta distributions taking values on compact domains, as well as one- and two-sided exponential random variable taking values on unbounded domains. In particular, our analysis covers the case of Hermite polynomials (normally-distributed random variables) and Laguerre polynomials (exponential random variables). Our framework provides a non-intrusive way to construct generalized Polynomial Chaos expansions for very general classes of distributions.

APPENDIX A. PROOFS

In this appendix we collect results which imply the results given by the three parts of Lemma 4.2. These are essentially well-known results in the theory of orthogonal polynomials. Our analysis uses these well-known results in fairly straightforward ways.

We treat individually the proofs of each of the three Lemmas 4.2.A, 4.2.B, and 4.2.C. The strategy for each proof is identical: First we characterize the interval in which the univariate orthogonal polynomial zeros z_j^i are located. Then we use established bounds on weighted orthogonal polynomial families on those intervals.

Note that our main reference for bounds on orthogonal polynomials is [21], which explicitly states these bounds in the form that we require. However, the conclusions in [21] are essentially a reshuffling of far deeper, more technical, and elegant results in [22, 32, 23, 24].

After proving the three Lemmas 4.2, we end with the proof of our main result, Theorem 4.1.

A.1. Proof of Lemma 4.2.A. Assume the setup of Lemma 4.2.A: that X^i is a scalar random variable taking values on $[-1, 1]$ with a Beta distribution with shape parameters $\gamma, \delta \geq \frac{1}{2}$. We first establish intervals in which the zeros of orthogonal polynomial lie, and then use boundedness results on these intervals. This bounded case is easiest, for which the n zeros z_j^i for $j = 1, \dots, n$ all lie inside the n -independent compact interval of orthogonality.

Lemma A.1. *For all $n \in \mathbb{N}$, we have $\{z_j^i\}_{j=1}^n \subset [-1, 1]$.*

See, e.g., Theorem 3.3.1 of [38]. We can now use results from [21].

Lemma A.2 ([21]). *For all $n \in \mathbb{N}$ and $0 \leq k \leq n-1$, we have*

$$\sup_{z^i \in [-1, 1]} |\psi_{k,n}(z^i)|^2 \leq C(\gamma^i, \delta^i)$$

Lemma 4.2.A now follows easily:

$$(43) \quad L_i(n) = \max_{0 \leq k \leq n-1} \max_{j=1, \dots, n} |\psi_{k,n}(z_j^i)|^2 \leq \max_{0 \leq k \leq n-1} \sup_{z^i \in [-1, 1]} |\psi_{k,n}(z^i)|^2 \leq C(\gamma^i, \delta^i).$$

A.2. Proof of Lemma 4.2.B. Both of the exponential cases are more subtle. Assume the setup of Lemma 4.2.B, that the random variable X^i has a two-sided exponential distribution, $\rho^i(x^i) = \exp(-|x^i|^\alpha)$ with $\alpha > \frac{3}{2}$. In order to characterize intervals containing the zeros of the associated orthogonal polynomials, we will need the numbers a_n^W , which for $n \geq 1$ are given by

$$a_n^W = \left(n \frac{\sqrt{\pi} \Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha}{2} + \frac{1}{2})} \right)^{1/\alpha}.$$

Note that, modulo an α -dependent constant, these numbers scale like $n^{1/\alpha}$. The a_n are the *Mhaskar-Rahkmanov-Saff* numbers associated to the weight $\sqrt{\rho^i}$ [27] and play an essential role in the analysis of weighted polynomials.

Lemma A.3 ([22]). *For each $n \in \mathbb{N}$, the n -point Gaussian quadrature nodes satisfy:*

$$\{z_j^i\}_{j=1}^n \subset [-\hat{a}_n^W, \hat{a}_n^W],$$

where \hat{a}_n^W satisfy

$$(44) \quad \hat{a}_n^W = a_n \left[1 + cn^{-2/3} \right],$$

with c a n -independent constant.

If α is an even integer, we can take $C = 0$ so that $\hat{a}_n^W = a_n^W$ [28]. We have bounds for weighted polynomials on the interval defined above.

Lemma A.4 ([21]). *Let $c > 0$ be as in (44) defining \widehat{a}_n^W . For all $n \in \mathbb{N}$ and $0 \leq k \leq n-1$, we have*

$$\sup_{z^i \in [-\widehat{a}_n^W, \widehat{a}_n^W]} |\psi_{k,n}(z^i)|^2 \leq C(\alpha)n^{2/3}.$$

With these two lemmas, essentially the same argument as in (43) can be used to establish,

$$L_i(n) = \max_{0 \leq k \leq n-1} \max_{j=1, \dots, n} |\psi_{k,n}(z_j^i)|^2 \leq Cn^{2/3},$$

which is Lemma 4.2.B.

A.3. Proof of Lemma 4.2.C. Assume the setup of Lemma 4.2.C, that the random variable X^i has a one-sided exponential distribution with density $\rho^i(x^i) \propto \exp(-|x^i|^\alpha)$ with $\alpha > \frac{3}{4}$.

In this half-line case, both the strategy and the results are essentially the same as with the whole real line in the previous section. One major change is in the constants a_K , which in this case are given by

$$a_n^H = \left(n \frac{\sqrt{\pi} \Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} \right)^{1/\alpha}.$$

Note again that these numbers scale like $n^{1/\alpha}$. These a_n^H are the *Mhaskar-Rahkmanov-Saff* numbers associated to the half-line weight $\sqrt{\rho^i}$.

Lemma A.5 ([22]). *For each $n \in \mathbb{N}$ with $n \geq 1$, the n -point Gaussian quadrature nodes satisfy:*

$$\{z_j^i\}_{j=1}^n \subset [0, \widehat{a}_n^H],$$

where \widehat{a}_n^H satisfy

$$(45) \quad \widehat{a}_n^H = a_n \left[1 + cn^{-2/3} \right],$$

with c a n -independent constant.

We have bounds for weighted polynomials on the interval defined above.

Lemma A.6 ([21]). *Let $C > 0$ be as in (45) defining \widehat{a}_n^H . For all $n \in \mathbb{N}$ and $0 \leq k \leq n-1$, we have*

$$\sup_{z^i \in [0, \widehat{a}_n^H]} |\psi_{k,n}(z^i)|^2 \leq C(\alpha)n^{2/3}.$$

Again the same argument as in (43) can be used to establish,

$$L_i(n) = \max_{0 \leq k \leq n-1} \max_{j=1, \dots, n} |\psi_{k,n}(z_j^i)|^2 \leq Cn^{2/3},$$

which is Lemma 4.2.C.

A.4. Proof of Theorem 4.1. Under the assumptions of this theorem, we perform the algorithm given in Section 3.3. Recalling the notation, we have a multi-index set $\Lambda \subset \Lambda_{\mathbf{n}-1}^P$ for some multi-index \mathbf{n} . The index set Λ has size N . We construct the tensor-product Gauss quadrature rule with n_i points in dimension i . Then the rectangular $\left(\prod_{i=1}^d n_i \right) \times N$ weighted Vandermonde-like matrix \mathbf{A} is defined, whose entries are

$$(46) \quad (A)_{p,q} = \sqrt{w_{\ell(p)}} \phi_{\mathbf{k}(q)}(\mathbf{z}_{\ell(p)}), \quad 1 \leq p, q \leq \prod_{i=1}^d n_i$$

where $\mathbf{k}(q)$ and $\ell(p)$ represent any enumeration of the elements in $\Lambda_{\mathbf{n}-1}^P$. The matrix \mathbf{A} is an orthogonal matrix. (See Lemmas 4.1 or 3.1, and Definition 3.1.) According to Section 3.3, \mathbf{D} is formed by subsampling rows from \mathbf{A} .

We subsample rows from \mathbf{A} without replacement. This does *not* sample iid from $\nu_{\mathbf{n}}$ since the samples are dependent. Thus, Theorem 2.2 cannot be used directly to analyze this procedure.

Nevertheless, the analysis may be amended to include this type of procedure; see Corollary 12.38 in Section 12.6 of [35] and the surrounding discussion.

Lemma A.7 ([35]). *Suppose \mathbf{A} is an orthogonal matrix whose entries are bounded by $\frac{1}{\sqrt{\prod_{i=1}^d n_i}} \sqrt{L}$. Suppose we subsample M rows from \mathbf{A} . Then under the sample count condition (15), the subsampled matrix satisfies the conclusions of Theorem 2.2.*

We then must determine how to bound the entries of the matrix \mathbf{A} . Assuming that the marginal components of the random variable \mathbf{X} each have a distribution satisfying any of the conditions in Lemmas 4.2, then the entries of \mathbf{A} , defined by (46), satisfy

$$\begin{aligned}
\left(\prod_{i=1}^d \sqrt{n_i} \right) |(A)_{p,q}| &= \left(\prod_{i=1}^d \sqrt{n_i} \right) \sqrt{w_{\ell(p)}} |\phi_{\mathbf{k}(q)}(\mathbf{z}_{\ell(p)})| \\
&\stackrel{(2),(27)}{=} \prod_{i=1}^d \sqrt{n_i w_{\ell(p)}^i} |\varphi_{\mathbf{k}(q)_i}^i(z_{\ell(p)}^i)| \\
&\stackrel{(22)}{=} \prod_{i=1}^d \sqrt{n_i \lambda_{n_i}^i(z_{\ell(p)}^i)} |\varphi_{\mathbf{k}(q)_i}^i(z_{\ell(p)}^i)| \\
&\stackrel{(36)}{=} \prod_{i=1}^d |\psi_{\ell(p)_i, n_i}| \\
&\stackrel{\text{Lemmas 4.2}}{\leq} \prod_{i=1}^d \sqrt{L_i(n_i)},
\end{aligned}$$

with the individual L_i factors given by the bounds in Lemmas 4.2 depending on the distribution of \mathbf{X} . Thus,

$$\sup_{p,q} \left(\prod_{i=1}^d n_i \right) (A)_{p,q}^2 \leq L(\mathbf{n})$$

Thus, under condition (39), then Lemma A.7 implies the conclusion of Theorem 4.1.

REFERENCES

- [1] N. Agarwal and N. R. Aluru. A domain adaptive stochastic collocation approach for analysis of MEMS under uncertainties. *Journal of Computational Physics*, 228(20):7662–7688, Nov. 2009.
- [2] I. Babuka, F. Nobile, and R. Tempone. A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data. *SIAM Review*, 52(2):317–355, Jan. 2010.
- [3] M. Bieri. A Sparse Composite Collocation Finite Element Method for Elliptic SPDEs. *SIAM Journal on Numerical Analysis*, 49(6):2277–2301, Jan. 2011.
- [4] M. Bieri, R. Andreev, and C. Schwab. Sparse Tensor Discretization of Elliptic sPDEs. *SIAM Journal on Scientific Computing*, 31(6):4281–4304, Dec. 2009.
- [5] T. Cai, L. Wang, and G. Xu. New Bounds for Restricted Isometry Constants. *IEEE Transactions on Information Theory*, 56(9):4388–4394, Sept. 2010.
- [6] E. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, Dec. 2005.
- [7] E. Candes and T. Tao. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, Dec. 2006.
- [8] E. J. Cands, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, Aug. 2006.
- [9] 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. *ESAIM:M2AN*, 49(3):815–837, 2015.
- [10] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Anal. Appl.*, 9(01):1–10, 2011.
- [11] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Apr. 2006.

- [12] D. L. Donoho and M. Elad. On the stability of the basis pursuit in the presence of noise. *Signal Processing*, 86(3):511–532, Mar. 2006.
- [13] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *Journal of Computational Physics*, 230(8):3015–3034, Apr. 2011.
- [14] 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.
- [15] 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.
- [16] J. Foo, X. Wan, and G. E. Karniadakis. The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications. *Journal of Computational Physics*, 227(22):9572–9595, Nov. 2008.
- [17] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, July 2007.
- [18] W. Gautschi. The condition of Vandermonde-like matrices involving orthogonal polynomials. *Linear Algebra and its Applications*, 52-53:293–300, July 1983.
- [19] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., 1991.
- [20] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363–386, Jan. 2015.
- [21] J. Jakeman, A. Narayan, and T. Zhou. A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions. arXiv:1602.06879.
- [22] A. L. Levin and D. S. Lubinsky. Christoffel functions, orthogonal polynomials, and Nevai’s conjecture for Freud weights. *Constructive Approximation*, 8(4):463–535, Dec. 1992.
- [23] E. Levin and D. Lubinsky. Orthogonal polynomials for exponential weights on. *Journal of Approximation Theory*, 134(2):199–256, June 2005.
- [24] E. Levin and D. Lubinsky. Orthogonal polynomials for exponential weights on , II. *Journal of Approximation Theory*, 139(1?):107–143, Mar. 2006.
- [25] H. Li and D. Zhang. Probabilistic collocation method for ow in porous media: comparisons with other stochastic methods. *Water Resources Research*, 43:W09409, 2007.
- [26] X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084–3113, May 2009.
- [27] H. N. Mhaskar and E. B. Saff. Where does the sup norm of a weighted polynomial live? *Constructive Approximation*, 1(1):71–91, Dec. 1985.
- [28] A. MT, P. Nevai, and V. Totik. Asymptotics for the Greatest Zeros of Orthogonal Polynomials. *SIAM Journal on Mathematical Analysis*, 17(3):745–751, May 1986.
- [29] A. Narayan, J. Jakeman, and T. Zhou. A christoffel function weighted least squares algorithm for collocation approximations. *to appear in Math. Comput.*, 2016.
- [30] 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.
- [31] A. Narayan and T. Zhou. Stochastic Collocation on Unstructured Multivariate Meshes. *Communications in Computational Physics*, 18(01):1–36, July 2015.
- [32] P. Nevai, T. Erdlyi, and A. Magnus. Generalized Jacobi Weights, Christoffel Functions, and Jacobi Polynomials. *SIAM Journal on Mathematical Analysis*, 25(2):602–614, Mar. 1994.
- [33] 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, Jan. 2008.
- [34] H. Rauhut. Compressive Sensing and Structured Random Matrices .: In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, pages 1–92. 2010.
- [35] H. Rauhut and S. Foucart. *A Mathematical Introduction to Compressive Sensing*. Birkhuser, New York, 2013 edition, Aug. 2013.
- [36] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1 -minimization. *Journal of Approximation Theory*, 164(5):517–533, May 2012.
- [37] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545–555, Feb. 2003.
- [38] G. Szegő. *Orthogonal Polynomials*. American Mathematical Soc., 4th edition, 1975.
- [39] G. Tang and G. Iaccarino. Subsampled Gauss Quadrature Nodes for Estimating Polynomial Chaos Expansions. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):423–443, Jan. 2014.
- [40] T. Tang and T. Zhou. On Discrete Least-Squares Projection in Unbounded Domain with Random Evaluations and its Application to Parametric Uncertainty Quantification. *SIAM Journal on Scientific Computing*, pages A2272–A2295, Jan. 2014.

- [41] R. Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, Jan. 1996.
- [42] R. A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA Journal of Numerical Analysis*, 27(2):232–261, Apr. 2007.
- [43] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [44] E. van den Berg and M. P. Friedlander. Probing the Pareto Frontier for Basis Pursuit Solutions. *SIAM Journal on Scientific Computing*, 31(2):890, 2009.
- [45] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, July 2010.
- [46] D. Xiu and G. E. Karniadakis. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, Jan. 2002.
- [47] 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.
- [48] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1 -minimization. *International Journal for Uncertainty Quantification*, 2(3):279–293, 2012.
- [49] J. Yang and Y. Zhang. Alternating Direction Algorithms for ℓ_1 -Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, Jan. 2011.
- [50] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman Iterative Algorithms for ℓ^1 -Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, Jan. 2008.
- [51] 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, 2014.