RESEARCH ARTICLE

Analysis of sparse recovery for Legendre expansions using envelope bound

Hoang Tran1 | Clayton Webster2,3

1Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
2Oden Institute for Computational Engineering & Sciences, The University of Texas at Austin, Austin, Texas, USA
3Behavioral Reinforcement Learning Lab, Lirio, LLC, Knoxville, Tennessee, USA

Correspondence
Hoang Tran, Computer Science and Mathematics, Division, Oak Ridge National Laboratory, PO Box 2008, MS6013 Oak Ridge, TN 37831-6013, USA. Email: tranha@ornl.gov

Funding information
Office of Science, Grant/Award Number: ERKJ314, ERKJ331, ERKJ345, AEOLUS, FASTMath Instit; U.S. Department of Energy, Grant/Award Number: DE-AC05-00OR22725; Laboratory Directed Research and Development program at the Oak Ridge National Laboratory; FASTMath Institute, Grant/Award Number: DE-AC02-05CH11231; Scientific Discovery through Advanced Computing (SciDAC); Department of Energy Mathematical Multifaceted Capabilities Center; AEOLUS (Advances in Experimental Design, Optimization and Learning for Uncertain Complex Systems); U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program, Grant/Award Numbers: ERKJ345, ERKJ331, ERKJ314.

Abstract
We provide novel sufficient conditions for the uniform recovery of sparse Legendre expansions using ℓ1 minimization, where the sampling points are drawn according to orthogonalization (uniform) measure. So far, conditions of the form $m \gtrsim \Theta^2 s \times \log$ factors have been relied on to determine the minimum number of samples $m$ that guarantees successful reconstruction of $s$-sparse vectors when the measurement matrix is associated to an orthonormal system. However, in case of sparse Legendre expansions, the uniform bound $\Theta$ of Legendre systems is so high that these conditions are unable to provide meaningful guarantees. In this paper, we present an analysis which employs the envelop bound of all Legendre polynomials instead, and prove a new recovery guarantee for $s$-sparse Legendre expansions, $m \gtrsim s^2 \times \log$ factors, which is independent of $\Theta$. Arguably, this is the first recovery condition established for orthonormal systems without assuming the uniform boundedness of the sampling matrix. The key ingredient of our analysis is an extension of chaining arguments, recently developed in Bourgain and Chkifa et al., to handle the envelop bound. Furthermore, our recovery condition is proved via restricted eigenvalue property, a less demanding replacement of restricted isometry property which is perfectly suited to the considered scenario. Along the way, we derive simple criteria to detect good sample sets. Our numerical tests show that sets of uniformly sampled points...
that meet these criteria will perform better recovery on average.

KEYWORDS
approximation theory, compressed sensing, Legendre expansions

1 INTRODUCTION

Compressed sensing (CS), originating as a signal processing technique [3, 4], has become an appealing approach for function approximation. Theory of CS indicates that if a function possesses a sparse representation in a known basis, it can be reconstructed from a limited number of suitable sampling points using nonlinear techniques such as convex optimization or greedy algorithms. In general, this number scales linearly with the sparsity level and only logarithmically with the size of representation, being much smaller than those required by traditional methods such as projection and interpolation. As the solutions of many parameterized partial differential equations (PDEs) are known to be sparse in the sense that they are well approximated by a sparse expansion in an orthonormal system such as Legendre or Chebyshev polynomials, it is no surprise that the interest in applying CS techniques to the polynomial approximation of high-dimensional functions and parameterized PDEs has been growing rapidly in recent years, see [2, 5–13] and the reference therein.

Let us consider a function \( g : \mathcal{U} \rightarrow \mathbb{R} \) defined on the domain \( \mathcal{U} \subset \mathbb{R}^d \), endowed with a probability measure \( \varrho \). We assume \( g \) is well represented by the finite expansion

\[
g(y) \approx \sum_{j \in J} c_j \Psi_j(y), \quad y \in \mathcal{U},
\]

where \( \{ \Psi_j \}_{j \in J} \) is a pre-determined orthonormal system associated with \( \varrho \) and indexed by a finite set \( J \). We define the corresponding polynomial subspace \( P_J := \text{span} \{ \Psi_j(y) : j \in J \} \). In CS approach, one aims to recover \( g \) by reconstructing unknowns \( \{ c_j \}_{j \in J} \) from \( m \) samples \( g(y_1), \ldots, g(y_m) \), which are drawn independently from the orthogonalization measure \( \varrho \). Denote by \( N \) the cardinality of \( J \), that is, \( N = \#(J) \). \( A \) the normalized sampling matrix and \( g \) the normalized observation of the target function, that is,

\[
A := \left( \frac{\Psi_j(y_i)}{\sqrt{m}} \right)_{1 \leq i \leq m \atop j \in J} \in \mathbb{R}^{m \times N}, \quad \text{and} \quad g := \left( \frac{g(y_i)}{\sqrt{m}} \right)_{1 \leq i \leq m} \in \mathbb{R}^m,
\]

this task amounts to solving for the coefficient vector \( c = (c_j)_{j \in J} \) that satisfies \( Ac = g \) (or, to be exact, \( ||Ac - g|| \leq \eta \) where \( \eta \) is the truncation error accounting for the representation of \( g \) by a finite series). Since the samples \( g(y_i) \) are costly to be acquired in many applications, it is desirable to reconstruct \( c \) with as small number of samples as possible, in particular, smaller than \( N \). Although the system \( Az = g \) becomes underdetermined, this is possible assuming that the expansion (1) is sparse, in which case \( c \) can be reconstructed by several efficient algorithms, for instance, solving the convex optimization problem promoting sparsity

\[
\arg\min_{z \in \mathbb{R}^N} ||z||_1 \quad \text{subject to} \quad Az = g.
\]
The sparse recovery for the orthonormal expansions via $\ell_1$ minimization has shown to be very promising. It is supported by well-known theoretical estimates [1, 2, 14–16] that $s$-sparse expansions\footnote{Expansions of the form (1) that have only $s$ nonzero coefficients} can be accurately recovered given the number of sampling points satisfying

$$m \geq \Theta^2 s \times \log \text{ factors},$$

where $\Theta$ is the uniform bound of the underlying basis: $\Theta = \sup_{j \in J} \|\Psi_j\|_{L^\infty(U)}$. For many orthonormal systems of interest such as trigonometric polynomials, Chebyshev polynomials, and Fourier transforms, $\Theta$ is a small constant, implying $m$ grows linearly with the sparsity level and at most logarithmically with the size of the expansion. These are indeed the favorable settings for sparse recovery. However, the case of Legendre expansions has proven to be problematic. There, the uniform bound $\Theta$ is large and $m$ determined as in (4) may be prohibitive. It is widely agreed that reconstructing directly from Legendre system and its orthogonalization (uniform) measure is not a good strategy; and several advanced techniques have been developed over the years to deal with this challenge, most notably, preconditioning approach and weighted $\ell_1$ minimization. The former method multiplies a suitable weight function to Legendre polynomials to form a preconditioned orthonormal system with significantly reduced $\Theta$ [9, 10, 17], while the latter takes advantage of small uniform bounds of low-order Legendre polynomials to enhance the recovery of sparse vectors with some structures [2, 12, 18].

That being said, it can be argued that the condition (4) is too pessimistic for sparse Legendre expansions, and uniform bound should not be entirely relied on as the single tool to determine the number of samples. To illustrate this, let us conduct a simple test on the reconstruction of sparse Legendre polynomials in one-dimensional domain, $U = [-1, 1]$, via $\ell_1$ minimization (3). In this example, $\Psi_j \equiv L_j$ (the univariate Legendre polynomial of order $j$), $A$ is the sampling matrix associated with $L_j$ as in (2), where $y_1, \ldots, y_m$ are drawn according to uniform distribution in $U$. As $\|L_j\|_{L^\infty(U)} = \sqrt{2j+1}$ and $\Theta \geq \|L_{N-1}\|_{L^\infty(U)}$, condition (4) yields

$$m \geq Ns \times \log \text{ factors},$$

a trivial estimate since the required number of samples is now greater than $N$. In other words, there is no sound theory for the sparse recovery from underdetermined Legendre systems with uniform sampling points. Further, it is worth remarking that $m$ actually depends linearly on the maximal degree of the sparse expansions, which can only be greater than, if not equal to, $N$. To test the sharpness of condition (4), we consider three different cases of $J$:

i. $J = \{j \in \mathbb{N} : 1 \leq j \leq 200\},$

ii. $J = \{j \in \mathbb{N} : 301 \leq j \leq 500\},$

iii. $J = \{j \in \mathbb{N} : 1801 \leq j \leq 2000\}.$

In all these cases, the size of the finite Legendre expansions is fixed at 200; however, the maximal degree is varied from 200 (case i.), 500 (case ii.), to 2000 (case iii.). We aim to reconstruct sparse vectors $c$ of coefficients, with sparsity ranging from 5 to 40, given noiseless observations $g = Ac$. For each sparsity level, we run 100 trials, in each of which the support of $c$ is selected randomly and its coefficients are drawn from a Gaussian distribution. We fix the number of sample points $m = 100$ throughout the test, but generate different random set of samples $\{y_i\}_{1 \leq i \leq m}$ for each trial. The numerical results on averaged $\ell^2$ errors and successful recovery rates are presented in Figure 1. Two important observations can be made. First, sparse vectors with sparsity 15 or less are accurately reconstructed, indicating...
that it is possible to recover sparse vectors directly from underdetermined Legendre systems. Second, the required number of samples should not depend on the maximal degree of Legendre expansions, evidenced by the fact that the errors and successful rates of sparse \( \ell_1 \) recovery are similar for three choices of \( J \) (even slightly worse for case i. where the maximum degree is smaller). This simple test suggests that there exists a gap between the performance of sparse Legendre recovery in practice and existing theory, in particular, condition (4).

In this paper, we establish novel estimates on the sufficient number of samples for the sparse recovery of Legendre expansions using uniform sampling and \( \ell_1 \) minimization. In one-dimensional case, we prove that the uniform recovery of all \( s \)-sparse Legendre polynomials is guaranteed with high probability provided that

\[
m \gtrsim s^2 \log^2(s) \log^2(N).
\]

We also prove the uniform recovery condition for the \( d \)-dimensional case, that is, \( U = [-1, 1]^d \),

\[
\frac{m}{\log^{4d-4}(m)} \gtrsim C_d s^2 \log^2(s) \log^2(N),
\]

where \( C_d \) is a universal constant depending only on the dimension. Note that here \( N \) is the size of the polynomial expansion, which is not necessarily related to its maximal degree. The aforementioned estimates reveal that

1. The number of samples grows quadratically in sparsity level, but only depends weakly on the size of expansions via log factors. Therefore, sparse recovery from underdetermined Legendre systems is provably possible,
2. The number of samples is completely independent on the uniform bound \( \Theta \) or maximal degree of Legendre expansions.

As such, they provide a mathematical justification for what we observed in the preliminary test. More importantly, these results bring about the first rigorous theory for \( \ell_1 \) recovery of Legendre approximations with uniform sampling.

We remark that the conditions (5)–(6) apply not only for exactly sparse and noiseless setting, but also for reconstruction of compressible unknown vectors (i.e., well-approximated by sparse vectors) from noisy measurements. While the quadratic growth of number of samples herein seems to resemble...
other conditions in discrete least square approximations [19] and CS using coherence-based arguments [15]; section 5.4 there are fundamental differences between those results with ours. First, in least square approximations, \( m \) scales quadratically with \( N \) (instead of \( s \)), which is much simpler to derive. On the other hand, the quadratic dependence of \( m \) on the sparsity using coherence-based arguments only holds for a very small class of random matrices called equiangular tight frames, which does not contain most orthonormal systems of interest, including the Legendre systems considered here. Finally, there were efforts for establishing recovery guarantees for Legendre systems with uniform sampling, see [9]. This work however assumes fixed polynomial subspace and relies on the uniform bound \( \Theta \); thus, while the derived condition features a linear dependence of \( m \) on \( s \), the associated constant also grows exponentially with respect to the maximal degree of polynomial subspaces.

Our results are established via certain versions of null space property and incoherence condition of the sampling matrix. To guarantee successful reconstruction, it is sufficient for the Legendre matrix to fulfill these conditions. Restricted isometry property (RIP) has typically been used to obtain the uniform recovery. Herein, however, we do not assume a fixed, known maximal degree for the polynomial subspace \( P_J \). Without this boundedness, we suspect that it is very challenging, if possible at all, to prove the eigenvalues of “restricted” matrices are bounded from above as required by RIP. To overcome this difficulty, we instead establish the uniform recovery via a variant of restricted eigenvalue property [20–22], which is a related condition but less demanding than RIP. In fact, restricted eigenvalue property only assumes some kind of positive definiteness of the sampling matrix, and many important matrices, such as correlated Gaussian and sub-Gaussian designs, have been shown to satisfy the restricted eigenvalue property while violate the RIP [23, 24]. As we shall see, null space property can be implied directly from restricted eigenvalue property. A major part of our paper therefore is devoted to deriving precise bounds on the number of samples such that the restricted eigenvalue property is fulfilled.

For this goal, our analysis employs chaining techniques, which provide probabilistic bounds for stochastic processes via careful constructions of progressively finer nets upon which union bounds can be applied in an efficient manner. This approach has been used to establish the RIP estimates of several bounded orthonormal systems [1, 2, 16]. However, unlike all previous works on the topic, we do not rely on the uniform bound, but rather apply and combine the envelope (i.e., pointwise) bound of Legendre systems to the chaining process. As illustrated in Figure 2, envelope bound is more efficient than uniform bound and invariant to the selection of polynomial subspaces. Thus, our new arguments using envelope bound lead to the improved estimates (5)–(6), which do not depend on the polynomial subspace. Independently of the main results, we believe that this approach should be of interest on its own, and can be extended to other random matrices which can be bounded precisely pointwise, but whose uniform bound is bad. As a by-product, we derive some simple criteria to predict good (preferable) sample sets. We conduct numerical tests showing that sets of uniformly sampled points that meet these criteria will perform better recovery on average.

Our paper is organized as follows. In Section 2, we review the restricted eigenvalue property and its relation to null space property and exact recovery guarantees. In Section 3, we prove a new bound for restricted eigenvalue property of subsampled one-dimensional Legendre matrices. The restricted eigenvalue property estimate for multi-dimensional Legendre matrices will be discussed in Section 4. Section 5 provides numerical illustrations on our criteria to select sample sets. Concluding remarks are given in Section 6. For the rest of this paper, \( C \) represents a universal constant, whose value may change from place to place but is independent of any parameter.
In this section, we present an overview of restricted eigenvalue property and the implication of null space property and uniform recovery from restricted eigenvalue property. The discussion in this section does not assume $A$ being a subsampled Legendre system, but is also applicable for a generic sampling matrix. First, the definition of restricted eigenvalue property is provided below.

Let $z \in \mathbb{R}^N$ and be indexed in $J$. For $S \subset J$, we denote by $S^c$ the complement of $S$ in $J$ and $z_S$ the restriction of $z$ to $S$. For $\alpha > 1$ and $0 < s \leq N$, define

$$C(S; \alpha) := \left\{ z \in \mathbb{R}^N : \|z_{S^c}\|_1 \leq \alpha \sqrt{s} \|z_S\|_2 \right\},$$

and $C(s; \alpha) := \bigcup_{S \subset J, \#(S) = s} C(S; \alpha)$. \hfill (7)

Let $B_1$ and $\partial B_1$ be the unit $\ell_1$-ball and sphere in $\mathbb{R}^N$ correspondingly

$$B_1 = \left\{ z \in \mathbb{R}^N : \|z\|_1 \leq 1 \right\}, \quad \text{and} \quad \partial B_1 = \left\{ z \in \mathbb{R}^N : \|z\|_1 = 1 \right\}.$$  

Similarly, denote by $B_2$ and $\partial B_2$ the unit $\ell_2$-ball and sphere in $\mathbb{R}^N$.

**Definition 1** (Restricted eigenvalue property). Given $A \in \mathbb{R}^{m \times N}$, $\alpha > 1$ and $0 < \delta < 1$, we say $A$ satisfies the restricted eigenvalue property of order $s$ with parameters $(\alpha, \delta)$ if

$$\|Az\|_2^2 \geq (1 - \delta)\|z\|_2^2, \quad \text{for all} \quad z \in C(s; \alpha),$$

or equivalently, $\|Az\|_2^2 \geq 1 - \delta, \quad \text{for all} \quad z \in \partial B_2 \cap C(s; \alpha)$. \hfill (8)

We note that the restricted eigenvalue property stated above is slightly different from those often found in the literature, for example, [20, 23, 24]. In particular, the inequality $\|z_S\|_1 \leq \sqrt{s} \|z_S\|_2$ yields that the set $C(s; \alpha)$ defined in (7) is slightly larger than

$$\bigcup_{S \subset J, \#(S) = s} \left\{ z \in \mathbb{R}^N : \|z_S\|_1 \leq \alpha \|z_S\|_1 \right\}$$

as considered in those works.

The theory of restricted eigenvalue property, as a sufficient condition for the success of $\ell_1$ minimization methods, is well-developed in the fields of statistical inference and machine learning, where
the RIP was proven to be severe for many random matrices of interest, for example, random designs with substantial correlation among rows/columns. In this paper, we develop recovery guarantees for a class of random matrices associated with Legendre orthonormal systems. As we do not assume the uniform boundedness, the fact that restricted eigenvalue property does not require an upper bound of \( \|Az\|_2 \) is also extremely useful in our analysis.

Next, we will present the implication of null space property and uniform recovery from restricted eigenvalue property. We also include discussions on estimates for restricted eigenvalue property of orthonormal systems in case uniform bound is used. These derivations are quite simple and based on the corresponding developments for the RIP. However, as far as we know, they are not very well-spread in the CS literature.

**Definition 2** (Restricted isometry property). Given a matrix \( A \in \mathbb{R}^{m \times N} \), we say \( A \) satisfies the restricted isometry property of order \( s \) if there exists \( 0 < \delta < 1 \) such that

\[
(1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta)\|z\|_2^2, \quad \forall z \in \mathbb{R}^N \text{ satisfying } \#(\text{supp}(z)) \leq s,
\]

or equivalently,

\[
1 - \delta \leq \|Az\|_2^2 \leq 1 + \delta, \quad \forall z \in \mathcal{E}_s,
\]

where \( \mathcal{E}_s := \{ z \in \mathbb{R}^N : \#(\text{supp}(z)) \leq s, \|z\|_2 = 1 \} \).

Compared to restricted eigenvalue property, RIP needs \( \|Az\|_2 \) to be bounded from both above and below, but only on the set of exactly sparse vectors in \( \mathbb{R}^N \). In particular, it is easy to see that \( \mathcal{E}_s \) is a strict subset of \( \partial B_2 \cap C(s; a) \) required in (8). However, in available RIP estimates, (9) is typically established for a larger set, that is, \( \sqrt{s}B_1 \), the \( \ell_1 \)-ball that contains \( \mathcal{E}_s \) [1, 2, 14, 25, 26].

Observe that \( \partial B_2 \cap C(s; a) \) is contained in the \( \ell_1 \)-ball of radius \( (1 + a)\sqrt{s} \).

**Lemma 1** For \( a > 1 \), \( 0 \leq s \leq N \), there follows \( \partial B_2 \cap C(s; a) \subset (1 + a)\sqrt{s}B_1 \).

**Proof.** Let \( z \in \partial B_2 \cap C(s; a) \). Then, \( \|z_S\|_1 \leq a\sqrt{s} \|z_S\|_2 \) for some \( S \subset J \), \( \#(S) = s \).

We have \( \|z\|_2 = \|z_S\|_1 + \|z_{\bar{S}}\|_1 \leq \sqrt{s} \|z_S\|_2 + a\sqrt{s} \|z_{\bar{S}}\|_2 = (1 + a)\sqrt{s} \|z\|_2 \), therefore, \( z \in (1 + a)\sqrt{s}B_1 \).

One can easily derive from available RIP proofs that for orthonormal systems with uniform bound \( \Theta \), RIP holds over the extended set \( (1 + a)\sqrt{s}B_1 \) under a modified number of samples where the sparsity \( s \) is replaced by \( (1 + a)^2 s \). This property automatically implies the restricted eigenvalue property as a consequence of Lemma 1 and therefore, the restricted eigenvalue property holds under the same sample complexity.

For a specific example, we can obtain the following result by slightly modifying [2] (theorem 2.2.)

**Proposition 1** (Restricted eigenvalue estimate for bounded orthonormal systems). Let \( \delta, \gamma \) be fixed parameters with \( 0 < \delta < 1, 0 < \gamma < 1 \) and \( \{ \psi_j \}_{j \in J} \) be an orthonormal system of finite size \( N = \#(J) \) and \( \Theta = \sup_{j \in J} \|\psi_j\|_{\ell_\infty} \). Let \( s_\alpha = (1 + \alpha)^2 s \). Assume that

\[
m \geq \frac{C}{\delta} \log \left( \frac{\Theta s}{\delta} \right) \max \left\{ \frac{\gamma}{\delta} \log \left( \frac{4 \Theta s}{\delta} \right), \frac{1}{\gamma} \log \left( \frac{\Theta s}{\delta} \right) \right\},
\]

and \( y_1, y_2, \ldots, y_m \) are drawn independently from the orthogonalization measure \( \mu \) associated to \( \{ \psi_j \} \). Then with probability exceeding \( 1 - \gamma \), the normalized sampling matrix
\( A \in \mathbb{R}^{m\times N} \) (defined as in (2)) satisfies

\[
\|Az\|_2^2 > (1 - \delta)\|z\|_2^2, \quad \forall z \in C(s; \alpha).
\]

Interestingly, restricted eigenvalue property of order \( s \) straightforwardly leads to null space property of order \( s \).²

**Proposition 2** Assume that \( A \in \mathbb{R}^{m\times N} \) satisfies the restricted eigenvalue property of order \( s \) with parameters \((\alpha, \delta)\). Then, \( A \) satisfies the null space property of order \( s \), that is, for all \( z \in \mathbb{R}^N \) and all \( S \subset J \) with \( |S| = s \),

\[
\|z_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|z_{\bar{S}}\|_1 + \tau\|Az\|_2, \quad \text{where} \quad \rho = \frac{1}{\alpha} \quad \text{and} \quad \tau = \frac{1}{\sqrt{1 - \delta}}.
\]  

**Proof.** For any \( z \in \mathbb{R}^N \), we only need to prove (10) for \( S \) chosen as the index set corresponding to \( s \) largest components of \( z \). Consider two cases:

If \( z \in C(S; \alpha) \), by (8), we have

\[
\|z_S\|_2 \leq \|z\|_2 \leq \frac{1}{\sqrt{1 - \delta}} \|Az\|_2 \leq \frac{\rho}{\sqrt{s}} \|z_S\|_1 + \tau\|Az\|_2.
\]

Otherwise, \( z \not\in C(S; \alpha) \) yields

\[
\|z_S\|_2 \leq \frac{1}{\alpha\sqrt{s}} \|z_{\bar{S}}\|_1 \leq \frac{\rho}{\sqrt{s}} \|z_S\|_1 + \tau\|Az\|_2.
\]

Recall \( c \) and \( g \) are the vectors of exact coefficients and observations such that \( \|Ac - g\|_2 \leq \eta \). Let \( c^\# \) be an approximation of \( c \) determined by

\[
e^\# = \arg\min_{\tilde{z} \in \mathbb{R}^s} \|\tilde{z}\|_1 \quad \text{subject to} \quad \|Az - g\|_2 \leq \eta.
\]  

Denote by \( \sigma_s(c) \), the error of best \( s \)-term approximation of \( c \) in \( \ell_1 \) norm, that is, \( \sigma_s(c)_1 = \inf_{A, J, |J|=s} \|c - c_J\|_1 \). The reconstruction error estimates follow from Proposition 2 using standard analysis, for example, [15] (theorems 4.19 and 4.22).

**Proposition 3** Let \( A \in \mathbb{R}^{m\times N} \), \( \alpha > 1 \), \( 0 < \delta < 1 \) and \( \delta_0 = \frac{1}{1+\alpha} < \frac{1}{2} \). Assume that \( A \) satisfies (10) for all \( z \in \mathbb{R}^N \) and all \( S \subset J \) with \( |S| = s \). For \( c \in \mathbb{R}^N \) and \( g \in \mathbb{R}^m \) satisfying \( \|Ac - g\|_2 \leq \eta \), let \( c^\# \) be the solution of (11). Then

\[
\|c - c^\#\|_1 \leq \frac{2}{1 - 2\delta_0} \sigma_s(c)_1 + \frac{4}{\sqrt{1 - \delta}} \cdot \frac{1 - \delta_0}{1 - 2\delta_0}\eta\sqrt{s}.
\]  

In combining Propositions 1–3, for bounded orthonormal systems, the robust and stable recovery (12) is guaranteed with high probability given the number of samples

\[
m \gtrsim \delta_0^{-2} C_\delta \Theta^2 s \times \log \text{ factors},
\]  

where \( C_\delta \) represents the dependence of \( m \) on \( \delta \). As we derive Proposition 1 from RIP estimates in [1, 2], \( C_\delta = \delta^{-6} \). However, better dependence on \( \delta \) is possible, for example, \( C_\delta = \delta^{-2} \) if applying the RIP estimates in [16, 26]. Also, the log factor terms can take several different forms. Generally, large \( \delta \) and

²Therefore, restricted eigenvalue property of order \( s \) is comparable to RIP of order \( 2s \).
\( \delta_0 \) are desirable for smaller number of samples, but this also lead to worse constants in reconstruction error estimates.

At first glance, the above number of samples established via restricted eigenvalue property seems to be more demanding than that obtained from the RIP \((m \gtrsim C_\delta \Theta^2 s \times \log factors)\) in the sense that besides \( \delta, m \) additionally depends on \( \delta_0 \). However, we can see that the dependence of error estimate constants on \( \delta \) is different for these two approaches. In RIP-based error estimates, \( \delta \) needs to be strictly bounded away from 1 and the error grows quite fast with moderately large \( \delta \), for example, with \( \delta < 4/\sqrt{41}, \) one has

\[
\|c - c^\#\|_1 \leq \frac{2 \left( \sqrt{1 - \delta^2} + 3\delta/4 \right)}{\sqrt{1 - \delta^2 - 5\delta/4}} \sigma_s(c) + \frac{4\sqrt{1 + \delta}}{\sqrt{1 - \delta^2 - 5\delta/4}} \cdot \eta \sqrt{s},
\]

see [15] (theorem 6.12); hence, an optimal dependence of \( m \) on \( \delta \) is critical. For the restricted eigenvalue property, as shown in (12), \( \delta \) appears only in the noise term, mildly affects the errors via the multiplication factor \( 1/\sqrt{1 - \delta} \) and can be chosen close to 1 without increasing this term much. As such, \( C_\delta \) becomes less important. A more profound penalty arises from \( \delta_0 \), which more or less plays the role of \( \delta \) in RIP approach. However, from (13), we have seen that the number of measurements \( m \) grows at rate \( \delta_0^{-2} \) as \( \delta_0 \) decreases. This agrees with the best known dependence of \( m \) on \( \delta \) using RIP.

As uniform recovery follows directly from restricted eigenvalue property (Propositions 2–3), the rest of this paper is devoted to derive precise bounds on the number of samples such that the restricted eigenvalue property holds for random Legendre systems. The advantage of restricted eigenvalue property in refraining us from the upper bound of \( \|A\|_2 \) is not clear in the above discussion, where the uniform bound of orthonormal systems is available. In the next part, we do not assume this information and exploit this advantage to its full strength.

3 | RESTRICTED EIGENVALUE PROPERTY FOR UNIVARIATE LEGENDRE SYSTEMS

We present an estimate for restricted eigenvalue property for one-dimensional Legendre basis. In this section, \( U := [-1, 1] \), \( \rho \) is the uniform probability measure on \( U \), \( J \) is a subset of \( \mathbb{N} \), and \( \{\Psi_j\}_{j \in J} \equiv \{L_j\}_{j \in J} \), the system of univariate Legendre polynomials orthonormal with respect to \( \rho \). We denote

\[
\psi(y, z) := \sum_{j \in J} z_j L_j(y), \quad y \in U, \quad z \in \mathbb{R}^N.
\]

The envelope bound of Legendre polynomials is central to our analysis. It is well known that

\[
|L_j(y)| \leq \frac{2/\sqrt{\pi}}{\sqrt{1 - y^2}}, \quad \forall y \in (-1, 1), \quad j \in \mathbb{N},
\]

see [12, 27]. This estimate is sharp in the sense that the constant \( 2/\sqrt{\pi} \) cannot be improved (see also an illustration in Figure 2). Let us define \( \Omega(y) := \frac{2/\sqrt{\pi}}{\sqrt{1 - y^2}} \).

We introduce our notion of “good sample sets”, which we refer to as preferable sets throughout.

**Definition 3** For \( m > 0, 0 < \gamma_0 < 1 \), let \( Q = \{y_1, \ldots, y_m\} \) be a set of samples where \( y_1, \ldots, y_m \) are drawn independently from the uniform distribution on \([-1, 1]\). Define \( m \)...
i.i.d random variables

\[ Z_i = \Omega \left( y_i \right) \exp \left( -\frac{1}{2\Omega^2 \left( y_i \right)} \sqrt{\frac{m}{y_0}} \right), \quad \forall i \in \{1, \ldots, m\}. \quad (15) \]

We name \( T(Q) = \sum_{i=1}^{m} Z_i \) the test value of \( Q \), and call \( Q \) a preferable sample set according to \( \gamma_0 \) if

\[ T(Q) \leq \frac{32\sqrt{2}}{\pi \sqrt{\pi}} m^{1/4} y_0^{3/4} + \frac{4\sqrt{2}}{\pi} m^{1/4} y_0^{-1/4}. \quad (16) \]

Loosely speaking, preferable set conceptualizes our preference to sample sets with which \( \sum_{i=1}^{m} Z_i \) is small. Observe that \( Z_i \) is bigger when \( y_i \) approaches the endpoints \((\pm 1)\), these sets slightly push sample points toward the center of the domain. Preferable sets according to large \( \gamma_0 \) (and small test value) should be less common. This intuition will be verified in Lemma 5, where we prove that \( Q \) is a preferable sample set according to \( \gamma_0 \) with probability exceeding \( 1 - \gamma_0 \).

The concept of preferable sets play a pivotal role in our analysis and is also useful in practice. We will show in both theory and computation that if we limit the sparse reconstruction to random Legendre matrices associated with preferable sample sets, better accuracy can be achieved. Enforcing this restriction is simple, as the computation cost of checking whether a sample set is preferable is negligible. Indeed, we can always generate multiple sets of \( m \) random samples, and pick a preferable one to commence the reconstruction process. We could follow Definition 3 strictly and choose a set that fulfills criterion (16), but empirical approach is also equally effective: first ranking all the sample sets according to their test values, and then selecting the preferable set among those in lower percentile of test value distribution, say, among \((100(1-\gamma_0))\%\) of sample sets with smallest \( \sum_{i=1}^{m} Z_i \).

### 3.1 Main results

Our first main theorem is stated as follows.

**Theorem 1** Let \( \delta, \gamma, \) and \( \gamma_0 \) be fixed parameters with \( 0 < \delta < 1, 0 < \gamma + \gamma_0 < 1 \) and \( \{L_j\}_{j \in J} \) be a univariate Legendre orthonormal system of finite size \( N = \#(J) \). Denote \( s_{a,\delta} = (1+a)^2 \frac{s}{\delta^{a/2}} \) and assume that

\[ m \geq C(1+\alpha)^2 s^2 \max \left\{ \delta^{-12} y_0^{-1} \log^2 \left( s_{a,\delta} \right) \log^2(2N), \quad \delta^{-\frac{15}{2}} \log^2 \left( s_{a,\delta} \right) \log(2N), \right\}. \quad (17) \]

Let \( y_1, y_2, \ldots, y_m \) be drawn independently from the uniform distribution on \([-1, 1]\). Then with probability exceeding \( 1 - (\gamma + \gamma_0) \), the normalized sampling matrix \( A \in \mathbb{R}^{m \times N} \) (defined as in (2)) satisfies

\[ \|Az\|_2^2 > (1-\delta)\|z\|_2^2, \quad \forall z \in C(s; \alpha). \quad (18) \]

The above result shows that for Legendre matrices to satisfy restricted eigenvalue property, the number of samples needs to scale quadratically with the sparsity and only logarithmically with the size of the coefficient vectors, therefore, verifying successful uniform recovery from underdetermined Legendre systems and uniform sampling. The dependence of sample complexity on \( \delta \) is severe \((\delta^{-12})\); however, as discussed in Section 2, \( \delta \) is allowed to approach 1 with our use of restricted eigenvalue property, and the growth of error bound when \( \delta \) tends to 1 is mild. Recall \( \delta_0 = 1/(1+\alpha) \), the dependence of the number of samples on \( \delta_0 \) is \( \delta_0^{-4} \).
A large part of our analysis will be conducted under the condition that the sample sets are preferable, giving rise to an additional parameter \( \gamma_0 \). In particular, the restricted eigenvalue property is guaranteed under a reduced probability \( 1 - (\gamma + \gamma_0) \). It is possible to restore the rate \( 1 - \gamma \) (which is normally seen in the literature), if we limit the sparse reconstruction to random matrices associated with preferable sets (which is simple to do).

**Proposition 4**  Let \( \delta, \gamma \) and \( \gamma_0 \) be fixed parameters in \((0, 1)\) and \( \{L_j\}_{j \in J} \) be a univariate Legendre orthonormal system of finite size \( N = \#(J) \). Denote \( s_{a, \delta} = \frac{(1 + a)^2}{\delta} \) and assume that

\[
m \geq C(1 + a)^4 s^2 \max \left\{ \delta^{-12} \gamma_0^{-1} \log^2 \left( s_{a, \delta} \right) \log^2(2N), \quad \delta^{-\frac{15}{2}} \log^2 \left( s_{a, \delta} \right) \log(2N), \quad \delta^{-4} \log \left( s_{a, \delta} \right) \log \left( \frac{\log(s_{a, \delta})}{(1 - \gamma_0) \gamma \delta} \right) \right\}. \tag{19}
\]

Let \( Q = \{y_1, y_2, \ldots, y_m\} \) be a set of samples drawn independently from the uniform distribution on \([-1, 1]\). If \( Q \) is a preferable set according to \( \gamma_0 \), then with probability exceeding \( 1 - \gamma \), the normalized sampling matrix \( A \in \mathbb{R}^{m \times N} \) (defined as in (2)) satisfies

\[
\|Az\|_2^2 > (1 - \delta)\|z\|_2^2, \quad \forall z \in C(s; \alpha). \tag{20}
\]

In Proposition 4, \( \gamma_0 \) can be considered a tuning parameter. Overall, the sample complexity is inversely proportional to \( \gamma_0 \); therefore, preferable sets according to high \( \gamma_0 \) allow reconstructions with lower number of samples and should be truly preferable. This fact is illustrated with numerical tests in Section 5. These sets are however increasingly uncommon as \( \gamma_0 \) tends to 1. Also, \( \gamma_0 \) very close to 1 may affect the third term of (19) unfavorably.

We will prove Theorem 1 and Proposition 4 in Sections 3.3 and 3.4 correspondingly. First, some preparatory lemmas are needed.

### 3.2 Supporting lemmas

We prove the following estimates involving the envelope bound \( \Omega \).

**Lemma 2**  For any \( \mu > 0 \), there holds \( \phi(y \in \mathcal{U} : |\Omega(y)| \geq \mu) \leq \frac{16}{\pi^2 \mu^2} \).

**Proof.**  We have \( |\Omega(y)| \geq \mu \) if and only if \( y^2 \geq 1 - \frac{16}{\pi^2 \mu^2} \). There follows \( \phi(y \in \mathcal{U} : |\Omega(y)| \geq \mu) \leq 1 - \sqrt{1 - \frac{16}{\pi^2 \mu^2}} \leq \frac{16}{\pi^2 \mu^2} \), as desired.

**Lemma 3**  For any \( 0 < \mu \leq 1 \) and \( \bar{T} \subset \mathcal{U} \) measurable with \( \phi(\bar{T}) \leq \mu \), there holds

\[
\int_{\bar{T}} |\Omega(y)|^2 d\phi \leq 2 \sqrt{\mu}.
\]

**Proof.**  We have \( \int_{\bar{T}} |\Omega(y)|^2 d\phi \leq \int_{1-\mu}^1 |\Omega(y)|^2 dy = \frac{4}{\pi} \left( \frac{\pi}{2} - \arcsin(1 - \mu) \right) \). Denoting \( \kappa = (\pi/2 - \arcsin(1 - \mu)) \), then \( 1 - \mu = \sin(\pi/2 - \kappa) = \cos(\kappa) \leq 1 - 4\kappa^2/\pi^2 \), which implies \( \kappa \leq \frac{\pi}{2} \sqrt{\mu} \).

**Lemma 4**  For all \( \beta > 0 \) and \( 0 \leq \tau < 1 \), there holds

\[
\int_{-1}^{1} \exp \left( -\beta \sqrt{1 - y^2} \right) \frac{dy}{(1 - y^2)^{\tau}} \leq 2\Gamma(2 - 2\tau)\beta^{2\tau - 2}. \tag{21}
\]
Here, $\Gamma$ denotes the gamma function.

**Proof.** Let us define $g(y) := \frac{\exp(-y\sqrt{1-y^2})}{(1-y^2)}$. We have

$$
\int_{-1}^{1} g(y) \, dq = \int_{0}^{\infty} q(y \in U^t : g(y) \geq t) \, dt = \int_{0}^{\exp(-\beta)} dt + \int_{\exp(-\beta)}^{\infty} q(y \in U^t : g(y) \geq t) \, dt,
$$

since $g(y) \geq \exp(-\beta), \forall y \in U$. By substituting $t = \exp(-\beta \sqrt{u})/u^r$, it gives

$$
\int_{-1}^{1} g(y) \, dq = \exp(-\beta) + \int_{1}^{0} q \left( y \in U^t : g(y) \geq \frac{\exp(-\beta \sqrt{u})}{u^r} \right) \left( \frac{\exp(-\beta \sqrt{u})}{u^r} \right)' \, du
$$

$$
= \exp(-\beta) + \int_{1}^{0} q(y \in U^t : |y| \geq \sqrt{1-u}) \left( \frac{\exp(-\beta \sqrt{u})}{u^r} \right)' \, du. \tag{22}
$$

With notice that $q(y \in U^t : |y| \geq \sqrt{1-u}) = 1 - \sqrt{1-u} \leq u$, there follows

$$
\int_{-1}^{1} g(y) \, dq \leq \exp(-\beta) + \int_{1}^{0} u \left( \frac{\exp(-\beta \sqrt{u})}{u^r} \right)' \, du = \int_{0}^{1} \exp(-\beta \sqrt{u}) \, du,
$$

where the second equality comes from integration by parts. Substituting $v = \beta \sqrt{u}$ yields

$$
\int_{-1}^{1} g(y) \, dq \leq \int_{0}^{\beta} \frac{2^r e^{-v}}{v^{2r}} \frac{2v}{\beta^2} \, dv = 2^r \beta^{2r-2} \int_{0}^{\beta} v^{2r+1} e^{-v} \, dv \leq 2 \Gamma(2 - 2\tau) \beta^{2r-2}. \quad \blacksquare
$$

In the next lemma, we verify the probability of a set of samples from uniform distribution on $[-1, 1]$ to be preferable.

**Lemma 5** Let $m > 0, 0 < \gamma_0 < 1$, and $Q = \{y_1, \ldots, y_m\}$ be a set of samples where $y_1, \ldots, y_m$ are drawn independently from the uniform distribution on $[-1, 1]$. Then, $Q$ is a preferable sample set according to $\gamma_0$ with probability exceeding $1 - \gamma_0$.

**Proof.** Let $Z_i$ be defined as in (60). By Lemma 4,

$$
\mathbb{E}Z_i \leq \frac{211/2}{\pi \sqrt{\pi}} \left( \frac{\gamma_0}{m} \right)^{3/4}, \quad \text{Var}(Z_i) \leq \mathbb{E}|Z_i|^2 \leq \frac{2^5}{\pi^2} \left( \frac{\gamma_0}{m} \right)^{1/2}. \tag{23}
$$

Referring $\mathbb{E}Z$ and $\text{Var}(Z_i)$ as $\mathbb{E}Z$ and $\text{Var}(Z)$, we apply Chebyshev inequality to obtain

$$
\mathbb{P}\left( \left| \sum_{i=1}^{m} Z_i - m\mathbb{E}Z \right| \geq \lambda \right) \leq \frac{m}{\lambda^2} \text{Var}(Z).
$$

Setting $\lambda = \sqrt{\frac{m}{\gamma_0}} \left( \mathbb{E}|Z|^2 \right)^{1/2}$, then with probability exceeding $1 - \gamma_0$,

$$
\sum_{i=1}^{m} Z_i \leq m\mathbb{E}Z + \sqrt{\frac{m}{\gamma_0}} \left( \mathbb{E}|Z|^2 \right)^{1/2}. \tag{24}
$$

Substituting (23)–(24) yields the conclusion. \quad \blacksquare

We also need some tail estimates, the proofs of which can be found in [2].
Let $X_1, \ldots, X_M$ be $M$ independent identically distributed real-valued random variables satisfying $|X_k| \leq a$ and $\mathbb{E} [X_k] = X$ for all $k$. We denote $\overline{X} = \frac{1}{M} \sum_{k=1}^{M} X_k$. For every $\mu > 0$,
\[
\mathbb{P}(|\overline{X} - X| \geq \mu) \leq 2 \exp \left(-\frac{M\mu^2}{2a^2}\right). \tag{25}
\]

Let $X_1, \ldots, X_M$ be $M$ independent identically distributed Bernoulli random variables with $\mathbb{E} [X_k] = X$ for all $k$. Denote $\overline{X} = \frac{1}{M} \sum_{k=1}^{M} X_k$. Then, for every $0 < \mu_1 < 1$, $\mu_2 > 0$ and $M \geq \frac{16e}{\mu_1 \mu_2}$, there holds
\[
\mathbb{P} \left(|\overline{X} - X| \geq \mu_1 X + \mu_2\right) \leq \exp \left( -\frac{M\mu_1 \mu_2}{16e}\right). \tag{26}
\]

Let $\mathcal{E}_{s,a} := \partial B_2 \cap C(s; \alpha)$, we conclude this subsection by proving the following important result involving covering number for $\mathcal{E}_{s,a}$. This step resembles but considerably extends the previous covering number estimates for $\mathcal{E}_1$ balls, for example, [1, 2, 14, 26], in the sense that a much more complicated pseudo-metric is considered herein.

**Lemma 8** For $0 < \zeta < 1$, $0 < \gamma_0 < 1$, $\mu > 0$, and $m > 0$, there exists a set $D \subset \mathbb{R}^N$ such that the following hold:

i. If $Q = \{y_1, \ldots, y_m\}$ is a set of $m$ i.i.d. samples drawn from $(U, \phi)$, preferable according to $\gamma_0$, then for all $z \in \mathcal{E}_{s,a}$, there exists $z' \in D$ depending on $z$ and $Q$ satisfying
\[
|\psi(y, z - z')| \leq \mu \quad \text{for all } y \text{ in a subset } U^* \text{ of } U \text{ with } \phi(U^*) > 1 - \zeta. \tag{27}
\]
\[
|\psi(y, z - z')| \leq \mu \quad \text{for all } y_i \in Q. \tag{28}
\]

ii. The cardinality of $D$ satisfies
\[
\log(\#(D)) \leq C(1 + a)^2 \frac{s}{\mu^2} \log(2N) \max \left\{ \frac{1}{\sqrt{\zeta}}, \sqrt{\frac{m}{\gamma_0}} \right\}. \tag{29}
\]

**Proof.** We will find $D$ using the empirical method of Maurey. First, we observe that $\mathcal{E}_{s,a} \subset \{(1 + a)\sqrt{s}B_1\}$ (see Lemma 1), hence if we denote $\mathcal{P} = \left\{ \pm e_j(1 + a)\sqrt{s} \right\}_{1 \leq j \leq N}$, where $(e_j)$ are canonical unit vectors in $\mathbb{R}^N$, we have $\mathcal{E}_{s,a} \subset \text{conv}(\mathcal{P})$.

Fix a $z \in \mathcal{E}_{s,a}$, $z$ can be represented as $z = \sum_{r=1}^{2N} \lambda_r v_r$, for some $\lambda_r \geq 0$ such that $\sum_{r=1}^{2N} \lambda_r = 1$ and $v_r$ listing $2N$ elements of $\mathcal{P}$. There exists a probability measure $\lambda$ on $\mathcal{P}$ that takes the values $v_r \in \mathcal{P}$ with probability $\lambda_r$. Let $z_1, \ldots, z_M$ be i.i.d random variables with law $\lambda$. Note that $\mathbb{E} z_k = z$, for all $k = 1, \ldots, M$. For each $y \in U$, $\psi(y, z_k)$ is also a real-valued random variable on probability space $(\psi(y, \mathcal{P}), \lambda)$ with
\[
|\psi(y, z_k)| \leq \Omega(y)(1 + a)\sqrt{s}, \quad \text{and} \quad \mathbb{E} \psi(y, z_k) = \psi(y, z).
\]

Denote $\overline{z} = \frac{1}{M} \sum_{k=1}^{M} z_k$. Let $D$ be the set of all possible outcomes of $\overline{z}$ and $\overline{\lambda}$ be the probability measure on $D$ according to $\overline{z}$. Note that $\#(D) \leq (2N)^M$, we derive a sufficient condition on $M$ so that $D$ satisfies the assertion (i).
First, let us define a characteristic function $\chi$ on $(U \times D, \varnothing \otimes \overline{\lambda})$, given by

$$\chi(y, z) = \begin{cases} 1, & \text{if } |\psi(y, z - z)| \geq \mu, \\ 0, & \text{if } |\psi(y, z - z)| < \mu. \end{cases}$$

Applying Lemma 6 yields for all $y \in U$,

$$\int_D \chi(y, z) d\overline{\lambda} = \mathbb{P}_{\overline{\lambda}}(|\psi(y, z - z)| \geq \mu) \leq 2 \exp \left( -\frac{M\mu^2}{2(1 + \alpha)\Omega^2(y)Ms} \right).$$

From Lemma 4, we have

$$\int_D \left( \int_U \chi(y, z) d\omega \right) d\overline{\lambda} = \int_U \left( \int_D \chi(y, z) d\overline{\lambda} \right) d\omega \leq \frac{C(1 + \alpha)^4 s^2}{M^2\mu^4},$$

which yields that with probability exceeding $2/3$, $\overline{\lambda} \in D$ satisfying

$$\mathbb{P}_y(|\psi(y, z - z)| \geq \mu) = \int_U \chi(y, z) d\omega \leq \frac{3C(1 + \alpha)^4 s^2}{M^2\mu^4} \leq \zeta,$$

provided that $M \geq C(1 + \alpha)^2 \frac{s}{\mu^2\sqrt{\zeta}}$.

To deal with the set $Q = \{y_1, \ldots, y_m\}$, we develop a different argument based on extending the approach in [15] (lemma 12.37). First, by symmetrization,

$$\mathbb{E}_{\overline{\lambda}} \left( \max_{y \in Q} |\psi(y, z - z)| \right) \leq \frac{2}{M} \mathbb{E}_{\overline{\lambda}} \max_{y \in Q} \sum_{k=1}^M \epsilon_k \psi(y, z_k),$$

where $\epsilon$ is a Rademacher sequence independent of $(z_1, \ldots, z_M)$. Since $|\psi(y, z_k)| \leq \Omega(y) (1 + \alpha)\sqrt{s}$ for all $y \in Q, k \in [M]$, we have $\|\psi(y, z_k)\|_2 \leq \Omega(y) (1 + \alpha)\sqrt{Ms}$. Standard calculations yield

$$\mathbb{P}_\epsilon \left( \sum_{k=1}^M \epsilon_k \psi(y, z_k) \geq t \right) \leq 2 \exp \left( -\frac{t^2}{2(1 + \alpha)^2\Omega^2(y)Ms} \right), \quad \forall t > 0.$$

By the union bound,

$$\mathbb{P}_\epsilon \left( \max_{y \in Q} \sum_{k=1}^M \epsilon_k \psi(y, z_k) \geq t \right) \leq \sum_{i=1}^m 2 \exp \left( -\frac{t^2}{2(1 + \alpha)^2\Omega^2(y)Ms} \right), \quad \forall t > 0.$$

We have for every $\kappa > 0$.

$$\mathbb{E}_\epsilon \max_{y \in Q} \left| \sum_{k=1}^M \epsilon_k \psi(y, z_k) \right| = \int_0^\kappa \mathbb{P}_\epsilon \left( \max_{y \in Q} \left| \sum_{k=1}^M \epsilon_k \psi(y, z_k) \right| \geq t \right) dt$$

$$\leq \int_0^\kappa 1 dt + 2 \sum_{i=1}^m \int_0^\infty \exp \left( -\frac{t^2}{2(1 + \alpha)^2\Omega^2(y)Ms} \right) dt$$

$$\leq \kappa + \sqrt{2\pi Ms(1 + \alpha)\sum_{i=1}^m \Omega(y)_i} \exp \left( -\frac{\kappa^2}{2(1 + \alpha)^2\Omega^2(y)Ms} \right).$$

Choosing $\kappa = (1 + \alpha)\sqrt{Ms}\sqrt{m/\gamma_0}$ leads to

$$\mathbb{E}_\epsilon \max_{y \in Q} \left| \sum_{k=1}^M \epsilon_k \psi(y, z_k) \right| \leq (1 + \alpha)\sqrt{Ms}\sqrt{m/\gamma_0} + \sqrt{2\pi Ms(1 + \alpha)\sum_{i=1}^m \Omega(y)_i} \exp \left( -\frac{1}{2\Omega^2(y)\sqrt{m/\gamma_0}} \right).$$

(33)
Proof of Theorem 1

Since $Q$ is a preferable sample set, we obtain from (16)

$$
\mathbb{E}_\xi \max_{y \in Q} \left| \sum_{k=1}^{M} \xi_k \psi \left( y, z_k \right) \right| \leq C(1 + \alpha) \sqrt{Ms} \left( m^{1/2} \gamma_0^{-1/2} + m^{1/2} \gamma_0^{3/2} \right) \leq C(1 + \alpha) \sqrt{Ms} \frac{m}{\gamma_0}.
$$

There follows $\mathbb{E}_\xi \left( \max_{y \in Q} |\psi \left( y, z - \bar{z} \right)| \right) \leq C(1+\alpha) \sqrt{\frac{s}{M} \frac{m}{\gamma_0}}$, yielding with probability exceeding 2/3, $\bar{z} \in D$ satisfies

$$
\max_{y \in Q} |\psi \left( y, z - \bar{z} \right)| \leq 3C(1 + \alpha) \sqrt{\frac{s}{M} \frac{m}{\gamma_0}} \leq \mu, \quad (34)
$$

provided that $M \geq \frac{C(1+\alpha)^2s}{\mu^2\sqrt{\gamma_0}}$. Thus, for $M \geq \frac{C(1+\alpha)^2s}{\mu^2\sqrt{\gamma_0}} \max \left\{ \frac{1}{\sqrt{\xi}}, \sqrt{\frac{m}{\gamma_0}} \right\}$, there exists $\bar{z} \in D$ that fulfills both (31) and (34). As $\log(\#(D)) \leq M \log(2N)$, the proof is concluded. ■

3.3 | Proof of Theorem 1

It is enough to prove (58) for all $z \in \mathcal{E}_{s,a}$ (i.e., $\partial B_2 \cap C(s;a)$). We conduct the analysis under assumption $\delta < 1/13$ for convenience; this parameter will be rescaled at the end. For $m \in \mathbb{N}$, let $Q = \{y_1, \ldots, y_m\}$ be a set of samples where $y_1, \ldots, y_m$ are drawn independently from the probability space $(U, q)$. From Lemma 5, $Q$ is preferable according to $\gamma_0$ with probability exceeding 1 – $\gamma_0$. We define the set of integers

$$
\mathcal{L} = Z \cap \left( \frac{\log(\delta)}{\log(1 + \delta)} + 1, \frac{\log(\Xi)}{\log(1 + \delta)} + 1 \right), \quad \text{where} \quad \Xi = \frac{4(1 + \alpha)^2s}{\sqrt{3} \pi \delta},
$$

and denote by $\underline{l}, \overline{l}$ the minimum and maximum of $\mathcal{L}$ respectively. $l$ and $\overline{l}$ then satisfy

$$(1 + \delta)^{\underline{l} - 2} \leq \delta \quad \text{and} \quad (1 + \delta)^{\overline{l} - 1} \leq \Xi \leq (1 + \delta)^{\overline{l}}. \quad (36)$$

**Step 1:** Under the condition that $Q$ is a preferable sample set according to $\gamma_0$ (pref. a.t. $\gamma_0$), for $0 < \zeta < 1$ (whose exact value will be set later), we seek to construct $\widetilde{\psi}$ approximating $\psi$ such that.

1. For all $z \in \mathcal{E}_{s,a}$, the following holds with probability exceeding 1 - $\zeta$ in $(U, q)$ and for all $y \in Q$

$$
(1 - 3\delta/2)\widetilde{\psi}(y, z) < |\psi(y, z)| < (1 + 3\delta/2)\widetilde{\psi}(y, z), \quad \text{if} \quad \widetilde{\psi}(y, z) > 0,
$$

$$
|\psi(y, z)| < 6\delta/5 \quad \text{or} \quad |\psi(y, z)| > \Xi, \quad \text{if} \quad \widetilde{\psi}(y, z) = 0. \quad (37)
$$

2. For all $z \in \mathcal{E}_{s,a}$, there exists a pairwise disjoint family of subsets $(\mathcal{I}_l)_{l \in \mathcal{L}} = \left( I_{l_{(z, q)}} \right)_{l \in \mathcal{L}}$ of $U$ depending on $z$ and $Q$ such that

$$
\widetilde{\psi}(\cdot, z) = \sum_{l \in \mathcal{L}} (1 + \delta)^l \chi_{l_l}. \quad (38)
$$

3. For every $l \in \mathcal{L}$, $\left( I_{l_{(z, q)}} \right)_{z \in \mathcal{E}_{s,a}}$ belongs to a finite class $F_l$ of subsets of $U$ satisfying $Q$ pref.a.t.$\gamma_0$

$$
\log(\#F_l) \leq C \frac{(1 + \alpha)^2s}{\delta^4(1 + \delta)^{2l-2}} \log(2N) \max \left\{ \sqrt{\frac{\log(\Xi/\delta)}{\zeta \log(1 + \delta)}}, \sqrt{\frac{m}{\gamma_0}} \right\}. \quad (39)
$$
First, for \( l \in \mathcal{L} \), let \( D_l \) be a finite subset of \( \mathbb{R}^N \) determined as in Lemma 8 with given \( m, y_0 \), as well as \( \mu = \delta(1 + \delta)^{-1}/2 \) and \( 0 < \zeta' < 1 \) (to be set accordingly to meet our needs). We have

\[
\log(\#D_l) \leq C \frac{(1 + \alpha)^2 s}{\delta^2(1 + \delta)^{2l-2}} \log(2N) \max \left\{ \frac{1}{\sqrt{s}}, \sqrt{\frac{m}{\gamma_0}} \right\}.
\] (40)

For a fixed \( z \in \mathcal{E}_{s,a} \), there exist \( z_l \in D_l \) and a measurable set \( U_l \subset U \) with \( q(U_l) \geq 1 - \zeta' \) such that

\[
|\psi(y, z - z_l)| \leq \delta(1 + \delta)^{-1}/2, \quad \forall y \in U_l,
\]

\[
|\psi(y, z - z_l)| \leq \delta(1 + \delta)^{-1}/2, \quad \forall y \in Q.
\]

Without loss of generality, we can assume \( Q \subset U_l \). We construct a pairwise disjoint family of subsets \( (I_l)_{l \in \mathcal{L}} \) and mapping \( \tilde{\psi}(\cdot, z) : U \to \mathbb{R} \) which depend on \( z \) and \( Q \), inductively for the integers \( l > \ldots > l \) according to

\[
I'_l = \{ y \in U : (1 + \delta)^{l-1} < |\psi(y, z_l)| < (1 + \delta)^{l+1} \},
\]

\[
I_l = I'_l \setminus \bigcup_{r \neq l} I'_r,
\]

\[
\tilde{\psi}(\cdot, z) = \sum_{l \in \mathcal{L}} (1 + \delta)^l \chi_l.
\] (41)

We proceed to prove that \( \tilde{\psi} \) satisfies (37)–(39), following closely the argument in\(^2\), Theorem 2.2. First, consider \( y \in \bigcap_{l \in \mathcal{L}} U_l \). If \( y \in I_l \) for some \( l \in \mathcal{L} \), then

\[
\tilde{\psi}(y, z) = (1 + \delta)^l > 0 \text{ and } (1 + \delta)^{l-1} < |\psi(y, z_l)| < (1 + \delta)^{l+1}.
\]

Since \( |||\psi(y, z)| - |\psi(y, z_l)||| \leq |\psi(y, z) - \psi(y, z_l)| \leq \delta(1 + \delta)^{-1}/2 \), we have

\[
|\psi(y, z)| < (1 + \delta)^{l+1} + \delta(1 + \delta)^{l-1} < \left( 1 + \frac{3}{2} \delta \right) \tilde{\psi}(y, z),
\]

and

\[
|\psi(y, z)| > (1 + \delta)^{l-1} - \delta(1 + \delta)^{-1} > \left( 1 - \frac{3\delta}{2} \right) \tilde{\psi}(y, z).
\]

If \( y \notin \bigcup_{l \in \mathcal{L}} I_l \), then \( \tilde{\psi}(y, z) = 0 \) and for every \( l \in \mathcal{L} \),

\[
|\psi(y, z_l)| \notin ((1 + \delta)^{l-1}, (1 + \delta)^{l+1}).
\]

With notice that \( |||\psi(y, z)| - |\psi(y, z_l)||| < \delta(1 + \delta)^{-1}/2 \), there follows

\[
|\psi(y, z)| \notin \bigcup_{l \in \mathcal{L}} \left( 1 + \frac{\delta}{2} \right) (1 + \delta)^{l-1}, \left( 1 + \frac{3\delta}{2} + \delta^2 \right) (1 + \delta)^{l-1}.
\]

Observe that \( \left( 1 + \frac{3\delta}{2} + \delta^2 \right) (1 + \delta)^{l-1} > \left( 1 + \frac{\delta}{2} \right) (1 + \delta)^l \), the previous intervals intersect for any two consecutive values of \( l \). We infer

\[
|\psi(y, z)| \leq \left( 1 + \frac{\delta}{2} \right) (1 + \delta)^{l-1}, \quad \text{or} \quad |\psi(y, z)| \geq \left( 1 + \frac{3\delta}{2} + \delta^2 \right) (1 + \delta)^{l-1}.
\]

It implies by (36) and assumption \( \delta < 1/13 \) that \( |\psi(y, z)| \leq \delta(1 + \delta/2)(1 + \delta) < 6\delta/5 \) (if the first inequality occurs) or \( |\psi(y, z)| > \Xi \) (if the second inequality occurs). We complete this case by emphasizing \( Q \subset \bigcap_{l \in \mathcal{L}} U_l \).
Next, consider $y \not\in \bigcap_{l \in \mathcal{L}} U_l$, (37) is not guaranteed. However, this only holds with probability not exceeding
\begin{equation}
q \left( \mathcal{U} \setminus \bigcap_{l \in \mathcal{L}} U_l \right) \leq \sum_{l \in \mathcal{L}} q(\mathcal{U} \setminus U_l) \leq \zeta'(#\mathcal{L}) \leq \frac{\log(\Xi/\delta)}{\log(1+\delta)} \zeta' = \zeta, \tag{42}
\end{equation}
for the last equality we set $\zeta' = \frac{\log(1+\delta)}{\log(\Xi/\delta)} \zeta$.

To summarize, we partitioned $\mathcal{U}$ into three sets
\[ I := \left( \bigcap_{l \in \mathcal{L}} U_l \right) \cap \left( \bigcup_{l \in \mathcal{L}} I_l \right), \quad \hat{I} := \left( \bigcap_{l \in \mathcal{L}} U_l \right) \setminus \left( \bigcup_{l \in \mathcal{L}} I_l \right), \]
\[ \mathcal{U}' := \mathcal{U} \setminus \left( \bigcap_{l \in \mathcal{L}} U_l \right), \]
and constructed $\tilde{\psi}(\cdot, z)$ depending on $z$ and $Q$ approximating $\psi(\cdot, z)$ satisfying
\begin{equation}
(1 - 3\delta/2)\tilde{\psi}(y, z) < |\psi(y, z)| < (1 + 3\delta/2)\tilde{\psi}(y, z), \quad \forall y \in I, \tag{43}
\end{equation}
\begin{equation}
(0 \leq |\psi(y, z)| < 6\delta/5 \lor |\psi(y, z)| > \Xi) \quad \text{and} \quad \tilde{\psi}(y, z) = 0, \quad \forall y \in \hat{I}, \tag{44}
\end{equation}
\begin{equation}
q(\mathcal{U}') \leq \zeta, \quad Q \subset I \cap \hat{I}, \quad \text{i.e.,} \quad Q \subset \mathcal{U}' = \emptyset. \tag{45}
\end{equation}

We further divide $\hat{I}$ into two subsets $I$ and $\hat{I}$ according to
\begin{equation}
I := \{ y \in \hat{I} : |\psi(y, z)| < 6\delta/5 \}, \quad \hat{I} := \{ y \in \hat{I} : |\psi(y, z)| > \Xi \}, \tag{46}
\end{equation}
as they will need different treatments. Note that for all $y \in \hat{I}$, $|\Omega(y)| \geq \frac{|\psi(y, z)|}{(1+\alpha)\sqrt{s}} > \frac{\Xi}{(1+\alpha)\sqrt{s}}$.

It remains to verify (39). For any $l \in \mathcal{L}$, $# \left\{ l^{(i, l)} z \in \mathcal{E}_{s, \alpha}, Q \text{ is preferable} \right\} \leq \#D_l$ and
\begin{equation}
\#F_l \leq \prod_{r \geq l} \#D_r. \quad \text{From (40), it gives}
\log(\#F_l) \leq \sum_{r \geq l} \log(\#D_r) \leq C\frac{(1+\alpha)^2s}{\delta^3(1+\delta)^{2l-2}} \log(2N) \max \left\{ \sqrt{\frac{\log(\Xi/\delta)}{\zeta \log(1+\delta)}}, \sqrt{\frac{m}{\gamma_0}} \right\}. \tag{43}
\end{equation}

**Step 2:** We derive estimates of $\|Az\|_2$ and $\|z\|_2$ in terms of $\tilde{\psi}(\cdot, z)$. The following bounds will be useful for this task. First, from (43),
\begin{equation}
(1 - 3\delta)|\tilde{\psi}(y, z)|^2 < |\psi(y, z)|^2 < (1 + 4\delta)|\tilde{\psi}(y, z)|^2, \quad \forall y \in I. \tag{47}
\end{equation}
(47) gives $|\tilde{\psi}(y, z)|^2 < (1 + 4\delta)|\psi(y, z)|^2$, which implies
\begin{equation}
\int_{I} |\tilde{\psi}(y, z)|^2 \, dq \leq (1 + 4\delta) \int_{I} |\psi(y, z)|^2 \, dq \leq (1 + 4\delta) \int_{\mathcal{U}} |\psi(y, z)|^2 \, dq = 1 + 4\delta. \tag{48}
\end{equation}
A lower bound of $\|Az\|_2$ in terms of $\tilde{\psi}(\cdot, z)$ is straightforward. Since $Q \cap \mathcal{U}' = \emptyset$ and $\tilde{\psi}(y, z) = 0, \forall y \in \hat{I}$, we have
\begin{equation}
\|Az\|_2^2 = \frac{1}{m} \sum_{i=1}^{m} |\psi(y_i, z)|^2 \geq (1 - 3\delta) \sum_{i=1}^{m} \frac{|\tilde{\psi}(y_i, z)|^2}{m}. \tag{49}
\end{equation}
It is worth mentioning that to obtain a reasonable upper estimate of $\|Az\|_2$ in similar manner is not easy, due to the lack of uniform bound for the considered orthonormal systems.
However, unlike RIP, upper estimate of $\|Az\|_2$ is not needed for restricted eigenvalue property.

For $\|z\|_2$, we decompose

$$\|z\|_2^2 = \int_I |\psi(y, z)|^2 dQ + \int_I |\psi(y, z)|^2 dQ + \int_I |\psi(y, z)|^2 dQ + \int_{I^c} |\psi(y, z)|^2 dQ,$$

and seek to produce upper bounds for each term in the RHS of (50). To begin with, by (47),

$$\int_I |\psi(y, z)|^2 dQ \leq (1 + 4\delta) \int_I |\tilde{\psi}(y, z)|^2 dQ. \quad (51)$$

By (46) and Lemma 3, note that $q(I) \leq 1$ and $q(I) \leq \frac{9\delta^2}{16(1+\alpha)^{1/2}}$, it gives

$$\int_I |\psi(y, z)|^2 dQ \leq (6\delta/5)^2 q(I) \leq \delta/6,$$

$$\int_I |\psi(y, z)|^2 dQ \leq (1 + \alpha^2 s) \int_I |\Omega(y)|^2 dQ \leq \frac{3\delta}{2}. \quad (52)$$

Similarly, setting $\zeta = \frac{\delta^2}{36(1+\alpha)^{1/2}}$, we have

$$\int_{I^c} |\psi(y, z)|^2 dQ \leq (1 + \alpha^2 s) \int_{I^c} |\Omega(y)|^2 dQ < \frac{\delta}{3}. \quad (53)$$

We combine (50)–(53) to get

$$\|z\|_2^2 < 2\delta + (1 + 4\delta) \int_I |\tilde{\psi}(y, z)|^2 dQ,$$

which in reference to (48) and (49) implies

$$\|z\|_2^2 - \|Az\|_2^2 < 2\delta + (1 + 4\delta) \int_I |\tilde{\psi}(y, z)|^2 dQ - (1 - 3\delta) \sum_{i=1}^m \frac{|\tilde{\psi}(y_i, z)|^2}{m} \leq 2\delta + 7\delta \int_I |\tilde{\psi}(y, z)|^2 dQ + (1 - 3\delta) \left( \int_{I^c} |\tilde{\psi}(y, z)|^2 dQ - \sum_{i=1}^m \frac{|\tilde{\psi}(y_i, z)|^2}{m} \right) \leq \frac{34\delta}{3} + (1 - 3\delta) \left( \int_{I^c} |\tilde{\psi}(y, z)|^2 dQ - \sum_{i=1}^m \frac{|\tilde{\psi}(y_i, z)|^2}{m} \right). \quad (54)$$

**Step 3:** we derive a positive upper bound of $\|z\|_2^2 - \|Az\|_2^2$ via (54), by employing the tail estimate in Lemma 7 and union bound. From the definition of $\tilde{\psi}$, it is easy to see that

$$\int_{I^c} |\tilde{\psi}(y, z)|^2 dQ - \sum_{i=1}^m \frac{|\tilde{\psi}(y_i, z)|^2}{m} = \sum_{i \in \tilde{Q}} (1 + \delta) \left( q(I_i) - \frac{\#(Q \cap I_i)}{m} \right) \quad (55)$$

Let $(\kappa_i)_{i \in \tilde{Q}}$ be a sequence of positive numbers. For any set $\Delta$ in the class $F_1$, for $Q$ being a set of samples $y_1, \ldots, y_m$ drawn independently from the probability space $(I^c, q)$ (not necessarily preferable), applying Lemma 7 yields with probability of $Q$ exceeding $1 - \exp\left(-\frac{m\kappa_i \delta^2}{16e}\right)$,

$$q(\Delta) - \frac{\#(Q \cap \Delta)}{m} \leq \delta q(\Delta) + \kappa_i. \quad (56)$$
By the union bound, with probability exceeding \(1 - \sum_{i \in \mathcal{L}} \exp\left(-\frac{mk_i \delta}{16e}\right) (#F_i)\), the previous inequality holds uniformly for all sets \(\Delta \in \bigcup_{i \in \mathcal{L}} F_i\). Therefore, with probability exceeding \(1 - \gamma_0 - \sum_{i \in \mathcal{L}} \exp\left(-\frac{mk_i \delta}{16e}\right) (#F_i)\), \(Q\) is preferable according to \(\gamma_0\) and (56) holds with \(Q\) uniformly for \(\Delta \in \bigcup_{i \in \mathcal{L}} F_i\). In this scenario, we can apply (56) with \(\Delta = I_i^{(\gamma_0)} (l \in \mathcal{L})\) to the sum in (55) and combine with (54) to infer that for all \(z \in \mathcal{E}_{s,a}\),

\[
||z||^2_2 - ||Az||^2_2 < \frac{34\delta}{3} + (1 - 3\delta) \sum_{i \in \mathcal{L}} (1 + \delta)^{2l} (\delta Q(l) + k_i)
= \frac{34\delta}{3} + \delta(1 - 3\delta) \int_U |\tilde{\psi}(y,z)|^2 dQ + (1 - 3\delta) \sum_{i \in \mathcal{L}} (1 + \delta)^{2l} k_i.
\] (57)

Note that \(\tilde{\psi}(y,z) = 0, \forall y \in I\) and \(\tilde{\psi}(y,z) \leq (1 + \delta)^3 \leq (1 + \delta)^3 \Xi, \forall y \in U\), we estimate

\[
\int_U |\tilde{\psi}(y,z)|^2 dQ = \int_I |\tilde{\psi}(y,z)|^2 dQ + \int_{U \setminus I} |\tilde{\psi}(y,z)|^2 dQ \leq (1 + 4\delta)
\times \int_I |\psi(y,z)|^2 dQ + q(U')(1 + \delta)^2 \Xi^2 < 1 + 5\delta
\]

and obtain

\[
||z||^2_2 - ||Az||^2_2 < \frac{25\delta}{2} + (1 - 3\delta) \sum_{i \in \mathcal{L}} (1 + \delta)^{2l} k_i.
\]

Finally, in order to obtain Theorem 1, we need to assign appropriate values for \(k_i\) and derive conditions on \(m\) such that

\[
\sum_{i \in \mathcal{L}} (1 + \delta)^{2l} k_i \leq \delta/2, \quad \text{and} \quad \sum_{i \in \mathcal{L}} \exp\left(-\frac{mk_i \delta}{16e} + \log (#F_i)\right) \leq \gamma.
\]

The two inequalities can be fulfilled if the numbers \(k_i\) and the integer \(m\) are chosen as follows

\[
k_i := \frac{\delta/2}{(#\mathcal{L})(1 + \delta)^{2l}}, \quad -\frac{mk_i \delta}{16e} + \log (#F_i) \leq \log \left(\frac{\gamma}{#\mathcal{L}}\right), \quad l \in \mathcal{L}.
\]

This implies that

\[
m \geq 32e (#\mathcal{L}) \frac{(1 + \delta)^{2l}}{\delta^2} \left[ \log (#F_i) + \log \left(\frac{#\mathcal{L}}{\gamma}\right) \right], \quad l \in \mathcal{L}.
\]

We have in view of (39) and assumption \(\delta < 1/13\) that

\[
32e (#\mathcal{L}) \frac{(1 + \delta)^{2l}}{\delta^2} \log (#F_i) \leq C \frac{\log(\Xi/\delta)}{\log(1 + \delta)} \cdot \frac{(1 + \delta)^{2l}}{\delta^2} \cdot \frac{(1 + \alpha)^2 s}{\delta^3 (1 + \delta)^{2l - 2}}
\times \log(2N) \max \left\{ \frac{\sqrt{\log(\Xi/\delta)}}{\gamma \log(1 + \delta)}, \sqrt{\frac{m}{\gamma_0}} \right\}
\leq C\delta^{-6}(1 + \alpha)^2 s \log(\Xi/\delta) \log(2N)
\times \max \left\{ (1 + \alpha)^2 s \sqrt{\frac{\log(\Xi/\delta)}{\delta^3}}, \sqrt{\frac{m}{\gamma_0}} \right\},
\]

\[
\times 32e (#\mathcal{L}) \frac{(1 + \delta)^{2l}}{\delta^2} \log \left(\frac{#\mathcal{L}}{\gamma}\right) \leq C \frac{\log(\Xi/\delta)}{\log(1 + \delta)}
\times \frac{(1 + \delta)^2 \Xi^2}{\delta^2} \cdot \frac{\log(\Xi/\delta)}{\gamma \log(1 + \delta)}
\leq C\delta^{-4}(1 + \alpha)^4 s^2 \log(\Xi/\delta) \log \left(\frac{\log(\Xi/\delta)}{\gamma \delta}\right).
\]
Combining the two estimates shows that $m$ in (17) guarantees
\[ \|Az\|_2^2 > (1 - 13\delta)\|z\|_2^2, \quad \forall z \in E_{s,a}, \]
with probability exceeding $1 - (\gamma + \gamma_0)$. Rescaling $\delta$ concludes the proof.

### 3.4 Proof of Proposition 4

The proof follows closely that of Theorem 1, with one key modification in Step 3. First, we note that Steps 1 and 2, in particular (54) and (55), were established under the condition that $Q$ is preferable according to $\gamma_0$, which holds with probability exceeding $1 - \gamma_0$. Independently, for $Q$ being a set of $m$ samples drawn uniformly from $[-1, 1]$ (not necessarily preferable), with probability exceeding $1 - \gamma$, (56) holds uniformly for all sets $\Delta \in \cup_{l \in L} F_l$, assuming $m$ satisfying \( \sum_{l \in L} \exp \left( -\frac{mk_l\delta}{16e} \right) (\# F_l) \leq \gamma \). Hence, given $Q$ preferable according to $\gamma_0$, (56) holds uniformly for all sets $\Delta \in \cup_{l \in L} F_l$ with probability exceeding $1 - \gamma_1 - \gamma_0$. With (54)–(56) in hand, proceeding as in Section 3.3, we can derive that under the condition (17), with probability exceeding $1 - (\gamma_1 - \gamma_0)$,
\[ \|Az\|_2^2 > (1 - \delta)\|z\|_2^2, \quad \forall z \in C(s; a). \quad (58) \]

By a change of variable, the proposition is concluded.

### 4 Restricted Eigenvalue Property for Multivariate Legendre Systems

We present an estimate for restricted eigenvalue property for multi-dimensional Legendre matrices. The overall strategy in this section is very similar to those in Section 3.1, but complete analysis involves intensive multivariate calculations. We will focus on such calculations and skip repetitive details whenever possible. Let $U := [-1, 1]^d$, $\rho$ be the uniform probability measure on $U$, $J$ be a subset of $\mathbb{N}^d$, and $\{\Psi_j\}_{j \in J} \equiv \{L_j\}_{j \in J}$, the system of multivariate Legendre polynomials orthonormal with respect to $\rho$. We again denote
\[ \psi(y, z) := \sum_{j \in J} z_j L_j(y), \quad y \in U, \quad z \in \mathbb{R}^N. \]

From the envelope bound of univariate Legendre polynomials, we have
\[ |L_j(y)| \leq \frac{2^d}{\pi^d/2 \Pi_{k=1}^d (1 - y_k^2)^{1/4}}, \quad (59) \]
for all $y = (y_1, \ldots, y_d) \in (-1, 1)^d$, $j \in \mathbb{N}^d$. Therefore, we define
\[ \Omega(y) := \frac{2^d}{\pi^d/2 \Pi_{k=1}^d (1 - y_k^2)^{1/4}}. \]

Let us introduce the notion of preferable set in multi-dimensional domain.

**Definition 4** For $m > 0$, $0 < \gamma_0 < 1$, let $Q = \{y_1, \ldots, y_m\}$ be a set of samples where $y_1, \ldots, y_m$ are drawn independently from the uniform distribution on $[-1, 1]^d$. Define $m$ i.i.d random variables
\[ Z_i = \Omega(y_i) \exp \left( -\frac{1}{2\Omega^2(y_i)} \sqrt{\frac{m}{\gamma_0}} \right), \quad \forall i \in \{1, \ldots, m\}, \quad (60) \]
and test value function $T(Q) = \sum_{i=1}^{m} Z_i$. We call $Q$ a preferable sample set according to $\gamma_0$ if $T(Q)$ is below $(100(1-\gamma_0))$-th percentile of the distribution of $T$.

It is straightforward from the above definition that $Q$ is preferable according to $\gamma_0$ with probability $1 - \gamma_0$. In Lemma 14, we show there exists a universal constant $C$ such that if $Q$ is preferable according to $\gamma_0$ then

$$\sum_{i=1}^{m} Z_i \leq C \left( \frac{m}{\gamma_0} \right)^{1/4} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d}{4^d \cdot \sqrt{m/\gamma_0}} \right) \right)^{d-1}.$$

### 4.1 Main result

Our main theorem in this part is stated as follows.

**Theorem 2** Let $\delta, \gamma$ and $\gamma_0$ be fixed parameters with $0 < \delta < 1$, $0 < \gamma + \gamma_0 < 1$ and $\{L_j\}_{j \in J}$ be a $d$-dimensional Legendre orthonormal system of finite size $N = \#(J)$. Denote $s_a = (1 + \alpha)^2 s$ and assume that

$$m \geq s_a^2 \max \left\{ \frac{C_{d,1}}{\delta^4 \gamma_0} \log^2 \left( \frac{C_{d,2} S_a^{3/2}}{\delta^2} \right) \log^2(2N) \left( \log \left( \frac{\pi^d}{4^d \sqrt{m/\gamma_0}} \right) \right)^{4d-4}, \right.$$

$$\frac{C_{d,2}}{\delta^4} \log^3 \left( \frac{C_{d,3} S_a^{3/2}}{\delta^2} \right) \left( \log \left( \frac{C_{d,4} S_a^{3/2}}{\delta^2} \right) \right)^{d-1},$$

$$\left. \frac{C_{d,3}}{\delta^4} \log \left( \frac{C_{d,4} S_a^{3/2}}{\delta^2} \right) \log \left( \frac{C_{d,4} S_a^{3/2}}{\delta^2} \right) \right\}. \quad (61)$$

Let $y_1, y_2, \ldots, y_m$ be drawn independently from uniform distribution on $[-1, 1]^d$. Then with probability exceeding $1 - (\gamma + \gamma_0)$, the normalized sampling matrix $A \in \mathbb{R}^{m \times N}$ (defined as in (2)) satisfies

$$\|Az\|_2^2 > (1-\delta)\|z\|_2^2, \quad \forall z \in C(s; \alpha). \quad (62)$$

Here, $C_{d,1} = C \left( \frac{4}{\pi} \right)^{4d}$, $C_{d,2} = C(d+1) \left( \frac{64 \sqrt{2}}{\pi^2} \right)^d$, $C_{d,3} = C(d+1) \left( \frac{4}{\pi} \right)^{2d}$, and $C_{d,4} = C(d+1)(d-1)^{d-1} \cdot \left( \frac{4}{\pi} \right)^{4d/3}$. We observe that for $d = 1$, condition (61) retrieves the estimate (17) in the univariate case (up to a minor change of the order of $\delta$ inside log factors). The number of log factors for multi-dimensional setting however scales linearly in $d$ and is at most $4d$; therefore, the sample complexity grows exponentially in dimension. Similarly to the univariate case, random matrices associated with preferable sets are more likely to satisfy restricted eigenvalue property. Also, preferable sets according to moderate and high $\gamma_0$ reduce the required number of samples for sparse reconstructions, as shown in the following proposition.

**Proposition 5** Let $\delta, \gamma$ and $\gamma_0$ be fixed parameters in $(0, 1)$ and $\{L_j\}_{j \in J}$ be a $d$-dimensional Legendre orthonormal system of finite size $N = \#(J)$. Denote $s_a =$
(1 + α)²s and assume that
\[
m \geq s^2 \max \frac{C_{d,1}}{\delta^{12 \gamma_0}} \log^2 \left( \frac{C_{d,s} \delta^{3/2}}{\delta^2} \right) \log^2(2N) \left( \log \left( \frac{\pi d}{4^d \sqrt{m}} \right) \right)^{4d-4} ,
\]
\[
\frac{C_{d,2}}{\delta^{15/2}} \log^{3/2} \left( \frac{C_{d,s} \delta^{3/2}}{\delta^2} \right) \log(2N) \left( \log \left( Cm \left( \frac{\pi d}{4} \right) \right) \right)^{d-1} ,
\]
\[
\frac{C_{d,3}}{\delta^4} \log \left( \frac{C_{d,s} \delta^{3/2}}{\delta^2} \right) \log \left( \frac{\log \left( C_{d,s} \frac{\delta^{3/2}}{\delta^2} \right)}{(1 - \gamma_0) \gamma \delta} \right) \left( \log \left( C\pi^d m \right) \right)^{d-1} \right} .
\]

Let \( Q = \{ y_1, y_2, \ldots, y_m \} \) be a set of samples drawn independently from the uniform distribution on \([-1, 1]^d\). If \( Q \) is a preferable set according to \( \gamma_0 \), then with probability exceeding \( 1 - \gamma \), the normalized sampling matrix \( A \in \mathbb{R}^{m \times N} \) (defined as in (2)) satisfies
\[
\|Az\|_2^2 > (1 - \delta)\|z\|_2^2 , \quad \forall z \in C(s; \alpha).
\]

Here, \( C_{d,1}, C_{d,2}, C_{d,3} \) and \( C_{d,4} \) are defined as in Theorem 2.

We will prove Theorem 2 in Sections 4.3. Proposition 5 can be derived directly from the proof of Theorem 2 and will be skipped. First, we need some preparatory lemmas.

4.2 Supporting lemmas

For \( d \geq 2 \), we define the function
\[
H_d(\beta) = \beta \left( \log \left( \frac{1}{\beta} \right) \right)^{d-1} .
\]
In particular, the case \( d = 2 \) gives \( H_2(\beta) = \beta \log \left( \frac{1}{\beta} \right) \). Note that \( \beta \mapsto \beta \log \left( \frac{1}{\beta} \right) \) is strictly increasing for \( 0 < \beta < 1/e \), and \( \beta \mapsto H_d(\beta) \) is strictly increasing for \( 0 < \beta < \frac{1}{e^{d-1}} \), the inverse function \( K_d(M) := H_d^{-1}(M) \) is well-defined for \( 0 < M < H_d \left( \frac{1}{e^{d-1}} \right) = \left( \frac{d-1}{e} \right)^{d-1} \). We prove a basic identity involving \( H_d \) and \( K_d \).

**Lemma 9** Let \( 0 < \beta < \left( \frac{d-1}{e} \right)^{d-1} \), there follows \( H_d \left( K_d^2(\beta) \right) = 2^{d-1} \beta K_d(\beta) \).

**Proof.** By definition of \( H_d \),
\[
H_d \left( K_d^2(\beta) \right) = K_d^2(\beta) \left( \log \left( \frac{1}{K_d(\beta)} \right) \right)^{d-1} = 2^{d-1} K_d^2(\beta) \left( \log \left( \frac{1}{K_d(\beta)} \right) \right)^{d-1} = 2^{d-1} K_d(\beta) H_d \left( K_d(\beta) \right) = 2^{d-1} \beta K_d(\beta) \).
\]

While an analytical formula of \( K_d(M) \) is not known to us, the following lemma provides an estimate of \( K_d(M) \).

**Lemma 10** For all \( d \geq 2 \), \( 0 < M < \left( \frac{d-1}{e} \right)^{d-1} \), there holds
\[
K_d(M) \geq \frac{M}{(d - 1)^{d-1}} \exp \left( -(d - 1) \log^2 \left( \frac{d - 1}{M \pi^{d-1}} \right) \right) .
\]

Consequently, \( \log \left( K_d(M) \right) \geq -\log \left( \frac{(d-1)^{d-1}}{M} \right) - (d - 1) \frac{1}{2} \log^2 \left( \frac{(d-1)^{d-1}}{M} \right) \).
Proof. We first prove (65) for $d = 2$. It is enough to show that
\[
\beta \log \left( \frac{1}{\beta} \right) \leq M, \quad \text{for} \quad \beta = M \exp \left( -\log^{1/2} \left( \frac{1}{M} \right) \right).
\] (66)

Denote $\beta_0 = \sqrt{\log \left( \frac{1}{M} \right)}$. Since $0 < M < 1/e$, then $1 < \beta_0 < \infty$. We have $\beta \log \left( \frac{1}{\beta} \right) = e^{-\beta_0 - \beta_0} (\beta_0^2 + \beta_0)$ and $M = e^{-\beta_0}$, thus $\beta \log \left( \frac{1}{\beta} \right) \leq M$ is equivalent to $\beta_0^2 + \beta_0 \leq e^{\beta_0}$, $\forall \beta_0 > 1$. An inspection of the map $\beta_0 \mapsto e^{\beta_0} - \beta_0^2 - \beta_0$ over $(1, \infty)$ proves the desired inequality.

For $d > 2$, from (66) and by change of variables, one can show that
\[
H_d(\beta) \leq M, \quad \text{for} \quad \beta = \frac{M}{(d - 1)^{d-1}} \exp \left( -(d - 1) \log^{1/2} \left( \frac{d - 1}{M^{1/(d-1)}} \right) \right),
\]
concluding the proof.

Lemma 11 Let $d \geq 1$. Define $v_d(r) := \mathbb{Q} \left( y \in U : \prod_{k=1}^{d} \left( 1 - y_k^2 \right) \leq r \right)$. Then, for any $0 < r \leq 1$,
\[
\frac{r}{2(d - 1)!} \left( \log \left( \frac{1}{\sqrt{r}} \right) \right)^{d-1} < v_d(r) \leq \frac{r}{(d - 1)!} \left( \log \left( \frac{2^d e}{\sqrt{r}} \right) \right)^{d-1},
\]
or equivalently,
\[
\frac{H_d(r)}{2^d (d - 1)!} < v_d(r) \leq \frac{2^{d+1} e^2}{(d - 1)!} H_d \left( \frac{r}{4^{d+2}} \right).
\] (67)

Proof. First, we prove the upper bound in (67) by induction on $d$. Observe that $v_d(r) = 1, \forall d \geq 1, r \geq 1$. Consider $r < 1$. For $d = 1$, we have
\[
v_1(r) = \mathbb{Q} \left( y \in [-1, 1] : (1 - y^2) \leq r \right) = 1 - \sqrt{1 - r} \leq r.
\]

For $d \geq 1$, applying Fubini’s theorem gives the recurrence relation
\[
v_{d+1}(r) = \int_{-1}^{1} v_d \left( \frac{r}{1 - y_{d+1}^2} \right) \frac{dy_{d+1}}{2} = \int_{r}^{1} v_d \left( \frac{r}{1 - y_{d+1}^2} \right) dy_{d+1} + \int_{\sqrt{1-r}}^{1} 1 dy_{d+1}.
\] (68)

By a change of variable and induction hypothesis, it gives
\[
\int_{0}^{\sqrt{1-r}} v_d \left( \frac{r}{1 - y_{d+1}^2} \right) dy_{d+1} = \frac{r}{2} \int_{r}^{1} v_d(u) du \sqrt{1 - \frac{r}{u}} \leq \frac{r}{(d - 1)!} \int_{r}^{1} \left( \log \left( \frac{2^d e}{\sqrt{u}} \right) \right)^{d-1} du.
\]

For all $u \in [r, 1]$, we observe that $\log \left( \frac{2^{d+1} e}{\sqrt{u} + \sqrt{u - r}} \right) \geq \log \left( \frac{2^d e}{\sqrt{u}} \right) \geq 0$, thus
\[
\left( \log \left( \frac{2^{d+1} e}{\sqrt{u} + \sqrt{u - r}} \right) \right)^{d-1} \geq \left( \log \left( \frac{2^d e}{\sqrt{u}} \right) \right)^{d-1}, \quad \forall d \geq 1.
\]

There follows
\[
\int_{0}^{\sqrt{1-r}} v_d \left( \frac{r}{1 - y_{d+1}^2} \right) dy_{d+1} \leq \frac{r}{(d - 1)!} \int_{r}^{1} \frac{\left( \log \left( \frac{2^{d+1} e}{\sqrt{u} + \sqrt{u - r}} \right) \right)^{d-1} du}{2 \sqrt{u(u-r)}}.
\]

(69)
Combining (68) and (69) and applying the estimate $1 - \sqrt{1 - r} \leq r$, we arrive at

$$v_{d+1}(r) \leq \frac{r}{d!} \left( \log \left( \frac{2^{d+1}e}{\sqrt{r}} \right) \right)^d - \left( \log \left( \frac{2^{d+1}e}{1 + \sqrt{1 - r}} \right) \right)^d + r.$$ 

It is enough to prove $\left( \log \left( \frac{2^{d+1}e}{1 + \sqrt{1 - r}} \right) \right)^d / d! \geq 1$. Applying Stirling’s approximation $d! < \sqrt{2\pi d} \left( \frac{d}{e} \right)^d e^{1/(12d)}$, [28] we will show that $\log \left( \frac{2^{d+1}e}{1 + \sqrt{1 - r}} \right) \geq (2\pi d)^{1/2} \left( \frac{d}{e} \right) e^{1/(12d)}$, or equivalently, after some rearrangement,

$$\log \left( \frac{2\left( \frac{2e}{1 + \sqrt{1 - r}} \right)^{1/d}}{e^{1/d}} \right) \geq \sqrt{2\pi e^{1/(12d)}} \left( \frac{d}{e} \right) e^{1/(2d)}.$$  

We observe

$$\log \left( \frac{2\left( \frac{2e}{1 + \sqrt{1 - r}} \right)^{1/d}}{e^{1/d}} \right) \geq \frac{\log (2e^{1/d})}{e^{1/d}} \geq \frac{\log(2e)}{e}, \quad \forall d \geq 1, 0 \leq r \leq 1. \quad (70)$$ 

On the other hand, it is easy to check that

$$\left[ \sqrt{2\pi e^{1/(12d)}} \right]^{1/d} \leq \frac{\sqrt{2\pi e^{1/12}}}{e}.$$  

A combination of (71) and (72) gives (70).

Next, we establish the lower bound of $v_d(r)$, again by induction on $d$. For $d = 1$, $v_1(r) = 1 - \sqrt{1 - r} \geq \frac{r}{2}$. For $d \geq 1$, recall

$$v_{d+1}(r) = \int_0^{\sqrt{1-r}} v_d \left( \frac{r}{1 - \sqrt{1-r}^2} \right) dy + \int_0^{\sqrt{1-r}} 1 dy + \int_0^{\sqrt{1-r}} v_d \left( \frac{r}{1 - \sqrt{1-r}^2} \right) dy.$$

By induction hypothesis, it gives

$$v_{d+1}(r) > \frac{r}{2} \int_0^1 \frac{v_d}{u \sqrt{1 - \frac{r}{u}}} du > \frac{r}{2(d-1)!} \int_0^1 \left( \log \left( \frac{1}{\sqrt{u}} \right) \right)^{d-1} du > \frac{r}{2(d-1)!} \int_0^1 \left( \log \left( \frac{1}{\sqrt{u}} \right) \right)^{d-1} du = \frac{r}{2(d-1)!} \left( \log \left( \frac{1}{\sqrt{v}} \right) \right)^d,$$

as desired.

We define $S_{\mu} := \Omega^{-1}(\mu, \infty) = \{ y \in U : |\Omega(y)| \geq \mu \}$. With notice that $q(S_{\mu}) = v_d \left( \frac{2^{2d-1}e}{\pi^{2d}\mu^4} \right)$, the following result is a direct consequence of Lemma 11 and a multivariate version of Lemma 2.

**Corollary 1**  
For $d \geq 2$, $\mu \geq \frac{2^d}{\pi^d \mu^2}$, there holds

$$\frac{1}{2^d(d-1)!} H_d \left( \frac{2^{2d+1}e^2}{\pi^2 \mu^4} \right) \leq q(S_{\mu}) \leq \frac{2^{d+1}e^2}{(d-1)!} H_d \left( \frac{2^{2d}}{\pi^2 \mu^4} \right). \quad (73)$$

We proceed to generalize estimates in Lemmas 3 and 4 for multi-dimensional setting.
Lemma 12  For \( d \geq 2, \mu \geq \frac{2^d}{\pi^{d/2}} \), there holds
\[
\int_{\Delta_x} |\Omega(y)|^2 dQ \leq \frac{2^{4d}(d+1)}{\pi^{2d}(d-1)!} \cdot \left( \frac{\log (e\mu^2)}{\mu^2} \right)^{d-1} = \frac{2^{4d}e(d+1)}{\pi^{2d}(d-1)!} \cdot H_d \left( \frac{1}{e\mu^2} \right). \tag{74}
\]

Proof.  We have \( \int_{\Delta_x} |\Omega(y)|^2 dQ = \int_0^\infty Q(y \in \Delta_x : |\Omega(y)|^2 \geq t) dt = \int_0^\infty h Q(\Delta_x) dt + \int_\Delta_x Q(\Delta_x) dt \). Applying the upper estimates of \( Q(\Delta_x) \) and \( Q(\Delta_{\sqrt{t}}) \) in Corollary 1, there follows
\[
\int_{\Delta_x} |\Omega(y)|^2 dQ \leq \frac{2^{4d}}{\pi^{2d}(d-1)!} \left( \frac{\log (e\mu^2)}{\mu^2} \right)^{d-1} + \int_0^\infty \left( \frac{\log (e\mu^2)}{t^2} \right)^{d-1} dt \]
\[
= \frac{2^{4d}}{\pi^{2d}(d-1)!} \left( \frac{\log (e\mu^2)}{\mu^2} \right)^{d-1} + \sum_{k=0}^{d-1} \frac{(d-1)!}{k!} \cdot \left( \frac{\log (e\mu^2)}{\mu^2} \right)^{d-1} \tag{75}
\]
For \( \mu \geq 1 \), observe that \( d < \log (e\mu^2) \), thus \( \frac{(d-1)!}{k!} \leq \left( \frac{\log (e\mu^2)}{\mu^2} \right)^{d-1-k} \), \( \forall 0 \leq k \leq d-1 \). Substituting these estimates to (75) yields (74).

Lemma 13  Let \( d \geq 2 \), there exists a universal constant \( C > 0 \) such that for all \( \beta \geq 10 \) and \( \tau \in \left\{ 0, \frac{1}{3}, \frac{1}{2} \right\} \),
\[
\int_{U^d} \frac{\exp \left( -\beta \prod_{k=1}^d (1-y_k^2)^{1/2} \right)}{\prod_{k=1}^d (1-y_k^2)^{1/2}} dQ \leq C\beta^{2\tau-2} (\log(\beta))^{d-1}. \tag{76}
\]
Proof.  Denote \( B_r = [-r, r]^d \) for \( 0 < r < 1 \), and represent \( Q \) as \( Q = \prod_{k=1}^d Q_k \), where \( Q_k \) is the univariate uniform measure on \([-1, 1]\). First, for any \( r \in (0, 1) \),
\[
\int_{U^d} \frac{\exp \left( -\beta \prod_{k=1}^d (1-y_k^2)^{1/2} \right)}{\prod_{k=1}^d (1-y_k^2)^{1/2}} dQ = \int_{B_r \cup (B_1 \setminus B_r)} \left( \int_{-1}^1 \frac{\exp \left( -\beta \prod_{k=1}^d (1-y_k^2)^{1/2} \right)}{\prod_{k=1}^d (1-y_k^2)^{1/2}} dQ_1 \right) dQ_2 \ldots dQ_d
\]
\[
= \int_{B_r} \prod_{k=1}^d (1-y_k^2)^{-r} \left( \int_{-1}^1 \frac{\exp \left( -\beta \prod_{k=1}^d (1-y_k^2)^{1/2} \right)}{(1-y_k^2)^{1/2}} dQ_1 \right) dQ_2 \ldots dQ_d
\]
\[
+ \int_{B_1 \setminus B_r} \prod_{k=2}^d (1-y_k^2)^{-r} \left( \int_{-1}^1 \frac{\exp \left( -\beta \prod_{k=1}^d (1-y_k^2)^{1/2} \right)}{(1-y_k^2)^{1/2}} dQ_1 \right) dQ_2 \ldots dQ_d
\]
\[
\leq C\beta^{2\tau-2} \int_{B_r} \prod_{k=2}^d (1-y_k^2)^{-r} dQ_2 \ldots dQ_d + \int_{-1}^1 (1-y_1^2)^{-r} dQ_1 \int_{B_1 \setminus B_r} \prod_{k=2}^d (1-y_k^2)^{-r} dQ_2 \ldots dQ_d
\]
\[
= C\beta^{2\tau-2} \left( \int_0^r \frac{dy}{1-y^2} \right)^{d-1} + \int_0^1 \frac{dy}{(1-y^2)^r} \left[ \left( \int_0^1 \frac{dy}{1-y^2} \right)^{d-1} - \left( \int_0^r \frac{dy}{1-y^2} \right)^{d-1} \right]
\]
\[
\leq C\beta^{2\tau-2} \left( \int_0^r \frac{dy}{1-y^2} \right)^{d-1} + (d-1) \left( \int_0^1 \frac{dy}{(1-y^2)^r} \right)^{d-1} \int_{r}^1 \frac{dy}{(1-y^2)^r},
\]
where the first inequality is derived from Lemma 4 and \( \exp \left( -\beta \prod_{k=1}^{d} (1 - y_k^2)^{\frac{1}{2}} \right) \leq 1. \)

Define the constants \( C_r = 1, 2F_1 \left( \frac{1}{4}, \frac{1}{2}; \frac{3}{2} ; 1 \right) \) \((\approx 1.2)\) and \( \frac{\pi}{2} \), respectively for \( \tau = 0, \frac{1}{4}, \frac{1}{2} \), where \( 2F_1 \) is the hypergeometric function, observe that

\[
\int_{0}^{r} \frac{dy}{1 - y^2} = \frac{1}{2} \log \left( \frac{1 + y}{1 - y} \right) \bigg|_{y=0}^{r} = \frac{1}{2} \log \left( \frac{1 + r}{1 - r} \right),
\]

\[
\int_{r}^{1} \frac{dy}{(1 - y^2)^{1/2}} = \begin{cases} 
1 - r, & \text{if } \tau = 0, \\
\left( y_2 F_1 \left( \frac{1}{4}, \frac{1}{2}; \frac{3}{2} ; y^2 \right) \right) \bigg|_{y=r}^{1} \leq C_r (1 - r)^{3/4}, & \text{if } \tau = 1/4, \\
\pi/2 - \arcsin(r) \leq \frac{\pi}{2} \sqrt{1 - r}, & \text{if } \tau = 1/2.
\end{cases}
\]

Therefore, for \( \tau \in \{0, 1/4, 1/2\} \) and \( 0 < r < 1 \), we have

\[
\int_{\mathcal{U}'} \frac{\exp \left( -\beta \prod_{k=1}^{d} (1 - y_k^2)^{1/2} \right)}{\prod_{k=1}^{d} (1 - y_k^2)^{1/2}} dQ \leq C \beta^{2r - 2} \frac{2^{d-1}}{d-1} \left( \log \left( \frac{1 + r}{1 - r} \right) \right)^{d-1} + (d - 1) C_r (1 - r)^{1-r}. \tag{77}
\]

Choose \( r = 1 - \frac{2}{\beta^2} \in (0, 1) \), then

\[
(1 - r)^{1-r} = 2^{1-r} \beta^{2r-2}, \quad \log \left( \frac{1 + r}{1 - r} \right) \leq \log \left( \frac{2}{1 - r} \right) = 2 \log(\beta). \tag{78}
\]

Substituting (78) to (77) yields

\[
\int_{\mathcal{U}'} \frac{\exp \left( -\beta \prod_{k=1}^{d} (1 - y_k^2)^{1/2} \right)}{\prod_{k=1}^{d} (1 - y_k^2)^{1/2}} dQ \leq C \beta^{2r - 2} (\log(\beta))^{d-1} + (d - 1) 2^{1-r} C_r \beta^{2r-2} \\
\leq C \beta^{2r - 2} (\log(\beta))^{d-1} + 2^{1-r} C_r \left( (d - 1) \frac{1}{2^{d}} C_r \right)^{d-1} \beta^{2r-2} \leq C \beta^{2r - 2} (\log(\beta))^{d-1},
\]

as desired, where the last inequality comes from observation \((d - 1) \frac{1}{2^{d}} C_r < \log(10) \leq \log(\beta). \)

The next lemma establishes that preferable sets according to \( \gamma_0 \) have their test values bounded from above by

\[
C \left( \frac{m}{\gamma_0} \right)^{1/4} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{4 \gamma_0} \right) \right)^{d-1}.
\]

**Lemma 14** For \( m > 0, \gamma_0 \in (0,1) \) satisfying \( \gamma_0 \leq \frac{m}{2^{d}} \left( \frac{4}{\pi} \right)^{2d} \), let \( y_1, y_2, \ldots, y_m \) be sampling points drawn independently from the uniform distribution on \([-1, 1]^d\) and

\[
Z_i = \Omega (y_i) \exp \left( -\frac{1}{2\Omega^2 (y_i)} \frac{m}{\gamma_0} \right), \quad \forall i \in \{1, \ldots, m\}.
\]

be \( m \) i.i.d random variables. With probability exceeding \( 1 - \gamma_0 \), there holds

\[
\sum_{i=1}^{m} Z_i \leq C \left( \frac{m}{\gamma_0} \right)^{1/4} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{4 \gamma_0} \right) \right)^{d-1}. \tag{79}
\]
Proof. By Lemma 13,
\[
\mathbb{E}Z_i \leq C \left( \frac{4}{\pi} \right)^{2d} \left( \frac{\gamma_0}{m} \right)^{3/4} \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{2^{2d+1} \cdot \sqrt{\gamma_0}} \right) \right)^{d-1},
\]
\[
\text{Var}(Z_i) \leq \mathbb{E}|Z_i|^2 \leq C \left( \frac{4}{\pi} \right)^{2d} \left( \frac{\gamma_0}{m} \right)^{1/2} \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{2^{2d} \cdot \sqrt{\gamma_0}} \right) \right)^{d-1},
\]
(80)

Referring \( \mathbb{E}Z_i \) and \( \text{Var}(Z_i) \) as \( \mathbb{E}Z \) and \( \text{Var}(Z) \), we apply Chebyshev inequality to obtain with probability exceeding \( 1 - \gamma_0 \):
\[
\sum_{i=1}^{m} Z_i \leq m\mathbb{E}Z + \sqrt{m \gamma_0 (\mathbb{E}|Z|^2)^{1/2}}.
\]
(81)

Substituting (80) to (81) yields
\[
\sum_{i=1}^{m} Z_i \leq Cm^{\frac{1}{4}} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{4^d \cdot \sqrt{\gamma_0}} \right) \right)^{d-1} + C \left( \frac{m}{\gamma_0} \right)^{1/4} \left( \frac{4}{\pi} \right)^d \left( \log \left( \frac{\pi^d \cdot \sqrt{m}}{4^d \cdot \sqrt{\gamma_0}} \right) \right)^{d-1}.
\]
(82)

Recall \( E_{s,\alpha} := \partial B_2 \cap C(s; \alpha) \), our last lemma in this section involves an estimate of covering number for \( E_{s,\alpha} \). This is an extension of Lemma 8 to the setting where \( \left(U^*, \varrho\right) \) is multi-dimensional sample space.

**Lemma 15** For \( 0 < \zeta < 1, \mu > 0, m > 0, 0 < \gamma_0 < \min \left\{ 1, \frac{m}{20^r} \left( \frac{\pi}{4} \right)^{2d} \right\} \), there exists a set \( D \subset \mathbb{R}^N \) such that the following hold:

i. If \( Q = \{ y_1, \ldots, y_m \} \) is a set of \( m \) i.i.d. samples drawn from \( (U, \varrho) \), preferable according to \( \gamma_0 \), then for all \( z \in E_{s,\alpha} \), there exists \( z' \in D \) depending on \( z \) and \( Q \) satisfying
\[
|\psi(y, z - z')| \leq \mu \quad \text{for all } y \text{ in a subset } U^* \text{ of } U \text{ with } \rho(U^*) > 1 - \zeta,
\]
(82)

\[
|\psi(y_i, z - z')| \leq \mu \quad \text{for all } y_i \in Q.
\]
(83)

ii. The cardinality of \( D \) satisfies
\[
\log(\#(D)) \leq C \left( 1 + \frac{\alpha}{\mu^2} \frac{\gamma_0^{1/2}}{s} \right) \left( \frac{4}{\pi} \right)^d K_d \left( \frac{2^{d-1} \zeta}{C} \right)^{-1/2} \left( \frac{\pi^d \cdot \sqrt{m}}{4^d \cdot \sqrt{\gamma_0}} \right)^{2d-2}.
\]
(84)

Proof. The proof follows closely that of Lemma 8, with all estimates involving the univariate envelope bound of Legendre polynomials replaced by those with the multivariate bound. For brevity, we only show the critical changes and refer the readers to Lemma 8 for the full arguments. The first change occurs in (30) where we now have
by Lemma 13
\[
\int_{D} \left( \int_{U} \chi(y, z) dQ \right) d\hat{x} = \int_{D} \left( \int_{U} \chi(y, \bar{z}) d\bar{x} \right) dQ
\leq \int_{U} 2 \exp \left( -\frac{M\mu^2}{2(1+\alpha)^2\Omega^2(y)\gamma} \right) dQ \leq C\beta^{-2}(\log(\beta))^{d-1}.
\]

Here, \( \beta = \frac{M\mu^2}{2(1+\alpha)^2\Omega^2(y)\gamma} \). Applying Markov’s inequality, with probability exceeding 2/3, \( \bar{z} \in D \) satisfies
\[
\mathbb{P}_y(|\psi(y, \bar{z})| \geq \mu) = \int_{U} \chi(y, \bar{z}) dQ \leq C\beta^{-2}(\log(\beta))^{d-1} = \frac{C}{2^{d-1}}H_d \left( \frac{1}{\beta^2} \right).
\] (85)

From the definition of \( K_d \), one has \( \mathbb{P}_y(|\psi(y, \bar{z})| \geq \mu) \leq \varsigma \), assuming the trivial condition \( \varsigma < C(d-1/2e)^{d-1} \) and
\[
\beta \geq \left[ K_d \left( \frac{2^{d-1}\varsigma}{C} \right) \right]^{-\frac{1}{d}}, \text{ or equivalently, } M \geq \frac{2(1+\alpha)^2\Omega^2(y)}{\mu^2} \left[ K_d \left( \frac{2^{d-1}\varsigma}{C} \right) \right]^{-\frac{1}{d}}.
\] (86)

Next, consider a preferable sample set \( Q = \{y_1, \ldots, y_m\} \) according to \( \gamma_0 \). Similar to Lemma 8, we have
\[
\mathbb{E}_x \max_{y_i \in Q} \left| \sum_{k=1}^{M} \epsilon_k \psi \left( y_i, z_k \right) \right| \leq (1+\alpha)\sqrt{Ms} \sqrt{\frac{m}{\gamma_0}} \sqrt{2\pi MS(1+\alpha)\sum_{i=1}^{m} \Omega(y_i)} \exp \left( -\frac{1}{2\Omega^2(y_i)} \sqrt{\frac{m}{\gamma_0}} \right).
\] (87)

Applying Lemma 14 gives
\[
\mathbb{E}_x \max_{y_i \in Q} \left| \sum_{k=1}^{M} \epsilon_k \psi \left( y_i, z_k \right) \right| \leq C(1+\alpha)\sqrt{Ms} \sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^d \left( \log \left( \frac{\pi^d}{4^d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{d-1}.
\]

Since \( \mathbb{E}_x \left( \max_{y_i \in Q} |\psi(y_i, z - \bar{z})| \right) \leq \frac{2}{M} \mathbb{E}_x \sum_{k=1}^{M} \epsilon_k \psi \left( y_i, z_k \right) \), applying Markov’s inequality, with probability exceeding 2/3, \( \bar{z} \in D \) satisfies
\[
\max_{y \in Q} |\psi(y, z - \bar{z})| \leq C(1+\alpha)\sqrt{\frac{s}{M}} \sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^d \left( \log \left( \frac{\pi^d}{4^d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{d-1}.
\]

Thus, \( \max_{y \in Q} |\psi(y, z - \bar{z})| \leq \mu \) provided that
\[
M \geq \frac{C(1+\alpha)^2s}{\mu^2} \sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^{d-2} \left( \log \left( \frac{\pi^d}{4^d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{2d-2}.
\] (88)

Finally, for \( M \) satisfying (86) and (88), there exists \( \bar{z} \in D \) that fulfills both (82) and (83). As \( \log(#(D)) \leq M \log(2N) \), the proof is concluded. \( \blacksquare \)

4.3 Proof of Theorem 2

We conduct the analysis under assumption \( \delta < 1/13 \); this parameter will be rescaled at the end. The proof follows the same logic to that of Theorem 1. For brevity, we only elaborate major changes specific to multi-dimensional setting and refer the readers to Theorem 1 for the rest of the arguments.

For \( m \in \mathbb{N} \), let \( Q = \{y_1, \ldots, y_m\} \) be a set of samples where \( y_1, \ldots, y_m \) are drawn independently from the probability space \((U, Q)\). Let \( \Xi \in \mathbb{R} \) whose value will be set later to meet our need, we define...
the set of integers

\[ L = \mathbb{Z} \cap \left( \frac{\log(\delta)}{\log(1 + \delta)} + 1, \frac{\log(\Xi)}{\log(1 + \delta)} + 1 \right). \]  

**(Step 1):** under the condition that \( Q \) is a preferable sample set according to \( \gamma_0 \), by following the same procedure in Step 1, Section 3.3, with an application of Lemma 8 instead of Lemma 8, we can construct \( \tilde{\psi} \) approximating \( \psi \) that satisfies exactly three properties in the aforementioned step, except that the cardinality of class \( F_i \) is now.

\[
\log(\#F_i) \leq C\frac{(1 + \alpha)^2 s}{\delta^3(1 + \delta)^{2d-2}} \log(2N) \max \left\{ \left( \frac{4}{\pi} \right)^d \left[ \frac{K_d}{2^d - 1} \frac{\zeta(1 + \delta)}{\log(2N/\delta)} \right]^{d-1}, \sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d}{4^d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{2d-2} \right\}. 
\]

Partitioning \( U' \) into \( I, \bar{I}, I \) and \( U' \) as in Section 3.3, then

\[
(1 - 3\delta/2)\psi(y, z) < |\psi(y, z)| \leq (1 + 3\delta/2)\psi(y, z), \quad \forall y \in I, 
\]

\[
0 \leq |\psi(y, z)| < 6\delta/5 \quad \text{and} \quad \tilde{\psi}(y, z) = 0, \quad \forall y \in \bar{I},
\]

\[
|\psi(y, z)| > \Xi \quad \text{and} \quad \tilde{\psi}(y, z) = 0, \quad \forall y \in I,
\]

\[
\nu(\nu') \leq \zeta, \quad \text{and} \quad Q \cap U' = \emptyset.
\]

Let \( \nu = \frac{\Xi}{(1 + \alpha)^{\sqrt{s}}} \), note that for all \( y \in \bar{I} \), \( |\Omega(y)| \geq \frac{\|y\psi(y)\|^2}{(1 + \alpha)\sqrt{s}} > \nu \), thus by Corollary 1,

\[
\nu(\bar{I}) \leq \frac{2^{d+1}e^2}{d!} H_d \left( \frac{2^{d(1 + \alpha)/2} \Xi}{e^{\pi^d/4} \sqrt{s}} \right), \quad \text{assuming that} \quad \frac{\Xi}{(1 + \alpha)^{\sqrt{s}}} \geq \left( \frac{4}{\pi} \right)^{d/2}.
\]

**(Step 2):** We derive estimates of \( ||Az||_2 \) and \( ||z||_2 \) in terms of \( \tilde{\psi}(. , z) \). Similarly to (47) and (49),

\[
||Az||_2^2 \geq (1 - 3\delta) \sum_{i=1}^{m} \frac{\tilde{\psi}(y_i, z)^2}{m}.
\]

For \( ||z||_2 \), we decompose

\[
||z||_2^2 = \int_I |\psi(y, z)|^2 d\nu + \int_I |\psi(y, z)|^2 d\bar{\nu} + \int_I |\psi(y, z)|^2 d\nu + \int_{U'} |\psi(y, z)|^2 d\nu.
\]  

The first two RHS terms can be bounded in the same way as in (51) and (52). For the third term, from Lemma 12 and the fact that \( \bar{I} \subset S_{\nu} \),

\[
\int_I |\psi(y, z)|^2 d\nu \leq (1 + \alpha)^2 s \int_S |\Omega(y)|^2 d\nu \leq (1 + \alpha)^2 s \frac{2^d e \theta d + 1}{\pi^d (d - 1)!} H_d \left( \frac{(1 + \alpha)^2 s}{e^{\pi^d/4} \Xi^2} \right) = \frac{11\delta}{12},
\]

if we set \( \Xi = \frac{(1 + \alpha)^{\sqrt{s}}}{\pi^{d/2} \sqrt{e}} \left[ \frac{11d}{12(1 + \alpha)^2 s} \frac{e^{d(d-1)/2}}{2^d} \right]^{d-1} \). Note that for \( \Xi \) and \( K_d \) to be well-defined, we need \( \frac{11\delta}{12(1 + \alpha)^2 s} \frac{e^{d(d-1)/2}}{2^d} < \left( \frac{d - 1}{2} \right)^{d-1} \). It can be checked that this condition is trivial. Also, \( \Xi \) defined above satisfies the condition \( \frac{\Xi}{(1 + \alpha)^{\sqrt{s}}} \geq \left( \frac{4}{\pi} \right)^{d/2} \) posed in Step 1.
For the last term, we have
\[
\int_{U'} |\psi(y, z)|^2 \, d\mathcal{Q} \leq (1 + \alpha)^2 s \int_{U'} |\Omega(y)|^2 \, d\mathcal{Q} \leq \frac{11\delta}{12}.
\] (98)

if setting \(\zeta := \xi(s')\). From Corollary 1, note that
\[
\zeta > \frac{1}{2^d(d - 1)!} H_d \left( 2^{d+2} K_d^2 \left( \frac{11\delta}{12} \frac{\pi^d(d - 1)!}{(1 + \alpha)^2 s} \cdot \frac{\pi^d(d - 1)!}{2^d e(d + 1)} \right) \right).
\] (99)

Combining all above estimates yields
\[
\|z\|_2^2 - \|Az\|_2^2 \leq \frac{34\delta}{3} + (1 - 3\delta) \left( \int_{U'} |\tilde{\psi}(y, z)|^2 \, d\mathcal{Q} - \sum_{i=1}^m \frac{|\tilde{\psi}(y, z)|^2}{m} \right).
\] (100)

**Step 3:** We derive a positive upper bound of \(\|z\|_2^2 - \|Az\|_2^2\). Let \((\kappa_l)_{l \in \mathcal{L}}\) be a sequence of positive numbers. From (100), analogously to Section 3.3, we can prove that with probability exceeding 1 - \(\gamma_0 - \sum_{l \in \mathcal{L}} \exp\left( -\frac{m\kappa_l}{16e} \right) \) \(#F_l\), \(Q\) is preferable according to \(\gamma_0\) and for all \(z \in \mathcal{E}_{s,a}\),
\[
\|z\|_2^2 - \|Az\|_2^2 \leq \frac{34\delta}{3} + \delta(1 - 3\delta) \int_{U'} |\tilde{\psi}(y, z)|^2 \, d\mathcal{Q} + (1 - 3\delta) \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l.
\]

Note that \(\tilde{\psi}(y, z) = 0, \forall y \notin I \cup U'\) and \(\tilde{\psi}(y, z) \leq (1 + \delta)\Xi, \forall y \in U'\), combining with (97), we estimate
\[
\int_{U'} |\tilde{\psi}(y, z)|^2 \, d\mathcal{Q} = \int_{I} |\tilde{\psi}(y, z)|^2 \, d\mathcal{Q} + \int_{U'} |\tilde{\psi}(y, z)|^2 \, d\mathcal{Q}
\]
\[
\leq (1 + 4\delta) \int_{I} |\psi(y, z)|^2 \, d\mathcal{Q} + \int_{U'} (1 + \delta)^2 \Xi^2 \, d\mathcal{Q}
\]
\[
< 1 + 4\delta + (1 + \delta)^2 (1 + \alpha)^2 s \int_{\mathcal{S}} |\Omega(y)|^2 \, d\mathcal{Q} < 1 + \frac{56\delta}{11},
\]
which yields
\[
\|z\|_2^2 - \|Az\|_2^2 < \frac{25\delta}{2} + (1 - 3\delta) \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l.
\]

To obtain Theorem 2, we derive conditions on \(m\) such that
\[
m \geq C \left( \frac{\#\mathcal{L}}{\delta^2} \right) \left( \frac{\#\mathcal{L}}{\gamma^2} \right) \log \left( \frac{\#\mathcal{L}}{\gamma} \right), \quad l \in \mathcal{L}.
\] (101)

We have in view of (90) and assumption \(\delta < 1/13\) that
\[
C(\#\mathcal{L}) \left( \frac{1 + \delta)^{2l}}{\delta^2} \right) \log \left( \frac{\#\mathcal{L}}{\gamma} \right) \leq C\delta^{-6} (1 + \alpha)^2 s \log (\Xi/\delta) \log (2N)
\]
\[
\times \max \left\{ \left( \frac{4}{\pi} \right)^d K_d \left( \frac{2d-1}{C} \cdot \frac{\zeta \log(1 + \delta)}{\log(\Xi/\delta)} \right) \right\}^{\frac{1}{2}},
\]
\[
\sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d}{4d} \cdot \frac{\sqrt{m}}{\gamma_0} \right) \right)^{2d-2},
\]
\[
C(\#\mathcal{L}) \left( \frac{1 + \delta)^{2l}}{\delta^2} \right) \log \left( \frac{\#\mathcal{L}}{\gamma} \right) \leq C \log (\Xi/\delta) \log (1 + \delta) \cdot \frac{(1 + \delta)^2 \Xi^2}{\delta^2} \cdot \log \left( \frac{\log (\Xi/\delta)}{\gamma \log (1 + \delta)} \right)
\] (102)
\[
\leq C\delta^{-3}\pi^{-d}(1 + \alpha)^2s \left[ K_d \left( \frac{11\delta}{12(1 + \alpha)^2s^2} \cdot \frac{\pi^d(d - 1)!}{2^d d(e + 1)} \right) \right]^{-1} \log(\Xi/\delta) \log \left( \frac{\log(\Xi/\delta)}{\gamma\delta} \right).
\]

To fulfill (101), we set \( m \) greater than the right hand sides in (102). First,

\[
C\delta^{-6}(1 + \alpha)^2 s \log(\Xi/\delta) \log(2N) \left( \frac{4}{\pi} \right)^{d} \left[ K_d \left( \frac{2^{d-1} \cdot \zeta \log(1 + \delta)}{C \cdot \log(\Xi/\delta)} \right) \right]^{-1} \leq m
\]

if

\[
\frac{C}{m^2} \delta^{-12}(1 + \alpha)^4 s^2 \log^2(\Xi/\delta) \log^2(2N) \left( \frac{4}{\pi} \right)^{2d} \leq K_d \left( \frac{2^{d-1} \cdot \zeta \log(1 + \delta)}{C \cdot \log(\Xi/\delta)} \right).
\]

Since \( H_d \) is strictly increasing, the above inequality is equivalent to

\[
H_d \left( \frac{C}{m^2} \delta^{-12}(1 + \alpha)^4 s^2 \log^2(\Xi/\delta) \log^2(2N) \left( \frac{4}{\pi} \right)^{2d} \right) \leq \frac{\gamma^{-d-1} \cdot \zeta \log(1 + \delta)}{C \cdot \log(\Xi/\delta)}.
\]

From (99) and Lemma 9, the right hand side is bounded from below by

\[
\frac{\pi^d}{2^{3d}(d + 1)} \cdot \frac{\log(1 + \delta)}{\delta^2} \cdot \frac{\delta}{C \log(\Xi/\delta)} \cdot \left( 1 + \alpha \right)^2 s^2 \cdot K_d \left( \frac{11\delta}{12(1 + \alpha)^2s} \cdot \frac{\pi^d(d - 1)!}{2^d d(e + 1)} \right),
\]

and we set \( m \) such that

\[
H_d \left( \frac{2^{3d}(d + 1)}{\pi^d} \cdot \frac{C(1 + \alpha)^2 s \log \left( \frac{\Xi}{\delta} \right)}{\delta^2} \cdot \frac{\delta}{H_d \left( \frac{C(1 + \alpha)^4 s^2 \log^2 \left( \frac{\Xi}{\delta} \right) \log^2(2N) \left( \frac{4}{\pi} \right)^{2d} \right)} \right)
\]

\[
\leq \frac{11\delta}{12(1 + \alpha)^2s} \cdot \frac{\pi^d(d - 1)!}{2^{3d} d(e + 1)}.
\]

Rearranging with notice that \( H_d(\beta) \geq \beta, \forall \beta \leq \frac{1}{e^{d-1}} \), it is sufficient that

\[
m^2 \geq \frac{C(d + 1)^2}{(d - 1)!} \cdot \frac{2^{11d}}{\delta^{4d}} \cdot \frac{(1 + \alpha)^8 s^4}{\delta^{15}} \cdot \frac{\log^3 \left( \frac{\Xi}{\delta} \right)}{\log^2(2N)} \left( \frac{\pi}{4} \right)^{2d}
\]

\[
\times \log^2(2N) \left( \log \left( \frac{m^2}{C(d + 1)} \cdot \frac{\delta^{12}}{(1 + \alpha)^4 s^2} \cdot \frac{\log^2 \left( \frac{\Xi}{\delta} \right)}{\log^2(2N)} \left( \frac{\pi}{4} \right)^{2d} \right) \right)^{d-1}
\]

\[
\times \log \left( \log \left( \frac{m^2}{C(d + 1)} \cdot \frac{\delta^{14}}{(1 + \alpha)^6 s^3} \cdot \frac{\log^3 \left( \frac{\Xi}{\delta} \right)}{\log^2(2N) \cdot 2^{3d}} \right) \right)^{d-1}.
\]

This condition can be replaced by the following shortened (yet slightly more restrictive) condition

\[
m \geq \frac{C(d + 1)^2}{(d - 1)!} \cdot \left( \frac{64\sqrt{2}}{\pi^2} \right)^{d} \cdot \frac{(1 + \alpha)^4 s^2}{\delta^{15/2}} \cdot \frac{\log^3 \left( \frac{\Xi}{\delta} \right)}{\log(2N)} \log \left( \log \left( Cm \left( \frac{\pi}{4} \right)^d \right) \right)^{d-1}.
\]

Secondly, we need

\[
m \geq C \left( \frac{1 + \alpha}{\delta^6} s \log(\Xi/\delta) \log(2N) \sqrt{\frac{m}{\gamma_0}} \left( \frac{4}{\pi} \right)^{2d} \left( \log \left( \frac{\pi^d}{4d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{2d-2},
\]

which is guaranteed given that

\[
m \geq C \left( \frac{4}{\gamma_0} \right)^{4d} \frac{(1 + \alpha)^4 s^2}{\delta^{12}} \log^2(\Xi/\delta) \log^2(2N) \left( \log \left( \frac{\pi^d}{4d} \cdot \sqrt{\frac{m}{\gamma_0}} \right) \right)^{4d-4}.
\]
The last condition is
\[
m \geq C \delta^{-3} \pi^{-d} (1 + \alpha)^{2s} K_d \left( \frac{11 \delta}{12(1 + \alpha)^2 s \cdot 2^{4d} e(d + 1)} \right)^{-1} \log(\Xi/\delta) \log \left( \frac{\log(\Xi/\delta)}{\gamma \delta} \right),
\]
i.e.,
\[
\frac{11 \delta}{12(1 + \alpha)^2 s \cdot 2^{4d} e(d + 1)} \geq H_d \left( \frac{C}{m} \cdot \frac{(1 + \alpha)^2 s}{\delta^3 \pi^d} \log(\Xi/\delta) \log \left( \frac{\log(\Xi/\delta)}{\gamma \delta} \right) \right).
\]
Rearranging and simplifying, we have
\[
m \geq \frac{C(d + 1)(d - 1)!}{(d - 1)!} \cdot \left( \frac{4}{\pi} \right)^{2d} \cdot \frac{(1 + \alpha)^{4s^2}}{\delta^4} \log(\Xi/\delta) \log \left( \frac{\log(\Xi/\delta)}{\gamma \delta} \right) \left( \log(C \pi^d m) \right)^{d-1}.
\]

Finally, we observe from the definition of \( \Xi \) and Lemma 10 that
\[
\log(\Xi/\delta) = \frac{1}{2} \log \left( \frac{(1 + \alpha)^2 s}{\delta^2 \pi^d e} \right) - \frac{1}{2} \log \left( K_d \left( \frac{11 \delta}{12(1 + \alpha)^2 s \cdot 2^{4d} e(d + 1)} \right) \right) \\
\leq \frac{1}{2} \log \left( \frac{(1 + \alpha)^{2s}}{\delta^2 \pi^d e} \right) + \log \left( \frac{C(d + 1)(d - 1)^{d-1}}{(d - 1)!} \cdot \frac{2^{4d}}{\pi^d} \cdot \frac{(1 + \alpha)^2 s}{\delta^4} \right) \\
\leq \log \left( \frac{C(d + 1)(d - 1)^{d-1}}{(d - 1)!} \cdot \frac{2^{4d}}{\pi^{3d/2}} \cdot \frac{(1 + \alpha)^{3s^2}}{\delta^2} \right). \quad (103)
\]

Subsuming (103) into the three above conditions on \( m \), we deduce that (61) guarantees
\[
\|Az\|_2^2 > (1 - 13\delta)\|z\|_2^2, \quad \forall z \in \mathcal{E}_{s,a},
\]
with probability exceeding 1 - (\( \gamma + \gamma_0 \)). Rescaling \( \delta \) concludes the proof.

5 | NUMERICAL ILLUSTRATIONS

In this section, we present some numerical experiments to illustrate an important observation from our theory, that is, restricting the sparse recovery on random Legendre matrices associated with preferable sets (in particular, sample sets with small test values) can improve the reconstruction performance. Here, the \( \ell_1 \) minimization problem is solved using SPGL1 package [29].

These tests are conducted for 1d Legendre approximations, where we set \( J := \{1, \ldots, 360\} \). For each experiment, we generate a large number of sets of samples (drawn independently from the uniform distribution in \([-1, 1]\)) and rank them according to their test values \( \sum_{i=1}^{m} Z_i \), where \( \gamma_0 \) is set to 0.8. We divide the sample sets into five equal-sized groups: the first group contains 20% of the sets with lowest test values (roughly, preferable sets according to 0.8), the second group contains next 20% of the sets with lowest test values and not being in the first group, and so on. The fifth group includes 20% of the sets with highest test values.

In the first experiment, we generate 1000 sample sets, each of which has 180 samples; hence, random matrices have fixed size 180 \times 360. For each matrix, we form a sparse coefficient vector \( c \) by randomly selecting the support of \( c \) from \( J \) and drawing its coefficients from a Gaussian distribution, and aim to reconstruct \( c \) from noiseless observations \( g = Ac \). The average errors and successful rates associated with five groups of sample sets are plotted in Figure 3 for two different sparsity levels of \( c \) (30 and 40). We observe that on average, reconstruction on sample sets with smaller test values results in reduced recovery errors and better successful rates, thus confirming our theory.

In the second experiment, we generate 1500 sample sets and again seek to reconstruct sparse vector \( c \) from noiseless observations \( g = Ac \). The sparsity of \( c \) is fixed at 15, and the number of samples increases with a step size 10 from 10 to 180. In Figure 4, we show the average errors and successful
Figure 3  Averaged recovery errors and successful rates associated with five different groups of sample sets with increasing test values in the reconstruction of 30- and 40-sparse coefficient vectors using random Legendre matrices.

Figure 4  Averaged recovery errors and successful rates associated with 20% of sample sets with lowest test values and 20% of sample sets with highest test values in the reconstruction of 15-sparse coefficient vectors using random Legendre matrices. The recovery performances over all uniform sample sets (solid black), as well as using preconditioned Legendre matrices with Chebyshev sampling (dotted blue) are shown for comparison.

rates associated with 20% of sample sets with lowest test values (roughly, preferable sets according to 0.8) and 20% of sets with highest test values. For comparison, we also plot the average errors and successful rates associated with all 1500 sample sets, as well as those from preconditioned Legendre matrices with Chebyshev sampling [9, 10, 17]. This technique is widely-accepted to be the optimal sampling strategy to reconstruct coefficient vectors from underdetermined Legendre matrices. To produce the results for preconditioning technique, 1500 sets of samples drawn from Chebyshev distribution in [−1, 1] are generated separately. Figure 4 reveals that preferable sets according to 0.8 are superior to general uniform sample sets and sets with high test values in sparse recovery. Although not outperforming Chebyshev sampling, preferable sets provide a simple selective strategy to improve the recovery property with uniform sampling.

The third test is similar to our second one; however, we fix the number of samples at 180 and plot the average errors and successful rates with increasing sparsity (from 5 to 60 at step size 5). Again,
sample sets with low test values show much better reconstruction, compared to general uniform sample sets and sets with high test values (Figure 5).

Finally, we perform the fourth test to illustrate our theory in case of noisy observations. Figure 6 plots the evolution of recovery errors with number of samples varying from 10 to 180 and fixed sparsity (left), as well as with sparsity varying from 5 to 60 and fixed number of samples (right). The setup here is similar to the previous tests, except that the observation $g$ is noisy: $g = Ac + \epsilon$, where each coefficient of $\epsilon$ is generated from Gaussian distribution with SD 0.001. We see that preferable sets according to 0.8 are superior to general uniform sample sets and sets with high test values as they produce smaller average recovery errors. However, unlike the noiseless experiments, all investigated approaches stagnate at errors roughly 0.005, when increasing the number of samples cannot improve the accuracy of reconstruction.
6 | CONCLUDING REMARKS

This paper provides a theoretical justification for the sparse reconstruction from underdetermined Legendre systems with uniform samples via $\ell_1$ minimization. Our analysis employs the envelop bound (rather than the prohibitive uniform bounds) of all Legendre polynomials, and by extending recent chaining arguments [4, 8] to deal with this bound, allows us to establish a new uniform recovery guarantee for sparse, multi-dimensional Legendre expansions, which is independent of the polynomial subspaces. To the best of our knowledge, this is the first time recovery condition is established for orthonormal systems without assuming the uniform boundedness of the sampling matrix. Extending the present results to related scenarios, such as non-uniform recovery and discrete least squares, to relax the dependence of sample complexity on Legendre uniform bound would be the next logical step. Also, we believe that the analysis approach herein should be of interest on its own, and can be applied elsewhere, to other random systems which can be bounded precisely pointwise, but whose uniform bound is bad. Our approach of combining the chaining techniques with envelop bound is general; however, our analysis relies on the close-form envelop bound of Legendre polynomials, and extension to other orthogonal systems requires adapting many estimates in Sections 3.2 and 4.2 to the new envelop bounds. Whether it is possible to derive a general analysis which covers more abstract forms of envelop bounds is open to us, and is one direction we would like to pursue next.

ACKNOWLEDGMENTS

This material is based upon work supported in part by: the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contracts and awards ERKJ314, ERKJ331, ERKJ345; the AEOLUS (Advances in Experimental Design, Optimization and Learning for Uncertain Complex Systems) Department of Energy Mathematical Multifaceted Capabilities Center; the Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute under Contract No. DE-AC02-05CH11231; and by the Laboratory Directed Research and Development program at the Oak Ridge National Laboratory, which is operated by UT-Battelle, LLC., for the U.S. Department of Energy under contract DE-AC05-00OR22725.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Hoang Tran https://orcid.org/0000-0001-7068-5621
Clayton Webster https://orcid.org/0000-0002-1375-0359

REFERENCES

[1] J. Bourgain, “An improved estimate in the restricted Isometry problem,” Lecture notes in mathematics, Springer International Publishing, Switzerland, 2014, pp. 65–70.
[2] A. Chkifa, N. Dexter, H. Tran, and C. Webster, Polynomial approximation via compressed sensing of high-dimensional functions on lower sets, Math. Comput. 87 (2018), no. 311, 1415–1450.
[3] E. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006), no. 1, 489–509.
[4] D. L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006), no. 4, 1289–1306.
[5] L. Mathelin and K. Gallivan, A compressed sensing approach for partial differential equations with random input data, Commun. Comput. Phys. 12 (2012), 919–954.

[6] L. Yan, L. Guo, and D. Xiu, Stochastic collocation algorithms using $\ell_1$-minimization, Int. J. Uncertain. Quantif. 2 (2012), no. 3, 279–293.

[7] H. Rauhut and C. Schwab, Compressive sensing Petrov-Galerkin approximation of high dimensional parametric operator equations, Math. Comput. 86 (2017), 661–700.

[8] J. Peng, J. Hampton, and A. Doostan, A weighted $\ell_1$-minimization approach for sparse polynomial chaos expansions, J. Comput. Phys. 267 (2014), 92–111.

[9] J. Hampton and A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, J. Comput. Phys. 280 (2015), 363–386.

[10] J. Jakeman, A. Narayan, and T. Zhou, A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions, SIAM J. Sci. Comput. 39 (2017), no. 3, A1114–A1144.

[11] S. Brugiapaglia, F. Nobile, S. Perotto, and S. Micheletti, A theoretical study of COMPRESSED SolutING for advection-diffusion-reaction problems, Math. Comput. 87 (2018), no. 309, 1–38.

[12] B. Adcock, Infinite-dimensional compressed sensing and function interpolation, Found. Comput. Math. 18 (2018), no. 3, 661–701.

[13] N. Dexter, H. Tran, and C. Webster, A mixed $\ell_1$ regularization approach for sparse simultaneous approximation of parameterized PDEs, ESAIM: M2AN. 53 (2019), no. 6, 2025–2045.

[14] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Commun. Pure Appl. Math. 61 (2008), 1025–1045.

[15] S. Foucart and H. Rauhut, A mathematical introduction to compressive sensing, Applied and Numerical Harmonic Analysis, Birkhäuser, New York, NY, 2013.

[16] I. Haviv, O. Regev, The restricted isometry property of subsampled Fourier matrices (Proc. Twenty-Seventh Annu. ACM-SIAM Symp. Discret. Algor.), 2016, pp. 288–297.

[17] H. Rauhut and R. Ward, Sparse Legendre expansions via $\ell_1$-minimization, J. Approx. Theory 164 (2012), 517–533.

[18] H. Rauhut and R. Ward, Interpolation via weighted $\ell_1$-minimization, Appl. Comput. Harmon. Anal. 40 (2016), no. 2, 321–351.

[19] A. Cohen, M. A. Davenport, and D. Leviatan, On the stability and accuracy of least squares approximations, Found. Comput. Math. 13 (2013), 819–834.

[20] P. Bickel, Y. Ritov, and A. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, Ann. Stat. 37 (2009), no. 4, 1705–1732.

[21] V. Koltchinskii, Sparsity in penalized empirical risk minimization, Ann. Inst. Henri Poincare (B) Probab. Stat. 45 (2009), 7–57.

[22] V. Koltchinskii, The Dantzig selector and sparsity oracle inequalities, Bernoulli 15 (2009), 799–828.

[23] S. van de Geer and P. Bühlmann, On the conditions used to prove oracle results for the lasso, Electron. J. Stat. 3 (2009), 1360–1392.

[24] G. Raskutti, M. J. Wainwright, and B. Yu, Restricted eigenvalue properties for correlated Gaussian designs, J. Mach. Learn. Res. 11 (2010), 2241–2259.

[25] E. Candès and T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inf. Theory 52 (2006), no. 12, 5406–5425.

[26] H. Rauhut, “Compressive sensing and structured random matrices,” Theoretical foundations and numerical methods for sparse recovery, Radon Series on Computational and Applied Mathematics, Vol 9, M. Fornasier (ed.), de Gruyter, Berlin, New York, 2010, pp. 1–92.

[27] V. Antonov and K. Holshhevnikov, An estimate of the remainder in the expansion of the generating function for the Legendre polynomials (generalization and improvement of Bernstein’s inequality), Vestnik Leningrad. Univ. Mat. 13 (1981), 163–166.

[28] H. Robbins, A remark on Stirling’s formula, Am. Math. Mon. 62 (1955), no. 1, 26–29.

[29] E. van den Berg and M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2008), no. 2, 890–912. https://doi.org/10.1137/080714488

How to cite this article: H. Tran, and C. Webster, Analysis of sparse recovery for Legendre expansions using envelope bound, Numer. Methods Partial Differ. Eq. 38 (2022), 2163–2198. https://doi.org/10.1002/num.22877