International Journal for Uncertainty Quantification, 2 (3): 279–293 (2012)

STOCHASTIC COLLOCATION ALGORITHMS USING ℓ_1-MINIMIZATION

Liang Yan,[1] Ling Guo,[2] & Dongbin Xiu[2,*]

[1] *Department of Mathematics, Southeast University, Nanjing, 210096, China*

[2] *Department of Mathematics, Purdue University, West Lafayette, Indiana 47907, USA*

Original Manuscript Submitted: 8/31/2011; Final Draft Received: 2/27/2012

The idea of ℓ_1-minimization is the basis of the widely adopted compressive sensing method for function approximation. In this paper, we extend its application to high-dimensional stochastic collocation methods. To facilitate practical implementation, we employ orthogonal polynomials, particularly Legendre polynomials, as basis functions, and focus on the cases where the dimensionality is high such that one can not afford to construct high-degree polynomial approximations. We provide theoretical analysis on the validity of the approach. The analysis also suggests that using the Chebyshev measure to precondition the ℓ_1-minimization, which has been shown to be numerically advantageous in one dimension in the literature, may in fact become less efficient in high dimensions. Numerical tests are provided to examine the performance of the methods and validate the theoretical findings.

KEY WORDS: *stochastic collocation, Legendre polynomials, ℓ_1-minimization, multi-dimensional interpolation*

1. INTRODUCTION

Stochastic computation has received intensive attention in recent years, due to the pressing need to conduct uncertainty quantification (UQ) in practical computing. Various numerical methods have been developed, among which the most widely used ones are based on (generalized) polynomial chaos (gPC), c.f. [1, 2]. 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. The development of stochastic collocation algorithms became very active, after the introduction of high-order algorithms using sparse grids [3], and produced many different techniques, cf. [4–14] to name a few. Roughly speaking, the construction of the gPC approximations takes two major approaches. One is based on interpolation, where the simulation samples are interpreted by the approximation precisely. The other one is the regression type, which approximately matches the simulation samples.

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 approach—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 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. 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

*Correspond to Dongbin Xiu, E-mail: dxiu@purdue.edu

2152–5080/12/$35.00

terms in their frequency domain. Following the seminal work of [15–18], the theory of CS has generated an enormous amount of interest in many disciplines and resulted in many newer theoretical results and practical implementations. Numerous literature on various aspects of CS exists and will not be discussed here.

In a recent work [19], 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 [19], some key properties, such as the probability under which the sparse random response function can be recovered, are studied. Numerical tests demonstrate that sparsity does occur in practical stochastic problems and the method can be effective for problems in large dimensions.

This paper extends the work of [19]. The particular focus of this paper is on the recoverability of stochastic solutions in high-dimensional random spaces. This is relevant because in UQ simulations the dimensionality is often determined by the number of random parameters and can be very large. It is not uncommon to encounter practical stochastic problems with dimensions on the order of hundreds, in addition to the traditional space and time dimensions.

In this paper, we will adopt orthogonal polynomials such as Legendre polynomials as the basis functions, as opposed to trigonometric functions used in the traditional CS. Many of the current results are motivated by [20, 21], recent studies on ℓ_1-minimization using Legendre polynomials in one dimension ($d = 1$). Here we will provide theoretical analysis on the convergence of such an approach in high dimensions $d \gg 1$. In this case, the cardinality of the polynomial space, M, grows very fast (often exponentially fast) when its degree (P) is increased. We will focus on the case where the polynomial degree is less than the dimensionality, i.e., $P < d$. The condition under which the ℓ_1-minimization approach can recover the unknown stochastic function is presented, and a set of numerical tests are provided to examine the performance of the method. An interesting variation of the approach, Chebyshev preconditioning, is also studied. This approach has been shown to be more efficient in one dimension with $d = 1$ [20, 21]. Here we demonstrate that in high dimensions when $d > P$, the Chebyshev preconditioned ℓ_1-minimization may become less effective than the direct ℓ_1-minimization.

We remark that the multi-dimensional polynomial approximation, particularly polynomial interpolation, has been intensively studied for a long time. It remains an active and challenging field. The use of ℓ_1-minimization represents a drastically different approach. And for this reason we will not review the traditional approximation theories and methods. Also, even though we present the results in the context of high-dimensional stochastic computation, they can be equally viewed from the more traditional approximation theory perspective.

Finally we note that one of the advantages of the ℓ_1-minimization is that it allows one to conduct polynomial approximations on unstructured nodes. (Any inherent structure of the nodes would usually incur fast growth of the number of the nodes in high dimensions, e.g., the sparse grids.) The traditional approximation method on unstructured nodes is polynomial regression. However, two of the recent developments are worth mentioning. One is the low-rank approximation in high dimensions [22] and the other is the least orthogonal interpolation [23].

The paper is organized as follows. After presenting the setup of the problem in Section 2, we present the main results in Section 3, where recoverability of both the direct ℓ_1-minimization and Chebyshev preconditioned ℓ_1-minimization are discussed. Numerical tests are provided in Section 4 to verify the theoretical findings, before the conclusions in Section 5.

2. STOCHASTIC COLLOCATION: THE SETUP

Here we adopt the standard setting for stochastic collocation methods. Let $Z = (Z_1, \ldots, Z_d) \in \mathbb{R}^d$, $d \geq 1$, be a set of (independent) random variables modeling the random inputs for a partial differential equation (PDE),

$$\begin{cases} u_t(x, t, Z) = \mathcal{L}(u), & D \times (0, T] \times I_Z, \\ \mathcal{B}(u) = 0, & \partial D \times [0, T] \times I_Z, \\ u = u_0, & D \times \{t = 0\} \times I_Z, \end{cases} \tag{1}$$

where $I_Z \subseteq \mathbb{R}^d$, $d \geq 1$, is the range of Z, $D \in \mathbb{R}^k$, $k = 1, 2$, or 3, is the physical domain, and $T > 0$. Here $\mathcal{L}$ stands for a (nonlinear) differential operator and $\mathcal{B}$ a boundary condition operator. We equip Z with a distribution function $\nu(z) = \text{Prob}(Z \leq z)$, where $z \in \mathbb{R}^d$ is real. For each component of Z, let $\nu_i(z_i) = \text{Prob}(Z_i \leq z_i)$, $z_i \in \mathbb{R}$, be its marginal distribution function. Obviously, when mutual independence among the components is assumed, we have $\nu(z) = \prod_{i=1}^{d} \nu_i(z_i)$. We remark that independence is a common assumption used in stochastic computing and is adopted here as well.

In the stochastic collocation method, one first chooses a set of nodes $\Theta_N = \{z^{(j)}\}_{j=1}^N \subset I_Z$, where $N \geq 1$ is the number of nodes. And then for each $j = 1, \ldots, N$, one solves a *deterministic problem* (1) at the node $z^{(j)}$,

$$\begin{cases} u_t(x,t,z^{(j)}) = \mathcal{L}(u), & D \times (0,T], \\ \mathcal{B}(u) = 0, & \partial D \times [0,T], \\ u = u_0, & D \times \{t=0\}, \end{cases} \tag{2}$$

and obtains $u^{(j)} \triangleq u(x,t,z^{(j)})$. Hereafter we will suppress the notions of x and t whenever possible, with the understanding that our statements are made for all fixed x and t.

Finally, once the pairings $\left(z^{(j)}, u^{(j)}\right)$, $j = 1, \ldots, N$, are obtained, the task is to construct a function $w(Z)$, usually residing in a proper polynomial space in term of Z, such that $w(Z) \approx u(Z)$ in a proper sense.

In the construction of the approximating function $w(Z)$, the pairing information can be enforced exactly by requiring $w(z^{(j)}) = u^{(j)}$ for all $j = 1, \ldots, N$. This usually leads to a (polynomial) interpolation problem. Alternatively, one can adopt a regression type approach which does not require precise matching of the function values at each node.

In this paper, we will adopt the interpolation type approach and focus exclusively on polynomial interpolation. Also we assume that I_Z is bounded, and with a proper scaling we confine ourselves to a hypercube, i.e., $Z \in [-1,1]^d, d \geq 1$.

2.1 Multi-dimensional Polynomial Spaces

We will use the standard polynomial space to construct our interpolation. To this end, let us first define a one-dimensional polynomial space. For each $i = 1, \ldots, d$, let W^{i,k_i} be the space of polynomials of degree up to k_i. That is,

$$W^{i,k_i} \triangleq \left\{ p : [-1,1] \to \mathbb{R} : p \in \text{span}\{(z_i)^m\}_{m=0}^{k_i} \right\}. \tag{3}$$

For multi-dimensional cases $d > 1$, we adopt multi-index $\mathbf{k} = (k_1, \ldots, k_d)$ with norm $|\mathbf{k}| = k_1 + \cdots + k_d$. The standard polynomial space is the space of all d-dimensional polynomials of degree up to an integer number P. We denote such a space the *total degree polynomial space* W_P^d, i.e.,

$$W_P^d = \bigotimes_{|\mathbf{k}| \leq P} W^{i,k_i}. \tag{4}$$

The cardinality of the space is

$$\dim W_P^d = \binom{P+d}{d} = \frac{(P+d)!}{P!d!}. \tag{5}$$

Another construction often employed in theoretical analysis is to let the polynomial degree in each variable go up to P. This results in the *full tensor polynomial space*,

$$Z_P^d = \bigotimes_{k_i \leq P} W^{i,k_i}. \tag{6}$$

The cardinality of this space is

$$\dim Z_P^d = (P+1)^d. \tag{7}$$

Note that when $d \gg 1$ the cardinality of both polynomial spaces grows very fast when the degree (P) is increased, resulting in the so-called "curse of dimensionality." The growth for Z_P^d is much faster than that of W_P^d. Therefore in practice the full tensor polynomial space Z_P^d is rarely used for $d > 5$.

2.2 Multi-dimensional Orthogonal Polynomials

We employ orthogonal polynomials as basis functions. For each $i = 1, \ldots, d$, let $\{l_m(z_i)\}$ be a set of orthogonal polynomials in variable z_i that satisfy the following orthogonality condition

$$\int_{-1}^{1} l_m(z_i) l_n(z_i)\, d\nu_i(z_i) = \delta_{mn}, \quad m, n, = 0, 1, 2, \ldots, \tag{8}$$

where m and n are the degrees of the polynomials and ν_i is the probability measure. Note here the polynomials have been normalized. Here we focus on the continuous case where $d\nu_i(z_i) = w_i(z_i)\, dz_i$, where $w(x_i)$ is the weight function satisfying the usual conditions to admit the existence of the orthogonal polynomials. The connection between the polynomial weights and the probability measure is well established, cf. [2].

In the multi-dimensional case, the orthogonal polynomials are constructed as tensor products of the one-dimensional polynomials in each variable.

$$L_{\mathbf{k}}(z) = \prod_{\mathbf{k}\in\Lambda} l_{k_i}(z_i), \tag{9}$$

where Λ is an index set, determined by the polynomial space (4) or (6). The orthogonality relation becomes

$$\int_{[-1,1]^d} L_{\mathbf{m}}(z) L_{\mathbf{n}}(z)\, d\nu(z) = \delta_{\mathbf{mn}}, \qquad \forall \mathbf{m}, \mathbf{n} \in \Lambda,$$

where the measure $d\nu(z) = d\nu_1(z_1) \cdots d\nu_d(z_d)$ is the product of the one-dimensional measures, and $\delta_{\mathbf{mn}} = \delta_{m_1 n_1} \cdots \delta_{m_d n_d}$ is the d-dimensional Kronecker delta function satisfying $\delta_{\mathbf{mn}} = 1$ if $\mathbf{m} = \mathbf{n}$, and $\delta_{\mathbf{mn}} = 0$ otherwise.

Upon choosing a proper ordering scheme for the multi-index, we can order the multi-dimensional polynomials via a single index. The orthogonality becomes

$$\int_{[-1,1]^d} L_m(z) L_n(z)\, d\nu(z) = \delta_{mn}, \qquad 1 \le m, n \le M, \tag{10}$$

where each single-index m corresponds uniquely to a multi-index $(m_1, \ldots, m_d)$, and M is the cardinality of the underlying polynomial space (5) or (7). (A more detailed discussion on the ordering can be found in [24].)

In this paper we will focus on Legendre polynomials, whose weight function is a constant $d\nu(z) = (1/2)^d dz$.

3. STOCHASTIC COLLOCATION VIA ℓ_1-MINIMIZATION

We now focus on the interpolation approach. We remark that multi-dimensional polynomial interpolation on arbitrary nodes is a fundamentally difficult problem. Here we adopt the ℓ_1-minimization approach to circumvent the difficulty. Upon choosing a degree P for the polynomial approximation, the target function $u(Z)$ can be approximated by

$$w(Z) = \sum_{m=1}^{M} c_m L_m(Z). \tag{11}$$

The interpolation condition of $w(z^{(i)}) = u^{(i)}$, $i = 1, \ldots, N$, results in the following problem

$$\mathbf{Ac} = \mathbf{f}, \tag{12}$$

where $\mathbf{c} = (c_1, \ldots, c_M)^T \in \mathbb{R}^M$ is the coefficient vector, $\mathbf{f} = (u^{(1)}, \ldots, u^{(N)})^T$ is the vector for the function samples, and $\mathbf{A} = (a_{n,m})$ is the Vandermonde-type interpolation matrix whose entries are

$$(a_{n,m}) = L_m(z^{(n)}), \qquad n = 1, \ldots, N, \quad m = 1, \ldots, M. \tag{13}$$

The problem is determined when $N = M$, overdetermined when $N > M$, and underdetermined when $N < M$.

It is the underdetermined case that is considered here. This is often encountered in practice, especially in high dimensions. When $d \gg 1$, the cardinality of the polynomial spaces (M) becomes extremely large, see (5) or (7), even when the order of the polynomials is moderate. On the other hand, in many practical applications the evaluation of the target function u is expensive—it requires large-scale numerical simulation of the system (2). Consequently one often has only a limited number of samples. In such cases, $N \ll M$ and the problem (12) becomes severely underdetermined.

One way to circumvent the difficulty is to employ the idea of compressive sensing. Define the ℓ_1-norm of the coefficient vector as

$$\|\mathbf{c}\|_1 = \sum_{m=1}^{M} |c_m|. \tag{14}$$

The coefficient vector can be solved by the following ℓ_1-minimization problem

$$\min \|\mathbf{c}\|_1 \qquad \text{subject to} \quad \mathbf{Ac} = \mathbf{f}. \tag{15}$$

This is a common approach in the context of CS, where a large amount of literature exists. Other forms of minimization exist. For example, instead of the ℓ_1-norm, one can adopt the ℓ_0-norm, defined as $\|\mathbf{c}\|_0 := \#\{m : c_m \neq 0\}$. Also, the interpolation condition $\mathbf{Ac} = \mathbf{f}$ can be relaxed to $\|\mathbf{Ac} - \mathbf{f}\| \leq \epsilon$, for some tolerance value ϵ, resulting in a regression type "de-noising" approach.

In what follows we will focus on the traditional ℓ_1-minimization form (15). The distinct features of our problem are (i) high dimensionality $d \gg 1$ and (ii) the use of orthogonal polynomials, particularly the Legendre polynomials. Despite the large amount of the literature, such a problem has not been well studied.

3.1 Auxiliary Results

Here we summarize some existing results that are useful in deriving the main results in this paper. For a vector $\mathbf{v} = (v_1, \ldots, v_M)^T \in \mathbb{R}^M$, we equip it with ℓ_p-norm

$$\|\mathbf{v}\|_p = \left(\sum_{m=1}^{M} |v_m|^p \right)^{1/p}, \; 1 \leq p < \infty,$$

with $\|\mathbf{v}\|_\infty = \max_{m=1,\ldots,M} |v_m|$. The vector is called s-sparse if $\|\mathbf{v}\|_0 := \#\{m : v_m \neq 0\} \leq s$.

Definition 3.1. The error of the best s-term approximation of a vector $\mathbf{v} \in \mathbb{R}^M$ in ℓ_p-norm is defined as

$$\sigma_{s,p}(\mathbf{v}) = \inf_{\|\mathbf{y}\|_0 \leq s} \|\mathbf{y} - \mathbf{v}\|_p. \tag{16}$$

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

Definition 3.2 ([15, 16])**.** Let **A** be an $N \times M$ matrix. Define the restricted isometry constant (RIC) $\delta_s < 1$ to be the smallest positive number, such that the inequality

$$(1 - \delta_s)\|\mathbf{c}\|_2^2 \leq \|\mathbf{Ac}\|_2^2 \leq (1 + \delta_s)\|\mathbf{c}\|_2^2 \tag{17}$$

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

Theorem 1 (Sparse recovery for RIP-matrices [16, 21, 25])**.** *Let* $\mathbf{A} \in \mathbb{R}^{N \times M}$ *be a matrix with RIC* δ_s *such that*

$$\delta_s < 0.307. \tag{18}$$

For a given $\widetilde{\mathbf{c}} \in \mathbb{R}^M$*, let* $\mathbf{c}$ *be the solution of the* ℓ_1*-minimization*

$$\min \|\mathbf{y}\|_1 \qquad \textit{subject to} \quad \mathbf{A}\mathbf{y} = \mathbf{A}\widetilde{\mathbf{c}}. \tag{19}$$

Then the reconstruction error satisfies

$$\|\mathbf{c} - \widetilde{\mathbf{c}}\|_2 \leq C \frac{\sigma_{s,1}(\widetilde{\mathbf{c}})}{\sqrt{s}} \tag{20}$$

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

Following [20], we call the orthonormal polynomial system $\{L_m(z)\}$ defined by (10) a *bounded orthonormal system* if it is uniformly bounded,

$$\sup_m \|L_m\|_\infty = \sup_m \sup_z |L_m(z)| \leq K, \tag{21}$$

for some $K \geq 1$.

Theorem 2 (RIP for bounded orthonormal systems [20, 21])**.** *Let* $\boldsymbol{A} \in \mathbb{R}^{N \times M}$ *be the interpolation matrix with entries* $\{a_{n,m} = L_m(z^{(n)})\}_{1 \leq n \leq N, 1 \leq m \leq M}$ *from* (13)*, where* $\{L_m\}$ *is a bounded orthonormal system satisfying* (21) *and orthogonality* (10)*, and the points* $z^{(n)}, n = 1, \ldots, N$*, are i.i.d. random samples drawn from the measure* ν *in* (10)*. Assume that*

$$N \geq C\delta^{-2} K^2 s \log^3(s) \log(M), \tag{22}$$

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

3.2 Main Result: Recoverability for $d \geq P$

We are now ready to discuss the high-dimensional case $d \gg 1$. In particular, we consider the total degree polynomial space W_P^d (4) and the case where its cardinality becomes so large that one can not afford a high-degree polynomial P. In such a case, we have $d \geq P$. The following result applies to Legendre polynomial construction, i.e., $d\nu(z) = (1/2)^d \, dz$ in (10).

Theorem 3 (Recoverability of direct ℓ_1-minimization for $d \geq P$)**.** *Let* $z^{(1)}, \ldots, z^{(N)}$ *be independent random samples drawn from the uniform distribution on* $[-1, 1]^d$*, and*

$$N \gtrsim 3^P s \log^3(s) \log(M), \tag{23}$$

where M *is the cardinality of the polynomial space* W_P^d (5) *with* $d \geq P$*, and* s *is the sparsity level of a given vector* $\widetilde{\mathbf{c}} \in \mathbb{R}^M$*. Let* $w \in W_P^d$ *be the polynomial approximation in the form of* (11) *using Legendre polynomials, i.e.,*

$$w(z) = \sum_{m=1}^{M} c_m L_m(z),$$

where the coefficient vector $\mathbf{c}$ *is solved by the* ℓ_1*-minimization problem* (15) *with the data vector* $\mathbf{f} = \mathbf{A}\widetilde{\mathbf{c}}$ *for the given* $\widetilde{\mathbf{c}}$*.*

Then with probability at least $1 - M^{-\gamma \log^3(s)}$*, where* γ *is a universal constant, the vector* $\widetilde{\mathbf{c}}$ *is recoverable to within a factor of its best* s*-term approximation in the sense that*

$$\|\mathbf{c} - \widetilde{\mathbf{c}}\|_2 \lesssim \frac{\sigma_{s,1}(\widetilde{\mathbf{c}})}{\sqrt{s}}. \tag{24}$$

Proof. The m_k-th degree one-dimensional Legendre polynomial l_{m_k} satisfies ([26])

$$\|l_{m_k}\|_\infty \leq (2m_k + 1)^{1/2}. \tag{25}$$

Then the d-dimensional Legendre polynomials satisfy

$$\|L_{\mathbf{m}}\|_\infty \leq \prod_{k=1}^{d} (2m_k + 1)^{1/2}. \tag{26}$$

Since the polynomials are in W_P^d, $|\mathbf{m}| \leq P$, then when $d \geq P$, the right-hand-side is maximized when P of the m_k's are one and the rest are zero. We then obtain the bound on the Legendre polynomials

$$\|L_{\mathbf{m}}\|_\infty \leq 3^{P/2}. \tag{27}$$

Using Theorem 2 for the bounded orthonormal system $\{L_{\mathbf{m}}\}$ with constant $K = 3^{P/2}$ and with uniform distribution from the orthogonality of Legendre polynomials, we obtain the asymptotic estimate (23) to ensure the RIC is less than any fixed value. The conclusion then follows by using Theorem 1. □

This result gives a bound on the number of samples required to accurately recover the unknown stochastic function u if it is in polynomial form.

A similar estimate can be applied to the full tensor polynomial space Z_P^d, whose cardinality is $M = (P+1)^d$. Using the one-dimensional bound (25), we obtain

$$\|L_{\mathbf{m}}\|_\infty \leq (2P + 1)^{d/2}, \qquad \max_{1 \leq i \leq d} m_i \leq P.$$

Then the required number of samples N obtained by Theorem 2 is

$$N \gtrsim M \delta^{-2} s \log^3(s) \log(M). \tag{28}$$

This bound is not useful, because the required number of samples N is now larger than M. And the interpolation problem becomes over-determined. This indicates that in multi-dimensional spaces $d > 1$, the full tensor polynomial space Z_P^d is not proper for the direct ℓ_1-minimization.

3.3 Main Result: Chebyshev Preconditioning

The interpolation condition $\mathbf{Ac} = \mathbf{f}$ can be preconditioned. Though mathematically equivalent, the preconditioned version can be advantageous in practice and result in more accurate solutions. It has been reported in [20, 21] that Chebyshev preconditioning works well in $d = 1$ for the Legendre polynomial approximation. Here we analyze the property of the Chebyshev preconditioning in multi-dimensional case ($d > 1$).

Consider the following one-dimensional function

$$q_m(z_i) = \sqrt{\frac{\pi}{2}} (1 - z_i^2)^{1/4} l_m(z_i), \qquad z_i \in [-1, 1], \tag{29}$$

where $l_m(z_i)$ is the orthonormal Legendre polynomial in variable $z_i, i = 1, \ldots, d$. The system $\{q_m\}$ is then orthonormal with respect to the Chebyshev probability measure $d\omega(z_i) = \pi^{-1}(1 - z_i^2)^{-1/2}\, dz_i$ on $[-1, 1]$, i.e.,

$$\int_{-1}^{1} q_m(z_i) q_n(z_i)\, d\omega(z_i) = \delta_{mn}. \tag{30}$$

For dimension $d > 1$, let

$$Q_{\mathbf{k}}(z) = \prod_{i=1}^{d} q_{k_i}(z_i), \tag{31}$$

where the multi-index $\mathbf{k}$ can be chosen, according to the total-degree polynomial space W_P^d (4), by

$$|\mathbf{k}| = \sum_{i=1}^{d} k_i \leq P. \tag{32}$$

Alternatively, one can use

$$\max_i k_i = P, \tag{33}$$

to correspond to the full tensor polynomial space Z_P^d (6).

Lemma 4. *The system $\{Q_\mathbf{k}\}$ is uniformly bounded on $[-1,1]^d$ and satisfies $\|Q_\mathbf{k}\|_\infty \leq 2^{d/2}$. Moreover, it is orthonormal with respect to the product Chebyshev probability measure*

$$d\mu(z) = \prod_{i=1}^{d} \pi^{-1}(1-(z_i)^2)^{-1/2}\, dz_i, \qquad z \in [-1,1]^d. \tag{34}$$

Note the result holds for both choices for the index set, (32) *and* (33)*.*

Proof. At degree k_i, the following function can be bounded ([26])

$$(1-z_i^2)^{1/4}|l_{k_i}| < 2\pi^{-1/2}, \qquad \forall z_i \in [-1,1],$$

where l_{k_i} is the one-dimensional Legendre polynomials of order k_i. Then the functions (29) satisfy

$$\|q_{k_i}(z_i)\|_\infty = \left\|\sqrt{\frac{\pi}{2}}(1-z_i^2)^{1/4} l_{k_i}(z_i)\right\|_\infty \leq \sqrt{\frac{\pi}{2}} \times 2\pi^{-1/2} = \sqrt{2}.$$

The system $\{Q_\mathbf{k}\}$ from (31) is then uniformly bounded and satisfies $\|Q_\mathbf{k}\|_\infty \leq 2^{d/2}$. The orthogonality of Q follows directly from (30). □

We now formulate the preconditioned version of the Legendre ℓ_1-minimization problem. Recall the original ℓ_1-minimization problem (15) and define an $N \times N$ diagonal matrix $\mathbf{W} = (w_{n,m})$, whose entries $w_{n,m} = (\pi/2)^{d/2} \prod_{i=1}^{d}(1-(z_i^{(n)})^2)^{1/4}\delta_{mn}$. The preconditioned ℓ_1-minimization problem takes the following form

$$\min \|\mathbf{c}\|_1 \qquad \text{subject to} \quad \mathbf{WAc} = \mathbf{Wf}. \tag{35}$$

Note that using the definition (31), the $(N \times M)$ matrix $\mathbf{WA}$ has entries precisely $Q_m(z^{(n)}), 1 \leq n \leq N, 1 \leq m \leq M$, where N is the number of sample points and M is the cardinality of the basis functions $\{Q\}$. The following result then holds.

Theorem 5 (Recoverability of preconditioned ℓ_1-minimization)**.** *Let $z^{(1)}, \ldots, z^{(N)}$ be independent random samples drawn from the Chebyshev measure* (34)*. Consider the Chebyshev preconditioned ℓ_1-minimization problem* (35)*, where the data vector $\mathbf{f} = \mathbf{A}\widetilde{\mathbf{c}}$ for a given vector $\widetilde{\mathbf{c}} \in \mathbb{R}^M$ with sparsity level s. Let $\mathbf{c}$ be the solution of the preconditioned minimization* (35)*. Then there exists*

$$N \gtrsim 2^d s \log^3(s) \log(M), \tag{36}$$

where M is the cardinality of the polynomial space W_P^d or Z_P^d, such that with probability at least $1 - M^{-\gamma \log^3(s)}$, where γ is a universal constant, the vector $\widetilde{\mathbf{c}}$ can be recovered in the sense that

$$\|\mathbf{c} - \widetilde{\mathbf{c}}\|_2 \lesssim \frac{\sigma_{s,1}(\widetilde{\mathbf{c}})}{\sqrt{s}}. \tag{37}$$

Proof. The proof is a trivial exercise of using the bound $\|Q_{\mathbf{k}}\|_\infty \le 2^{d/2}$ derived above. $\square$

This result establishes the recoverability of the Chebyshev preconditioned ℓ_1-minimization. The estimate of the required number of points N in (36) applies to both spaces W_P^d and Z_P^d, and for arbitrary values of d and P. Compared to the result of direct ℓ_1-minimization (23) for $d \ge P$, where N scales with 3^P, the number of points for the preconditioned ℓ_1-minimization scales with 2^d and can be larger, for sufficiently large dimensions d with moderate polynomial order P. Therefore, in very high dimensional spaces the Chebyshev preconditioned ℓ_1-minimization may be less efficient than the direct ℓ_1-minimization because it requires more sample points. This is in contrast to the established result for $d = 1$, where the preconditioned ℓ_1-minimization is more effective [20, 21].

4. NUMERICAL TESTS

In this section we provide numerical tests to verify the theoretical findings. While there exist a large number of studies on the performance of ℓ_1-minimization, we here focus exclusively on the Legendre polynomial case. For the implementation of the minimization, we employ the available tools such as Spectral Projected Gradient algorithm (SPGL1) from [27] that was implemented in the MATLAB package SPGL1 [28].

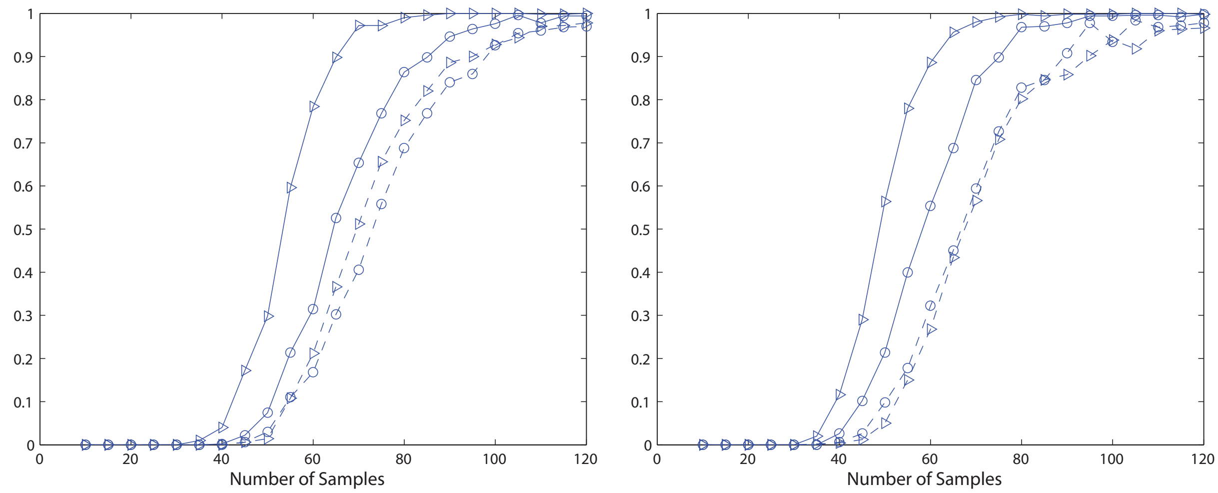
We conduct two groups of tests, each with a given target function in polynomial form serving as the exact solution. The first group is at relatively low dimension of $d = 3$, which allows us to reach higher polynomial degrees and also to conduct tests in the full tensor product space Z_P^d, as well as the total degree space W_P^d. Another group of tests are conducted in a higher dimension of $d = 10$, where the condition $d \ge P$ in Theorem 3 is satisfied. The goal of these tests is to examine the recoverability results from the theorems of the paper. We remark that the dimensionality of $d = 10$ bears no special meaning, as the results from other dimensions such as $d = 15, 20$, demonstrate similar behavior. We choose to demonstrate the results of $d = 10$ largely because it allows us to employ sufficiently high order polynomial spaces to examine the asymptotic theoretical estimates. For more practical simulations at larger dimensions, see [19]. Another remark is that in the tests we did not employ PDE. In practical stochastic collocation computing, the PDE solver is usually a "black-box" to provide function evaluations at the sample points. Here we employ analytically known functions to provide the sample values, and this is merely for benchmarking purpose.

In what follows we will use the term "*direct-uniform*" for the direct ℓ_1-minimization from Theorem 3 and "*pre-Chebyshev*" for the preconditioned ℓ_1-minimization from Theorem 5, where the former utilizes samples from uniform distribution with the direct ℓ_1-minimization (15) and the latter Chebyshev distribution with the preconditioned ℓ_1-minimization (35). We also introduce two other variants of the implementation. One is "*direct-Chebyshev*" where we draw samples from the Chebyshev distribution and apply the direct ℓ_1-minimization (15). The other is "*pre-uniform*" where we draw samples from the uniform distribution and then apply the preconditioned ℓ_1-minimization (35). Though no theoretical results are available for these two variants, they produce comparable results numerically. We include them here for the sake of completeness.

4.1 Low-dimensional Tests in $d = 3$

In this set of tests we examine both the total-degree polynomial space W_P^d (4) and the full tensor polynomial space Z_P^d (6). The target (exact) function is in polynomial form. We first 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 from a Gaussian distribution with zero mean and unit variance. The procedure produces target coefficients $\widetilde{\mathbf{c}}$ that we seek to recover using the ℓ_1-minimization algorithms.

We first examine the frequency of successful recoveries. 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 $\|\mathbf{c} - \widetilde{\mathbf{c}}\|_\infty \le 10^{-3}$. As both Theorem 3 and 5 indicate, the ℓ_1-minimization will reach a success probability of $1 - M^{-\gamma \log^3(s)}$ only when a sufficient number of sample points are used. This is clearly demonstrated in Fig. 1, where the success rate is plotted against increasing numbers of sample points, with a fixed sparsity level of $s = 10$. All four implementations, *direct-uniform*, *pre-Chebyshev*, *direct-Chebyshev*, and *pre-uniform*, are examined in both polynomial spaces W_P^d and Z_P^d. The results clearly show that if the number of points does not reach a critical value,



(a) (b)

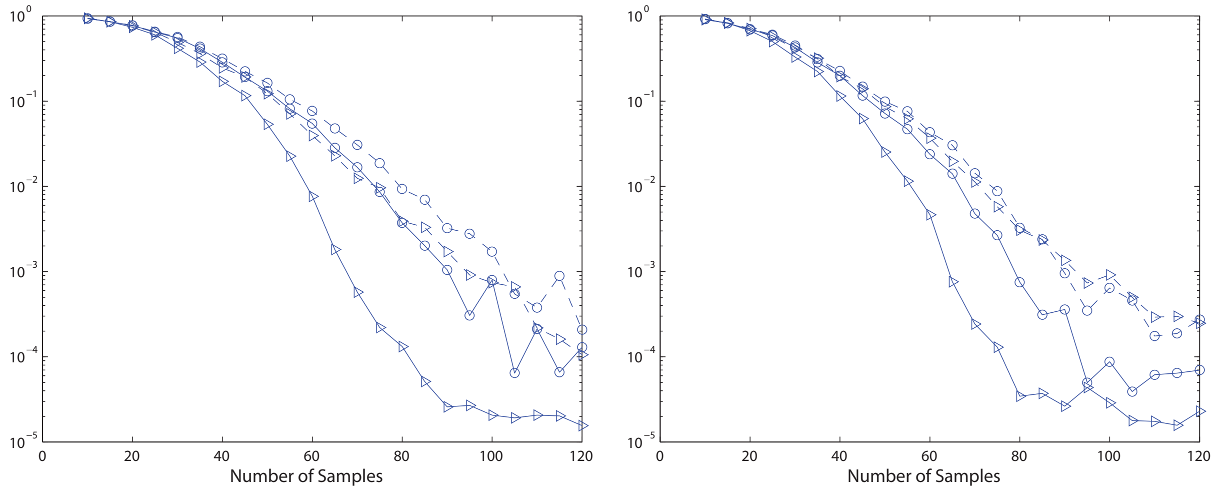
FIG. 1: Probability of successful recovery *vs.* number of sample points ($d = 3$ and $s = 10$). Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*. (a) Total degree polynomial space W_d^P with $P = 10$ ($M = 286$). (b) Full tensor polynomial space Z_d^P with $P = 5$ ($M = 216$).

there will be no recovery at all. A large portion of successful recovery can be achieved only if the number of samples is sufficiently big. And this holds true for all four variants of the implementations and in both polynomial spaces.

We also observe that the best success rate is achieved by the *pre-Chebyshev*, i.e., the preconditioned ℓ_1-minimization of Theorem 5. And the least effective approach appears to be the *direct-uniform*, i.e., the direct ℓ_1-minimization of Theorem 3. This is consistent with the earlier studies of [20, 21], where the Chebyshev preconditioned ℓ_1-minimization is shown to be advantageous in $d = 1$. Our current results indicate that its advantage holds true in low dimensions. The other two variants, *pre-uniform* and *direct-Chebyshev*, produce better results than the direct ℓ_1-minimization. This suggests that the introduction of the Chebyshev measure, either via the sampling distribution or the preconditioning matrix, can help the performance. And clearly the Chebyshev preconditioned ℓ_1-minimization (*pre-Chebyshev*) produces the best results. In fact, in order to achieve near-one success probability, the required number of samples for the *pre-Chebyshev*, the preconditioned ℓ_1-minimization, is nearly half of that of the *direct-uniform*—the direct ℓ_1-minimization.

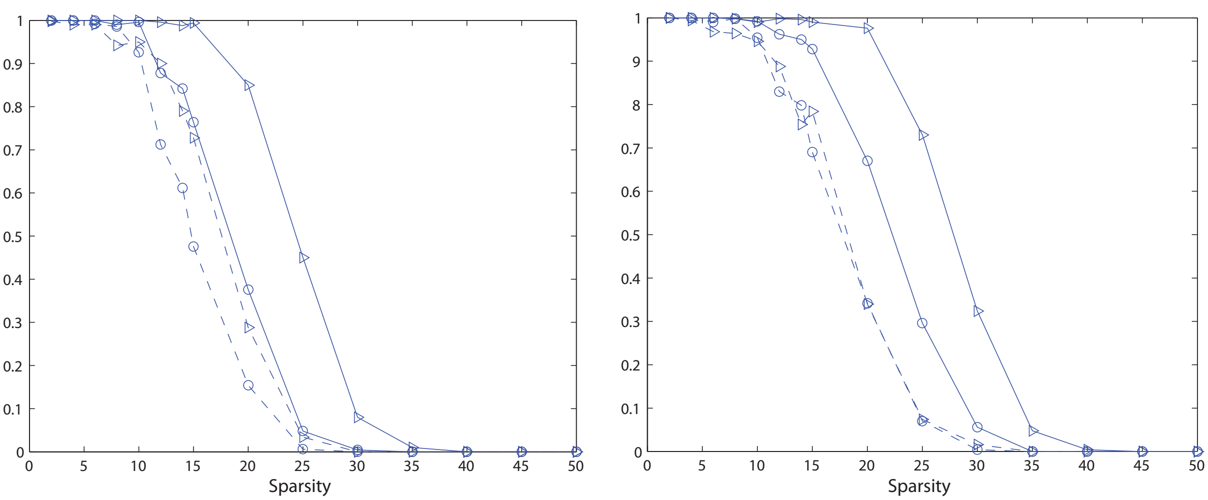
Next we examine the errors in the reconstructed polynomial interpolation. The results are plotted in Fig. 2. We observe the same trend for all four implementations—the errors decay as the number of samples is increased. While this is consistent with the theoretical prediction, we again observe that the preconditioned ℓ_1-minimization, the *pre-Chebyshev*, produces superior results than the direct ℓ_1-minimization, the *direct-uniform*. The two variants, the *pre-uniform* and *direct-Chebyshev*, offer some improvements over the *direct-uniform*. These results are consistent with those in Fig. 1. From these we conclude that in low dimensions the preconditioned ℓ_1-minimization should be favored in practice, similar to the established results in one dimension ($d = 1$) [20, 21].

We now consider the effect of the sparsity s of the target function on the performance of the algorithms. In Fig. 3, the probability of successful recovery is plotted against increasing level of sparsity s, when the number of sample points is fixed at $N = 100$. It is obvious that with the fixed number of points, the algorithms can recover the target function only up to a limited sparsity level, beyond which no recovery can be achieved. While this is certainly as expected, we observe that once again the preconditioned ℓ_1-minimization offers the most effective results. A similar trend is also observed in Fig. 4, where the reconstruction errors are plotted against increasing level of sparsity s. While all implementations become less accurate and eventually lose accuracy at larger values of s, the preconditioned version offers the most accurate results. The two variants, the *pre-uniform* and the *direct-Chebyshev*, are again slightly better than the direct ℓ_1-minimization.



(a) (b)

FIG. 2: Reconstruction error *vs.* number of sample points $(d = 3$ and $s = 10)$. Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*. (a) Total degree polynomial space W_d^P with $P = 10$ $(M = 286)$. (b) Full tensor polynomial space Z_d^P with $P = 5$ $(M = 216)$.



(a) (b)

FIG. 3: Probability of successful recovery *vs.* sparsity s $(d = 3$ and $N = 100)$. Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*. (a) Total degree polynomial space W_d^P with $P = 10$ $(M = 286)$. (b) Full tensor polynomial space Z_d^P with $P = 5$ $(M = 216)$.

4.2 High-dimensional Tests in $d = 10$

We now provide a set of tests similar to those in the previous section but in a higher dimension. In particular, we present the results in $d = 10$. Though high dimensional, $d = 10$ is certainly not exceedingly large. It is chosen mostly for demonstration purpose, for it allows us to use reasonably high degree (P) for the basis polynomials. In this case ($d = 10$), the full tensor space Z_P^d is not considered because its cardinality is too high to generate the interpolation matrix with reasonable size. Therefore we focus on the total degree polynomial space W_P^d (4).

For $d = 10$, the cardinality (M) of the space is shown in Table 1, for degrees up to five. Despite the rapid growth of the cardinality, another practical concern is the "gap" between degrees. That is, in order to construct a polynomial

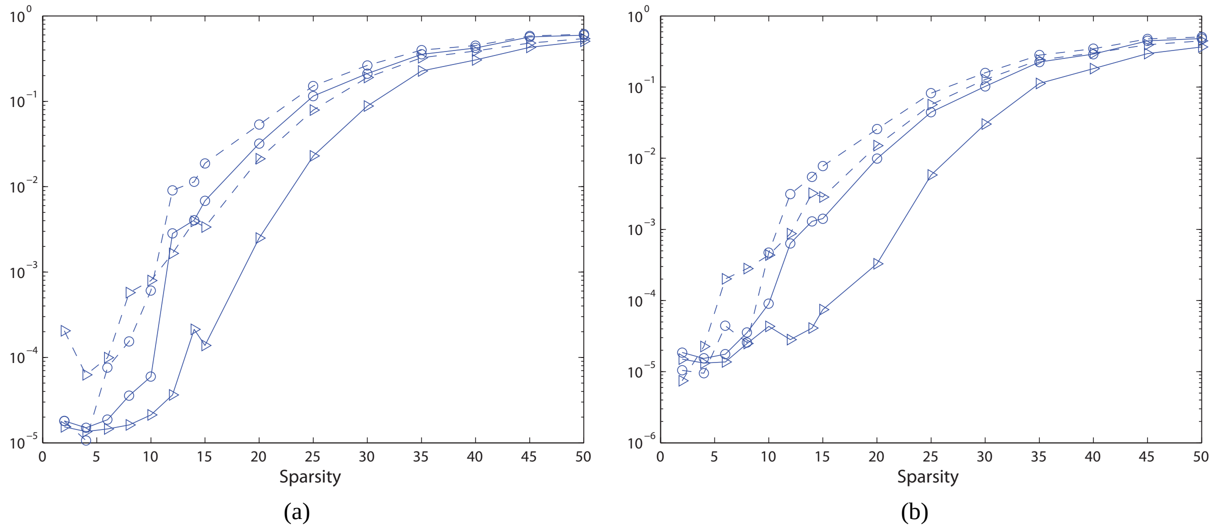


FIG. 4: Reconstruction error *vs.* sparsity s $(d = 3$ and $N = 100)$. Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*. (a) Total degree polynomial space W_d^P with $P = 10$ $(M = 286)$. (b) Full tensor polynomial space Z_d^P with $P = 5$ $(M = 216)$.

TABLE 1: The cardinality (M) of total-degree polynomial space W_P^d (4) at $d = 10$

Degree P	1	2	3	4	5
Cardinality M	11	66	286	1001	3003

interpolation of certain degree, the number of points must be at least equal to the corresponding cardinality. For example, if one is able to produce $N = 100$ samples in a particular application, then the best approximation via the traditional approaches is of second degree, which requires $M \geq 66$ samples. The "extra" number of samples ($N > M$ in this case) will usually produce at most marginal, if any, improvement over the second-degree approximation. The next notable improvement in the approximation, the third-degree approximation, would require at least $M = 286$ samples, a significant investment over the available $N = 100$ samples if the samples are expensive to produce. On the other hand, the ℓ_1-minimization is not restricted to this constraint. Since it works with arbitrary number of samples, any additional samples could, in principle, produce progressively better approximation.

We first fix the sparsity level s and then randomly generate a coefficient vector $\widetilde{\mathbf{c}}$ which in turn determines the target function. The four variations of the ℓ_1-minimization are applied to recover the target function. The degree of the polynomial space is fixed at $P = 4$, whose cardinality is $M = 1001$. Note the tests now belong to the high-dimensional case of $d > P$ discussed in Section 3.2. In Fig. 5 we plot the errors in the reconstruction versus the number of sample points, with a fixed sparsity level of $s = 10$. All four variants of the implementation demonstrate same trend—errors decay with increasing number of samples. A notable error reduction occurs at around $N \sim 100$, which corresponds to the required number of samples to produce successful recovery with large probability. These are consistent with the theoretical findings. It also can be seen that the two dotted curves, both computed by the direct ℓ_1-minimization algorithm (15), produce notably more accurate results than the two solid curves, which are computed by the preconditioned ℓ_1-minimization algorithm (35). This is *drastically different* from the previous low-dimensional tests, where the preconditioned algorithm (35) is more accurate.

More tests are conducted and the results at $s = 20$ are shown in Fig. 6. A similar trend can be seen here, with the notable error reduction occurring at larger number of samples. This is reasonable because the sparsity level is higher in this case. Once again, the direct ℓ_1-minimization algorithm (15) produces notably more accurate results than the preconditioned algorithm (35). We also remark that the choice of sampling distribution, uniform or Chebyshev, does not produce sufficient differences in the results.

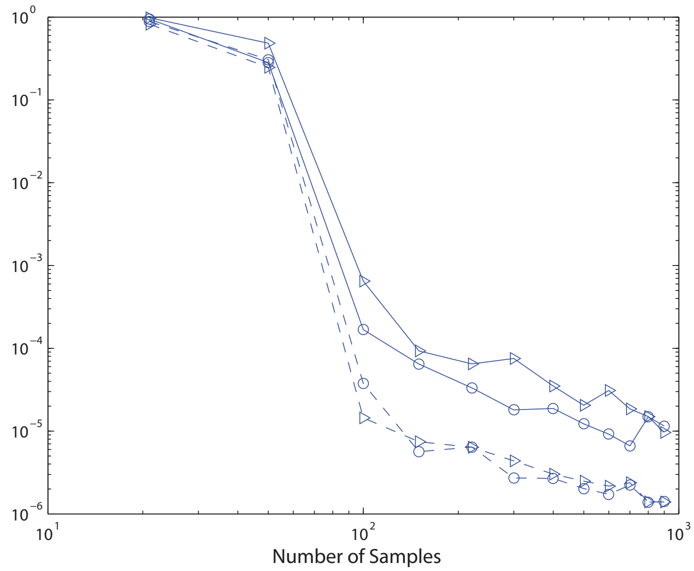


FIG. 5: Reconstruction error *vs.* number of sample points $(d = 10$ and $s = 10)$. Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*.

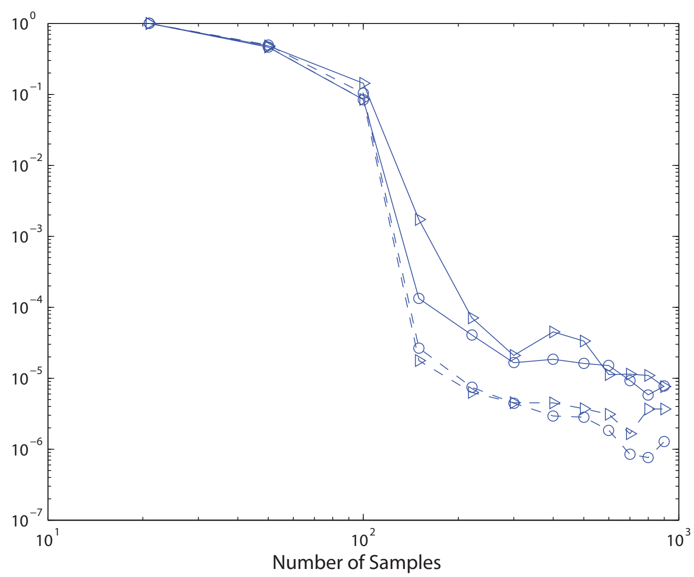


FIG. 6: Reconstruction error *vs.* number of sample points $(d = 10$ and $s = 20)$. Line patterns: dotted-circle, *direct-uniform*; dotted-triangle, *direct-Chebyshev*; solid-circle, *pre-uniform*; solid-triangle, *pre-Chebyshev*.

We therefore conclude that in the high-dimensional cases of $d \geq P$ the direct ℓ_1-minimization (15) should be preferred over the preconditioned algorithm (35). This is also consistent with the theoretical findings. Note that for the preconditioned ℓ_1-minimization, the required number of samples scales as 2^d from (36). And the estimate holds true for all dimensionality. On the other hand, for the direct ℓ_1-minimization, the required number of samples scales as 3^P from (23) when $d \geq P$. In the test conducted here, $d = 10$ and $P = 4$. It is clear that $3^P < 2^d$ and consequently the direct algorithm requires fewer samples for successful recovery. This explains its better accuracy than the preconditioned algorithm with the same number of samples.

5. SUMMARY

In this paper we study the ℓ_1-minimization method for stochastic collocation in high-dimensional random space. In particular we focus on a polynomial interpolation type approach using Legendre polynomials, which is a topic of few studies. We derive the recoverability of the ℓ_1-minimization of both direct minimization and Chebyshev preconditioned minimization. The results are largely extensions of the general theories on ℓ_1-minimization and some existing ones on the Legendre approach in one dimension. Our results establish the validity of the approach in high-dimensional space. Moreover, we demonstrate that in low dimensions, the Chebyshev preconditioned ℓ_1-minimization is more efficient than the direct ℓ_1-minimization, consistent with the existing studies in one dimension. On the other hand, in high dimensions, the opposite result holds true—the direct algorithm becomes more efficient than the Chebyshev preconditioned algorithm. Extensive numerical tests verify these theoretical findings. Note that the current work focuses exclusively on the polynomial recoverability of the method. Our ongoing work is to extend the results to approximation of general stochastic functions in high-dimensional random spaces.

ACKNOWLEDGMENT

This work is in part supported by AFOSR, DOE/NNSA, NSF.

REFERENCES

1. Ghanem, R. and Spanos, P., *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, 1991.
2. Xiu, D. and Karniadakis, G., The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.
3. Xiu, D. and Hesthaven, J., High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.
4. Agarwal, N. and Aluru, N., A domain adaptive stochastic collocation approach for analysis of mems under uncertainties, *J. Comput. Phys.*, 228(20):7662–7688, 2009.
5. Babuska, I., Nobile, F., and Tempone, R., A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.
6. Bieri, M., A sparse composite collocation finite element method for elliptic sPDEs, Tech. Rep. 2009-08, Seminar for Applied Mathematics, ETHZ, 2009.
7. Bieri, M. and Schwab, C., Sparse high order FEM for elliptic sPDEs, *Comput. Meth. Appl. Math. Eng.*, 198:1149–1170, 2009.
8. Constantine, P., Doostan, A., and Iaccarino, G., A hybrid collocation/Galerkin scheme for convective heat transfer problems with stochastic boundary conditions, *Int. J. Numer. Methods Eng.*, 80:868–880, 2009.
9. Foo, J., Wan, X., and Karniadakis, G., The multi-element probabilistic collocation method (ME-PCM): Error analysis and applications, *J. Comput. Phys.*, 227(22):9572–9595, 2008.
10. Ma, X. and Zabaras, N., An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations, *J. Comput. Phys.*, 228:3084–3113, 2009.
11. Nobile, F., Tempone, R., and Webster, C., An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data, *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.
12. Nobile, F., Tempone, R., and Webster, C., A sparse grid stochastic collocation method for partial differential equations with random input data, *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.
13. Xiu, D., Efficient collocational approach for parametric uncertainty analysis, *Comm. Comput. Phys.*, 2(2):293–309, 2007.
14. Ganapathysubramanian, B. and Zabaras, N., Sparse grid collocation methods for stochastic natural convection problems, *J. Comput. Phys.*, 225(1):652–685, 2007.
15. Candès, E. and Tao, T., Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
16. Candès, E., Romberg, J., and Tao, T., Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

17. Candès, E. and Tao, T., Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
18. Donoho, D. L., Compressed sensing, *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
19. Doostan, A. and Owhadi, H., A non-adapted sparse approximation of PDEs with stochastic inputs, *J. Comput. Phys.*, 230(8):3015–3034, 2011.
20. Rauhut, H., Compressive sensing and structured random matrices, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, Fornasier, M. (Ed.), pp. 1–92. Berlin, New York (DE GRUYTER), 2010.
21. Rauhut, H. and Ward, R., Sparse legendre expansions via ℓ_1-minimization, Submitted, 2010.
22. A.Doostan, and Iaccarino, G., A least-squares approximation of partial differential equations with high-dimensional random inputs, *J. Comput. Phys.*, 228(12):4332–4345, 2009.
23. Narayan, A. and Xiu, D., Stochastic collocation methods on unstructured grids in high dimensions via interpolation, *SIAM J. Sci. Comput.*, submitted, 2011.
24. Xiu, D., *Numerical methods for stochastic computations*, Princeton Univeristy Press, Princeton, New Jersey, 2010.
25. Cai, T., Wang, L., and Xu, G., New bounds for restricted isometry constants, *IEEE Trans. Infor. Theory*, 56(9):4388–4394, 2010.
26. Szegö, G., *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1939.
27. van den Berg, E. and Friedlander, M., Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.*, 31(2):890–912, 2008.
28. van den Berg, E. and Friedlander, M., SPGL1: A solver for large-scale sparse reconstruction, http://www.cs.ubc.ca/labs/scl/spgl1, 2007.