

## POLYNOMIAL APPROXIMATION VIA COMPRESSED SENSING OF HIGH-DIMENSIONAL FUNCTIONS ON LOWER SETS

ABDELLAH CHKIFA, NICK DEXTER, HOANG TRAN, AND CLAYTON G. WEBSTER

**ABSTRACT.** This work proposes and analyzes a compressed sensing approach to polynomial approximation of complex-valued functions in high dimensions. In this context, the target function is often smooth and characterized by a rapidly decaying orthonormal expansion, whose most important terms are captured by a lower (or downward closed) set. Motivated by this fact, we present an innovative weighted  $\ell_1$ -minimization procedure with a precise choice of weights for imposing the downward closed preference. Theoretical results reveal that our computational approaches possess a provably reduced sample complexity compared to existing compressed sensing techniques presented in the literature. In addition, the recovery of the corresponding best approximation using these methods is established through an improved bound for the restricted isometry property. Our analysis represents an extension of the approach for Hadamard matrices by J. Bourgain [*An improved estimate in the restricted isometry problem*, Lecture Notes in Math., vol. 216, Springer, 2014, pp. 65–70] to the general bounded orthonormal systems, quantifies the dependence of sample complexity on the successful recovery probability, and provides an estimate on the number of measurements with explicit constants. Numerical examples are provided to support the theoretical results and demonstrate the computational efficiency of the novel weighted  $\ell_1$ -minimization strategy.

### 1. INTRODUCTION

Compressed sensing (CS) is an appealing approach for reconstructing signals from underdetermined systems, with far smaller number of measurements compared to the signal length [6, 17]. Under the sparsity or compressibility assumption of the signals, this approach enjoys a significant improvement in sample complexity in contrast to traditional methods such as discrete least squares, projection, and interpolation. As the solutions of many parameterized partial differential equations (PDEs) are known to be compressible in the sense that they are well approximated by a sparse expansion in an orthonormal system (see, e.g., [14] and the references therein), it is no surprise that the interest in applying compressed sensing techniques

---

Received by the editor February 17, 2016, and, in revised form, November 8, 2016.

2010 *Mathematics Subject Classification.* Primary 35R60, 52C17, 94A08, 94A12, 15B52; Secondary 60H25, 62M40, 68W20, 03D32, 11M50.

*Key words and phrases.* Compressed sensing, high-dimensional methods, polynomial approximation, convex optimization, downward closed (lower) sets.

This material is based upon work supported in part by: the U.S. Defense Advanced Research Projects Agency, Defense Sciences Office under contract and award numbers HR0011619523 and 1868-A017-15; the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract number ERKJ259; and 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.

to the approximation of high-dimensional functions and parameterized systems has been growing rapidly in recent years [18, 24, 30, 31, 34, 35, 39, 47, 48].

In these works, the target function is a quantity of interest (QoI) associated with the solution of a parameterized PDE of the form

$$(1.1) \quad \mathcal{D}(u, \mathbf{y}) = 0,$$

where  $\mathcal{D}$  is a differential operator and  $\mathbf{y} := (y_1, \dots, y_d)$  is a parameter vector in a compact tensor product domain  $\mathcal{U} = \prod_{k=1}^d \mathcal{U}_k \subset \mathbb{R}^d$ , e.g.,  $\mathcal{U} = [-1, 1]^d$ . The solution  $u$  to such PDEs is therefore a map  $\mathbf{y} \in \mathcal{U} \mapsto u(\mathbf{y}) \in \mathcal{V}$  where  $\mathcal{V}$  is the solution space, typically a Sobolev space, e.g.,  $\mathcal{V} = H_0^1$ . The algorithms proposed in the previously cited works are designed to approximate a QoI consisting of a function  $g : \mathbf{y} \in \mathcal{U} \mapsto G(u(\mathbf{y}))$  which, e.g., is either the evaluation of  $u$  at a fixed point of the space/time domain or a linear functional in  $u$ . Introducing  $\mathcal{F} := \mathbb{N}_0^d = \{\boldsymbol{\nu} = (\nu_k)_{k=1}^d : \nu_k \in \mathbb{N}_0\}$ , and a measure  $\varrho : \mathcal{U} \rightarrow \mathbb{R}_+$  with  $\varrho(\mathbf{y}) = \prod_{k=1}^d \varrho_k(y_k)$ , the resulting functions are smooth, complex-valued, and can be expanded in an  $L^2(\mathcal{U}, d\varrho)$ -orthonormal basis  $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{F}}$  according to

$$(1.2) \quad g(\mathbf{y}) = \sum_{\boldsymbol{\nu} \in \mathcal{F}} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}(\mathbf{y}),$$

where  $\Psi_{\boldsymbol{\nu}} = \prod_{k=1}^d \Psi_{\nu_k}$  are tensor products of  $L^2(\mathcal{U}_k, d\varrho_k)$ -orthonormal polynomials, and the coefficients  $c_{\boldsymbol{\nu}}$  belong to  $\mathbb{C}$ . The series (1.2) is generally referred to as the polynomial chaos (PC) expansion of  $g$  (see, e.g., [22, 46]), whose convergence rates are well understood [39]. The orthonormal systems of particular interest in this work consist of Legendre and Chebyshev expansions. The polynomial approximation of the function  $g$  in the CS setting is fairly straightforward. First, one truncates the expansion (1.2) in the multivariate polynomial space

$$(1.3) \quad \mathbb{P}_{\mathcal{J}} := \text{span}\{\Psi_{\boldsymbol{\nu}}(\mathbf{y}) : \boldsymbol{\nu} \in \mathcal{J}\}$$

with  $\mathcal{J} := \{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_N\}$  a finite set of indices whose cardinality  $N := \#\mathcal{J}$  is large enough to yield  $g \simeq \sum_{\boldsymbol{\nu} \in \mathcal{J}} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ . Then, for some  $m \leq N$ , generate  $m$  samples  $\mathbf{y}_1, \dots, \mathbf{y}_m$  in the parametric domain  $\mathcal{U}$  independently from the orthogonalization measure  $\varrho$  associated with  $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{F}}$ , and find an approximation  $g^{\#}$  of  $g$  of the form

$$(1.4) \quad g^{\#} = \sum_{\boldsymbol{\nu} \in \mathcal{J}} c_{\boldsymbol{\nu}}^{\#} \Psi_{\boldsymbol{\nu}},$$

where  $\mathbf{c}^{\#} := (c_{\boldsymbol{\nu}}^{\#})_{\boldsymbol{\nu} \in \mathcal{J}}$  is the sparsest signal with an inherent interpolatory aspect, i.e., among solutions  $\mathbf{z}$  of underdetermined system  $\Psi \mathbf{z} = \mathbf{g}$ . Here, the matrix  $\Psi \in \mathbb{R}^{m \times N}$  contains the samples of the PC basis and the vector  $\mathbf{g}$  is the observation of the target function, i.e.,

$$(1.5) \quad \Psi := \left( \Psi_{\boldsymbol{\nu}_j}(\mathbf{y}_i) \right)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq N}} \quad \text{and} \quad \mathbf{g} := (g(\mathbf{y}_1), \dots, g(\mathbf{y}_m)),$$

respectively. In practice, noisy formulations of this problem are also considered by investigating the expansion tail  $\sum_{\boldsymbol{\nu} \notin \mathcal{J}} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}$ .

To date, the sparse recovery of the polynomial expansion (1.2) via CS has shown to be very promising. However, this approach requires a low uniform bound of the underlying basis, given by

$$\Theta = \sup_{\boldsymbol{\nu} \in \mathcal{J}} \|\Psi_{\boldsymbol{\nu}}\|_{L^\infty(\mathcal{U})},$$

as the sample complexity  $m$  required to recover the best  $s$ -term approximation (up to a multiplicative constant) scales with the following bound (see, e.g., [20]):

$$(1.6) \quad m \gtrsim \Theta^2 s \times \log \text{ factors.}$$

This poses a challenge for many multivariate polynomial approximation strategies as  $\Theta$  is prohibitively large in high dimensions. In particular, for  $d$ -dimensional problems,  $\Theta = 2^{d/2}$  for Chebyshev systems and preconditioned Legendre systems [40]. Moreover, when using the standard Legendre expansion, the theoretical number of samples can exceed the cardinality of the polynomial subspace, unless the subspace a priori excludes all terms of high total order (see, e.g., [24, 47]). Therefore, the advantages of sparse polynomial recovery methods, coming from reduced query complexity, are eventually overcome by the curse of dimensionality, in that such techniques require at least as many samples as traditional sparse interpolation techniques in high dimensions [23, 32, 33].

Nevertheless, in many engineering and science applications, the target functions, despite being high-dimensional, are smooth and often characterized by a rapidly decaying polynomial expansion, whose most important coefficients are of low order [12, 14, 15, 26]. In such situations, the quest for finding the approximation containing the largest  $s$  terms can be restricted to polynomial spaces associated with lower (or downward closed) sets.

**Definition 1.1** (Lower set). An index set  $\Lambda \subset \mathcal{F}$  is called a lower set (also called *downward closed set*) if and only if

$$(1.7) \quad \nu \in \Lambda \quad \text{and} \quad \mu \leq \nu \implies \mu \in \Lambda,$$

where  $\mu \leq \nu$  if and only if  $\mu_k \leq \nu_k$  for all  $1 \leq k \leq d$ .

The practicality of downward closed sets is mainly computational, and has been demonstrated in different approaches such as quasi-optimal strategies [3], Taylor expansion [9], interpolation methods [11], and discrete least squares [10]. For instance, in the context of parametric PDEs such as (1.1), it was shown in [12] that for a large class of operators  $\mathcal{D}$  with a certain type of anisotropic dependence on  $\mathbf{y}$ , the solution map  $\mathbf{y} \mapsto u(\mathbf{y})$  can be approximated by best  $s$ -term PC expansions associated with index sets of cardinality  $s$ , resulting in algebraic rates  $s^{-\alpha}$ ,  $\alpha > 0$  in the uniform and/or mean average sense. The same rates are preserved with index sets that are lower. In addition, for  $\mathcal{U} = [-1, 1]^d$ , such lower sets of cardinality  $s$  also enable the equivalence property  $\|\cdot\|_{L^2(\mathcal{U}, d_{\mathcal{Q}})} \leq \|\cdot\|_{L^\infty} \leq s^\gamma \|\cdot\|_{L^2(\mathcal{U}, d_{\mathcal{Q}})}$  in arbitrary dimensions  $d$  with, e.g.,  $\gamma = 2$  for the uniform measure and  $\gamma = \frac{\log 3}{\log 2}$  for Chebyshev measure.

This paper is focused on developing and analyzing CS approximations confined to downward closed sets, used to overcome the curse of dimensionality in the sampling complexity bound (1.6). As such, our work also provides a fair comparison with existing numerical polynomial approaches in high dimensions [3, 9–11, 43]. To achieve our goal, we study a sparse recovery approach for imposing the downward closed structure, namely a weighted  $\ell_1$ -minimization with the specific choice of weights  $\omega_\nu = \|\Psi_\nu\|_{L^\infty(\Omega)}$ , and also briefly survey an iterative hard thresholding method constrained to lower sets. In addition, we develop a rigorous theoretical framework that provides the analytic evidence for the improved performance of our proposed methods in reconstructing smooth functions.

In the context of CS, it is a well-established fact that sparse recovery is strongly tied to the concept of the restricted isometry property (RIP) of the (normalized) sampling matrix  $\Psi$ . However, motivated by the fact that the best  $s$ -term approximation is typically associated with a lower set, herein we adapt a weaker version of the RIP which we call the *lower RIP*. Unlike the standard RIP which requires all submatrices formed by  $s$  columns of  $\Psi$  to be well conditioned, the lower RIP involves only  $s$ -tuples of columns whose indices form a lower set. Given the lower RIP assumption, we establish stable and robust reconstruction guarantees for the *best lower  $s$ -term approximation* of  $g$ , which is the best among all approximations of  $g$  supported on lower index sets of cardinality  $s$ . It is reasonable to expect that this approximation, while weaker, is close to best  $s$ -term approximation for smooth functions  $g$  considered throughout this effort.

More importantly, the improved sample complexity for high-dimensional function recovery, using our methods, can be deduced directly from the sufficient condition for lower RIP. For clarification, a complete technical description of (1.6) is given by the condition

$$(1.8) \quad m \geq C\Theta^2 s \log^3(s) \log(N),$$

which was developed in [7, 37, 42], and is often cited in the case of the standard RIP, used to guarantee uniform recovery with probability exceeding  $1 - N^{-\log^3(s)}$ . In this work, we develop three critical components that enable us to systematically reduce the number of samples given by (1.8):

- (1) The lower RIP is associated with downward closed sets which allows us to employ efficient bounds of basis functions defined on those sets, derived in [10, 13] for discrete least squares, and replace  $\Theta^2 s$  by

$$(1.9) \quad K(s) = \sup_{\substack{\Lambda \subset \mathcal{J}, \Lambda \text{ lower} \\ \#\Lambda = s}} \left\| \sum_{\nu \in \Lambda} |\Psi_\nu|^2 \right\|_{L^\infty(\mathcal{U})},$$

which is significantly smaller.

- (2) We can reasonably choose  $\mathcal{J}$  as the Hyperbolic Cross index set

$$(1.10) \quad \mathcal{H}_s := \left\{ \nu \in \mathcal{F} : \prod_{k=1}^d (\nu_k + 1) \leq s \right\},$$

which is the smallest set that surely contains the best lower  $s$  indices (i.e., the union of all lower sets of cardinality  $s$ ). The cardinality of  $\mathcal{H}_s$  grows mildly in  $s$  and  $d$ , compared to other common choices such as tensor product and total degree. Indeed, from [8, Theorem 3.7] and [29], we have  $N := \#\mathcal{H}_s \leq \min\{2s^3 4^d, e^2 s^{2+\log_2(d)}\}$ , which facilitates both slow growth of  $m$  with respect to the dimension  $d$  and acceleration of matrix-vector multiplication.

- (3) We extend the chaining arguments, recently developed in [5, 25] for unitary matrices, to general bounded orthonormal systems, so as to decrease the logarithm factor in (1.8) by one unit. Following the approach in [5], we modify the covering argument for this task. In addition, we provide the technical details necessary to quantify the universal constants, and the constraint of the number of samples  $m$  on the success probability. It is worth noting that our analysis shows a success probability slightly weaker than that associated with (1.8).

By combining all the above ingredients, the analysis herein (see Theorems 2.2 and 3.3) details the improvements to (1.8), by showing that the sufficient condition required to reconstruct the best lower  $s$ -term approximation, with probability exceeding  $1 - N^{-\log(s)}$  is given by

$$(1.11) \quad m \geq CK(s) \log^2(s) \min \left\{ \log(s) + d, \log(s) \log(d) \right\}.$$

As shown in Lemma 3.5, for  $\mathcal{J} = \mathcal{H}_s$  in high dimensions, i.e.,  $2^d > s$ ,

$$(1.12) \quad \Theta^2 s \geq \begin{cases} s^2/2, & \text{if } (\Psi_\nu) \text{ is Chebyshev basis,} \\ s^{\frac{\log 3}{\log 2} + 1}/3, & \text{if } (\Psi_\nu) \text{ is Legendre basis,} \end{cases}$$

while as indicated in Lemma 3.7,

$$(1.13) \quad K(s) \leq \begin{cases} s^{\frac{\log 3}{\log 2}}, & \text{if } (\Psi_\nu) \text{ is Chebyshev basis,} \\ s^2, & \text{if } (\Psi_\nu) \text{ is Legendre basis.} \end{cases}$$

Therefore, the advantage of our sample complexity, given by (1.11), compared to the well-known condition (1.8), is that our sufficient requirements for recovery possess: lower order of  $s$ ; lower order in the logarithmic factor; and an efficient and explicit definition of  $\log(N)$  given by  $\min\{\log(s) + d, \log(s) \log(d)\}$ .

**1.1. Related works.** Our lower RIP is a specific case of the weighted RIP concept introduced in [41], several results on which carry over into our context. The weighted RIP, while applicable to general weights, only gives the recovery guarantee for the best weighted  $s$ -term approximation, which is incomparable to and, in the case of large weights, much weaker than the best  $s$ -term approximation in regard to the number of terms to be recovered. The advantage of using weights in reducing the computational complexity is therefore inconclusive. Furthermore, the selection of multi-index set  $\mathcal{J}$  in [41] depends on the weight, can be restrictive, and is not suitable in certain cases, e.g., the Chebyshev weight advocated in this paper. However, it can be beneficial to consider best weighted  $s$ -term approximations given that a finite upper bound of the coefficients on weighted  $\ell_p$ -norms is available (as in the context of parametric PDEs, [39]). By Stechkin-type estimates, e.g., [41, Eq. (5)], one can obtain an error bound under the linear scaling of the number of samples in the weighted sparsity, instead of the superlinear complexity proposed in the present work.

The idea of using the weights to boost the recovery performance of  $\ell_1$ -minimization has appeared elsewhere, e.g., in the context of regularization or removing aliasing [2, 41], as well as incorporating a priori information related to the support set or the decay of the polynomial coefficients [1, 21, 35, 48, 49]. On the contrary, our approach does not require any such a priori knowledge; for an improved recovery performance, the generic requirement on the target functions is that the multi-indices of best (largest) polynomial coefficients are captured in a lower set. Moreover, we are concerned with the RIP-based uniform recovery of high-dimensional functions on lower set, which requires knowledge of the expansion tail, whereas similar weighted  $\ell_1$  strategies have been studied in [1] using the RIPless approach, in which the tail estimates are not needed. Similar sample complexity for sparse Legendre and Chebyshev approximations has been achieved; furthermore, it has been remarked in [1] that the required number of measurements agrees with the best known conditions for approximations on a fixed lower set via discrete least

squares [10] and cannot be improved in case of Legendre expansions [36]. In this sense, our choice of weights is optimal.

The RIP estimate herein extends the strategy in [5], introduced to improve the standard RIP for Hadamard matrices, to the general case of continuous bounded orthonormal systems. Upon completion of this work, we became aware of the work [25], in which a different strategy of net constructions was introduced, leading to a reduction in sample complexity (1.8) by one logarithmic factor, as well as an improved dependence on restricted isometry constant. While [25] is only concerned with asymptotic estimates for Fourier matrices, we believe that one might extend such arguments to the setting in the present effort.

**1.2. Notation and preliminaries.** Throughout this paper, we use  $C$  to denote a generic positive constant whose value may be different from place to place but which is independent of any parameters. For  $\Lambda \subset \mathcal{F}$ ,  $\Lambda^c$  denotes the complement of  $\Lambda$ ,  $\mathbf{z}_\Lambda$  is the restriction of  $\mathbf{z} = (z_\nu)_{\nu \in \mathcal{F}}$  to  $\Lambda$ . For convenience,  $\|\cdot\|_{L^\infty} := \|\cdot\|_{L^\infty(\mathcal{U})}$ . Given the multi-index notation  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_d) \in \mathcal{F}$ , we define

$$\text{supp}(\boldsymbol{\nu}) := \{k : \nu_k \neq 0\}, \quad \|\boldsymbol{\nu}\|_0 := \#(\text{supp}(\boldsymbol{\nu})).$$

For  $g = \sum_{\boldsymbol{\nu} \in \mathcal{F}} c_\nu \Psi_\nu$  and  $\Lambda$  a set of indices, we denote  $g_\Lambda := \sum_{\boldsymbol{\nu} \in \Lambda} c_\nu \Psi_\nu$ . We normalize the sampling matrix and observation in (1.5) as

$$(1.14) \quad \mathbf{A} := \frac{\boldsymbol{\Psi}}{\sqrt{m}} = \left( \frac{\Psi_{\nu_j}(\mathbf{y}_i)}{\sqrt{m}} \right)_{\substack{1 \leq i \leq m, \\ 1 \leq j \leq N}}, \quad \tilde{\mathbf{g}} := \frac{\mathbf{g}}{\sqrt{m}} = \left( \frac{g(\mathbf{y}_i)}{\sqrt{m}} \right)_{1 \leq i \leq m}.$$

Also, the normalized expansion tail is referred to as

$$(1.15) \quad \boldsymbol{\xi} := \left( \frac{g_{\mathcal{J}^c}(\mathbf{y}_i)}{\sqrt{m}} \right)_{1 \leq i \leq m}.$$

Under the newly introduced notation, the exact coefficients  $\mathbf{c} = (c_\nu)_{\nu \in \mathcal{J}}$  satisfy  $\mathbf{A}\mathbf{c} + \boldsymbol{\xi} = \tilde{\mathbf{g}}$ . Assuming that  $\boldsymbol{\xi}$  is small (whose a priori upper bound is assumed if  $\ell_1$ -minimization is used), we approximate  $g$  via  $g^\# = \sum_{\boldsymbol{\nu} \in \mathcal{J}} c_\nu^\# \Psi_\nu$ , where  $\mathbf{c}^\# = (c_\nu^\#)_{\nu \in \mathcal{J}}$  is among the solutions  $\mathbf{z}$  of  $\mathbf{A}\mathbf{z} \approx \tilde{\mathbf{g}}$ .

**1.3. Organization.** Our paper is organized as follows. First, using the recently developed chaining technique, in Section 2 a new RIP estimate for general bounded orthonormal systems is provided. To avoid unimportant technicalities, the discussion in this section will focus on standard RIP, however, the analysis is general and does not depend on whether standard RIP or lower RIP is considered. In Section 3, we describe the new mathematical tools necessary to establish the concept of the lower RIP and the sparse recovery on lower sets. Section 4 is devoted to presenting the innovative theoretical results and analysis for polynomial approximation using our versions of weighted  $\ell_1$ -minimization. Several high-dimensional computational experiments supporting the theory are given in Section 5. Finally, the complete technical details of our RIP estimate proofs and a discussion on an alternative recovery approach on lower sets based on iterative hard thresholding can be found in Appendix A and Appendix B, respectively.

## 2. IMPROVED RIP ESTIMATE FOR BOUNDED ORTHONORMAL SYSTEM

The restricted isometry property (RIP) is an important ingredient for sparse recovery guarantees, which is given by the following definition.

**Definition 2.1** (RIP). Given a matrix  $\mathbf{A} \in \mathbb{C}^{m \times N}$ , we say that  $\mathbf{A}$  satisfies the restricted isometry property of order  $s$  if there exists  $0 < \delta_s < 1$  such that

$$(2.1) \quad (1 - \delta_s) \|\mathbf{z}\|_2^2 \leq \|\mathbf{Az}\|_2^2 \leq (1 + \delta_s) \|\mathbf{z}\|_2^2,$$

for all  $\mathbf{z} \in \mathbb{C}^N$  satisfying  $\#(\text{supp}(\mathbf{z})) \leq s$ . The smallest positive number  $\delta_s$  for (2.1) to hold is referred to as the restricted isometry constant of  $\mathbf{A}$ .

We prove the following RIP estimate for bounded orthonormal system, inspired by the approach in [5].

**Theorem 2.2.** Let  $\delta, \gamma$  be fixed parameters with  $0 < \delta < 1/13$ ,  $0 < \gamma < 1$  and let  $\{\Psi_\nu\}_{\nu \in \mathcal{J}}$  be an orthonormal system of finite size  $N = \#(\mathcal{J})$ . Assume that

$$(2.2) \quad m \geq 2^6 e \frac{\Theta^2 s}{\delta^2} \log\left(\frac{\Theta^2 s}{\delta^2}\right) \max\left\{\frac{2^5}{\delta^4} \log\left(40 \frac{\Theta^2 s}{\delta^2} \log\left(\frac{\Theta^2 s}{\delta^2}\right)\right) \log(4N), \frac{1}{\delta} \log\left(\frac{1}{\gamma \delta} \log\left(\frac{\Theta^2 s}{\delta^2}\right)\right)\right\},$$

and  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$  are drawn independently from the orthogonalization measure  $\varrho$  associated to  $\{\Psi_\nu\}$ . Then with probability exceeding  $1 - \gamma$ , the normalized sampling matrix  $\mathbf{A} \in \mathbb{C}^{m \times N}$  satisfies

$$(2.3) \quad (1 - 13\delta) \|\mathbf{z}\|_2^2 \leq \|\mathbf{Az}\|_2^2 \leq (1 + 13\delta) \|\mathbf{z}\|_2^2,$$

for all  $\mathbf{z} \in \mathbb{C}^N$ ,  $\#(\text{supp}(\mathbf{z})) \leq s$ .

*Remark 2.3.* We remark that the results of Theorem 2.2 is not limited to the Legendre and Chebyshev systems on the unit hypercube, which are of main interest in this paper, but also holds for any bounded orthonormal system on any underlying domain, discrete or continuous.

The complete detailed proof of Theorem 2.2 is given in Appendix A.2. However, to assist the reader in better understanding the logic of our proof we next provide a sketch that explains the essential features on how we achieved the improved RIP estimate.

*Sketch of proof.* To begin, let us denote

$$\begin{aligned} \psi(\mathbf{y}, \mathbf{z}) &:= \sum_{\nu \in \mathcal{J}} z_\nu \Psi_\nu(\mathbf{y}), \quad \forall \mathbf{y} \in \mathcal{U}, \mathbf{z} \in \mathbb{C}^N, \\ \text{and } \mathcal{E}_s &:= \{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_2 = 1, \#(\text{supp}(\mathbf{z})) \leq s\}. \end{aligned}$$

Our goal is to derive conditions on  $m$  such that for a set of  $m$  random samples  $\{\mathbf{y}_i\}_{i=1}^m \subset \mathcal{U}$ , drawn according to  $\varrho$ , then with high probability, there holds for all  $\mathbf{z} \in \mathcal{E}_s$ :

$$(2.4) \quad \frac{1}{m} \sum_{i=1}^m |\psi(\mathbf{y}_i, \mathbf{z})|^2 \approx \int_{\mathcal{U}} |\psi(\mathbf{y}, \mathbf{z})|^2 d\varrho.$$

Note that the LHS of the above expression equals  $\|\mathbf{Az}\|_2^2$ , and the RHS is  $\|\mathbf{z}\|_2^2 = 1$ ; thus (2.4) is nothing but (2.3) rewritten with the new notation. We construct a finite set  $\mathcal{S}$  of piecewise constant functions on  $\mathcal{U}$ , which provides a good approximation

of  $\psi(\cdot, \mathbf{z})$ , denoted by  $\tilde{\psi}(\cdot, \mathbf{z})$ , for all  $\mathbf{z} \in \mathcal{E}_s$ , and reduce the task of showing (2.4) holds for infinitely many  $\mathbf{z} \in \mathcal{E}_s$  to that of proving

$$(2.5) \quad \frac{1}{m} \sum_{i=1}^m |\tilde{\psi}(\mathbf{y}_i, \mathbf{z})|^2 \approx \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho$$

for all  $\tilde{\psi}(\cdot, \mathbf{z}) \in \mathcal{S}$ . In particular, the family of “discrete” approximations  $\tilde{\psi}(\cdot, \mathbf{z})$  is constructed in such a way that (see Appendix A.2, Step 1):

- (1) for any  $\mathbf{z} \in \mathcal{E}_s$ ,  $\tilde{\psi}(\cdot, \mathbf{z}) \approx \psi(\cdot, \mathbf{z})$ ;
- (2)  $\tilde{\psi}(\cdot, \mathbf{z})$  can be represented as a piecewise constant function on  $\mathcal{U}$ :  $\tilde{\psi}(\cdot, \mathbf{z}) = \sum_{l \in \mathcal{L}} \tilde{\psi}_l^{\mathbf{z}}(\cdot)$ , where each  $\tilde{\psi}_l^{\mathbf{z}}$  is a constant function, pairwise disjointly supported on a subset of  $\mathcal{U}$ , representing a scale of  $\tilde{\psi}(\cdot, \mathbf{z})$  and  $\mathcal{L}$  is a finite set of scale; and
- (3) for each  $l \in \mathcal{L}$ , the class  $\{\tilde{\psi}_l^{\mathbf{z}} : \mathbf{z} \in \mathcal{E}_s\}$  is finite, whose cardinality is as small as possible.

With the use of (1) and (2) one can establish the bound (see Appendix A.2, Step 2):

$$(2.6) \quad \left| \frac{1}{m} \sum_{i=1}^m |\psi(\mathbf{y}_i, \mathbf{z})|^2 - \int_{\mathcal{U}} |\psi(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| \lesssim \left| \frac{1}{m} \sum_{i=1}^m |\tilde{\psi}(\mathbf{y}_i, \mathbf{z})|^2 - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| \leq \sum_{l \in \mathcal{L}} \left| \frac{1}{m} \sum_{i=1}^m |\tilde{\psi}_l^{\mathbf{z}}(\mathbf{y}_i)|^2 - \int_{\mathcal{U}} |\tilde{\psi}_l^{\mathbf{z}}(\mathbf{y})|^2 d\varrho \right|.$$

Using the basic tail estimate given by Lemma A.2 yields for any  $l$  and  $\mathbf{z}$ , with high probability

$$(2.7) \quad \frac{1}{m} \sum_{i=1}^m \tilde{\psi}_l^{\mathbf{z}}(\mathbf{y}_i) \approx \int_{\mathcal{U}} \tilde{\psi}_l^{\mathbf{z}}(\mathbf{y}) d\varrho.$$

We can then obtain (2.4) by employing (2.6) and applying the union bound for (2.7) over all  $l, \mathbf{z}$  (see Appendix A.2, Step 3). For this argument to yield small  $m$ , it is critical to construct  $\tilde{\psi}(\cdot, \mathbf{z})$  so that the total number of functions  $\tilde{\psi}_l^{\mathbf{z}}$  (over  $l \in \mathcal{L}, \mathbf{z} \in \mathcal{E}_s$ ) is finite and minimal, justifying the requirement given by (3) above.

As an example, for each  $l \in \mathcal{L}$ , we can define  $\tilde{\psi}_l^{\mathbf{z}}$  according to a covering of  $\mathcal{E}_s$  under the pseudo-metric

$$d(\mathbf{z}, \mathbf{z}') = \sup_{\mathbf{y} \in \mathcal{U}} |\psi(\mathbf{y}, \mathbf{z} - \mathbf{z}')|,$$

so that  $\#\{\tilde{\psi}_l^{\mathbf{z}} : \mathbf{z} \in \mathcal{E}_s\}$  is roughly a covering number of  $\mathcal{E}_s$ . However, one can check that in our high-dimensional setting, this covering number grows exponentially in the dimension of  $\mathcal{U}$ . Fortunately, an inspection of  $|\psi(\cdot, \mathbf{z} - \mathbf{z}')|$  reveals that these functions often have tall spikes in a small subregion of  $\mathcal{U}$ , while they are relatively small for the rest of the domain. This motivates us to consider a new “distance” between  $\mathbf{z}$  and  $\mathbf{z}'$ , which is significantly smaller than  $d(\mathbf{z}, \mathbf{z}')$ , given by an upper bound of  $|\psi(\mathbf{y}, \mathbf{z} - \mathbf{z}')|$  for *most*  $\mathbf{y} \in \mathcal{U}$ . More rigorously, we define

$$d_\zeta(\mathbf{z}, \mathbf{z}') := \inf_{\substack{\tilde{\mathcal{U}} \subset \mathcal{U} \\ \varrho(\tilde{\mathcal{U}}) = 1 - \zeta}} \sup_{\mathbf{y} \in \tilde{\mathcal{U}}} |\psi(\mathbf{y}, \mathbf{z} - \mathbf{z}')|.$$

Although  $d_\zeta$  is not a proper pseudo-metric, an adaptation of the covering number result can still be derived in this case (see Lemma A.3). This argument is similar



in spirit to [25, Lemma 3.5]. The approximation  $\tilde{\psi}$ , constructed with  $d_\varsigma$ , may not agree with  $\psi$  in a small subdomain of  $\mathcal{U}$ , but one can tune  $\varsigma$  so as to not affect the estimate (2.6).

This completes the sketch of the main proof. □

*Remark 2.4.* In brief, the RIP (and subsequently, best  $s$ -term reconstruction) occurs with probability exceeding  $1 - \gamma$  under the condition

$$(2.8) \quad m \geq C\Theta^2 s \max\{\log^2(\Theta^2 s) \log(N), \log(\Theta^2 s) \log(\log(\Theta^2 s)/\gamma)\}.$$

The first constraint in (2.8) therefore reduces the order of  $\log(s)$  in (1.8) by one unit. The second constraint, on the other hand, has an additional  $\log$  factor compared to the well-known one, i.e.,  $m \geq C\Theta^2 s \log(1/\gamma)$  (see [37]), after balancing leading to a weaker success probability, as discussed in Section 1.

### 3. SPARSE RECOVERY ON LOWER SETS

In this section we focus on a smooth  $g$ , given by (1.2), and exploit the fast decay of its polynomial expansion to further improve (2.8). Central to this task is the concept of lower or downward closed sets, given Definition 1.1. With this in mind, instead of best  $s$ -term approximations, we are interested in *best lower  $s$ -term approximation* of  $g$ , which is the best among all approximations of  $g$  supported on lower sets of cardinality  $s$ . More rigorously, let  $\Lambda^*$  be a lower subset of  $\mathcal{F}$  which realizes the infimum

$$(3.1) \quad \Lambda^* := \arg \min_{\substack{\Lambda \text{ lower} \\ \#(\Lambda) \leq s}} \|g - g_\Lambda\|,$$

where the norm to be specified later. Here  $\|g - g_{\Lambda^*}\|$  is the *best lower  $s$ -term error*, and our goal is to find approximations of  $g$  with error scaling linearly in  $\|g - g_{\Lambda^*}\|$ . We expect the best lower  $s$ -term error, while generally larger, is close to best  $s$ -term error in our setting. These quantities are particularly identical provided that  $g$  is  $s$ -sparse,  $\text{supp}(g)$  lower, represented by finite Legendre and Chebyshev expansions.

To achieve our goal, it is reasonable to consider a relaxed version of RIP that specifically involves  $s$ -tuples of columns associated with lower sets. Given a multi-index set  $\Lambda \subset \mathcal{F}$ , we introduce the quantity

$$(3.2) \quad K(\Lambda) := \left\| \sum_{\nu \in \Lambda} |\Psi_\nu|^2 \right\|_{L^\infty}$$

and, with an abuse of notation, denote

$$(3.3) \quad K(s) := \sup_{\substack{\Lambda \subset \mathcal{J}, \Lambda \text{ lower} \\ \#(\Lambda) = s}} K(\Lambda),$$

which has already been mentioned in (1.9). We define next the *lower restricted isometry property (lower RIP)*. This property is exclusive to the present setting and defined here for submatrices whose columns are associated with indices  $\nu \in \mathcal{F}$ .

**Definition 3.1** (Lower RIP). For  $\mathbf{A} \in \mathbb{C}^{m \times N}$  as in (1.14), we say that  $\mathbf{A}$  satisfies the lower restricted isometry property of order  $s$  if there exists  $0 < \delta_{\ell,s} < 1$  such that

$$(3.4) \quad (1 - \delta_{\ell,s})\|z\|_2^2 \leq \|\mathbf{A}z\|_2^2 \leq (1 + \delta_{\ell,s})\|z\|_2^2$$

for all  $\mathbf{z} \in \mathbb{C}^N$  satisfying  $K(\text{supp}(\mathbf{z})) \leq K(s)$ . The smallest positive number  $\delta_{\ell,s}$  for (3.4) to hold is referred to as the lower restricted isometry constant of  $\mathbf{A}$ .

*Remark 3.2.* The lower RIP is a specific case of the weighted RIP, introduced in [41] for general weights, here with the weights  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$ . By introducing the notation  $\|\mathbf{z}\|_{0,\omega} = \sum_{\nu \in \text{supp}(\mathbf{z})} \omega_\nu^2$  for  $\mathbf{z} \in \mathbb{C}^N$ , the weighted RIP constant  $\delta_{\omega,s}$  was defined as the smallest number  $\delta_{\omega,s}$  for which

$$(3.5) \quad (1 - \delta_{\omega,s})\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{z}\|_2^2 \leq (1 + \delta_{\omega,s})\|\mathbf{z}\|_2^2, \quad \|\mathbf{z}\|_{0,\omega} \leq s.$$

By (3.2), observe that  $K(\text{supp}(\mathbf{z})) \leq \|\mathbf{z}\|_{0,\omega}$ , hence given  $\mathbf{z}$  such that  $\|\mathbf{z}\|_{0,\omega} \leq K(s)$ , then  $K(\text{supp}(\mathbf{z})) \leq K(s)$  so that (3.4) is satisfied showing that  $\delta_{\omega,K(s)} \leq \delta_{\ell,s}$ . For the Chebyshev and Legendre systems, the polynomials  $\Psi_\nu$  all attain their supremums at  $(1, \dots, 1)$ , hence for any  $\mathbf{z} \in \mathbb{C}^N$ , then  $K(\text{supp}(\mathbf{z})) = \|\mathbf{z}\|_{0,\omega}$ , showing that

$$(3.6) \quad \delta_{\ell,s} = \delta_{\omega,K(s)}.$$

Note the change of order in this relation: loosely speaking, the lower RIP of order  $s$  corresponds to the weighted RIP of order  $K(s)$ .

An important subclass of  $\mathbf{z}$  satisfying (3.4) is  $\mathbf{z} \in \mathbb{C}^N$  with  $\#(\text{supp}(\mathbf{z})) \leq s$  and  $\text{supp}(\mathbf{z})$  lower. One may want to consider a more natural isometry property which requires (3.4) for only vectors  $\mathbf{z}$  in the above class. We can see from the following analysis that this property is weaker but requires the same sampling cost as (3.4). The sample complexity for lower RIP is established in the following theorem.

**Theorem 3.3.** *Let  $\delta, \gamma$  be fixed parameters with  $0 < \delta < 1/13$ ,  $0 < \gamma < 1$  and let  $\{\Psi_\nu\}_{\nu \in \mathcal{J}}$  be an orthonormal system of finite size  $N = \#(\mathcal{J})$ . Assume that*

$$(3.7) \quad m \geq 2^6 e \frac{K(s)}{\delta^2} \log\left(\frac{K(s)}{\delta^2}\right) \max\left\{\frac{2^5}{\delta^4} \log\left(40 \frac{K(s)}{\delta^2} \log\left(\frac{K(s)}{\delta^2}\right)\right) \log(4N), \frac{1}{\delta} \log\left(\frac{1}{\gamma\delta} \log\left(\frac{K(s)}{\delta^2}\right)\right)\right\},$$

and  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$  are drawn independently from the orthogonalization measure  $\rho$  associated to  $\{\Psi_\nu\}$ . Then with probability exceeding  $1 - \gamma$ , the normalized sampling matrix  $\mathbf{A} \in \mathbb{C}^{m \times N}$  satisfies

$$(3.8) \quad (1 - 13\delta)\|\mathbf{z}\|_2^2 < \|\mathbf{A}\mathbf{z}\|_2^2 < (1 + 13\delta)\|\mathbf{z}\|_2^2$$

for all  $\mathbf{z} \in \mathbb{C}^N$ ,  $K(\text{supp}(\mathbf{z})) \leq K(s)$ .

The proof of Theorem 3.3 is discussed in the Appendix A.3. This proof essentially follows the same path as the proof of Theorem 2.2 with few minor changes.

*Remark 3.4.* In brief, the random matrix  $\mathbf{A}$  satisfies the lower RIP of order  $s$  and, subsequently, guarantees lower reconstruction with probability exceeding  $1 - \gamma$  if the sample size  $m$  satisfies

$$(3.9) \quad m \geq CK(s) \max\{\log^2(K(s)) \log(N), \log(K(s)) \log(\log(K(s))/\gamma)\}.$$

Next, we present a theoretical comparison between the complexity bounds required by standard RIP (2.8) and lower RIP (3.9), showing the computational cost saving with our best lower approximations. Assuming that no information about the support set or the decay of the polynomial coefficients is a priori known, we

reasonably make the choice  $\mathcal{J} = \mathcal{H}_s$ , which is the smallest set that surely contains the best lower  $s$  indices. For the sake of notational clearness, we denote  $(L_\nu)_{\nu \in \mathcal{F}}$  and  $(T_\nu)_{\nu \in \mathcal{F}}$  the Legendre and Chebyshev basis of  $L^2(\mathcal{U}, d_\varrho)$  with  $\varrho$  being the uniform or Chebyshev measure, respectively. For such polynomials, we have for any  $\nu \in \mathcal{F}$

$$(3.10) \quad \|T_\nu\|_{L^\infty} = 2^{\|\nu\|_0/2} \quad \text{and} \quad \|L_\nu\|_{L^\infty} = \prod_{k=1}^d \sqrt{2\nu_k + 1}.$$

First, we have the following sharp estimates.

**Lemma 3.5.** *Let  $\mathcal{H}_s$  be defined as in (1.10) with  $s \leq 2^{d+1}$ . There holds*

$$(3.11) \quad s/2 \leq \sup_{\nu \in \mathcal{H}_s} \|T_\nu\|_{L^\infty}^2 \leq s \quad \text{and} \quad s^{\frac{\log 3}{\log 2}}/3 \leq \sup_{\nu \in \mathcal{H}_s} \|L_\nu\|_{L^\infty}^2 \leq s^{\frac{\log 3}{\log 2}}.$$

Moreover, the upper estimates hold for all  $s \geq 1$ .

*Proof.* For  $\nu \in \mathcal{H}_s$ , it is easy to see that  $\|T_\nu\|_{L^\infty}^2 = 2^{\|\nu\|_0} \leq \prod_{k=1}^d (\nu_k + 1) \leq s$ . Also, since  $b \mapsto \frac{\log(2b+1)}{\log(b+1)}$  is decreasing over  $[1, +\infty)$ ,  $(2b + 1) \leq (b + 1)^{\log 3 / \log 2}$  for any  $b \geq 1$  which implies  $\|L_\nu\|_{L^\infty}^2 = \prod_{k=1}^d (2\nu_k + 1) \leq \prod_{k=1}^d (\nu_k + 1)^{\frac{\log 3}{\log 2}} \leq s^{\frac{\log 3}{\log 2}}$ .

On the other hand, since  $s \leq 2^{d+1}$ , then  $2^{d'} < s \leq 2^{d'+1}$  for some  $d' \leq d$ , so that the index  $\nu = e_1 + \dots + e_{d'}$  belongs to  $\mathcal{H}_s$  and yields  $\|T_\nu\|_{L^\infty}^2 = 2^{d'} \geq s/2$  and  $\|L_\nu\|_{L^\infty}^2 = 3^{d'} = (2^{d'})^{\frac{\log 3}{\log 2}} \geq (s/2)^{\frac{\log 3}{\log 2}} = s^{\frac{\log 3}{\log 2}}/3$ .  $\square$

For the polynomial systems such as Chebyshev or Legendre (or more generally Jacobi systems),  $\log(\Theta) \lesssim \log(s)$  over the hyperbolic cross  $\mathcal{H}_s$  regardless of the dimension  $d$ . An immediate consequence of the previous lemma and condition (2.8) is that the RIP can be obtained for  $\mathcal{J} = \mathcal{H}_s$  with

$$(3.12) \quad m \geq C s^{1+\beta} \max \{ \log^2(s) \log(N), \log(s) \log(\log(s)/\gamma) \},$$

where  $\beta = 1$  for Chebyshev systems and  $\beta = \frac{\log 3}{\log 2} \simeq 1.58$  for Legendre systems, respectively. The lower bounds in (3.11) show that these orders of  $s$  are sharp in high dimension ( $d \geq \log_2(s) - 1$ ). Following from the estimate  $\#(\mathcal{H}_s) \leq \varepsilon^{-1} s^{1+1/\varepsilon} (1 - \varepsilon)^{-d/\varepsilon}$ ,  $\forall 0 < \varepsilon < 1$  (see [8, Theorem 3.7]), if we set  $\varepsilon = 1/2$ , it is easy to see that

$$(3.13) \quad N \leq 2s^3 4^d.$$

Another estimate of  $N$  can be found in [29, Theorem 4.9]

$$(3.14) \quad N \leq e^2 s^{2+\log_2(d)}.$$

Taking the logarithm of  $N$  and plugging into (3.12) give the sample complexity only depending linearly on  $d$  under (3.13) and logarithmically (with another log factor of  $s$ ) under (3.14):

$$(3.15) \quad m \geq C s^{1+\beta} \max \left\{ \min \left\{ \log^3(s) + d \log^2(s), \log^3(s) \log(d) \right\}, \log(s) \log(\log(s)/\gamma) \right\}.$$

Although we have eliminated the exponential growth on  $d$ , the condition (1.6) has not been broken up to this step. Rather, the bound (3.15) is merely acquired from (1.6) with an estimate of  $\Theta$  on the Hyperbolic Cross subspace. We proceed

to detail the complexity bound of lower RIP. As the Chebyshev and Legendre polynomials attain their supremum at the point  $\mathbf{1} = (1, \dots, 1)$  with the supremums given in (3.10), the value of  $K$  defined in (3.2)-(3.3) is then known and one can derive estimates for it. For these two systems, we use the notations  $K_T(\Lambda)$ ,  $K_T(s)$ ,  $K_L(\Lambda)$  and  $K_L(s)$ , respectively, where

$$(3.16) \quad K_T(\Lambda) = \sum_{\nu \in \Lambda} 2^{\|\nu\|_0} \quad \text{and} \quad K_L(\Lambda) = \sum_{\nu \in \Lambda} \prod_{k=1}^d (2\nu_k + 1).$$

The following estimates of  $K_T(\Lambda)$  and  $K_T(s)$  can be found in [10].

**Lemma 3.6.** *Let  $\Lambda \subset \mathcal{F}$  be a lower set with  $\#\Lambda \geq 2$ . There holds*

$$(3.17) \quad 2\#\Lambda - 1 \leq K_T(\Lambda) \leq (\#\Lambda)^{\frac{\log 3}{\log 2}},$$

$$(3.18) \quad 3\#\Lambda - 1 \leq K_L(\Lambda) \leq (\#\Lambda)^2.$$

We note that the left sides in (3.17) and (3.18) follow from  $\prod_{k=1}^d (2\nu_k + 1) \geq 3$  and  $2^{\|\nu\|_0} \geq 2$  for any index  $\nu \neq 0$ . We also note that the right side inequalities are sharp, equalities hold for lower sets of the form  $\{\mu \leq e_1 + \dots + e_{d'}\}$ . An immediate implication of the Lemma 3.6 are the bounds  $2s - 1 \leq K_T(s) \leq s^{\frac{\log 3}{\log 2}}$  and  $3s - 1 \leq K_L(s) \leq s^2$  for all  $s \geq 2$ . The upper bounds are actually sharp in high dimension. We indeed have the following.

**Lemma 3.7.** *Let  $s \leq 2^{d+1}$ . There holds*

$$(3.19) \quad \frac{s^{\frac{\log 3}{\log 2}}}{3} \leq K_T(s) \leq s^{\frac{\log 3}{\log 2}}, \quad \frac{s^2}{4} \leq K_L(s) \leq s^2.$$

The upper estimates hold for all  $s \geq 2$ .

*Proof.* For  $\nu \in \mathcal{H}_s$ , the rectangular block  $\mathcal{R}_\nu := \{\mu \leq \nu\} \subset \mathcal{H}_s$  is lower and has a tensor format, so that from identities  $1 + \sum_{b'=1}^b 2 = 1 + 2b$  and  $\sum_{b'=0}^b (2b' + 1) = (1 + b)^2$ , one infers

$$K_T(\mathcal{R}_\nu) = \prod_{k=1}^d (1 + 2\nu_k) \quad \text{and} \quad K_L(\mathcal{R}_\nu) = \prod_{k=1}^d (1 + \nu_k)^2.$$

Since  $s \leq 2^{d+1}$ , then  $2^{d'} < s \leq 2^{d'+1}$  for some  $d' \leq d$ . For  $\nu = e_1 + \dots + e_{d'} \in \mathcal{H}_s$ , one obtains  $K_T(\mathcal{R}_\nu) = 3^{d'} = (2^{d'})^{\frac{\log 3}{\log 2}}$  and  $K_L(\mathcal{R}_\nu) = (2^{d'})^2$ , which implies

$$K_T(\mathcal{R}_\nu) \geq (s/2)^{\frac{\log 3}{\log 2}} = s^{\frac{\log 3}{\log 2}}/3, \quad K_L(\mathcal{R}_\nu) \geq (s/2)^2 = s^2/4,$$

which completes the proof. □

Combining the complexity bound (3.9) with the estimate (3.19), we arrive at the following RIP-based recovery guarantee of best lower  $s$ -term approximations, with probability exceeding  $1 - \gamma$ ,

$$(3.20) \quad m \geq C s^{1+\beta'} \max \left\{ \min \left\{ \log^3(s) + d \log^2(s), \log^3(s) \log(d) \right\}, \right. \\ \left. \log(s) \log(\log(s)/\gamma) \right\},$$

where  $\beta' = \frac{\log 3}{\log 2} - 1 \simeq 0.58$  for Chebyshev system and  $\beta' = 1$  for Legendre system. These conditions eliminate the dependence on  $\Theta^2 s$  at the cost of a superlinear growth on  $s$ , yet are clearly weaker than those required by standard RIP; see (3.15).

We close this section by pointing out that the dependence on  $d$  in (3.20) can be fully eliminated if instead of  $\mathcal{J} = \mathcal{H}_s$ , we work with  $\mathcal{J} = \mathcal{A}_s$ , the union of all anchored sets  $\Lambda$  of cardinality smaller than  $s$ . Such sets are described in [16] and are characterized by  $\Lambda$  lower and  $e_k \in \Lambda$  if and only if  $e_{k'} \in \Lambda$  for any  $k' = 1, \dots, k - 1$ . Indeed, it is easy to see that  $\mathcal{A}_s$  is included in the projection of  $\mathcal{H}_s$  into an  $s$ -dimensional space, hence  $d$  in (3.20) can be replaced by  $s$ . Such a subclass of lower sets is also relevant in polynomial approximation of parametric PDEs (see, e.g., [14]). It should also be emphasized that other types of polynomial spaces, e.g., Total Degree, have been attempted to overcome the fast growth of query complexity in high-dimensional problems (see, e.g., [47]). However, these approaches impose an a priori choice of the polynomial subspace unrelated to the number of terms to be reconstructed and employ the standard RIP guaranteed by (1.8). In our work,  $\mathcal{J} = \mathcal{H}_s$  is optimally determined based on the need of recovering best lower  $s$  terms, and our lower RIP requires less samples than standard RIP, as discussed throughout.

4. WEIGHTED  $\ell_1$ -MINIMIZATION FOR POLYNOMIAL APPROXIMATION ON LOWER SETS

In this section, we study a weighted  $\ell_1$ -minimization approach with a precise choice of weights that enable us to realize sparse reconstruction under the lower RIP. To begin, let  $\omega = (\omega_\nu)_{\nu \in \mathcal{F}} \in (0, \infty)^{\mathcal{F}}$  be a sequence of weights. Given a vector  $\mathbf{z} = (z_\nu)_{\nu \in \mathcal{F}}$  of complex components or a function  $g = \sum_{\nu \in \mathcal{F}} z_\nu \Psi_\nu$ , we define the weighted  $\ell_1$ -norm of  $\mathbf{z}$  and  $g$  by

$$\|g\|_{\omega,1} = \|\mathbf{z}\|_{\omega,1} := \sum_{\nu \in \mathcal{F}} \omega_\nu |z_\nu|,$$

and the *best lower  $s$ -term* error in weighted  $\ell_1$ -norm by

$$\sigma_s^{(\ell)}(\mathbf{z})_{\omega,1} := \inf_{\substack{\Lambda \text{ lower} \\ \#(\Lambda) \leq s}} \|\mathbf{z} - \mathbf{z}_\Lambda\|_{\omega,1}, \quad \sigma_s^{(\ell)}(g)_{\omega,1} := \inf_{\substack{\Lambda \text{ lower} \\ \#(\Lambda) \leq s}} \|g - g_\Lambda\|_{\omega,1}.$$

Recall that we are working with  $\mathcal{J} = \mathcal{H}_s$ , unless otherwise stated.

Assuming that an estimate  $\eta$  of the tail  $g_{\mathcal{H}_s^c}$  is available (specified later), our weighted  $\ell_1$ -minimization procedure for recovering an approximation  $g^\#$  of  $g$ , defined by

$$(4.1) \quad g^\# = \sum_{\nu \in \mathcal{H}_s} c_\nu^\# \Psi_\nu,$$

is given in the following: *Given  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$ . Find  $g^\#$  from (4.1), with  $\mathbf{c}^\# = (c_\nu^\#)_{\nu \in \mathcal{H}_s}$  being a minimizer of the following constrained optimization problem*

$$(4.2) \quad \min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to} \quad \|\tilde{\mathbf{g}} - \mathbf{A}\mathbf{z}\|_2 \leq \frac{\eta}{\sqrt{m}}.$$

The recovery guarantees that using weighted  $\ell_1$ -minimization has been analyzed in [41] for general choice of  $(\omega_\nu)_{\nu \in \mathcal{J}}$ . As discussed in Section 1, the benefit of weighting in terms of query complexity is inconclusive therein. In this work, we specifically define the weights  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$  for use with  $\ell_1$ -minimization and justify how smooth functions can be reconstructed with a significantly reduced number of samples compared to the unweighted method (thanks to lower RIP). Such choice of weight has also been studied in [1] for the RIPless recovery approach while the condition  $\omega_\nu \geq \|\Psi_\nu\|_{L^\infty}$  has been imposed elsewhere [2, 41].

In what follows we consider and analyze two scenarios using our weighted  $\ell_1$ -minimization procedure (4.1)-(4.2). First, we assume only an upper estimate of the tail  $g_{\mathcal{H}_s^c}$  is available, and second, exact knowledge of the tail is assumed.

**4.1. Given an upper estimate of the tail.** In this case, the proof of the recovery guarantee of  $g^\#$  in (4.1), using the optimization procedure (4.2), follows the arguments of general weighted  $\ell_1$ -minimization analysis (see [41, Theorem 6.1]). However, we remove the condition of large  $s$  required in the work [41], based on an improvement of the null space property specific to our setting (see Proposition 4.4). We first need some intermediate estimates.

**Lemma 4.1.** *For any  $d \geq 1$  and any  $s \geq 2$ ,*

$$(4.3) \quad K_T(2s) \geq 2K_T(s) \quad \text{and} \quad K_L(2s) \geq 4K_L(s).$$

*In addition,*

$$(4.4) \quad K_T(s) \geq \frac{3}{2} \max_{\nu \in \mathcal{H}_s} \|T_\nu\|_{L^\infty}^2 \quad \text{and} \quad K_L(s) \geq \frac{4}{3} \max_{\nu \in \mathcal{H}_s} \|L_\nu\|_{L^\infty}^2.$$

*Proof.* For  $\nu = (\nu_k)_{1 \leq k \leq d}$ , we use the notation  $\hat{\nu} = (\nu_k)_{2 \leq k \leq d}$ . Let  $\Lambda$  be a lower set of cardinality  $s$ . We introduce

$$\Lambda' := \{(2\nu_1, \hat{\nu}), (2\nu_1 + 1, \hat{\nu}) : \nu = (\nu_1, \hat{\nu}) \in \Lambda\}.$$

It is easily checked that  $\Lambda' \subset \mathcal{F}$  is lower and  $\#\Lambda' = 2\#\Lambda = 2s$ . Therefore  $\Lambda' \subset \mathcal{H}_{2s}$ . Moreover, for the tensorized Chebyshev and Legendre systems,  $\omega_\nu = \prod_{k=1}^d \omega_{\nu_k}$ , where  $\omega_{\nu_k}$  denote the sup norm in one dimension. Hence

$$(4.5) \quad K(2s) \geq K(\Lambda') = \sum_{\mu \in \Lambda'} \omega_\mu^2 = \sum_{\nu \in \Lambda} (\omega_{2\nu_1}^2 + \omega_{2\nu_1+1}^2) \omega_{\hat{\nu}}^2 \geq \sum_{\nu \in \Lambda} 2\omega_{\nu_1}^2 \omega_{\hat{\nu}}^2 = 2K(\Lambda),$$

where we have used the increase of the weights which yields  $\omega_{2\nu_1}, \omega_{2\nu_1+1} \geq \omega_{\nu_1}$ . For the Legendre system, we have

$$(4.6) \quad (\omega_{2\nu_1}^2 + \omega_{2\nu_1+1}^2) = (4\nu_1 + 1 + 4\nu_1 + 3) = 4(2\nu_1 + 1) = 4\omega_{\nu_1}^2.$$

Since  $\Lambda$  is an arbitrary lower set included in  $\mathcal{H}_s$ , (4.5) and (4.6) imply (4.3).

Now let  $\nu \in \mathcal{H}_s$  be the index that maximizes  $\|T_\nu\|_{L^\infty}$  over  $\mathcal{H}_s$ . We have  $\mathcal{R}_\nu \subset \mathcal{H}_s$ , so that  $K_T(s) \geq K_T(\mathcal{R}_\nu) = \prod_{k=1}^d (1 + 2\nu_k) \geq 3^{\|\nu\|_0} = (3/2)^{\|\nu\|_0} \|T_\nu\|_{L^\infty}^2$ . Similarly, for  $\nu \in \mathcal{H}_s$  maximizing  $\|L_\nu\|_{L^\infty}$ ,  $K_L(s) \geq K_L(\mathcal{R}_\nu) = \prod_{k=1}^d (1 + \nu_k)^2 \geq (4/3)^{\|\nu\|_0} \prod_{k=1}^d (1 + 2\nu_k)$ , where we have used  $(1 + t)^2 \geq \frac{4}{3}(1 + 2t)$  for any  $t \geq 1$ . Since for  $s \geq 2$ , we get that  $\nu \neq 0$  and  $\|\nu\|_0 \geq 1$ . The proof is complete.  $\square$

*Remark 4.2.* In the previous proof, if we define  $\Lambda'$  by copying  $\Lambda$  using  $3\nu_1, 3\nu_1 + 1, 3\nu_1 + 2$ , we get  $K_T(3s) \geq 3K_T(s)$ . For convenience, in the next proposition, we will employ the estimates: for any  $d \geq 1$  and any  $s \geq 2$ ,

$$(4.7) \quad K_T(3s) \geq 3K_T(s) \quad \text{and} \quad K_L(2s) \geq 3K_L(s),$$

where the second one is slightly weaker than (4.3).

We also employ the weighted null space property result [41, Theorem 4.5], restated below with small changes in parameters, in the subsequent analysis.

**Lemma 4.3.** *Given  $\mathcal{J}$  a subset of  $\mathcal{F}$  (not necessarily  $\mathcal{H}_s$ ) and  $\omega = (\omega_\nu)_{\nu \in \mathcal{J}}$  a general weight. Let  $\mathbf{A} \in \mathbb{C}^{m \times N}$  be defined as in (1.14) and satisfy the weighted RIP (3.5) with constant*

$$\delta_{\omega,3s} < 1/5$$

for some  $s \geq (4/3) \sup_{\nu \in \mathcal{J}} \omega_\nu^2$ . Then,  $\mathbf{A}$  satisfies the weighted robust null space property of order  $s$  with constants  $\rho = \frac{4\delta}{1-\delta} < 1$ ,  $\tau = \frac{\sqrt{1+\delta}}{1-\delta}$ , and  $\delta = \delta_{\omega,3s}$ , i.e.,

$$\|\mathbf{z}_\Lambda\|_2 \leq \frac{\rho}{\sqrt{s}} \|\mathbf{z}_{\Lambda^c}\|_{\omega,1} + \tau \|\mathbf{A}\mathbf{z}\|_2 \text{ for all } \mathbf{z} \in \mathbb{C}^N \text{ and all } \Lambda \subset \mathcal{J} \text{ with } \sum_{\nu \in \Lambda} \omega_\nu^2 \leq s.$$

Now we are ready to provide a null space property result appropriate to the lower reconstruction, associated with Chebyshev and Legendre systems defined on the Hyperbolic Cross index set.

**Proposition 4.4.** *Let  $s \geq 2$ ,  $\mathcal{J} = \mathcal{H}_s$ , and  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$ , and let  $\mathbf{A} \in \mathbb{C}^{m \times N}$  be a normalized sampling matrix satisfying the lower RIP (3.4) with*

$$(4.8) \quad \delta_{\ell,\alpha s} < 1/5,$$

where  $\alpha = 2$  for Legendre system and  $\alpha = 3$  for Chebyshev system. Then, for any  $\Lambda \subset \mathcal{H}_s$  with  $K(\Lambda) \leq K(s)$  and any  $\mathbf{z} \in \mathbb{C}^N$ ,

$$(4.9) \quad \|\mathbf{z}_\Lambda\|_2 \leq \frac{\rho}{\sqrt{K(s)}} \|\mathbf{z}_{\Lambda^c}\|_{\omega,1} + \tau \|\mathbf{A}\mathbf{z}\|_2$$

with  $\rho = \frac{4\delta}{1-\delta} < 1$ ,  $\tau = \frac{\sqrt{1+\delta}}{1-\delta}$ , and  $\delta = \delta_{\ell,\alpha s}$ .

*Proof.* Since  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$ , we have  $K(\Lambda) = \sum_{\nu \in \Lambda} \omega_\nu^2$ , therefore proving (4.9) is equivalent to showing that  $\mathbf{A}$  satisfies the weighted robust null space property of order  $K(s)$  with constants  $\rho$  and  $\tau$  as in our proposition. By Lemma 4.3, this can follow if  $\mathbf{A}$  satisfies weighted RIP with  $\delta_{\omega,3K(s)} < 1/5$  and  $K(s) \geq (4/3) \sup_{\nu \in \mathcal{H}_s} \omega_\nu^2$ . In view of (4.7),  $3K(s) \leq K(\alpha s)$  for both Legendre and Chebyshev systems. Since  $\delta_{\omega,t}$  is increasing in  $t$ ,  $\delta_{\omega,3K(s)} \leq \delta_{\omega,K(\alpha s)} = \delta_{\ell,\alpha s} < 1/5$ ; see Remark 3.2 for the equality. We also have from (4.4) that  $K(s) \geq (4/3) \sup_{\nu \in \mathcal{H}_s} \omega_\nu^2$  for any  $s \geq 2$  for both systems, which completes the proof.  $\square$

Combining (3.20) and Proposition 4.4 yields the uniform recovery of  $g$  up to the best lower  $s$ -term error and the tail bound.

**Theorem 4.5.** *Let  $s \geq 2$ ,  $\mathcal{J} = \mathcal{H}_s$  and  $N = \#(\mathcal{H}_s)$ . Consider a number of samples*

$$(4.10) \quad m \geq C s^{1+\beta'} \max \left\{ \min \left\{ \log^3(s) + d \log^2(s), \log^3(s) \log(d) \right\}, \log(s) \log(\log(s)/\gamma) \right\},$$

where  $\beta' = \frac{\log 3}{\log 2} - 1$  if  $\{\Psi_\nu\}$  is a Chebyshev system and  $\beta' = 1$  if  $\{\Psi_\nu\}$  is a Legendre system. Let  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$  be drawn independently from the orthogonalization measure  $\rho$  associated to  $\{\Psi_\nu\}$  and  $\mathbf{A} \in \mathbb{C}^{m \times N}$  be the associated normalized sampling matrix as in (1.14). Then, with probability exceeding  $1 - \gamma$ , the following holds for all functions  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$ : Given  $\hat{\mathbf{g}}, \boldsymbol{\xi}$  as in (1.14)–(1.15) and  $\eta$  satisfying

$\|\boldsymbol{\xi}\|_2 \leq \frac{\eta}{\sqrt{m}}$ , the function  $g^\# = \sum_{\nu \in \mathcal{H}_s} c_\nu^\# \Psi_\nu$ , with  $\mathbf{c}^\# = (c_\nu^\#)_{\nu \in \mathcal{H}_s}$  a minimizer of (4.2), satisfies

$$\begin{aligned} \|g - g^\#\|_\infty &\leq \|g - g^\#\|_{\omega,1} \leq c_1 \sigma_s^{(\ell)}(g)_{\omega,1} + d_1 \eta \sqrt{\frac{K(s)}{m}}, \\ \|g - g^\#\|_2 &\leq c_2 \frac{\sigma_s^{(\ell)}(g)_{\omega,1}}{\sqrt{K(s)}} + d_2 \eta \sqrt{\frac{1}{m}}. \end{aligned}$$

Above,  $c_1, c_2, d_1$ , and  $d_2$  are universal constants.

*Proof.* According to Theorem 3.3 and the discussions that follow, for a fixed  $\alpha > 0$ ,  $\mathbf{A}$  satisfies the lower RIP of order  $\alpha s$  and  $\delta_{\ell, \alpha s} < 1/5$  with probability exceeding  $1 - \gamma$  under the complexity (4.10). Setting  $\alpha = 2$  for Legendre system and  $\alpha = 3$  for Chebyshev system and applying Proposition 4.4, (4.9) holds for all  $\mathbf{z} \in \mathbb{C}^N$  and all  $\Lambda \subset \mathcal{J}$  with  $K(\Lambda) \leq K(s)$ . Following the arguments in [41, Theorem 4.2], for  $s \geq 2$  and all  $\mathbf{z}, \tilde{\mathbf{z}} \in \mathbb{C}^N$ , we have

$$\begin{aligned} (4.11) \quad \|\mathbf{z} - \tilde{\mathbf{z}}\|_{\omega,1} &\leq \frac{1+\rho}{1-\rho} (\|\mathbf{z}\|_{\omega,1} - \|\tilde{\mathbf{z}}\|_{\omega,1} + 2\sigma_s^{(\ell)}(\tilde{\mathbf{z}})_{\omega,1}) + \frac{2\tau\sqrt{K(s)}}{1-\rho} \|\mathbf{A}(\mathbf{z} - \tilde{\mathbf{z}})\|_2, \\ \|\mathbf{z} - \tilde{\mathbf{z}}\|_2 &\leq \frac{c_2}{\sqrt{K(s)}} (\|\mathbf{z}\|_{\omega,1} - \|\tilde{\mathbf{z}}\|_{\omega,1} + 2\sigma_s^{(\ell)}(\tilde{\mathbf{z}})_{\omega,1}) + d_2 \|\mathbf{A}(\mathbf{z} - \tilde{\mathbf{z}})\|_2, \end{aligned}$$

where  $c_2, d_2$  are universal constants. Recall that  $\mathbf{c} = (c_\nu)_{\nu \in \mathcal{J}}$  denotes the exact coefficients, which satisfy  $\|\mathbf{A}\mathbf{c} - \tilde{\mathbf{g}}\|_2 = \|\boldsymbol{\xi}\|_2 \leq \frac{\eta}{\sqrt{m}}$ . Since  $\tilde{\mathbf{c}}$  is a minimizer of (4.2), then  $\|\mathbf{A}\mathbf{c}^\# - \tilde{\mathbf{g}}\|_2 \leq \frac{\eta}{\sqrt{m}}$  and  $\|\mathbf{c}^\#\|_{\omega,1} \leq \|\mathbf{c}\|_{\omega,1}$ . We set  $\mathbf{z} = \mathbf{c}^\#$  and  $\tilde{\mathbf{z}} = \mathbf{c}$  in (4.11) to obtain

$$\begin{aligned} (4.12) \quad \|\mathbf{c} - \mathbf{c}^\#\|_{\omega,1} &\leq \frac{2(1+\rho)}{1-\rho} \sigma_s^{(\ell)}(\mathbf{c})_{\omega,1} + \frac{4\tau}{1-\rho} \eta \sqrt{\frac{K(s)}{m}}, \\ \|\mathbf{c} - \mathbf{c}^\#\|_2 &\leq \frac{2c_2}{\sqrt{K(s)}} \sigma_s^{(\ell)}(\mathbf{c})_{\omega,1} + 2d_2 \eta \sqrt{\frac{1}{m}}. \end{aligned}$$

The bounds of  $g - g^\#$  in weighted  $\ell_1$ - and  $\ell_2$ -norms follow from (4.12), with an update on universal constants  $c_2$  and  $d_2$ . Finally, the bound on  $\|g - g^\#\|_\infty$  can be derived with observation

$$\|g - g^\#\|_\infty \leq \sum_{\nu \in \mathcal{J}} |c_\nu - c_\nu^\#| \|\Psi_\nu\|_{L^\infty} + \sum_{\nu \in \mathcal{F} \setminus \mathcal{J}} |c_\nu| \|\Psi_\nu\|_{L^\infty} = \|g - g^\#\|_{\omega,1}. \quad \square$$

*Remark 4.6.* In Theorem 4.5,  $\boldsymbol{\xi}$  is actually a random variable varying with the sampling points  $(\mathbf{y}_i)_{1 \leq i \leq m}$ . It can be shown, however, that  $\|\boldsymbol{\xi}\|_2 \leq \|g_{\mathcal{H}_s^\varepsilon}\|_{\omega,1}$  for every set of samples, thus  $\eta$  can be set deterministically as  $\sqrt{m} \|g_{\mathcal{H}_s^\varepsilon}\|_{\omega,1} \leq \eta$ .

**4.2. Given an exact estimate of the tail.** This assumption allows us to prove a stronger convergence rate. It should be mentioned that a result of this type has been derived in [41], however, applying this to the reconstruction of the best lower  $s$ -term approximations (which is comparable to best weighted  $K(s)$ -term approximations, see Remark 3.2) would lead to a choice of the index set depending on the weights as

$$(4.13) \quad \mathcal{J} = \{\nu \in \mathbb{N}_0^d : \omega_\nu^2 \leq CK(s)\}.$$

In our work,  $\mathcal{J}$  is specified instead to be the smallest set that contains the supports of all downward closed sets of cardinality  $s$ , i.e.,  $\mathcal{J} = \mathcal{H}_s$ . As shown in Lemma



4.1,  $\omega_\nu^2 \leq \frac{3}{4}K(s)$  for all  $\nu \in \mathcal{H}_s$ . The converse, i.e.,  $\omega_\nu^2 \geq CK(s)$  for all  $\nu \notin \mathcal{H}_s$ , however, does not hold. Indeed, for our Chebyshev weights, definition (4.13) would lead to a  $\mathcal{J}$  with infinite cardinality. Therefore,  $\mathcal{H}_s$  represents a significantly smaller index set than those considered in [41]. Nonetheless, the recovery of  $g$  up to the best lower  $s$ -term error is still available without condition (4.13), provided that  $c_\nu/\omega_\nu$  is small in  $\mathcal{H}_s^c$ . To clarify this assumption, for  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$ , we introduce a parameter  $\lambda \geq 0$  such that

$$(4.14) \quad \max_{\nu \in \mathcal{H}_s^c} \frac{|c_\nu|}{\omega_\nu} \leq (1 + \lambda) \min_{\nu \in \tilde{\mathcal{J}}} \frac{|c_\nu|}{\omega_\nu}$$

for some set  $\tilde{\mathcal{J}} \subset \mathcal{H}_s$  with  $K(\tilde{\mathcal{J}}) \geq 2K(s)$  or, equivalently,

$$(4.15) \quad \lambda := \min_{\substack{\tilde{\mathcal{J}} \subset \mathcal{H}_s \\ K(\tilde{\mathcal{J}}) \geq 2K(s)}} \left( \frac{\max_{\nu \in \mathcal{H}_s^c} |c_\nu|/\omega_\nu}{\min_{\nu \in \tilde{\mathcal{J}}} |c_\nu|/\omega_\nu} \right) - 1.$$

Many subsets  $\tilde{\mathcal{J}}$  of  $\mathcal{H}_s$  with cardinality  $2s$  satisfy  $K(\tilde{\mathcal{J}}) \geq 2K(s)$ ; see (4.3). Thus, the minimum in the right-hand side of (4.14) can be taken over only  $2s$  multi-indices in  $\mathcal{H}_s$ . On the other hand, for function  $g$  whose expansion coefficients decay fast, it is reasonable to assume small  $\max_{\nu \in \mathcal{H}_s^c} \frac{|c_\nu|}{\omega_\nu}$ , due to small  $c_\nu$  and possibly big  $\omega_\nu$  for  $\nu \in \mathcal{H}_s^c$ . As a result,  $\lambda$  is expected to be small in this effort.

We first need an estimate of  $\|\mathbf{c}_{\mathcal{H}_s^c}\|_2$ . The choice  $\mathcal{J} = \mathcal{H}_s$  is not essential in this development, for which reason we state the result for general  $\|\mathbf{c}_{\mathcal{J}^c}\|_2$ .

**Lemma 4.7.** *Let  $\mathbf{c} = (c_\nu)_{\nu \in \mathcal{F}}$  and  $\mathcal{J}$  be a subset of  $\mathcal{F}$ . For all  $s \geq 1$ , there holds*

$$(4.16) \quad \|\mathbf{c}_{\mathcal{J}^c}\|_2 \leq \frac{\|\mathbf{c}_{\mathcal{J}^c}\|_{\omega,1}}{\sqrt{K(s)}} + \sqrt{K(s)} \max_{\nu \in \mathcal{J}^c} \frac{|c_\nu|}{\omega_\nu}.$$

*Proof.* We introduce  $\mathcal{J}' = \{\nu \in \mathcal{J}^c : \omega_\nu^2 \geq K(s)\}$  and  $\mathcal{J}'' = \mathcal{J}^c \setminus \mathcal{J}'$ . By definition of  $\mathcal{J}'$ , we have

$$\|\mathbf{c}_{\mathcal{J}'}\|_2 \leq \sqrt{\sum_{\nu \in \mathcal{J}'} \frac{\omega_\nu^2 c_\nu^2}{K(s)}} \leq \frac{\|\mathbf{c}_{\mathcal{J}'}\|_{\omega,1}}{\sqrt{K(s)}}.$$

Since  $\|\mathbf{c}_{\mathcal{J}^c}\|_2 \leq \|\mathbf{c}_{\mathcal{J}'}\|_2 + \|\mathbf{c}_{\mathcal{J}''}\|_2$  and  $\|\mathbf{c}_{\mathcal{J}^c}\|_{\omega,1} = \|\mathbf{c}_{\mathcal{J}'}\|_{\omega,1} + \|\mathbf{c}_{\mathcal{J}''}\|_{\omega,1}$ , one only needs to show that

$$\|\mathbf{c}_{\mathcal{J}''}\|_2 \leq \frac{\|\mathbf{c}_{\mathcal{J}''}\|_{\omega,1}}{\sqrt{K(s)}} + \sqrt{K(s)} \max_{\nu \in \mathcal{J}''} \frac{|c_\nu|}{\omega_\nu}.$$

We order  $\mathcal{J}''$  according to a nonincreasing order of  $(|c_\nu|/\omega_\nu)_{\nu \in \mathcal{J}''}$  and then partition  $\mathcal{J}''$  as  $\mathcal{J}'' = \bigcup_{l=1}^L \mathcal{J}_l$  where we inductively choose the sets  $\mathcal{J}_l$  according to:  $\mathcal{J}_0 = \emptyset$ ; for  $l \geq 1$  and  $\mathcal{J}_0, \dots, \mathcal{J}_{l-1}$  having been built, we set  $\mathcal{M}_l = \mathcal{J}'' \setminus \{\mathcal{J}_0 \cup \dots \cup \mathcal{J}_{l-1}\}$ , and let  $\mathcal{J}_l = \mathcal{M}_l$  if  $K(\mathcal{M}_l) < K(s)$ , or be a subset containing the largest elements of  $\mathcal{M}_l$  such that  $K(s) \leq K(\mathcal{J}_l) \leq 4K(s)$ . If the induction does not terminate,  $L = \infty$  in which case

$$K(s) \leq K(\mathcal{J}_l) \leq 4K(s), \quad \forall l \geq 1.$$

If the induction terminates,  $L < \infty$  and we have only  $K(\mathcal{J}_L) < K(s)$ .

Now, we denote by  $r_l^+$  and  $r_l^-$  the largest and smallest entries in  $(|c_\nu|/\omega_\nu)_{\nu \in \mathcal{J}_l}$ . An easy extension of [20, Lemma 6.14] yields for all  $1 \leq l < L$ :

$$\|\mathbf{c}_{\mathcal{J}_l}\|_2 \leq \frac{\|\mathbf{c}_{\mathcal{J}_l}\|_{\omega,1}}{\sqrt{K(\mathcal{J}_l)}} + \frac{\sqrt{K(\mathcal{J}_l)}}{4}(r_l^+ - r_l^-) \leq \frac{\|\mathbf{c}_{\mathcal{J}_l}\|_{\omega,1}}{\sqrt{K(s)}} + \sqrt{K(s)}(r_l^+ - r_l^-).$$

In the case  $L < \infty$ , we have  $\|\mathbf{c}_{\mathcal{J}_L}\|_2 \leq r_L^+ \sqrt{K(\mathcal{J}_L)} \leq r_L^+ \sqrt{K(s)}$ . There follows in both cases  $L < \infty$  and  $L = \infty$ :

$$\|\mathbf{c}_{\mathcal{J}''}\|_2 \leq \sum_{l=1}^L \|\mathbf{c}_{\mathcal{J}_l}\|_2 \leq \frac{\|\mathbf{c}_{\mathcal{J}''}\|_{\omega,1}}{\sqrt{K(s)}} + \sqrt{K(s)}r_1^+.$$

Since  $r_1^+ \leq \max_{\nu \in \mathcal{J}''} \frac{|c_\nu|}{\omega_\nu}$  the proof is complete. □

We are now ready to state and prove the recovery guarantee, assuming an exact estimate of  $g_{\mathcal{H}_s^c}$  exists. The recovery holds with high probability for each fixed function and therefore is nonuniform.

**Theorem 4.8.** *Let  $s \geq 2$ ,  $\mathcal{J} = \mathcal{H}_s$ , and  $N = \#(\mathcal{H}_s)$ , and let  $m$  be as in (4.10). Consider a function  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$  with  $\|g\|_{\omega,1} < \infty$ , and  $\lambda$  defined as in (4.15). We introduce  $E_g := \max\{\sqrt{2}\|g_{\mathcal{H}_s^c}\|_2, \frac{\sqrt{2}\|g_{\mathcal{H}_s^c}\|_{\omega,1}}{\sqrt{K(s)}}\}$  and let  $\eta$  be such that*

$$(4.17) \quad E_g \leq \frac{\eta}{\sqrt{m}} \leq (1 + \varepsilon)E_g$$

for some  $\varepsilon > 0$ . Then, with probability exceeding  $1 - \gamma$ , the function  $g^\# = \sum_{\nu \in \mathcal{H}_s} c_\nu^\# \Psi_\nu$ , with the vector  $\mathbf{c}^\# = (c_\nu^\#)_{\nu \in \mathcal{H}_s}$  a minimizer of (4.2), satisfies

$$(4.18) \quad \begin{aligned} \|g - g^\#\|_\infty &\leq \|g - g^\#\|_{\omega,1} \leq (1 + \lambda + \varepsilon)c_3\sigma_s^{(\ell)}(g)_{\omega,1}, \\ \|g - g^\#\|_2 &\leq (1 + \lambda + \varepsilon)c_4 \frac{\sigma_s^{(\ell)}(g)_{\omega,1}}{\sqrt{K(s)}}. \end{aligned}$$

Here,  $c_3$  and  $c_4$  are universal constants.

*Proof.* We introduce  $\widehat{\boldsymbol{\xi}} = (\widehat{\xi}_i)_{1 \leq i \leq m} = (g_{\mathcal{H}_s^c}(\mathbf{y}_i))_{1 \leq i \leq m}$ . Since  $\mathbf{y}_i$  are i.i.d. random variables with respect to  $\varrho$ , by Lemma 4.7,

$$(4.19) \quad \mathbb{E}[\widehat{\xi}_i^2] = \|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2 \leq \frac{2\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}^2}{K(s)} + 2K(s)M_{\mathcal{H}_s^c}^2,$$

where we have defined  $M_{\mathcal{H}_s^c} := \max_{\nu \in \mathcal{H}_s^c} \frac{|c_\nu|}{\omega_\nu}$ . We consider two cases:

- Case 1:  $\|\mathbf{c}_{\mathcal{H}_s^c}\|_2 > \frac{\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}}{\sqrt{K(s)}}$ . Since  $|\widehat{\xi}_i| \leq \|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1} < \sqrt{K(s)}\|\mathbf{c}_{\mathcal{H}_s^c}\|_2$ , it holds that

$$\mathbb{E}\left[(\widehat{\xi}_i^2 - \mathbb{E}[\widehat{\xi}_i^2])^2\right] \leq \mathbb{E}[\widehat{\xi}_i^4] \leq K(s)\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2 \mathbb{E}[\widehat{\xi}_i^2] \leq K(s)\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^4.$$

Applying Bernstein's inequality with the mean-zero random variable  $\widehat{\xi}_i^2 - \mathbb{E}[\widehat{\xi}_i^2]$ ,

$$\mathbb{P}\left(\sum_{i=1}^m \frac{\widehat{\xi}_i^2}{m} - \|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2 \geq \kappa\right) \leq \exp\left(-\frac{m\kappa^2/2}{K(s)\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^4 + \frac{\kappa}{3}K(s)\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2}\right).$$

Choose  $\kappa = \|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2$ ; then there follows

$$\mathbb{P}\left(\sum_{i=1}^m \frac{\widehat{\xi}_i^2}{m} \geq 2\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2\right) \leq \exp\left(-\frac{3m}{8K(s)}\right).$$

• Case 2:  $\|\mathbf{c}_{\mathcal{H}_s^c}\|_2 \leq \frac{\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}}{\sqrt{K(s)}}$ . Similarly to [41], by Bernstein’s inequality,

$$\mathbb{P}\left(\sum_{i=1}^m \frac{\widehat{\xi}_i^2}{m} \geq \frac{2\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}^2}{K(s)}\right) \leq \mathbb{P}\left(\sum_{i=1}^m \frac{\widehat{\xi}_i^2}{m} - \|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2 \geq \frac{\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}^2}{K(s)}\right) \leq \exp\left(-\frac{3m}{8K(s)}\right).$$

In both cases, given that  $m \geq (8/3)s^{1+\beta'} \log(1/\gamma) \geq (8/3)K(s) \log(1/\gamma)$ , then with probability exceeding  $1 - \gamma$ ,

$$\frac{1}{m} \|\widehat{\boldsymbol{\xi}}\|_2^2 \leq \max\left\{2\|\mathbf{c}_{\mathcal{H}_s^c}\|_2^2, \frac{2\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}^2}{K(s)}\right\} \leq \frac{4\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1}^2}{K(s)} + 4K(s)M_{\mathcal{H}_s^c}^2,$$

where the second inequality follows from (4.19). Under the condition (4.17), an application of Theorem 4.5 yields

$$(4.20) \quad \|g - g^\# \|_{\omega,1} \leq c_1\sigma_s^{(\ell)}(\mathbf{c})_{\omega,1} + (1 + \varepsilon)d_1\|\mathbf{c}_{\mathcal{H}_s^c}\|_{\omega,1} + (1 + \varepsilon)d_1K(s)M_{\mathcal{H}_s^c}.$$

Let  $\Lambda^*$  denote the support of best lower  $s$ -term approximation of  $g$  in  $\ell_{\omega,1}$ -norm,  $\tilde{\mathcal{J}}$  be determined by (4.14)-(4.15). For every  $\nu \in \hat{\mathcal{J}} := \tilde{\mathcal{J}} \setminus \Lambda^*$ , there holds  $\omega_\nu^2 M_{\mathcal{H}_s^c} \leq (1 + \lambda)\omega_\nu^2 \frac{|c_\nu|}{\omega_\nu} = (1 + \lambda)\omega_\nu |c_\nu|$ . Summing these over  $\hat{\mathcal{J}}$  gives

$$(4.21) \quad K(s)M_{\mathcal{H}_s^c} \leq K(\hat{\mathcal{J}})M_{\mathcal{H}_s^c} \leq (1 + \lambda)\|\mathbf{c}_{\hat{\mathcal{J}}}\|_{\omega,1},$$

the first inequality coming from  $K(\tilde{\mathcal{J}}) \geq 2K(s)$ . We finally combine (4.20) and (4.21) to obtain (4.18), which completes the proof.  $\square$

### 5. NUMERICAL ILLUSTRATIONS

In this section, we provide several numerical examples to demonstrate the efficiency of our weighted  $\ell_1$ -minimization with  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$  for smooth multivariate function recovery. We focus here on the approximation of  $g$  in terms of orthonormal Legendre expansions by solving

$$(5.1) \quad \min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to} \quad \|\tilde{\mathbf{g}} - \mathbf{A}\mathbf{z}\|_2 \leq \frac{\eta}{\sqrt{m}},$$

for various choices of weights, where  $\eta = \sqrt{m}\|\mathbf{g}_{\mathcal{H}_s^c}\|_2$ . This tail bound complies with Theorem 4.5, based on an estimate of  $\|\boldsymbol{\xi}\|_2$  (or  $\|\tilde{\mathbf{g}} - \mathbf{A}\mathbf{c}\|_2$ ) via  $\|\mathbf{g}_{\mathcal{H}_s^c}\|_2$ , and is less conservative than those analytically studied in Theorem 4.8. As we test with simple functions, the expansion of  $g$  can be computed numerically with the use of a quadrature approximation yielding an estimate of the tail  $\eta$  a priori. The software code **SPGL1** [44, 45] is employed to solve (5.1).

In each example, we choose the polynomial subspace  $\mathbb{P}_{\mathcal{J}} \equiv \mathbb{P}_{\mathcal{H}_s}$ , and increase the number of samples  $m$  up to some  $m_{\max} < N = \#(\mathcal{H}_s)$ . For each ratio  $m/N$ , the set of random samples is fixed over various choices of weights for a performance comparison. We then run 50 trials for the averaged  $L_q^2$  error, setting the maximum number of iterations in **SPGL1** per trial to 10,000. Our results are shown in Figures 1–4 for several multivariate functions. The left panels represent the magnitudes of polynomial coefficients (computed with MATLAB via a sparse-grid algorithm) indexed in  $\mathcal{H}_s$  and sorted lexicographically. The center panels depict the decay of the coefficients once sorted by magnitude. The right panels show the corresponding convergence results. In Figure 5, we include some results of the approximation errors in the  $L^\infty$ -norm, showing little change in the performance of the investigated weights compared to the  $L_q^2$  case.

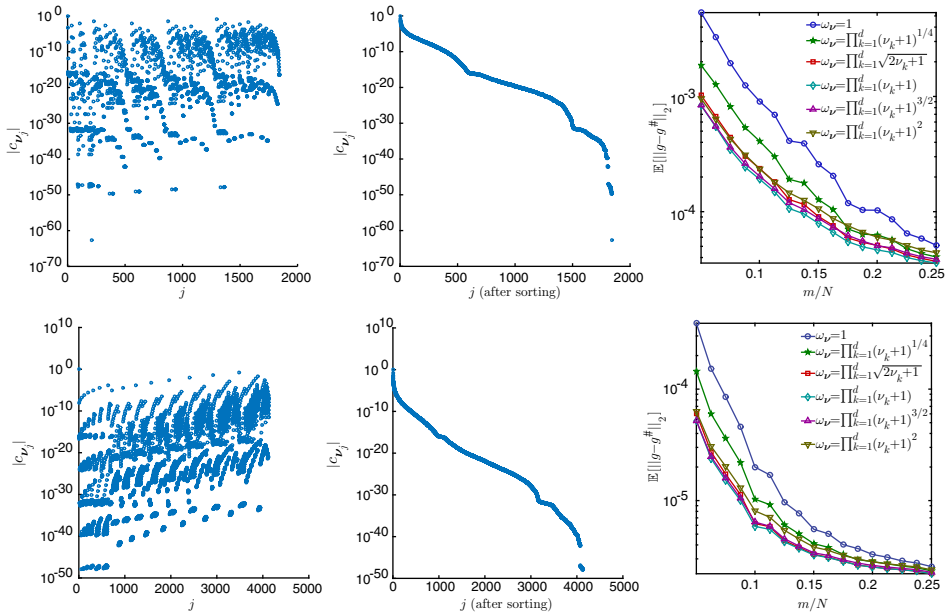


FIGURE 1. Comparison of the averaged  $L^2$  error in approximating  $g(\mathbf{y}) = \frac{\prod_{k=\lceil d/2 \rceil+1}^d \cos(16y_k/2^k)}{\prod_{k=1}^{\lceil d/2 \rceil} (1-y_k/4^k)}$  using weighted  $\ell^1$ -minimization with various choices of weights. **(Top)**  $d = 8$ ,  $N = 1843$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 1.5922e - 05$ . **(Bottom)**  $d = 16$ ,  $N = 4129$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 1.3388e - 06$ .

Our experiments indicate that for functions with very fast decaying polynomial expansions,  $\omega_{\nu} = \|\Psi_{\nu}\|_{L^\infty}$  is virtually the optimal weight. Indeed, for the function concerned in Figure 1, consisting of trigonometric and rational univariate functions, we see in both  $d = 8$  and  $d = 16$  that the weight  $\omega_{\nu} = \|\Psi_{\nu}\|_{L^\infty}$  significantly outperforms the unweighted  $\ell^1$  approach. We note that in  $d = 16$ , higher weights begin to have decreasing benefit, performing worse than our proposed weight.

The results in Figure 2 are related to a function that involves the exponent of a sum of univariate trigonometric functions. We note that only a small fraction of coefficients exceed  $10^{-16}$  in both  $d = 8$  and  $d = 16$ . In this case, we see that the weighted  $\ell^1$  with the weight  $\omega_{\nu} = \|\Psi_{\nu}\|_{L^\infty}$  performs the best out of all of the weights supplied. We also observe that increasing the weights beyond  $\|\Psi_{\nu}\|_{L^\infty}$  leads to a corresponding increase in the averaged  $L^2$  error.

In Figure 3, we test with a root of a rational function. Here again, the weight  $\omega_{\nu} = \|\Psi_{\nu}\|_{L^\infty}$  performs the best, and the approximation errors grow as the weights increase beyond this weight. The two highest weights even perform worse than unweighted  $\ell^1$  for higher values of  $m/N$ . Comparing Figures 1 and 3, the similar center panels suggest similar decay rates for both functions inside  $\mathcal{H}_s$ . The performance of the weights between the two examples, however, drastically differs, possibly due to the different support of the large coefficients. For this function, we were unable to test with  $d = 16$  and  $N = 4129$  due to the high expansion tail.

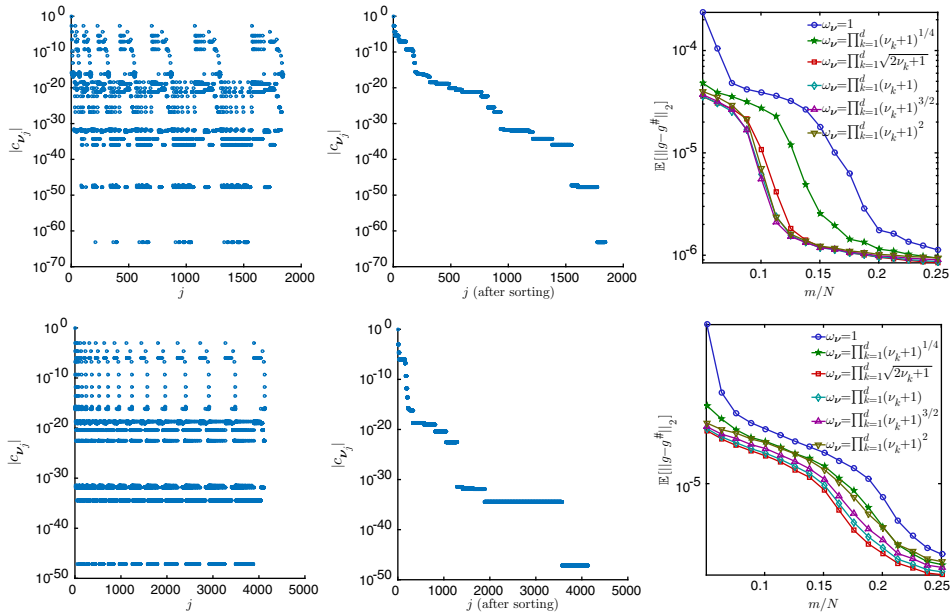


FIGURE 2. Comparison of the averaged  $L^2$  error in approximating  $g(\mathbf{y}) = \exp\left(\frac{-\sum_{k=1}^d \cos(y_k)}{8d}\right)$  using weighted  $\ell^1$ -minimization with various choices of weights. **(Top)**  $d = 8$ ,  $N = 1843$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 4.0232e - 07$ . **(Bottom)**  $d = 16$ ,  $N = 4129$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 2.0155e - 06$ .

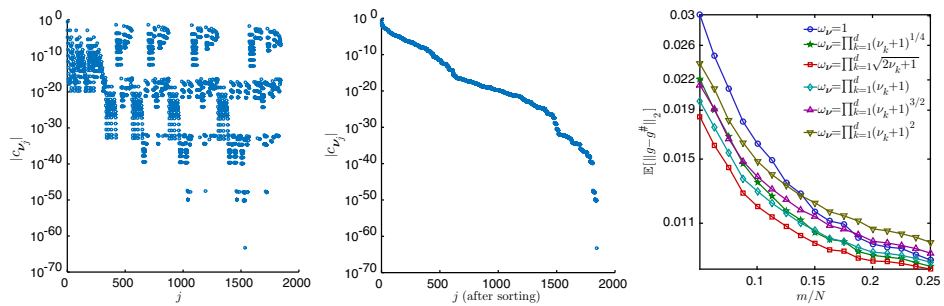


FIGURE 3. Comparison of the averaged  $L^2$  error in approximating  $g(\mathbf{y}) = \left[ \frac{\prod_{k=1}^{\lceil d/2 \rceil} (1+4^k y_k^2)}{\prod_{k=\lceil d/2 \rceil+1}^d (100+5y_k)} \right]^{1/d}$  using weighted  $\ell^1$ -minimization with various choices of weights.  $d = 8$ ,  $N = 1843$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 6.1018e - 3$ .

On the other hand, if the polynomial expansion is less sparse, our weight may not be optimal. Figure 4 shows the results for an exponential function of a linear combination of the 1- $d$  variables. For this function, over half of the coefficients in  $\mathcal{H}_s$  exceed  $10^{-8}$ , and  $\omega_\nu = \|\Psi_\nu\|_{L^\infty}$  performs worse than the larger weights. Still, we observe here, as in all tests for smooth functions in higher dimensions, i.e.,  $d = 8$  and  $d = 16$ , that our weight consistently provides improved accuracy compared with the unweighted  $\ell_1$ , thus confirming the theory presented throughout.

All of the results have been presented in terms of the  $L^2_\rho$  error thus far. To investigate how the approximations behave pointwise, we also measure all the errors in the  $L^\infty$ -norm. The achieved results show little or no change in the relative performance with respect to different choice of weights, compared to that of the  $L^2_\rho$  case, i.e., our weight is always more accurate than the unweighted approach and in cases of Figure 1–3 is the optimal weight. For brevity, we only show here the  $L^\infty$  error of the approximations considered in Figure 2. Figure 5 highlights a key advantage of our approach that focusing on best lower  $s$ -term recovery mitigates spurious oscillations that occur when using high-degree polynomial approximations.

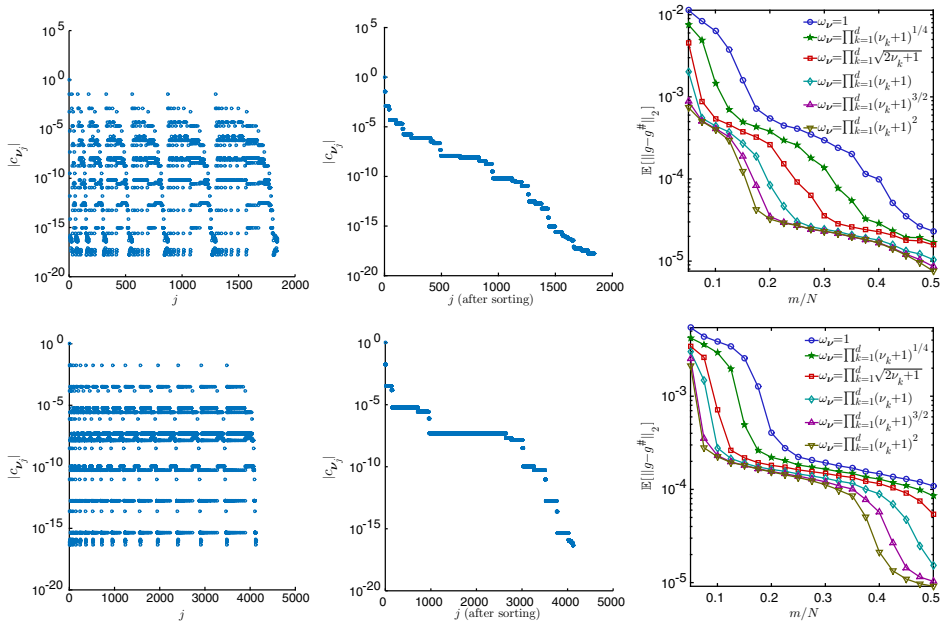


FIGURE 4. Comparison of the averaged  $L^2_\rho$  error in approximating  $g(\mathbf{y}) = \exp\left(-\frac{\sum_{k=1}^d y_k}{2d}\right)$  using weighted  $\ell^1$ -minimization with various choices of weights. **(Top)**  $d = 8$ ,  $N = 1843$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 7.2714e - 07$ . **(Bottom)**  $d = 16$ ,  $N = 4129$ , and  $\|g_{\mathcal{H}_s^c}\|_2 = 3.7412e - 07$ .

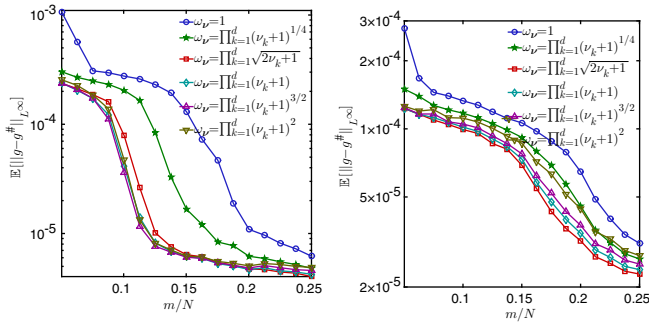


FIGURE 5. Comparison of the averaged  $L^\infty$  error of approximations studied in Figure 2. (**Left**)  $d = 8$  and (**Right**)  $d = 16$ .

## 6. CONCLUDING REMARKS

In this work we present a novel compressed sensing approach for sparse Legendre and Chebyshev approximations of real and complex functions in high dimensions. Motivated by the fact that the target function in many applications is smooth and characterized by a rapidly decaying orthonormal expansion, whose most important terms are captured by a lower set, we develop an  $\ell_1$ -minimization procedure with a precise choice of weights to impose the lower preference. Through rigorous analysis and numerical illustrations we demonstrate that the proposed method possesses a significantly reduced sample complexity compared to existing techniques. This advantage is established through the introduction of lower RIP, a weaker version of RIP that is associated with lower sets, and an optimal choice of polynomial subspace. In addition, we prove a generalized version of the result in [5] for bounded orthonormal systems and improve the RIP estimate by one logarithm factor.

Extending the theory and the procedures developed herein for approximating high-dimensional parameterized PDE systems is the next logical step, as it has been known that for a large class of such systems, the polynomial chaos expansions of parameterized solutions decay exponentially. A significant challenge associated with this problem is that the “signal”  $\mathbf{c} = (c_\nu)_{\nu \in \mathcal{J}}$  to be recovered is Hilbert-valued, rather than real or complex, components, i.e.,  $c_\nu \in \mathcal{V}$ . As shown in [18, 24, 30, 35, 39, 48], standard compressed sensing techniques only approximate a functional of PDE solutions, for instance,  $u(x^*, \mathbf{y})$  (via  $c_\nu(x^*)$ ) at a single location  $x^*$  in physical space. Although  $u(\cdot, \mathbf{y})$  can be constructed from these pointwise evaluations using piecewise polynomial interpolation, least square regression, etc., this practice faces several limitations. First, the decay of the polynomial coefficients may vary greatly over points in the physical domain, leading to different levels of accuracy among pointwise evaluations, rendering less efficient global evaluation. Second, a priori estimate of the tail expansion is required at every selected node, which may not be a realistic requirement. Also, estimates of  $c_\nu$  in energy norms, well known in the existing theory, cannot be exploited for improving the convergence of recovery algorithms. To address these difficulties, we aim to investigate new convex optimization and thresholding frameworks for Hilbert-valued functions, so that  $c_\nu \in \mathcal{V}$  can be directly computed. The mathematical analysis and computational aspect of this approach will be documented in future work.

APPENDIX A. PROOFS OF THE RIPs

We recall the following notations, which will be used throughout this section:

$$\begin{aligned}
 B_{1,N} &:= \{z \in \mathbb{C}^N : \|z\|_1 = 1\}, \\
 \mathcal{E}_s &:= \{z \in \mathbb{C}^N : \|z\|_2 = 1, \#(\text{supp}(z)) \leq s\}, \\
 \mathcal{E}_s^\ell &:= \{z \in \mathbb{C}^N : \|z\|_2 = 1, K(\text{supp}(z)) \leq K(s)\}, \text{ and} \\
 \psi(\mathbf{y}, z) &:= \sum_{\nu \in \mathcal{J}} z_\nu \Psi_\nu(\mathbf{y}), \quad \text{with } \mathbf{y} \in \mathcal{U}, z \in \mathbb{C}^N.
 \end{aligned}$$

**A.1. Supporting lemmas.** First, we derive a Chernoff-Hoeffding bound for complex random variables, as well as a tail bound for Bernoulli random variables.

**Lemma A.1.** *Let  $X_1, \dots, X_M$  be  $M$  independent identically distributed complex-valued random variables satisfying  $|X_k| \leq a$  and  $\mathbb{E}[X_k] = X$  for all  $k$ . We denote  $\bar{X} = \frac{1}{M} \sum_{k=1}^M X_k$ . For every  $\mu > 0$ ,*

$$(A.1) \quad \mathbb{P}\left(|\bar{X} - X| \geq \mu\right) \leq 4 \exp\left(-\frac{M\mu^2}{4a^2}\right).$$

*Proof.* We have

$$\mathbb{P}\left(|\bar{X} - X| \geq \mu\right) \leq \mathbb{P}\left(|\Re(\bar{X} - X)| \geq \frac{\mu}{\sqrt{2}}\right) + \mathbb{P}\left(|\Im(\bar{X} - X)| \geq \frac{\mu}{\sqrt{2}}\right).$$

Applying Hoeffding’s inequality [27] for two sequences of bounded real random variables  $\{\Re(X_k)\}$  and  $\{\Im(X_k)\}$ , there holds

$$\begin{aligned}
 \mathbb{P}\left(|\Re(\bar{X} - X)| \geq \frac{\mu}{\sqrt{2}}\right) &\leq 2 \exp\left(-\frac{M\mu^2}{4a^2}\right), \\
 \mathbb{P}\left(|\Im(\bar{X} - X)| \geq \frac{\mu}{\sqrt{2}}\right) &\leq 2 \exp\left(-\frac{M\mu^2}{4a^2}\right).
 \end{aligned}$$

The proof is then complete. □

**Lemma A.2.** *Let  $X_1, \dots, X_M$  be  $M$  independent identically distributed Bernoulli random variables with  $\mathbb{E}[X_k] = X$  for all  $k$ . Denote  $\bar{X} = \frac{1}{M} \sum_{k=1}^M X_k$ . Then, for every  $0 < \mu_1 < 1$ ,  $\mu_2 > 0$  and  $M \geq \frac{16e}{\mu_1\mu_2}$ , there holds*

$$(A.2) \quad \mathbb{P}\left(|\bar{X} - X| \geq \mu_1 X + \mu_2\right) \leq \exp\left(-\frac{M\mu_1\mu_2}{16e}\right).$$

*Proof.* Let  $\epsilon_1, \dots, \epsilon_M$  be a Rademacher sequence. Using symmetrization and Khintchine inequality [20, Section 8], we have for any  $q \geq 1$ ,

$$\begin{aligned}
 \mathbb{E}_X |\bar{X} - X|^q &\leq \frac{2^q}{M^q} \mathbb{E}_X \mathbb{E}_\epsilon \left| \sum_{k=1}^M \epsilon_k X_k \right|^q \\
 &\leq \frac{2^{q+\frac{3}{4}} e^{-q/2} q^{q/2}}{M^q} \mathbb{E}_X \left( \sum_{k=1}^M X_k^2 \right)^{q/2} = \left( \frac{2^{1+\frac{3}{4q}} \sqrt{q}}{\sqrt{eM}} \right)^q \mathbb{E}_X \bar{X}^{q/2}.
 \end{aligned}$$



Denote  $C = 2^{7/4}e^{-1/2}$ ; then there follows

$$\begin{aligned} \left(\mathbb{E}_X |\bar{X} - X|^q\right)^{\frac{1}{q}} &\leq \frac{2^{1+\frac{3}{4q}}\sqrt{q}}{\sqrt{eM}} \left(\mathbb{E}_X \bar{X}^{\frac{q}{2}}\right)^{\frac{1}{q}} \\ &\leq C\sqrt{\frac{q}{M}}\sqrt{\bar{X}} + C\sqrt{\frac{q}{M}} \left(\mathbb{E}_X |\bar{X} - X|^q\right)^{\frac{1}{2q}}. \end{aligned}$$

The last inequality implies

$$\left(\mathbb{E}_X |\bar{X} - X|^q\right)^{\frac{1}{q}} \leq C^2\frac{q}{M} + 2C\sqrt{\frac{q}{M}}\sqrt{\bar{X}}.$$

Applying Markov's inequality gives

$$\mathbb{P}\left(|\bar{X} - X| \geq eC^2\frac{q}{M} + 2eC\sqrt{\frac{q}{M}}\sqrt{\bar{X}}\right) \leq e^{-q}.$$

Finally, it is easy to see that

$$eC^2\frac{q}{M} + 2eC\sqrt{\frac{q}{M}}\sqrt{\bar{X}} \leq \mu_1 X + \frac{C^2qe}{M} \left(\frac{e}{\mu_1} + 1\right) < \mu_1 X + \frac{C^2qe^2\sqrt{2}}{M\mu_1}.$$

Defining  $q = \frac{M\mu_1\mu_2}{16e}$  so that  $\mu_2 = \frac{C^2qe^2\sqrt{2}}{M\mu_1}$ , we conclude the proof.  $\square$

Next, we state and prove an extended covering number result.

**Lemma A.3.** *For  $0 < \varsigma < 1$ ,  $\mu > 0$ , there exists a set  $D \subset \mathbb{C}^N$  such that*

(i) *For all  $\mathbf{z} \in \mathcal{E}_s$ ,  $\mathbf{y}_1, \dots, \mathbf{y}_m \in \mathcal{U}$ , there exists  $\mathbf{z}' \in D$  satisfying:*

(A.3)  $|\psi(\mathbf{y}, \mathbf{z} - \mathbf{z}')| \leq \mu$  with probability exceeding  $1 - \varsigma$  in  $(\mathcal{U}, \varrho)$ , and

(A.4)  $|\psi(\mathbf{y}_i, \mathbf{z} - \mathbf{z}')| \leq \mu$  for at least  $(1 - \varsigma)m$  indices  $i \in \{1, \dots, m\}$ .

(ii) *The cardinality of  $D$  satisfies*

(A.5)  $\log(\#(D)) \leq (8/\mu^2)\Theta^2s \log(4N) \log(12/\varsigma).$

*Proof.* We will find  $D$  using the empirical method of Maurey. First, we observe that  $\mathcal{E}_s \subset \sqrt{s}B_{1,N}$ , hence if we denote  $\mathcal{P} = \{\pm \mathbf{e}_j\sqrt{2s}, \pm i\mathbf{e}_j\sqrt{2s}\}_{1 \leq j \leq N}$ , where  $(\mathbf{e}_j)$  are canonical unit vectors in  $\mathbb{C}^N$ , we have  $\mathcal{E}_s \subset \text{conv}(\mathcal{P})$ . Every  $\mathbf{z} \in \mathcal{E}_s$  can be represented as  $\mathbf{z} = \sum_{r=1}^{4N} \lambda_r \mathbf{v}_r$ , for some  $\lambda_r \geq 0$ ,  $\sum_{r=1}^{4N} \lambda_r = 1$  and  $\mathbf{v}_r$  listing  $4N$  elements of  $\mathcal{P}$ . There exists a probability measure  $\lambda$  on  $\mathcal{P}$  that takes the values  $\mathbf{v}_r \in \mathcal{P}$  with probability  $\lambda_r$ .

Let  $\mathbf{z}_1, \dots, \mathbf{z}_M$  be i.i.d random variables with law  $\lambda$ . Note that  $\mathbb{E}\mathbf{z}_k = \mathbf{z}$ , for all  $k = 1, \dots, M$ . For each  $\mathbf{y} \in \mathcal{U}$ ,  $\psi(\mathbf{y}, \mathbf{z}_k)$  is also a complex-valued random variable on probability space  $(\psi(\mathbf{y}, \mathcal{P}), \lambda)$  with

$$|\psi(\mathbf{y}, \mathbf{z}_k)| \leq \Theta\sqrt{2s} \quad \text{and} \quad \mathbb{E}\psi(\mathbf{y}, \mathbf{z}_k) = \psi(\mathbf{y}, \mathbf{z}).$$

Denote  $\bar{\mathbf{z}} = \frac{1}{M} \sum_{k=1}^M \mathbf{z}_k$ , let  $D$  be the set of all possible outcomes of  $\bar{\mathbf{z}}$  and let  $\bar{\lambda}$  be the probability measure on  $D$  according to  $\bar{\mathbf{z}}$ . We define a characteristic function  $\chi$  on  $(\mathcal{U} \times D, \varrho \otimes \bar{\lambda})$  such that

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

Applying Lemma A.1 yields for all  $\mathbf{y} \in \mathcal{U}$ ,

$$\int_D \chi(\mathbf{y}, \bar{\mathbf{z}}) d\bar{\lambda} = \mathbb{P}_{\bar{\mathbf{z}}}(|\psi(\mathbf{y}, \bar{\mathbf{z}} - \mathbf{z})| \geq \mu) \leq 4 \exp\left(-\frac{M\mu^2}{8\Theta^2 s}\right).$$

There follows

$$\int_D \left( \int_{\mathcal{U}} \chi(\mathbf{y}, \bar{\mathbf{z}}) d\varrho \right) d\bar{\lambda} = \int_{\mathcal{U}} \left( \int_D \chi(\mathbf{y}, \bar{\mathbf{z}}) d\bar{\lambda} \right) d\varrho \leq 4 \exp\left(-\frac{M\mu^2}{8\Theta^2 s}\right),$$

which by Markov’s inequality yields that with probability exceeding  $2/3$ ,  $\bar{\mathbf{z}}$  satisfies

$$(A.6) \quad \mathbb{P}_{\mathbf{y}}(|\psi(\mathbf{y}, \bar{\mathbf{z}} - \mathbf{z})| \geq \mu) = \int_{\mathcal{U}} \chi(\mathbf{y}, \bar{\mathbf{z}}) d\varrho \leq 12 \exp\left(-\frac{M\mu^2}{8\Theta^2 s}\right).$$

Now, repeating the above arguments to the set  $\{\mathbf{y}_1, \dots, \mathbf{y}_m\}$  with discrete uniform distribution, one can also derive that with probability of  $\bar{\mathbf{z}}$  exceeding  $2/3$ , there holds

$$(A.7) \quad |\psi(\mathbf{y}_i, \bar{\mathbf{z}} - \mathbf{z})| \geq \mu, \text{ for at most } 12 \exp\left(-\frac{M\mu^2}{8\Theta^2 s}\right)m \text{ indices } i \in \{1, \dots, m\}.$$

Hence, there exists a realization  $\mathbf{z}'$  of  $\bar{\mathbf{z}}$  in  $D$  fulfilling both (A.6) and (A.7). Note that  $\#(D) \leq (4N)^M$ . By defining a new variable  $\varsigma = 12 \exp\left(-\frac{M\mu^2}{8\Theta^2 s}\right)$  and eliminating  $M$ , we conclude the proof.  $\square$

We observe that a sharper estimate of  $\log(\#(D))$  is possible. Indeed, one can bound  $\#(D)$  by  $\binom{4N+M}{M}$  instead of  $(4N)^M$ , under which the assertion (ii) is replaced by

$$(A.8) \quad \log(\#(D)) \leq \frac{8}{\mu^2} \Theta^2 s \log\left(e + \frac{eN\mu^2}{\Theta^2 s \log(12/\varsigma)}\right) \log(12/\varsigma).$$

Consequently, the RIP estimate can be improved with a slightly weaker logarithm factor. We will detail this point later in Remark A.5.

**A.2. Proof of Theorem 2.2.**

*Proof.* We define the set of integers

$$(A.9) \quad \mathcal{L} = \mathbb{Z} \cap \left( \frac{\log(\delta)}{\log(1 + \delta)} + 1, \frac{\log(\Theta\sqrt{s})}{\log(1 + \delta)} + 1 \right),$$

and denote by  $\underline{l}, \bar{l}$  the minimum and maximum of  $\mathcal{L}$ , respectively, where  $\mathcal{L}$  has been chosen so that the integers  $\underline{l}$  and  $\bar{l}$  satisfy

$$(A.10) \quad (1 + \delta)^{\underline{l}-2} \leq \delta \quad \text{and} \quad (1 + \delta)^{\bar{l}} \geq \Theta\sqrt{s}.$$

Let  $Q := \{\mathbf{y}_1, \dots, \mathbf{y}_m\}$  be the sample set contained in  $\mathcal{U}$  and denote by  $\varrho_Q$  the discrete uniform measure associated with  $Q$ .

*Step 1.* For  $0 < \varsigma < 1$  (exact value will be set later), we seek to construct  $\tilde{\psi}$  approximating  $\psi$  such that:

- (i) For all  $\mathbf{z} \in \mathcal{E}_s$ , the following holds with probability exceeding  $1 - \varsigma$  in  $(\mathcal{U}, \varrho)$ , as well as probability exceeding  $1 - \varsigma$  in  $(Q, \varrho_Q)$ :

$$(A.11) \quad \begin{aligned} (1 - 3\delta/2) \tilde{\psi}(\mathbf{y}, \mathbf{z}) < |\psi(\mathbf{y}, \mathbf{z})| < (1 + 3\delta/2) \tilde{\psi}(\mathbf{y}, \mathbf{z}), & \text{ if } \tilde{\psi}(\mathbf{y}, \mathbf{z}) > 0, \\ |\psi(\mathbf{y}, \mathbf{z})| < 6\delta/5, & \text{ if } \tilde{\psi}(\mathbf{y}, \mathbf{z}) = 0. \end{aligned}$$

(ii) For all  $\mathbf{z} \in \mathcal{E}_s$ , there exists a pairwise disjoint family of subsets  $(I_l^{(\mathbf{z})})_{l \in \mathcal{L}}$  of  $\mathcal{U}$  depending on  $\mathbf{z}$  such that

$$(A.12) \quad \tilde{\psi}(\cdot, \mathbf{z}) = \sum_{l \in \mathcal{L}} (1 + \delta)^l \chi_{I_l^{(\mathbf{z})}}.$$

(iii) For every  $l \in \mathcal{L}$ ,  $(I_l^{(\mathbf{z})})_{\mathbf{z} \in \mathcal{E}_s}$  belongs to a finite class  $F_l$  of subsets of  $\mathcal{U}$  satisfying

$$(A.13) \quad \log(\#F_l) \leq \frac{32}{\delta^3(1 + \delta)^{2l-2}} \Theta^2 s \log(4N) \log\left(\frac{12 \log(\delta^{-1} \Theta \sqrt{s})}{\varsigma \log(1 + \delta)}\right).$$

First, for  $l \in \mathcal{L}$ , let  $D_l$  be a finite subset of  $\mathbb{C}^N$  determined as in Lemma A.3 with  $\mu = \frac{\delta(1+\delta)^{l-1}}{2}$  and  $0 < \varsigma' < 1$  (to be set accordingly to meet our needs). We have

$$(A.14) \quad \log(\#D_l) \leq \frac{32}{\delta^2(1 + \delta)^{2l-2}} \Theta^2 s \log(4N) \log(12/\varsigma').$$

For a fixed  $\mathbf{z} \in \mathcal{E}_s$ , there exists  $\mathbf{z}_l \in D_l$  and a measurable set  $\mathcal{U}_l \subset \mathcal{U}$  with  $\varrho(\mathcal{U}_l) \geq 1 - \varsigma'$  such that

$$|\psi(\mathbf{y}, \mathbf{z} - \mathbf{z}_l)| \leq \frac{\delta(1 + \delta)^{l-1}}{2}, \forall \mathbf{y} \in \mathcal{U}_l,$$

and  $\mathbf{y}_i$ 's are contained in  $\mathcal{U}_l$  for at least  $(1 - \varsigma')m$  indices  $i \in \{1, \dots, m\}$ .

We construct a pairwise disjoint family of subsets  $(I_l^{(\mathbf{z})})_l$  and mapping  $\tilde{\psi}(\cdot, \mathbf{z}) : \mathcal{U} \rightarrow \mathbb{R}$  which depend on  $\mathbf{z}$  and  $Q$ , inductively for the integers  $\bar{l} > \dots > \underline{l}$  according to:

$$(A.15) \quad \begin{aligned} I_l^{(\mathbf{z})} &= \{\mathbf{y} \in \mathcal{U} : (1 + \delta)^{l-1} < |\psi(\mathbf{y}, \mathbf{z}_l)| < (1 + \delta)^{l+1}\}, \\ I_l^{(\mathbf{z})} &= I_l^{(\mathbf{z})} \setminus \bigcup_{r>l} I_r^{(\mathbf{z})}, \text{ and} \\ \tilde{\psi}(\cdot, \mathbf{z}) &= \sum_{l \in \mathcal{L}} (1 + \delta)^l \chi_{I_l^{(\mathbf{z})}}. \end{aligned}$$

In the following, we prove that  $\tilde{\psi}$  satisfies (A.11)–(A.13). First, consider  $\mathbf{y} \in \bigcap_{l \in \mathcal{L}} \mathcal{U}_l$ . If  $\mathbf{y} \in I_l^{(\mathbf{z})}$  for some  $l \in \mathcal{L}$ , then

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

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

$$\begin{aligned} |\psi(\mathbf{y}, \mathbf{z})| &< (1 + \delta)^{l+1} + \frac{\delta}{2}(1 + \delta)^{l-1} < \left(1 + \frac{3}{2}\delta\right) \tilde{\psi}(\mathbf{y}, \mathbf{z}), \\ |\psi(\mathbf{y}, \mathbf{z})| &> (1 + \delta)^{l-1} - \frac{\delta}{2}(1 + \delta)^{l-1} > \left(1 - \frac{3\delta}{2}\right) \tilde{\psi}(\mathbf{y}, \mathbf{z}). \end{aligned}$$

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

$$|\psi(\mathbf{y}, \mathbf{z}_l)| \notin ((1 + \delta)^{l-1}, (1 + \delta)^{l+1}).$$

We notice that  $||\psi(\mathbf{y}, \mathbf{z})| - |\psi(\mathbf{y}, \mathbf{z}_l)|| < \delta(1 + \delta)^{l-1}/2$ , and there follows

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

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

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

In view of the identities in (A.10) and  $\|\psi\|_{L^\infty} \leq \Theta\sqrt{s}$ , the second inequality cannot occur. As for the first, it implies by (A.10) and assumption  $\delta < 1/13$  that  $|\psi(\mathbf{y}, \mathbf{z})| \leq \delta(1 + \delta/2)(1 + \delta) < 6\delta/5$ .

Next, consider  $\mathbf{y} \notin \bigcap_{l \in \mathcal{L}} \mathcal{U}_l$ . Condition (A.11) is not guaranteed in this case. However, these only hold with probability not exceeding

$$(A.16) \quad \begin{aligned} \varrho\left(\mathcal{U} \setminus \bigcap_{l \in \mathcal{L}} \mathcal{U}_l\right) &\leq \sum_{l \in \mathcal{L}} \varrho(\mathcal{U} \setminus \mathcal{U}_l) \leq \varsigma'(\#\mathcal{L}) \leq \frac{\log(\delta^{-1}\Theta\sqrt{s})}{\log(1 + \delta)} \varsigma' = \varsigma, \\ \#\left\{i : \mathbf{y}_i \notin \bigcap_{l \in \mathcal{L}} \mathcal{U}_l\right\} &\leq \varsigma' m(\#\mathcal{L}) \leq \frac{\log(\delta^{-1}\Theta\sqrt{s})}{\log(1 + \delta)} \varsigma' m = \varsigma m, \end{aligned}$$

when setting  $\varsigma' = \frac{\log(1 + \delta)}{\log(\delta^{-1}\Theta\sqrt{s})} \varsigma$ .

In summary, we have proved that for all  $\mathbf{z} \in \mathcal{E}_s$ , the following is satisfied:

$$(A.17) \quad (1 - 3\delta/2) \tilde{\psi}(\mathbf{y}, \mathbf{z}) < |\psi(\mathbf{y}, \mathbf{z})| < (1 + 3\delta/2) \tilde{\psi}(\mathbf{y}, \mathbf{z}), \quad \text{for } \mathbf{y} \in I,$$

$$(A.18) \quad 0 \leq |\psi(\mathbf{y}, \mathbf{z})| < 6\delta/5 \quad \text{and} \quad \tilde{\psi}(\mathbf{y}, \mathbf{z}) = 0, \quad \text{for } \mathbf{y} \in \hat{I},$$

$$(A.19) \quad \varrho(\mathcal{U}') \leq \varsigma,$$

where the three sets  $I := (\bigcap_{l \in \mathcal{L}} \mathcal{U}_l) \cap (\bigcup_{l \in \mathcal{L}} I_l^{(\mathbf{z})})$ ,  $\hat{I} := (\bigcap_{l \in \mathcal{L}} \mathcal{U}_l) \setminus (\bigcup_{l \in \mathcal{L}} I_l^{(\mathbf{z})})$ , and  $\mathcal{U}' := \mathcal{U} \setminus (\bigcap_{l \in \mathcal{L}} \mathcal{U}_l)$  define a partition of  $\mathcal{U}$ , and depend on  $\mathbf{z}$  and  $Q$ .

It remains to verify (A.13). For any  $l \in \mathcal{L}$ ,  $\#\{I_l^{(\mathbf{z})} \mid \mathbf{z} \in \mathcal{E}_s\} \leq \#D_l$  and  $\#F_l \leq \prod_{r \geq l} \#D_r$ . From (A.14), we see that

$$\log(\#F_l) \leq \sum_{r \geq l} \log(\#D_r) \leq \frac{32}{\delta^3(1 + \delta)^{2l-2}} \Theta^2 s \log(4N) \log(12/\varsigma').$$

*Step 2.* Derive essential estimates of  $\|\mathbf{z}\|_2$  and  $\|\mathbf{Az}\|_2$  in terms of  $\tilde{\psi}(\cdot, \mathbf{z})$ . First, given  $\mathbf{z} \in \mathcal{E}_s$ , we observe that

$$(A.20) \quad 1 = \|\mathbf{z}\|_2^2 = \int_{\mathcal{U}} |\psi(\mathbf{y}, \mathbf{z})|^2 d\varrho \quad \text{and} \quad \|\mathbf{Az}\|_2^2 = \sum_{i=1}^m \frac{|\psi(\mathbf{y}_i, \mathbf{z})|^2}{m} = \int_Q |\psi(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q.$$

It is easy to check that if  $\delta \leq 1/13$ , one has for real numbers  $a, b > 0$ ,  $(1 - \frac{3\delta}{2})a < b < (1 + \frac{3\delta}{2})a$  implies  $(1 - 3\delta)a^2 < b^2 < (1 + 4\delta)a^2$ , which also implies  $(1 - 4\delta)b^2 < a^2 < (1 + 4\delta)b^2$ , so that  $|b^2 - a^2| < 4\delta \min(a^2, b^2)$ . Therefore, from (A.17) we get that

$$\begin{aligned} \left| |\psi(\mathbf{y}, \mathbf{z})|^2 - |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 \right| &< 4\delta |\psi(\mathbf{y}, \mathbf{z})|^2, \quad \text{for } \mathbf{y} \in I, \\ \left| |\psi(\mathbf{y}, \mathbf{z})|^2 - |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 \right| &< 4\delta |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2, \quad \text{for } \mathbf{y} \in \hat{I}. \end{aligned}$$

This, combined with (A.18) and (A.19) and the fact that  $|\psi(\cdot, \mathbf{z})|^2$  and  $|\tilde{\psi}(\cdot, \mathbf{z})|^2$  are uniformly bounded in  $\mathcal{U}$  by  $\Theta^2 s$  and by  $(1 + \delta)^2 \Theta^2 s \leq 2\Theta^2 s$ , implies

$$\begin{aligned} \left| \|\mathbf{z}\|_2^2 - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| &\leq \int_I \left| |\psi(\mathbf{y}, \mathbf{z})|^2 - |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 \right| d\varrho \\ &\quad + \int_{\widehat{I}} \left| |\psi(\mathbf{y}, \mathbf{z})|^2 - |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 \right| d\varrho + \int_{\mathcal{U}'} \left| |\psi(\mathbf{y}, \mathbf{z})|^2 - |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 \right| d\varrho \\ \text{(A.21)} \quad &\leq 4\delta \int_I |\psi(\mathbf{y}, \mathbf{z})|^2 d\varrho + \frac{36\delta^2}{25} \varrho(\widehat{I}) + 2\Theta^2 s\varsigma. \end{aligned}$$

By noticing that  $\varrho(\widehat{I}) \leq 1$  and setting  $\varsigma = \frac{\delta}{6\Theta^2 s}$ , we infer

$$\text{(A.22)} \quad \left| \|\mathbf{z}\|_2^2 - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| < 4\delta + \frac{\delta}{6} + \frac{\delta}{3} = \frac{9\delta}{2}.$$

Repeating the above argument for the probability space  $(Q, \varrho_Q)$  with notice that  $\varrho_Q(\mathcal{U}' \cap Q) \leq \varsigma$  yields

$$\text{(A.23)} \quad \left| \|\mathbf{Az}\|_2^2 - \int_Q |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q \right| \leq 4\delta \int_Q |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q + \frac{\delta}{2}.$$

From (A.22) and (A.23), we obtain

$$\begin{aligned} \left| \|\mathbf{Az}\|_2^2 - \|\mathbf{z}\|_2^2 \right| &\leq 4\delta \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho + 5\delta \\ \text{(A.24)} \quad &\quad + (1 + 4\delta) \left| \int_Q |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| \\ &\leq 4\delta \left( 1 + \frac{9\delta}{2} \right) + 5\delta + (1 + 4\delta) \left| \int_Q |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right|. \end{aligned}$$

*Step 3.* We derive an upper bound of  $\left| \|\mathbf{Az}\|_2^2 - \|\mathbf{z}\|_2^2 \right|$  via (A.24), by employing a basic tail estimate (Lemma A.2) and the union bound. From the definition of  $\tilde{\psi}$ , we have that

$$\text{(A.25)} \quad \left| \int_Q |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho_Q - \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \right| \leq \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \left| \frac{\#(Q \cap I_l^{(z)})}{m} - \varrho(I_l^{(z)}) \right|.$$

Let  $(\kappa_l)_{l \in \mathcal{L}}$  be a sequence of positive numbers. Applying Lemma A.2, for any set  $\Delta$  in the class  $F_l$ , with probability of  $Q$  exceeding  $1 - \exp\left(-\frac{m\kappa_l\delta}{16e}\right)$ , there holds

$$\text{(A.26)} \quad \left| \frac{\#(Q \cap \Delta)}{m} - \varrho(\Delta) \right| \leq \delta\varrho(\Delta) + \kappa_l.$$

By the union bound, with probability exceeding  $1 - \sum_{l \in \mathcal{L}} \exp\left(-\frac{m\kappa_l\delta}{16e}\right) (\#F_l)$ , the previous inequality holds uniformly for all sets  $\Delta \in \bigcup_{l \in \mathcal{L}} F_l$ . Therefore, with probability exceeding  $1 - \sum_{l \in \mathcal{L}} \exp\left(-\frac{m\kappa_l\delta}{16e}\right) (\#F_l)$ , we can apply (A.26) with  $\Delta = I_l^{(z)}$

( $l \in \mathcal{L}$ ) to the sum in (A.25) and combine with (A.24) to infer that for all  $\mathbf{z} \in \mathcal{E}_s$ ,

$$\begin{aligned} \left| \|\mathbf{A}\mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2 \right| &\leq 4\delta \left(1 + \frac{9\delta}{2}\right) + 5\delta + \delta(1 + 4\delta) \int_{\mathcal{U}} |\tilde{\psi}(\mathbf{y}, \mathbf{z})|^2 d\varrho \\ &\quad + (1 + 4\delta) \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l \\ &\leq 4\delta \left(1 + \frac{9\delta}{2}\right) + 5\delta + \delta(1 + 4\delta) \left(1 + \frac{9\delta}{2}\right) + (1 + 4\delta) \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l \\ &\leq 12\delta + \delta/3 + (1 + 4\delta) \sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l, \end{aligned}$$

for the last inequality we have used  $\delta < 1/13$ .

Finally, in order to obtain Theorem 2.2, we need to assign appropriate values for  $\kappa_l$  and derive conditions on  $m$  such that

$$\sum_{l \in \mathcal{L}} (1 + \delta)^{2l} \kappa_l \leq \delta/2 \quad \text{and} \quad \sum_{l \in \mathcal{L}} \exp\left(-\frac{m\kappa_l\delta}{16e} + \log(\#F_l)\right) \leq \gamma.$$

The two inequalities can be fulfilled if for example the numbers  $\kappa_l$  and the integer  $m$  are chosen as follows:

$$\kappa_l := \frac{\delta/2}{(\#\mathcal{L})(1 + \delta)^{2l}}, \quad -\frac{m\kappa_l\delta}{16e} + \log(\#F_l) \leq \log\left(\frac{\gamma}{\#\mathcal{L}}\right), \quad l \in \mathcal{L}.$$

This implies that

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

Observe that since  $\varsigma = \frac{\delta}{6\Theta^2s}$ , we have in view of (A.13) that

$$\begin{aligned} &32e (\#\mathcal{L}) \frac{(1 + \delta)^{2l}}{\delta^2} \log(\#F_l) \\ &\leq 2^{10} e \frac{(1 + \delta)^2}{\delta^5 \log(1 + \delta)} \Theta^2s \log(4N) \log\left(\frac{\Theta\sqrt{s}}{\delta}\right) \log\left(\frac{72\Theta^2s}{\delta} \cdot \frac{\log(\Theta\sqrt{s}/\delta)}{\log(1 + \delta)}\right) \\ &< 2^{10} e \frac{\Theta^2s}{\delta^6} \log(4N) \log\left(\frac{\Theta^2s}{\delta^2}\right) \log\left(40 \frac{\Theta^2s}{\delta^2} \log\left(\frac{\Theta^2s}{\delta^2}\right)\right), \\ &32e (\#\mathcal{L}) \frac{(1 + \delta)^{2l}}{\delta^2} \log\left(\frac{\#\mathcal{L}}{\gamma}\right) \\ &\leq 32e \frac{(1 + \delta)^2}{\delta^2 \log(1 + \delta)} \Theta^2s \log\left(\frac{\Theta\sqrt{s}}{\delta}\right) \log\left(\frac{1}{\gamma} \cdot \frac{\log(\Theta\sqrt{s}/\delta)}{\log(1 + \delta)}\right) \\ &< 2^5 e \frac{\Theta^2s}{\delta^3} \log\left(\frac{\Theta^2s}{\delta^2}\right) \log\left(\frac{1}{\gamma\delta} \log\left(\frac{\Theta^2s}{\delta^2}\right)\right). \end{aligned}$$

Here, we employed estimates  $(1 + \delta)^2 < (14/13)^2$  and  $\log^{-1}(1 + \delta) < 1.1/\delta$ , obtained from the small condition of  $\delta$ . Combining the two estimates and  $(a + b) \leq 2 \max(a, b)$  shows that  $m$  as in Theorem 2.2 is suitable.  $\square$

We conclude this subsection with two remarks detailing some slight technical improvements of Theorem 2.2.

*Remark A.4.* We can give a sharper approximation of  $\psi$  by refining the map  $\tilde{\psi}(\cdot, \mathbf{z})$ . For example, given an integer  $k \geq 1$ , we define  $\mathcal{L}$  as

$$(A.27) \quad \mathcal{L} = \frac{\mathbb{Z}}{k} \cap \left( \frac{\log(\delta)}{\log(1+\delta)} + \frac{1}{k}, \frac{\log(\Theta\sqrt{s})}{\log(1+\delta)} + \frac{1}{k} \right),$$

and construct the domains  $D_l$  using  $\mu = \frac{\delta(1+\delta)^{l-1/k}}{2k}$ . We replace the domains  $I_l'(z)$  in (A.15) by

$$I_l'(z) := \{\mathbf{y} \in \mathcal{U} : (1+\delta)^{l-1/k} < |\psi(\mathbf{y}, \mathbf{z}_l)| < (1+\delta)^{l+1/k}\}.$$

Using the elementary inequalities  $(1+\delta)^{1/k} \leq 1+\delta/k$  and  $(1+\delta)^{1/k} - \frac{\delta}{2k(1+\delta)^{1/k}} \geq 1 + \frac{\delta}{2k}$  and assuming  $\delta/k \leq 1/13$ , we verify that (A.17) can be improved as

$$(A.28) \quad \left(1 - \frac{3\delta}{2k}\right) \tilde{\psi}(\mathbf{y}, \mathbf{z}) < |\psi(\mathbf{y}, \mathbf{z})| < \left(1 + \frac{3\delta}{2k}\right) \tilde{\psi}(\mathbf{y}, \mathbf{z}), \text{ for } \mathbf{y} \in I,$$

while (A.18) and (A.19) are unchanged. An inspection of the proof shows that this yields (A.21) with  $4\delta/k$  instead of  $4\delta$  and with  $36\delta^2/25$  and  $2\Theta^2s\varsigma$  unchanged. We mention, however, that the cardinality  $\#\mathcal{L}$  and the bound in (A.13) on  $\log(\#(F_l))$  gets roughly multiplied by  $k$  and  $k^3$ , respectively.

*Remark A.5.* If (A.8) is applied to bound  $\log(\#D_l)$ , we can obtain

$$\begin{aligned} \log(\#D_l) &\leq \frac{32}{\delta^2(1+\delta)^{2l-2}} \Theta^2 s \log\left(e + \frac{eN\delta^2(1+\delta)^{2l-2}}{4\Theta^2 s \log(12/\varsigma')}\right) \log(12/\varsigma') \\ &\leq \frac{32}{\delta^2(1+\delta)^{2l-2}} \Theta^2 s \log\left(e + \frac{eN\delta^2}{4\log(12/\varsigma')}\right) \log(12/\varsigma'), \quad l \in \mathcal{L}, \end{aligned}$$

where  $\varsigma' = \frac{\delta \log(1+\delta)}{6\Theta^2 s \log(\delta^{-1}\Theta\sqrt{s})}$ , instead of (A.14). Subsequently, the term  $\log(4N)$  in sample complexity (2.2) and (3.7) can be replaced by  $\log\left(e + \frac{eN\delta^2}{4\log\left(36\frac{\Theta^2 s}{\delta^2} \log\left(\frac{\Theta^2 s}{\delta^2}\right)\right)}\right)$ .

### A.3. Proof of Theorem 3.3.

*Proof.* The proof of Theorem 3.3 follows closely that of Theorem 2.2 with one critical change: Instead of  $\mathcal{E}_s$ , we only approximate  $\|\mathbf{Az}\|_2$  on the set  $\mathcal{E}_s^\ell$  defined as

$$(A.29) \quad \mathcal{E}_s^\ell = \{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_2 = 1 \text{ and } K(\text{supp}(\mathbf{z})) \leq K(s)\}.$$

We have that  $\mathcal{E}_s^\ell \subset \text{conv}(\mathcal{P}^\ell)$ , where  $\mathcal{P}^\ell = \{\pm \frac{\mathbf{e}_j}{\omega_j} \sqrt{2K(s)}, \pm i \frac{\mathbf{e}_j}{\omega_j} \sqrt{2K(s)}\}_{1 \leq j \leq N}$ , with  $\omega_j = \|\Psi_j\|_{L^\infty}$  and  $(\mathbf{e}_j)$  being canonical unit vectors in  $\mathbb{C}^N$ . On the other hand,

$$(A.30) \quad |\psi(\mathbf{y}, \mathbf{z})| \leq \sqrt{K(s)}, \quad \forall \mathbf{z} \in \mathcal{E}_s^\ell, \mathbf{y} \in \mathcal{U}.$$

We thus can derive an extended covering number result for  $\mathcal{E}_s^\ell$  (similar to Lemma A.3 for  $\mathcal{E}_s$ ), then replace the bound  $\Theta\sqrt{s}$  by  $\sqrt{K(s)}$  throughout the previous proofs, resulting in Theorem 3.3. □

## APPENDIX B. ITERATIVE HARD THRESHOLDING ON LOWER SETS

Thresholding approach for finding best  $s$ -term approximations consists of solving problems of the form:

$$(B.1) \quad \min \|\mathbf{A}\mathbf{z} - \tilde{\mathbf{g}}\|_2 \quad \text{subject to} \quad \text{supp}(\mathbf{z}) \leq s.$$

In this appendix, we show that lower RIP is also relevant to this approach by studying a thresholding method for lower sparse recovery, which is guaranteed with the reduced query complexity (3.20). The method solves the following constrained minimization problem:

$$(B.2) \quad \min \|\mathbf{A}\mathbf{z} - \tilde{\mathbf{g}}\|_2 \quad \text{subject to} \quad \text{supp}(\mathbf{z}) \leq s, \quad \text{supp}(\mathbf{z}) \text{ lower.}$$

To achieve the minimum in (B.2) we first define the hard lower thresholding operator

$$(B.3) \quad H_s^\ell(\mathbf{z}) = \arg \min_{\substack{\text{supp}(\tilde{\mathbf{z}}) \text{ lower} \\ |\text{supp}(\tilde{\mathbf{z}})|=s}} \|\mathbf{z} - \tilde{\mathbf{z}}\|_2.$$

Our goal is to approximate a smooth function of the form  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$  by a sparse expansion supported in a predefined polynomial subspace  $\mathbb{P}_{\mathcal{J}}$ . We let  $s$  be the sparsity level,  $\mathcal{J} = \mathcal{H}_s$  defined as in (1.10), with  $\mathbf{A} \in \mathbb{C}^{m \times N}$  and  $\tilde{\mathbf{g}} \in \mathbb{C}^m$  the normalized sampling matrix and vector of observations, respectively. In what follows, we consider the following lower version of iterative hard thresholding algorithm [4].

**Algorithm B.1** (Iterative hard thresholding on lower sets).

- (1) **Initialization:** set the initial approximation  $\mathbf{c}^0$  as an  $s$ -sparse lower vector, e.g.,  $\mathbf{c}^0 = \mathbf{0}$ .
- (2) **Iteration:** repeat until a stopping criterion is met at  $n = \bar{n}$ :

$$\mathbf{c}^{n+1} = H_s^\ell(\mathbf{c}^n + \mathbf{A}^*(\tilde{\mathbf{g}} - \mathbf{A}\mathbf{c}^n)).$$

- (3) **Output:**  $\mathbf{c}^\# = \mathbf{c}^{\bar{n}}$ , and  $g^\# = \sum_{\nu \in \mathcal{H}_s} c_\nu^\# \Psi_\nu$ .

Our method is designed for function reconstruction with preference to low-indexed terms, similarly to a thresholding approach named *iterative hard weighted thresholding*, [28]. However, we focus on high-dimensional polynomial approximations and use the lower instead of weighted sparsity constraint for the thresholding operator. A surrogate of  $H_s^\ell$  can exploit the lower set structure and is independent of weights, whose optimal choice is an important problem for iterative hard weighted thresholding. Below we provide the convergence result for Algorithm B.1.

**Theorem B.2.** Let  $\mathcal{J} = \mathcal{H}_s$ . For  $s \geq 2$ , consider a number of samples as in (3.20). Let  $\mathbf{y}_1, \dots, \mathbf{y}_m$  be drawn independently from the orthogonalization measure  $\varrho$  associated to  $\{\Psi_\nu\}$  and  $\mathbf{A} \in \mathbb{C}^{m \times N}$  the normalized sampling matrix. Then:

- (i) with probability exceeding  $1 - \gamma$ , the following holds for all  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$ : the function  $g^n = \sum_{\nu \in \mathcal{H}_s} c_\nu^n \Psi_\nu$ , with  $\mathbf{c}^n = (c_\nu^n)_{\nu \in \mathcal{H}_s}$  solving (2), satisfies

$$(B.4) \quad \|g^n - g_{\Lambda^*}\|_2 \leq \rho^n \|g^0 - g_{\Lambda^*}\|_2 + \frac{\tau}{\sqrt{m}} \|\tilde{\boldsymbol{\xi}}\|_2,$$

where  $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_i)_{1 \leq i \leq m} = (g_{(\Lambda^*)^c}(\mathbf{y}_i))_{1 \leq i \leq m}$ .



(ii) for a fixed function  $g = \sum_{\nu \in \mathcal{F}} c_\nu \Psi_\nu$  with  $\|g\|_{\omega,1} < \infty$ , with probability exceeding  $1 - \gamma$ , for all  $n \in \mathbb{N}$ , the function  $g^n = \sum_{\nu \in \mathcal{H}_s} c_\nu^n \Psi_\nu$ , with  $\mathbf{c}^n = (c_\nu^n)_{\nu \in \mathcal{H}_s}$  solving (2), satisfies

$$(B.5) \quad \|g^n - g\|_2 \leq \rho^n \|g^0 - g_{\Lambda^*}\|_2 + (\tau\sqrt{2} + 1)\sigma_s^{(\ell)}(g)_2 + \frac{\tau\sqrt{2}}{\sqrt{K(s)}} \|g_{(\Lambda^*)^c}\|_{\omega,1}.$$

Here,  $\rho$  and  $\tau$  are universal constants with  $\rho < 1$ ,  $\Lambda^*$  is the support of best lower  $s$ -term approximation of  $g$  in  $\ell_2$ -norm, and  $\sigma_s^{(\ell)}(g)_2 = \|g_{(\Lambda^*)^c}\|_2$  is the best lower  $s$ -term error in  $\ell_2$ -norm.

*Proof.* We will prove that the assertion (i) holds provided that  $\delta_{\ell,\alpha s} < 1/2$  (where  $\alpha = 2$  for Legendre system and  $\alpha = 3$  for Chebyshev system), following the technique in [19] for standard iterative hard thresholding. First, let  $\mathbf{v}^n = \mathbf{c}^n + \mathbf{A}^*(\tilde{\mathbf{g}} - \mathbf{A}\mathbf{c}^n)$ . On one hand, by definition (2) of  $\mathbf{c}^{n+1}$  in Algorithm B.1, and since  $\Lambda^*$  is lower of cardinality smaller than  $s$ ,

$$\begin{aligned} \|\mathbf{v}^n - \mathbf{c}_{\Lambda^*}\|_2^2 &\geq \|\mathbf{v}^n - \mathbf{c}^{n+1}\|_2^2 = \|(\mathbf{v}^n - \mathbf{c}_{\Lambda^*}) - (\mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*})\|_2^2 \\ &= \|\mathbf{v}^n - \mathbf{c}_{\Lambda^*}\|_2^2 + \|\mathbf{c}_{\Lambda^*} - \mathbf{c}^{n+1}\|_2^2 - 2\Re\langle \mathbf{v}^n - \mathbf{c}_{\Lambda^*}, \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle. \end{aligned}$$

This yields that  $\|\mathbf{c}_{\Lambda^*} - \mathbf{c}^{n+1}\|_2^2 \leq 2\Re\langle \mathbf{v}^n - \mathbf{c}_{\Lambda^*}, \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle$ . On the other hand,

$$\begin{aligned} \langle \mathbf{v}^n - \mathbf{c}_{\Lambda^*}, \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle &= \langle (\mathbf{Id} - \mathbf{A}^* \mathbf{A})\mathbf{c}^n + \mathbf{A}^* \tilde{\mathbf{g}} - \mathbf{c}_{\Lambda^*}, \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle \\ &= \langle (\mathbf{Id} - \mathbf{A}^* \mathbf{A})(\mathbf{c}^n - \mathbf{c}_{\Lambda^*}), \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle + \frac{1}{\sqrt{m}} \langle \tilde{\boldsymbol{\xi}}, \mathbf{A}(\mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*}) \rangle. \end{aligned}$$

It is easy to see that  $\Lambda^{n,*} := \text{supp}(\mathbf{c}^n) \cup \text{supp}(\mathbf{c}^{n+1}) \cup \text{supp}(\mathbf{c}_{\Lambda^*})$  is a lower set which satisfies  $K(\Lambda^{n,*}) \leq 3K(s) \leq K(\alpha s)$  (see (4.7) for the second inequality), thus from Theorem 3.3,

$$\begin{aligned} |\langle (\mathbf{Id} - \mathbf{A}^* \mathbf{A})(\mathbf{c}^n - \mathbf{c}_{\Lambda^*}), \mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*} \rangle| &\leq \delta_{\ell,\alpha s} \|\mathbf{c}^n - \mathbf{c}_{\Lambda^*}\|_2 \|\mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*}\|_2, \\ |\langle \tilde{\boldsymbol{\xi}}, \mathbf{A}(\mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*}) \rangle| &\leq \sqrt{1 + \delta_{\ell,\alpha s}} \|\tilde{\boldsymbol{\xi}}\|_2 \|\mathbf{c}^{n+1} - \mathbf{c}_{\Lambda^*}\|_2. \end{aligned}$$

We then have

$$\|\mathbf{c}_{\Lambda^*} - \mathbf{c}^{n+1}\|_2 \leq 2\delta_{\ell,\alpha s} \|\mathbf{c}^n - \mathbf{c}_{\Lambda^*}\|_2 + \frac{2\sqrt{1 + \delta_{\ell,\alpha s}}}{\sqrt{m}} \|\tilde{\boldsymbol{\xi}}\|_2.$$

The inequality (B.4) follows given  $\delta_{\ell,\alpha s} < 1/2$ .

For (ii), note that we have  $\mathbb{E}(\tilde{\xi}_i^2) = \|\mathbf{c}_{(\Lambda^*)^c}\|_2^2$  and  $|\tilde{\xi}_i| \leq \|\mathbf{c}_{(\Lambda^*)^c}\|_{\omega,1}$  implying that

$$\mathbb{E}\left(\tilde{\xi}_i^2 - \mathbb{E}[\tilde{\xi}_i^2]\right)^2 \leq \mathbb{E}[\xi_i^4] \leq \|\mathbf{c}_{(\Lambda^*)^c}\|_{\omega,1}^2 \|\mathbf{c}_{(\Lambda^*)^c}\|_2^2.$$

We can consider two cases  $\|\mathbf{c}_{(\Lambda^*)^c}\|_2^2 > \frac{\|\mathbf{c}_{(\Lambda^*)^c}\|_{\omega,1}^2}{K(s)}$  and  $\|\mathbf{c}_{(\Lambda^*)^c}\|_2^2 \leq \frac{\|\mathbf{c}_{(\Lambda^*)^c}\|_{\omega,1}^2}{K(s)}$  and prove using Bernstein's inequality, similarly to Theorem 4.8, that with probability exceeding  $1 - \gamma$ , we always have

$$\frac{\|\tilde{\boldsymbol{\xi}}\|_2^2}{m} \leq 2\|\mathbf{c}_{(\Lambda^*)^c}\|_2^2 + \frac{2\|\mathbf{c}_{(\Lambda^*)^c}\|_{\omega,1}^2}{K(s)}.$$

Substituting the previous inequality to (B.4), we obtain

$$\|g^n - g_{\Lambda^*}\|_2 \leq \rho^n \|g^0 - g_{\Lambda^*}\|_2 + \tau\sqrt{2} \|c_{(\Lambda^*)^c}\|_2 + \frac{\tau\sqrt{2}}{\sqrt{K(s)}} \|c_{(\Lambda^*)^c}\|_{\omega,1}.$$

The inequality (B.5) can then be concluded by virtue of the triangle inequality.  $\square$

In general, it may not be feasible to find the optimal vector  $H_s^\ell(\mathbf{z})$  exactly. Greedy procedures can be used to explore a near optimal lower,  $s$ -sparse truncation of  $\mathbf{z}$  and provide an approximation  $\tilde{H}_s^\ell(\mathbf{z})$  of  $H_s^\ell(\mathbf{z})$  (see [9]). A significant advantage in our context is that we do not have to compute the components of  $\mathbf{z}$  inductively but have them all at hand. The exploration cost is therefore a fraction of that of matrix multiplication. Exploiting full knowledge of  $\mathbf{z}$ , new algorithms for constructing the surrogates can also be considered. However, we expect the theoretical recovery guarantees using the surrogates are weaker than those proved in Theorem B.2; see also the case of low rank tensor recovery [38]. The numerical realization of Algorithm B.1 and comparison with standard iterative hard thresholding and iterative hard weighted thresholding will be conducted in future work.

## REFERENCES

- [1] B. Adcock, *Infinite-dimensional compressed sensing and function interpolation*, preprint (2015).
- [2] B. Adcock, *Infinite-dimensional  $\ell_1$  minimization and function approximation from pointwise data*, preprint (2015).
- [3] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, *Convergence of quasi-optimal stochastic Galerkin methods for a class of PDES with random coefficients*, *Comput. Math. Appl.* **67** (2014), no. 4, 732–751, DOI 10.1016/j.camwa.2013.03.004. MR3163876
- [4] T. Blumensath and M. E. Davies, *Iterative hard thresholding for compressed sensing*, *Appl. Comput. Harmon. Anal.* **27** (2009), no. 3, 265–274, DOI 10.1016/j.acha.2009.04.002. MR2559726
- [5] J. Bourgain, *An improved estimate in the restricted isometry problem*, *Geometric aspects of functional analysis*, *Lecture Notes in Math.*, vol. 2116, Springer, Cham, 2014, pp. 65–70, DOI 10.1007/978-3-319-09477-9\_5. MR3364679
- [6] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, *IEEE Trans. Inform. Theory* **52** (2006), no. 2, 489–509, DOI 10.1109/TIT.2005.862083. MR2236170
- [7] M. Cheraghchi, V. Guruswami, and A. Velingker, *Restricted isometry of Fourier matrices and list decodability of random linear codes*, *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, Philadelphia, PA, 2012, pp. 432–442. MR3186765
- [8] A. Chernov and D. Dũng, *New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness*, *J. Complexity* **32** (2016), no. 1, 92–121, DOI 10.1016/j.jco.2015.09.001. MR3424640
- [9] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab, *Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs*, *Modél. Math. Anal. Numér.* **47** (2013), no. 1, 253–280.
- [10] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone, *Discrete least squares polynomial approximation with random evaluations—application to parametric and stochastic elliptic PDEs*, *ESAIM Math. Model. Numer. Anal.* **49** (2015), no. 3, 815–837, DOI 10.1051/m2an/2014050. MR3342229
- [11] A. Chkifa, A. Cohen, and C. Schwab, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs*, *Found. Comput. Math.* **14** (2014), no. 4, 601–633, DOI 10.1007/s10208-013-9154-z. MR3230010
- [12] A. Chkifa, A. Cohen, and C. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*, *J. Math. Pures Appl.* (9) **103** (2015), no. 2, 400–428, DOI 10.1016/j.matpur.2014.04.009. MR3298364

- [13] A. Cohen, M. A. Davenport, and D. Leviatan, *On the stability and accuracy of least squares approximations*, *Found. Comput. Math.* **13** (2013), no. 5, 819–834, DOI 10.1007/s10208-013-9142-3. MR3105946
- [14] A. Cohen and R. DeVore, *Approximation of high-dimensional parametric PDEs*, *Acta Numer.* **24** (2015), 1–159, DOI 10.1017/S0962492915000033. MR3349307
- [15] A. Cohen, R. DeVore, and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's*, *Anal. Appl. (Singap.)* **9** (2011), no. 1, 11–47, DOI 10.1142/S0219530511001728. MR2763359
- [16] A. Cohen, G. Migliorati, and F. Nobile, *Discrete least-squares approximations over optimized downward closed polynomial spaces in arbitrary dimension*, *Constructive Approximation* (2016), to appear.
- [17] D. L. Donoho, *Compressed sensing*, *IEEE Trans. Inform. Theory* **52** (2006), no. 4, 1289–1306, DOI 10.1109/TIT.2006.871582. MR2241189
- [18] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of PDEs with stochastic inputs*, *J. Comput. Phys.* **230** (2011), no. 8, 3015–3034, DOI 10.1016/j.jcp.2011.01.002. MR2774328
- [19] S. Foucart, *Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants*, *Approximation Theory XIII: San Antonio 2010*, *Springer Proc. Math.*, vol. 13, Springer, New York, 2012, pp. 65–77, DOI 10.1007/978-1-4614-0772-0\_5. MR3026204
- [20] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, *Applied and Numerical Harmonic Analysis*, Birkhäuser/Springer, New York, 2013. MR3100033
- [21] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, *Recovering compressively sampled signals using partial support information*, *IEEE Trans. Inform. Theory* **58** (2012), no. 2, 1122–1134, DOI 10.1109/TIT.2011.2167214. MR2918013
- [22] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991. MR1083354
- [23] M. D. Gunzburger, C. G. Webster, and G. Zhang, *Stochastic finite element methods for partial differential equations with random input data*, *Acta Numer.* **23** (2014), 521–650, DOI 10.1017/S0962492914000075. MR3202242
- [24] J. Hampton and A. Doostan, *Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies*, *J. Comput. Phys.* **280** (2015), 363–386, DOI 10.1016/j.jcp.2014.09.019. MR3273141
- [25] I. Haviv and O. Regev, *The restricted isometry property of subsampled fourier matrices*, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (2016), 288–297.
- [26] V. H. Hoang and C. Schwab, *Regularity and generalized polynomial chaos approximation of parametric and random second-order hyperbolic partial differential equations*, *Anal. Appl. (Singap.)* **10** (2012), no. 3, 295–326, DOI 10.1142/S0219530512500145. MR2948896
- [27] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, *J. Amer. Statist. Assoc.* **58** (1963), 13–30. MR0144363
- [28] J. Jo, *Iterative hard thresholding for weighted sparse approximation*, arXiv:1312.3582 (2013).
- [29] T. Kühn, W. Sickel, and T. Ullrich, *Approximation of mixed order Sobolev functions on the  $d$ -torus: asymptotics, preasymptotics, and  $d$ -dependence*, *Constr. Approx.* **42** (2015), no. 3, 353–398, DOI 10.1007/s00365-015-9299-x. MR3416161
- [30] L. Mathelin and K. A. Gallivan, *A compressed sensing approach for partial differential equations with random input data*, *Commun. Comput. Phys.* **12** (2012), no. 4, 919–954, DOI 10.4208/cicp.151110.090911a. MR2913445
- [31] A. Narayan and T. Zhou, *Stochastic collocation on unstructured multivariate meshes*, *Commun. Comput. Phys.* **18** (2015), no. 1, 1–36, DOI 10.4208/cicp.020215.070515a. MR3371550
- [32] F. Nobile, R. Tempone, and C. G. Webster, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, *SIAM J. Numer. Anal.* **46** (2008), no. 5, 2411–2442, DOI 10.1137/070680540. MR2421041
- [33] F. Nobile, R. Tempone, and C. G. Webster, *A sparse grid stochastic collocation method for partial differential equations with random input data*, *SIAM J. Numer. Anal.* **46** (2008), no. 5, 2309–2345, DOI 10.1137/060663660. MR2421037
- [34] J. Peng, J. Hampton, and A. Doostan, *On polynomial chaos expansion via gradient-enhanced  $\ell_1$ -minimization*, *J. Comput. Phys.* **310** (2016), 440–458, DOI 10.1016/j.jcp.2015.12.049. MR3457977

- [35] 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, DOI 10.1016/j.jcp.2014.02.024. MR3183283
- [36] R. B. Platte, L. N. Trefethen, and A. B. J. Kuijlaars, *Impossibility of fast stable approximation of analytic functions from equispaced samples*, SIAM Rev. **53** (2011), no. 2, 308–318, DOI 10.1137/090774707. MR2806639
- [37] H. Rauhut, *Compressive sensing and structured random matrices*, Theoretical foundations and numerical methods for sparse recovery, Radon Ser. Comput. Appl. Math., vol. 9, Walter de Gruyter, Berlin, 2010, pp. 1–92, DOI 10.1515/9783110226157.1. MR2731597
- [38] H. Rauhut, R. Schneider, and Ž. Stojanac, *Low rank tensor recovery via iterative hard thresholding*, Linear Algebra Appl. **523** (2017), 220–262, DOI 10.1016/j.laa.2017.02.028. MR3624675
- [39] H. Rauhut and C. Schwab, *Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations*, Math. Comp. **86** (2017), no. 304, 661–700, DOI 10.1090/mcom/3113. MR3584544
- [40] H. Rauhut and R. Ward, *Sparse Legendre expansions via  $\ell_1$ -minimization*, J. Approx. Theory **164** (2012), no. 5, 517–533, DOI 10.1016/j.jat.2012.01.008. MR2903116
- [41] H. Rauhut and R. Ward, *Interpolation via weighted  $\ell_1$  minimization*, Appl. Comput. Harmon. Anal. **40** (2016), no. 2, 321–351, DOI 10.1016/j.acha.2015.02.003. MR3440176
- [42] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math. **61** (2008), no. 8, 1025–1045, DOI 10.1002/cpa.20227. MR2417886
- [43] H. Tran, C. G. Webster, and G. Zhang, *Analysis of quasi-optimal polynomial approximations for parameterized pdes with deterministic and stochastic coefficients*, Numerische Mathematik (2016), to appear.
- [44] E. van den Berg and M. P. Friedlander, *SPGL1: A solver for large-scale sparse reconstruction*, June 2007, <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [45] E. van den Berg and M. P. Friedlander, *Probing the Pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput. **31** (2008/09), no. 2, 890–912, DOI 10.1137/080714488. MR2466141
- [46] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg. **191** (2002), no. 43, 4927–4948, DOI 10.1016/S0045-7825(02)00421-8. MR1932024
- [47] L. Yan, L. Guo, and D. Xiu, *Stochastic collocation algorithms using  $\ell_1$ -minimization*, Int. J. Uncertain. Quantif. **2** (2012), no. 3, 279–293, DOI 10.1615/Int.J.UncertaintyQuantification.2012003925. MR3044795
- [48] X. Yang and G. E. Karniadakis, *Reweighted  $\ell_1$ -minimization method for stochastic elliptic differential equations*, Journal of Computational Physics **248** (2013), 87–108.
- [49] X. Yu and S. Baek, *Sufficient conditions on stable recovery of sparse signals with partial support information*, IEEE Signal Processing Letters **20** (2013), no. 5, 539–542.

DEPARTMENT OF COMPUTATIONAL AND APPLIED MATHEMATICS, OAK RIDGE NATIONAL LABORATORY, 1 BETHEL VALLEY ROAD, P.O. BOX 2008, OAK RIDGE, TENNESSEE 37831-6164

*E-mail address:* `chkifam@ornl.gov`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TENNESSEE, KNOXVILLE, TENNESSEE 37996

*E-mail address:* `ndexter@utk.edu`

DEPARTMENT OF COMPUTATIONAL AND APPLIED MATHEMATICS, OAK RIDGE NATIONAL LABORATORY, 1 BETHEL VALLEY ROAD, P.O. BOX 2008, OAK RIDGE, TENNESSEE 37831-6164

*E-mail address:* `tranha@ornl.gov`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TENNESSEE, KNOXVILLE, TENNESSEE 37996  
 – AND – DEPARTMENT OF COMPUTATIONAL AND APPLIED MATHEMATICS, OAK RIDGE NATIONAL LABORATORY, 1 BETHEL VALLEY ROAD, P.O. BOX 2008, OAK RIDGE, TENNESSEE 37831-6164

*E-mail address:* `webstercg@math.utk.edu`