Interpolation via weighted ℓ_1 minimization

Holger Rauhut*, Rachel Ward †

August 3, 2013

Abstract

Functions of interest are often smooth and sparse in some sense, and both priors should be taken into account when interpolating sampled data. Classical linear interpolation methods are effective under strong regularity assumptions, but cannot incorporate nonlinear sparsity structure. At the same time, nonlinear methods such as ℓ_1 minimization can reconstruct sparse functions from very few samples, but do not necessarily encourage smoothness. Here we show that *weighted* ℓ_1 minimization effectively merges the two approaches, promoting both sparsity and smoothness in reconstruction. More precisely, we provide specific choices of weights in the ℓ_1 objective to achieve rates for functions with coefficient sequences in weighted ℓ_p spaces, $p \leq 1$. We consider the implications of these results for spherical harmonic and polynomial interpolation, in the univariate and multivariate setting. Along the way, we extend concepts from compressive sensing such as the restricted isometry property and null space property to accommodate weighted sparse expansions; these developments should be of independent interest in the study of structured sparse approximations and continuous-time compressive sensing problems.

Key words: bounded orthonormal systems, compressive sensing, interpolation, weighted sparsity, ℓ_1 minimization

Dedicated to the memory of Ted Odell

1 Introduction

This paper aims to merge classical smoothness-based methods for function interpolation with modern sparsity constraints and nonlinear reconstruction methods. We will focus on the classical interpolation problem, where given sampling points and associated function

*H. Rauhut is with RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Templergraben 55, 52062 Aachen, Germany, rauhut@mathc.rwth-aachen.de

†R. Ward is with the Mathematics Department at the University of Texas at Austin, 2515 Speedway, Austin, Texas, 78712, rward@ma.utexas.edu

values, we wish to find a function agreeing with the data which is suitably well-behaved. Classically, well-behavedness has been measured in terms of smoothness: the more derivatives a function has, the stronger the reconstruction rate obtained using linear methods. More recently, areas such as compressive sensing have focused on sparsity rather than smoothness as a measure of complexity. Results in compressive sensing imply that a function with sparse representation in a known basis can be reconstructed from a small number of suitably randomly distributed sampling points, and reconstructed using nonlinear techniques such as convex optimization or greedy methods. In reality, functions of interest may only be somewhat smooth and somewhat sparse. This is particularly apparent in high-dimensional problems, where sparse and low-degree tensor product expansions are often preferred according to the *sparsity-of-effects principle*, which states that most models are principally governed by main effects and low order interactions. We might hope to combine classical smoothness-based approaches with the nonlinear sparsity-based approaches to arrive at better interpolation methods for such functions.

Recall that the smoothness of a function tends to be reflected in the rapid decay of its Fourier series, and vice versa. Smoothness can then be viewed as a structured sparsity constraint, with low-order Fourier basis functions being more likely to contribute to the best s-term approximation. We will show that structured sparse expansions are imposed by weighted ℓ_p coefficient spaces in the range $0 < p \leq 1$. Accordingly, we will use weighted ℓ_1 minimization, a convex surrogate for weighted ℓ_p minimization with $p < 1$, as our reconstruction method of choice.

The contributions of this paper are as follows.

1. We provide the first rigorous analysis for function interpolation using weighted ℓ_1 minimization. We show that with the appropriate choice of weights, one obtains approximation rates for functions with coefficient sequences lying in weighted ℓ_p spaces with $0 < p \leq 1$. In the high-dimensional setting, our rates are better than those possible by classical linear interpolation methods, and require only mild smoothness assumptions. Indeed, the number of sampling points required by weighted ℓ_1 minimization to achieve a desired rate grows only linearly with the ambient dimension, rather than exponentially. We illustrate the power of our results through several specific examples, including spherical harmonic interpolation and tensorized Chebyshev and Legendre polynomial interpolation, and show how we are able to improve on previous estimates for unweighted ℓ_1 minimization for such problems. We expect that the results for polynomial interpolation should have applications in uncertainty quantification [25] and, in particular, in the computation of generalized polynomial chaos expansions [13, 14].

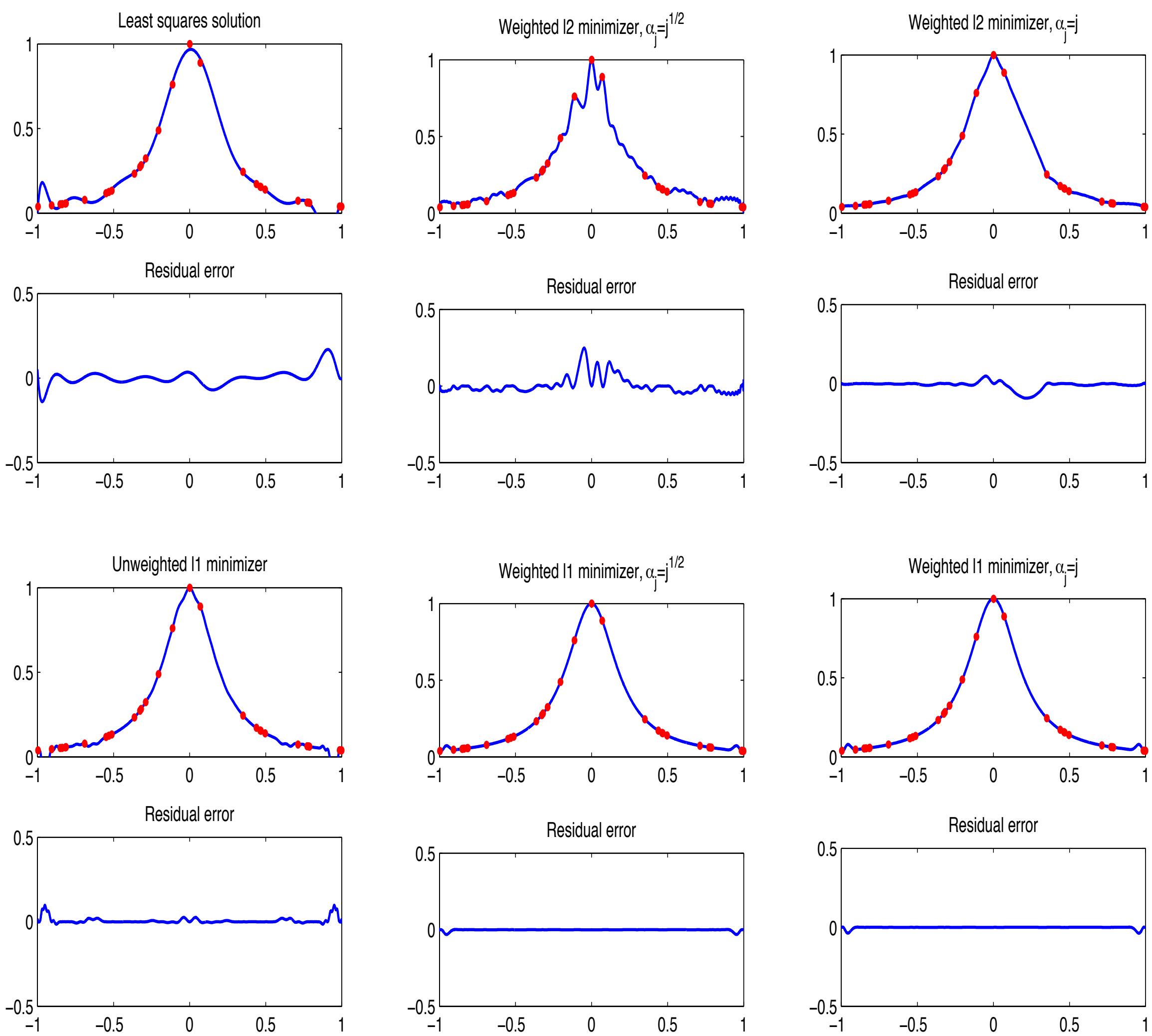


Figure 1: Illustration of the effectiveness of weighted ℓ_1 minimization for function interpolation. All approximations were obtained from the same set of $m = 30$ sampling points chosen i.i.d. from the Chebyshev distribution on $[-1, 1]$.

2. We show through numerical experiments that weighted ℓ_1 minimization tends to produce more accurate approximations which are less sensitive to the choice of interpolation points, compared to more classical reconstruction methods like least squares and weighted ℓ_2 minimization. As such, reconstructions using weighted ℓ_1 minimization appear to be more robust to *Runge's phenomenon*, or the tendency of high-degree

polynomial interpolants to oscillate near the edges of their domain. To illustrate these effects, Figure 1 plots several Legendre polynomial interpolations of the function

$$f_1(x) = \frac{1}{1+25x^2},$$

which was originally considered by Runge [34] to illustrate the instability of polynomial interpolation. Among all interpolation methods, we find that weighted ℓ_1 minimization results in smaller residual error. More details and extensive numerical experiments can be found in Section 2.3.

3. In order to derive stability and robustness recovery guarantees for weighted ℓ_1 minimization, we generalize the notion of restricted isometry property in compressive sensing to weighted restricted isometry property, and also develop notions of weighted sparsity which take into account prior information on the likelihood that any particular index contributes to the sparse expansion. These developments should be of independent interest as tools that can be used more generally for the analysis of structured sparsity models and continuous-time sparse approximation problems.

We will assume throughout that the sampling points for interpolation are drawn from a suitable probability distribution, and we focus only on the setting where interpolation points are chosen in advance, independent of the target function.

1.1 Organization

The remainder of the paper is organized as follows. In Sections 1.2 and 1.3, we introduce weighted ℓ_p spaces and discuss how they promote smoothness and sparsity. In Section 1.4 we state two of the main interpolation results, and in Section 1.5 we introduce the concept of weighted restricted isometry property for a linear map. Section 1.6 discusses previous work on weighted ℓ_1 minimization, and in Section 1.7 we compare our main results to those possible with linear reconstruction methods. In Section 2 we discuss the implications of our main results for spherical harmonic and tensorized polynomial bases, and provide several numerical illustrations. We further analyze concepts pertaining to weighted sparsity in Section 3, and in Section 4 we elaborate on the weighted restricted isometry property and weighted null space property. In Section 5 we show that matrices arising from orthonormal systems have the weighted restricted isometry property as long as the weights are matched to the L_∞ norms of the function system, and we finish in Section 6 by presenting our main results on interpolation via weighted ℓ_1 minimization.

1.2 Preliminaries on weighted sparsity

We will work with coefficient sequences $\boldsymbol{x}$ indexed by a set Λ which may be finite or countably infinite. We will associate to a vector $\boldsymbol{\omega} = (\omega_j)_{j\in\Lambda}$ of weights $\omega_j \geq 1$ the weighted ℓ_p spaces,

$$\ell_{\omega,p} := \left\{ \boldsymbol{x} = (x_j)_{j\in\Lambda}, \quad \|\boldsymbol{x}\|_{\omega,p} := \Big(\sum_{j\in\Lambda} \omega_j^{2-p}|x_j|^p\Big)^{1/p} < \infty \right\}, \quad 0 < p \leq 2. \tag{1}$$

Also central to our analysis will be the weighted ℓ_0-"norm",

$$\|\boldsymbol{x}\|_{\omega,0} = \sum_{\{j:x_j\neq 0\}} \omega_j^2,$$

which counts the squared weights of non-zero entries of $\boldsymbol{x}$. We also define the weighted cardinality of a set S to be $\omega(S) := \sum_{j\in S}\omega_j^2$. Since $\omega_j \geq 1$ by assumption, we always have $\omega(S) \geq |S|$, the cardinality of S. When $\boldsymbol{\omega} \equiv 1$, these weighted norms reproduce the standard ℓ_p norms, in which case we use the standard ℓ_p-notation $\|\cdot\|_p$. The exponent $2-p$ in the definition of the spaces $\ell_{\omega,p}$ is somewhat uncommon but turns out to be the most convenient definition for our purposes. For $\boldsymbol{x} \in \ell_{\omega,p}$ and for a subset S of the index set Λ, we define $\boldsymbol{x}_S \in \ell_{\omega,p}$ as the restriction of $\boldsymbol{x}$ to S.

For $s \geq 1$, the error of best weighted s-term approximation of the vector $\boldsymbol{x} \in \ell_{\omega,p}$ is defined as

$$\sigma_s(\boldsymbol{x})_{\omega,p} = \inf_{\boldsymbol{z}:\|\boldsymbol{z}\|_{\omega,0}\leq s} \|\boldsymbol{x}-\boldsymbol{z}\|_{\omega,p}. \tag{2}$$

Unlike unweighted best s-term approximations, weighted approximations of vectors are not straightforward to compute in general. Nevertheless, we will show in Section 3 how to approximate $\sigma_s(\boldsymbol{x})_{\omega,p}$ using a quantity that can easily computed from $\boldsymbol{x}$ by sorting and thresholding.

1.3 Weighted ℓ_p spaces for smoothness and sparsity

The weighted ℓ_p coefficient spaces introduced in the previous section can be used to define weighted function spaces. Let $\psi_j : \mathcal{D} \to \mathbb{C}$, $j \in \Lambda$, be a sequence of functions on some domain $\mathcal{D}$ indexed by the set Λ. For a probability measure ν on $\mathcal{D}$ we assume the orthonormality condition

$$\int_{\mathcal{D}} \psi_j(t)\overline{\psi_k(t)}d\nu(t) = \delta_{j,k} \quad \text{for all } j,k.$$

We will call ν the orthogonalization measure associated to the system $(\psi_j)_{j\in\Lambda}$. The function spaces we consider are the weighted quasi-normed spaces,

$$S_{\omega,p} := \left\{ f(t) = \sum_{j\in\Lambda} x_j\psi_j(t), \;\; t\in\mathcal{D}, \;\; |\!|\!|\, f \,|\!|\!|_{\omega,p} := \|\boldsymbol{x}\|_{\omega,p} < \infty \right\}, \quad 0 < p \leq 1,$$

again with $\omega_j \geq 1$ implicitly assumed. The best s-term approximation to $f \in S_{\omega,p}$ is the function

$$f_S = \sum_{j \in S} x_j \psi_j, \tag{3}$$

where $S \subset \Lambda$ is the set realizing the weighted best s-term approximation of $\boldsymbol{x}$, and the best weighted s-term approximation error is

$$\sigma_s(f)_{\omega,p} = \sigma_s(\boldsymbol{x})_{\omega,p}. \tag{4}$$

The following Stechkin-type estimate, described in more detail in Section 3, can be used to bound the best s-term approximation of a vector by an appropriate weighted vector norm:

$$\sigma_s(\boldsymbol{x})_{\omega,q} \leq \left(s - \|\boldsymbol{\omega}\|_\infty^2\right)^{1/q-1/p} \|\boldsymbol{x}\|_{\omega,p}, \quad p < q \leq 2, \quad \|\boldsymbol{\omega}\|_\infty^2 < s. \tag{5}$$

This estimate illustrates how small ℓ_p-norm for $p < 1$ supports small sparse approximation error. Conditions of the form $\|\boldsymbol{\omega}\|_\infty^2 < s$ are somewhat natural in the context of weighted sparse approximations, as those indices with weights $\omega_j^2 > s$ cannot possibly contribute to best weighted s-term approximations. Without loss of generality, we will usually replace a countably-infinite set Λ by the finite subset $\Lambda_0 \subset \Lambda$ corresponding to indices with weights $\omega_j^2 < s$ (or, for technical reasons, $\omega_j^2 \leq s/2$), if such a finite set exists.

1.4 Interpolation via weighted ℓ_1 minimization

In treating the interpolation problem, we first assume that the index set Λ is finite with $N = |\Lambda|$. Given sampling points $t_1, \ldots, t_m \in \mathcal{D}$ and $f = \sum_{j \in \Lambda} x_j \psi_j$ we can write the vector of sample values $\boldsymbol{y} = (f(t_\ell))_{\ell=1,\ldots,m}$ succinctly in matrix form as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is the *sampling matrix* with entries

$$A_{\ell,j} = \psi_j(t_\ell), \quad \ell = 1, \ldots, m, \quad j \in \Lambda.$$

Better sets of interpolation points are usually associated with better condition number for the sampling matrix $\boldsymbol{A}$. In our theorems, the sampling points are drawn independently from the orthogonalization measure ν associated to the orthonormal system (ψ_j); as a consequence, the random matrix $\boldsymbol{A}\boldsymbol{A}^*$, properly normalized, is the identity matrix in expectation.

We will consider the setting where the number of samples m is smaller than the ambient dimension N, in which case there are infinitely many functions $g \in S_{\omega,p}$ which interpolate the given data. From within this infinite set, we would like to pick out the function of minimal quasi-norm $|||\, g \,|||_{\omega,p}$. However, this minimization problem only becomes tractable once $p = 1$ whence the quasi-norm becomes a norm. As a convex relaxation to the weighted quasi-norm $p < 1$, we consider for interpolation the function $f^\sharp(t) = \sum_{j\in\Lambda} x_j^\sharp \psi_j(t)$ whose coefficient vector $\boldsymbol{x}^\sharp$ is the solution of the weighted ℓ_1 minimization program

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$$

The equality constraint in the ℓ_1 minimization ensures that the function $f^\sharp$ interpolates f at the points t_ℓ, that is, $f^\sharp(t_\ell) = f(t_\ell)$, $\ell = 1, \ldots, m$. Let us give the following result on interpolation via weighted ℓ_1 minimization with respect to $\|\cdot\|_{\omega,1}$.

Theorem 1.1. *Suppose $(\psi_j)_{j\in\Lambda}$ is an orthonormal system of finite size $|\Lambda| = N$, and consider weights $\omega_j \geq \|\psi_j\|_\infty$. For $s \geq 2\|\boldsymbol{\omega}\|_\infty^2$, fix a number of samples*

$$m \geq c_0 s \log^3(s) \log(N), \tag{6}$$

and suppose that t_ℓ, $\ell = 1, \ldots, m$, are sampling points drawn i.i.d. from the orthogonalization measure associated to $(\psi_j)_{j\in\Lambda}$. With probability exceeding $1 - N^{-\log^3(s)}$, the following holds for all functions $f = \sum_{j\in\Lambda} x_j \psi_j$: given samples $y_\ell = f(t_\ell)$, $\ell = 1, \ldots, m$, let $\boldsymbol{x}^\sharp$ be the solution of

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$$

and set $f^\sharp(t) = \sum_{j\in\Lambda} x_j^\sharp \psi_j(t)$. Then the following error rates are satisfied:

$$\begin{aligned} \|f - f^\sharp\|_{L_\infty} \leq \; & \vert\!\vert\!\vert f - f^\sharp \vert\!\vert\!\vert_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1}, \\ & \|f - f^\sharp\|_{L_2} \leq d_1 \sigma_s(f)_{\omega,1}/\sqrt{s}. \end{aligned}$$

Here $\sigma_s(f)_{\omega,1}$ is the best s-term approximation error of f defined in (4). The constants c_0,c_1, and d_1 are universal, independent of everything else.

This interpolation theorem is nonstandard in two respects: the number of samples m required to achieve a prescribed rate scales only logarithmically with the size of the system, and the error guarantees are given by best s-term approximations in weighted coefficient norms.

The constraint on the weights $\omega_j \geq \|\psi_j\|_\infty$ allows us to bound the L_∞ norm by the weighted ℓ_1 coefficient norm: for a function $f \in S_{\omega,p}$,

$$\|f\|_{L_\infty} = \sup_{t\in\mathcal{D}} \left| \sum_{n=-\infty}^{\infty} x_n \psi_n(t) \right| \leq \sup_{t\in\mathcal{D}} \sum_{n=-\infty}^{\infty} |x_n||\psi_n(t)| \leq \sum_{n=-\infty}^{\infty} |x_n|\omega_n = \vert\!\vert\!\vert f \vert\!\vert\!\vert_1,$$

and so if $f_0 = \sum_{j\in S} x_j \psi_j$ with $|S| = s$ is the best s-term approximation to f in the L_∞ norm, then by the Stechkin-type estimate (5) with $q = 1$ we have

$$\|f - f_0\|_{L_\infty} \leq \vert\!\vert\!\vert f - f_0 \vert\!\vert\!\vert_1 \leq (s - \|\boldsymbol{\omega}\|_\infty^2)^{1-1/p} \vert\!\vert\!\vert f \vert\!\vert\!\vert_{\omega,p}, \qquad p < 1.$$

By choosing weights so that $\omega_j \geq \|\psi_j\|_{L_\infty} + \|\psi_j'\|_{L_\infty}$, one may also arrive at bounds of the form $\|f\|_{L_\infty} + \|f'\|_{L_\infty} \leq |||\, f \,|||_1$, reflecting how steeper weights encourage more smoothness. We do not pursue such a direction in this paper, but this should be interesting for future research.

In Section 6 we will prove a more general version of Theorem 1.1 showing robustness of weighted ℓ_1 minimization to noisy samples,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\xi}.$$

Using this robustness to noise, we will be able to treat the case where the index set Λ is countably infinite, by regarding the values $f(t_\ell), \ell = 1, \ldots, m$, as noisy samples of a finite-dimensional approximation to f. For example, this will allow us to show the following result.

Theorem 1.2. *Suppose $(\psi_j)_{j\in\Lambda}$ is an orthonormal system, consider weights $\omega_j \geq \|\psi_j\|_\infty$, and for a parameter $s \geq 1$, let $N = |\Lambda_0|$ where $\Lambda_0 = \{j : \omega_j^2 \leq s/2\}$. Consider a number of samples*

$$m \geq c_0 s \log^3(s) \log(N).$$

Consider a fixed function $f = \sum_{j\in\Lambda} x_j \psi_j$ with $|||\, f \,|||_{\omega,1} < \infty$. Draw sampling points t_ℓ, $\ell = 1, \ldots, m$, independently from the orthogonalization measure associated to $(\psi_j)_{j\in\Lambda}$. Let $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ be the sampling matrix with entries $A_{\ell,j} = \psi_j(t_\ell)$. Let $\eta > 0$ and $\varepsilon \geq 0$ be such that $\eta \leq |||\, f - f_{\Lambda_0} \,|||_{\omega,1} \leq \eta(1+\varepsilon)$. From samples $y_\ell = f(t_\ell)$, $\ell = 1, \ldots, m$, let $\boldsymbol{x}^\sharp$ be the solution of

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq (m/s)^{1/2}\eta$$

and set $f^\sharp(t) = \sum_{j\in\Lambda_0} x_j^\sharp \psi_j(t)$. Then with probability exceeding $1 - N^{-\log^3(s)}$,

$$\begin{aligned} \|f - f^\sharp\|_{L_\infty} \leq |||\, f - f^\sharp \,|||_{\omega,1} &\leq c_1 \sigma_s(f)_{\omega,1}, \\ \|f - f^\sharp\|_{L^2} &\leq d_1 \sigma_s(f)_{\omega,1}/\sqrt{s}. \end{aligned}$$

Above, c_0 is an absolute constant and c_1, d_1 are constants which depend only on the distortion ε.

Several remarks should be made about Theorem 1.2.

1. The minimization problem in Theorem 1.2 requires knowledge of, or at least an estimate of, the tail bound $|||\, f - f_{\Lambda_0} \,|||_{\omega,1}$. It might be possible to avoid this using greedy or iterative methods, but this remains to be investigated. In subsequent corollaries of this result, we will assume exact knowledge of the tail bound, $\eta = |||\, f - f_{\Lambda_0} \,|||_{\omega,1}$, for simplicity of presentation.

2. If the size N of Λ_0 is polynomial in s, then the number of samples reduces to $m \geq Cs\log^4(s)$ to achieve reconstruction with probability $> 1 - s^{-\log^3(s)}$.

1.5 Weighted restricted isometry property

One of the main tools we use in the proofs of Theorems 1.1 and 1.2 is the *weighted restricted isometry property* ($\boldsymbol{\omega}$-RIP) for a linear map $\boldsymbol{A} : \mathbb{C}^N \to \mathbb{C}^m$, which generalizes the concept of restricted isometry property in compressive sensing.

Definition 1.3 ($\boldsymbol{\omega}$-RIP constants)**.** *For $\boldsymbol{A} \in \mathbb{C}^{m \times N}$, $s \geq 1$, and weight ω, the $\boldsymbol{\omega}$-RIP constant $\delta_{\omega,s}$ associated to $\boldsymbol{A}$ is the smallest number for which*

$$(1 - \delta_{\omega,s})\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1 + \delta_{\omega,s})\|\boldsymbol{x}\|_2^2 \tag{7}$$

for all $\boldsymbol{x} \in \mathbb{C}^N$ with $\|\boldsymbol{x}\|_{\omega,0} = \sum_{j \in supp(\boldsymbol{x})} \omega_j^2 \leq s$.

For weights $\boldsymbol{\omega} \equiv 1$, the $\boldsymbol{\omega}$-RIP reduces to the standard RIP, as introduced in [7, 6]. For general weights $\omega_j \geq 1$, the $\boldsymbol{\omega}$-RIP is a weaker assumption for a matrix than the standard RIP, as it requires the map to act as a near-isometry on a smaller set.

Example 1.4. If the weights grow like $\omega_j = j^{\alpha/2}$ with $\alpha > 0$, then without loss we may take $N = s^{1/\alpha}$, as even single indices $j > N$ have weighted cardinality exceeding s. One also calculates that if $\|\boldsymbol{x}\|_{\omega,0} \leq s$, then $\boldsymbol{x}$ is supported on an index set of cardinality at most $\alpha^{1/\alpha} s^{1/(\alpha+1)}$. Following the approach of [1], see also [2, 3], taking a union bound and applying covering arguments, one may argue that an $m \times N$ i.i.d. subgaussian random matrix has the $\boldsymbol{\omega}$-RIP with high probability once

$$m = \mathcal{O}\left(\alpha^{1/\alpha - 1} s^{1/(\alpha+1)} \log s\right).$$

This is a smaller number of measurements than the $m = \mathcal{O}\big(s \log(N/s)\big)$ lower bound required for the unweighted RIP. This observation should be of independent interest, but we focus in this paper on random matrices formed by sampling orthonormal systems.

1.6 Related work on weighted ℓ_1 minimization

Weighted ℓ_1 minimization has been analyzed previously in the compressive sensing literature. Weighted ℓ_1 minimization with weights $\omega_j \in \{0, 1\}$ was introduced independently in the papers [20, 37, 36] and extended further in [19]. The paper [18] seems to be the first to provide conditions under which weighted ℓ_1 minimization is stable and robust under weaker sufficient conditions than the analogous conditions for standard ℓ_1 minimization for general weights. Improved sufficient conditions were recently provided for this setting in [39]. The analysis in all of these works is based on the standard restricted isometry property and does not directly extend to the setting of function interpolation.

Weighted ℓ_1 minimization has also been considered under probabilistic models. In [38], the vector indices are partitioned into two sets, and indices on each set have different probabilities p_1, p_2 of being nonzero; the weights are partitioned into two ω_1, ω_2 accordingly.

The papers [20, 21] provide further analysis in this setting where the entries of the unknown vector fall into two or more sets, each with a different probability of being nonzero. Finally, the paper [28] considers a full Bayesian model, where certain probabilities are associated with each component of the signal in such a way that the probabilities vary in a "continuous" manner across the indices. All of these works take a Grassmann angle approach, and the analysis is thus restricted to the setting of Gaussian matrices and to the noiseless setting.

1.7 Comparison with classical interpolation results

Although weighted ℓ_1 minimization was recently investigated empirically in [15] for multivariate polynomial interpolation in the context of polynomial chaos expansions, weighted ℓ_p spaces, for $0 < p \leq 1$, are nonstandard in the interpolation literature. More standard spaces are the weighted ℓ_2 spaces (see e.g. [24, 29]) such as

$$S_\omega := \left\{ f = \sum_{j \in \Lambda} x_j \phi_j, \quad \|f\|_\omega^2 := \sum_{j \in \Lambda} \omega_j |x_j|^2 < \infty \right\}. \tag{8}$$

where (ϕ_j) is the tensorized Fourier basis on the torus $\mathbb{T}^d$. For the choice of weights $\omega_j = (1 + \|j\|_2^2)^r$, these spaces coincide with the Sobolev spaces $W^{r,2}(\mathbb{T}^d)$ of functions with r derivatives in $L_2(\mathbb{T}^d)$. Optimal interpolation rates for these Sobolev spaces are obtained using smooth and localized kernels (as opposed to polynomials). For example, from equispaced points on the d-dimensional torus with mesh size $h > 0$, [29] derives error estimates of the form

$$\|f - f^\#\|_\infty = \mathcal{O}(h^{r-d/2}) \|f\|_\omega.$$

Writing out this error rate in terms of the number of samples $m = (1/h)^d$, this is

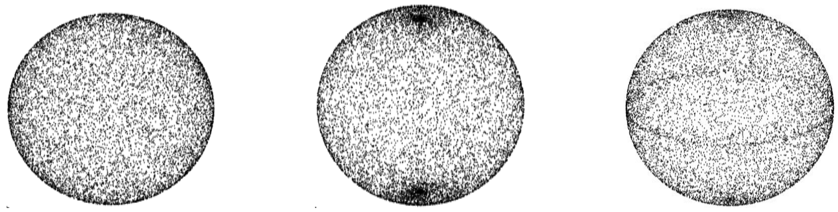
$$\|f - f^\#\|_\infty \leq \mathcal{O}(m^{1/2 - r/d}) \|f\|_\omega.$$

In contrast, Theorem 6.4 implies that weighted ℓ_1 minimization gives the rate

$$\|f - f^\#\|_\infty \leq \mathcal{O}\left(\frac{m}{\log^4(m)}\right)^{1-1/p} |\!|\!| f |\!|\!|_{\omega,p}.$$

A striking distinction between the two bounds is their behavior in high dimensions d: while the Sobolev bound deteriorates exponentially with increasing dimension, the weighted ℓ_1 minimization bound is essentially independent of the dimension. Asymptotically in the number of samples m, the rate provided in this paper is better when

$$p < \frac{1}{\frac{r}{d} + \frac{1}{2}},$$



(a) $\sin(\theta)\, d\theta d\varphi$ (b) $d\theta d\varphi$ (c) $|\tan(\theta)|^{1/3}\, d\theta d\varphi$

Figure 2: An illustration of i.i.d. samples from various spherical measures. $(\pi, \varphi) \in [0, \pi) \times [0, 2\pi]$. The distribution (c) is the most incoherent with respect to the spherical harmonic basis.

where we recall that r is the highest order of differentiability and d is the ambient dimension. While we made this comparison in the setting of interpolation on the d-dimensional torus, a similar comparison between classical results and the bounds provided in this paper could be made for interpolation on the sphere or the d-dimensional cube. In all settings, the interpolation theorems in this paper improve on smoothness-based bounds in the regime of high dimensions / mild smoothness.

2 Case studies

In this section we consider several examples and demonstrate how Theorem 1.2 gives rise to various sampling theorems for polynomial and spherical harmonic interpolation. One could derive similar results in weighted ℓ_p spaces using Theorem 6.4.

2.1 Spherical harmonic interpolation

The spherical harmonics Y_ℓ^k form an orthonormal system for square-integrable functions on the sphere $S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_2 = 1\}$, and serve as a higher-dimensional analog of the univariate trigonometric basis. They are orthogonal with respect to the uniform spherical measure. In spherical coordinates $(\varphi, \theta) \in [0, 2\pi) \times [0, \pi)$, $(x = \cos(\varphi)\sin(\theta), y = \sin(\varphi)\sin(\theta), z = \cos(\theta))$ for $(x, y, z) \in S^2$, the orthogonality reads

$$\int_0^{2\pi} \int_0^{\pi} Y_\ell^k(\varphi, \theta) \bar{Y}_{\ell'}^{k'}(\varphi, \theta) \sin(\theta) d\theta d\varphi = \delta_{\ell\ell', kk'}, \quad k, \ell \in \mathbb{Z}, \quad |k| \leq \ell. \tag{9}$$

The spherical harmonics are bounded according to $\|Y_\ell^k\|_{L_\infty} \leq \ell^{1/2}$, and this bound is realized at the poles of the sphere, $\theta = 0, \pi$. As shown in [32, 5], one can precondition the spherical harmonics to transform them into a system with smaller uniform bound, orthonormal with respect to a different measure. For example, the preconditioned function system

$$Z_\ell^k(\varphi,\theta) = \sin(\theta)^{1/2} Y_\ell^k(\varphi,\theta),$$

normalized by the proper constant, is orthonormal on the sphere with respect to the measure $d\mu = d\theta d\varphi$ by virtue of (9). The Z_ℓ^k are more uniformly bounded than the spherical harmonics Y_ℓ^k; as noted in [23],

$$\|Z_\ell^k\|_{L_\infty} \leq C\ell^{1/4}$$

for a universal constant C. A sharper preconditioning estimate was shown in [5] for the system

$$\tilde{Z}_\ell^k(\varphi,\theta) := (\sin^2(\theta)\cos(\theta))^{1/6} Y_\ell^k(\varphi,\theta). \tag{10}$$

Normalized properly, this system is orthonormal on the sphere with respect to the measure $d\nu = |\tan(\theta)|^{1/3} d\theta d\varphi$, which is nonstandard and illustrated in Figure 2. This system obeys the uniform bound

$$\|\tilde{Z}_\ell^k\|_\infty \leq C\ell^{1/6}, \tag{11}$$

with C a universal constant.

We consider implications of Theorem 1.2 for interpolation with spherical harmonic expansions. We state a result in the setting where sampling points are drawn from the measure $|\tan(\theta)|^{1/3} d\theta d\varphi$, but similar results (albeit with steeper weights) can be obtained for sampling from the measures $d\theta d\varphi$ and $\sin(\theta)\, d\theta d\varphi$.

Corollary 2.1 (Interpolation with spherical harmonics)**.** *Consider the preconditioned spherical harmonics $\tilde{Z}_\ell^k, |k| \leq \ell$, and associated orthogonalization measure $d\nu = |\tan(\theta)|^{1/3} d\theta d\varphi$. Fix weights $\omega_{\ell,k} = C\ell^{1/6}$ and index set $\Lambda_0 = \{(\ell,k) : |k| \leq \ell \leq s^3\}$ of size $N = s^6$, and fix a number of samples*

$$m \geq c_0 s \log^4(s).$$

Consider a fixed function $f(\varphi,\theta) = \sum_{\ell,k} x_{\ell,k} \tilde{Z}_\ell^k(\varphi,\theta) \in S_{\omega,1}$ and let $\eta = \|\!| f - f_{\Lambda_0} \|\!|_{\omega,1}$. Draw $(\varphi_j,\theta_j), j = 1,\dots,m$, i.i.d. from $d\nu$, and consider sample values $y_j = f(\varphi_j,\theta_j)$, $j = 1,\dots,m$. Then with probability exceeding $1 - N^{-\log^3(s)}$, the function $f^\sharp = \sum_{(\ell,k)\in\Lambda_0} x^\sharp_{\ell,k} \tilde{Z}_\ell^k$ formed from the solution $x^\sharp$ of the weighted ℓ_1 minimization program

$$\min_{u_{\ell,k}} \sum_{\ell,k\in\Lambda_0} \omega_{\ell,k}|u_{\ell,k}| \text{ subject to } \sum_{j=1}^m \Big(\sum_{\ell,k\in\Lambda_0} u_{\ell,k} \tilde{Z}_\ell^k(\varphi_j,\theta_j) - y_j \Big)^2 \leq (m/s)\eta^2$$

satisfies the error bounds

$$\begin{aligned} \|f - f^\sharp\|_\infty \leq \|\!| f - f^\sharp \|\!|_{\omega,1} &\leq c_1 \sigma_s(f)_{\omega,1}, \\ \|f - f^\sharp\|_{L_2} &\leq c_2 \sigma_s(f)_{\omega,1}/\sqrt{s}. \end{aligned}$$

It is informative to compare these results with previously available bounds for unweighted ℓ_1 minimization. Using the same sampling distribution $d\nu = |\tan(\theta)|^{1/3} d\theta d\varphi$ and number of basis elements $N = s^6$, existing bounds for unweighted ℓ_1 minimization (see [5]) require a number of samples

$$m \geq cN^{1/6} s \log^4(s) = cs^2 \log^4(s)$$

to achieve an error estimate of the form $\|f - f^\sharp\|_{L_\infty} \leq \sqrt{s}\sigma_s(f)_1$ (see [32] for more details). That is, *more* measurements m are required to achieve a *weaker* reconstruction rate. However, stronger assumptions on f are required in the sense that the result above requires the *weighted* best s-term approximation error to be small while the bound from [32] works with the unweighted best s-term approximation error. Expressed differently, our result requires more smoothness which is in line with the general philosophy of this paper.

2.2 Tensorized polynomial interpolation

The tensorized trigonometric polynomials on $\mathcal{D} = \mathbb{T}^d$ are given by

$$\psi_{\mathbf{k}}(\mathbf{t}) = \psi_{k_1}(t_1)\psi_{k_2}(t_2)\dots\psi_{k_d}(t_d), \quad \mathbf{k} \in \mathbb{Z}^d,$$

with $\psi_j(t) = e^{2\pi i j t}$. These functions are orthonormal with respect to the tensorized uniform measure. Because this system is uniformly bounded, Theorem 1.2 applies with constant weights $\omega_j \equiv 1$. Nevertheless, higher weights promote smoother reconstructions.

Other tensorized polynomial bases of interests are not uniformly bounded, but we can get reconstruction guarantees by considering weighted ℓ_1 minimization with properly chosen weights.

2.2.1 Chebyshev polynomials

Consider the tensorized Chebyshev polynomials on $\mathcal{D} = [-1, 1]^d$:

$$C_{\mathbf{k}}(\mathbf{t}) = C_{k_1}(t_1)C_{k_2}(t_2)\dots C_{k_d}(t_d), \quad \mathbf{k} \in \mathbb{N}^d, \tag{12}$$

where $C_k(t) = \sqrt{2}\cos\big((k-1)\arccos(t)\big)$. The Chebyshev polynomials form a basis for the real algebraic polynomials on $\mathcal{D}$, and are orthonormal with respect to the tensorized Chebyshev measure

$$d\mu = \frac{d\mathbf{t}}{(2\pi)^d \Pi_{j=1}^d (1 - t_j^2)^{1/2}}. \tag{13}$$

The tensorized Chebyshev polynomials are not uniformly bounded; since $\|C_k\|_\infty = 2^{1/2}$ we have $\|C_{\mathbf{k}}\|_\infty = 2^{\frac{\|\mathbf{k}\|_0}{2}}$. This motivates us to apply Theorem 1.2 with weights

$$\omega_{\mathbf{k}} = \prod_{\ell=1}^d (k_\ell + 1)^{1/2}, \tag{14}$$

noting that $\|C_{\mathbf{k}}\|_\infty \leq \omega_{\mathbf{k}}$. (More generally, one could also work with weights of the form $\omega_{\mathbf{k}} = 2^{\frac{\|\mathbf{k}\|_0}{2}} v_{\mathbf{k}}$, where $v_{\mathbf{k}}$ tends to infinity as $\|\mathbf{k}\|_{\mathbf{2}} \to \infty$.) Such weights encourage both sparse and low order tensor products of Chebyshev polynomials. The subset of indices

$$H_s^d = \{k \in \mathbb{N}_0^d, \omega_{\mathbf{k}}^2 \leq s\} = \left\{k \in \mathbb{N}_0^d, \prod_{\ell=1}^{d} (k_\ell + 1) \leq s\right\}$$

forms a hyperbolic cross. As argued in [12], the size of a hyperbolic cross can be bounded according to

$$|H_s^d| \leq Cs \min\{\log(s)^{d-1}, d^{\log(s)}\}.$$

Corollary 2.2. *Consider the tensorized Chebyshev polynomial basis* $(C_{\mathbf{k}})$ *for* $[-1,1]^d$, *and weights* $\omega_{\mathbf{k}}$ *as in* (14). *Let* $\Lambda_0 = \{\mathbf{k} \in \mathbb{N}_0^d : \omega_{\mathbf{k}}^2 \leq s/2\}$, *and let* $N = |\Lambda_0| \leq C(s \min\{\log(s)^{d-1}, d^{\log(s)}\})$. *Fix a number of samples*

$$m \geq c_0 s \log^3(s) \log(N). \tag{15}$$

Consider a function $f = \sum_{\mathbf{k}\in\Lambda} x_{\mathbf{k}} C_{\mathbf{k}}$, *and sampling points* $\mathbf{t}_\ell$, $\ell = 1, \dots, m$ *drawn i.i.d. from the tensorized Chebyshev measure on* $[-1,1]^d$. *Let* $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ *be the sampling matrix with entries* $A_{\ell,j} = \psi_j(t_\ell)$. *From samples* $y_\ell = f(\mathbf{t}_\ell)$, $\ell = 1, \dots, m$, *let* $\boldsymbol{x}^\sharp$ *be the solution of*

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \sqrt{m/s}\, \|\!| f - f_{\Lambda_0} \|\!|_{\omega,1}$$

and set $f^\sharp(\mathbf{t}) = \sum_{\mathbf{k}\in\Lambda_0} x_{\mathbf{k}}{}^\sharp C_{\mathbf{k}}(\mathbf{t})$. *Then with probability exceeding* $1 - N^{-\log^3(s)}$,

$$\begin{aligned} \|f - f^\sharp\|_{L_\infty} \leq \|\!| f - f^\sharp \|\!|_{\omega,1} &\leq c_1 \sigma_s(f)_{\omega,1}, \\ \|f - f^\sharp\|_{L^2} &\leq d_1 \sigma_s(f)_{\omega,1}/\sqrt{s}. \end{aligned}$$

Above, c_0, c_1, *and* d_1 *are universal constants.*

Note that with the stated estimate of N, m satisfies (15) once

$$m \geq c_0 \log(d) s \log^4(s).$$

This means that the required number of samples m above grows only logarithmically with the ambient dimension d as opposed to exponentially, as required for classical interpolation bounds using linear reconstruction methods.

2.2.2 Legendre polynomials

Consider now the tensorized Legendre polynomials on $\mathcal{D} = [-1,1]^d$:

$$L_{\mathbf{k}}(\mathbf{t}) = L_{k_1}(t_1) L_{k_2}(t_2) \dots L_{k_d}(t_d), \quad \mathbf{k} \in \mathbb{N}^d, \tag{16}$$

where L_k is the univariate orthonormal Legendre polynomial of degree k. The Legendre polynomials form a basis for the real algebraic polynomials on $\mathcal{D}$, and are orthonormal with respect to the tensorized uniform measure on $\mathcal{D}$. Since $\|L_k\|_\infty \le \sqrt{k}$ we have

$$\|L_{\mathbf{k}}\|_\infty \le \prod_{\ell=1}^{d}(k_\ell+1)^{1/2}, \tag{17}$$

and we may apply Theorem 1.2 with hyperbolic cross weights $\omega_{\mathbf{k}} = \prod_{\ell=1}^{d}(k_\ell+1)^{1/2}$ as in Corollary 2.2. In doing so, we arrive at the following result.

Corollary 2.3. *Consider the tensorized Legendre polynomial basis and weights $\omega_{\mathbf{k}}$ as in (14) and with Λ_0, N, and m as in Corollary 2.2. Consider a function $f = \sum_{\mathbf{k}\in\Lambda} x_{\mathbf{k}} L_{\mathbf{k}}$, and suppose that $\mathbf{t}_\ell$, $\ell = 1,\dots,m$, are drawn i.i.d. from the tensorized uniform measure on $[-1,1]^d$. Let $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ be the associated sampling matrix with entries $A_{\ell,j} = \psi_j(t_\ell)$. From samples $y_\ell = f(\mathbf{t}_\ell)$, $\ell = 1,\dots,m$, let $\boldsymbol{x}^\sharp$ be the solution of*

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \le \sqrt{m/s}\,|\!|\!|\, f - f_{\Lambda_0} \,|\!|\!|_{\omega,1}$$

and set $f^\sharp(\mathbf{t}) = \sum_{\mathbf{k}\in\Lambda_0} x^\sharp_{\mathbf{k}} L_{\mathbf{k}}(\mathbf{t})$. Then with probability exceeding $1 - N^{-\log^3(s)}$,

$$\begin{aligned} \|f - f^\sharp\|_{L_\infty} \le |\!|\!|\, f - f^\sharp \,|\!|\!|_{\omega,1} &\le c_1 \sigma_s(f)_{\omega,1}, \\ \|f - f^\sharp\|_{L^2} &\le d_1 \sigma_s(f)_{\omega,1}/\sqrt{s}. \end{aligned} \tag{18}$$

Above, c_0, c_1, and d_1 are universal constants.

Although the univariate orthonormal Legendre polynomials are not uniformly bounded on $[-1,1]$, they can be transformed into a bounded orthonormal system by considering the weight

$$v(t) = (\pi/2)^{1/2}(1-t^2)^{1/4}, \quad t \in [-1,1],$$

and recalling Theorem 7.3.3 from [35] which states that, for all $j \ge 1$,

$$\sup_{t\in[-1,1]} v(t)|L_j(t)| \le \sqrt{2 + 1/j} \le \sqrt{3}. \tag{19}$$

Then the preconditioned system $Q_j(t) = v(t)L_j(t)$ is orthonormal with respect to the Chebyshev measure, and is uniformly bounded on $[-1,1]$ with constant $K = \sqrt{3}$. A statement similar to Corollary 2.2 can also be applied to tensorized preconditioned Legendre polynomials, if sampling points are chosen from the tensorized Chebyshev measure. For further details, we refer the reader to [31].

2.3 Numerical illustrations

2.3.1 Polynomial interpolation

Polynomial interpolation usually refers to fitting the unique trigonometric or algebraic polynomial of degree $m-1$ through a given set of data of size m. When m is large, this problem becomes ill-conditioned, as illustrated for example by Runge's phenomenon, or the tendency of high-degree polynomial interpolants to oscillate at the edges of an interval (the analogous phenomenon for trigonometric polynomial interpolation is Gibb's phenomenon [16]). While Runge's phenomenon can be significantly minimized by carefully choosing interpolation nodes – Chebyshev nodes for algebraic polynomial interpolation or equispaced nodes for trigonometric interpolation – the effects cannot be completely eliminated. Two methods known to reduce the effects of Runge's phenomenon are the method of least squares, where one foregoes exact interpolation for a least squares projection of the data onto a polynomial of lower degree [11], or by doing weighted ℓ_2 regularization [24], e.g. use for interpolation the function $f^\sharp(t) = \sum_{j\in\Lambda} x_j^\sharp \psi_j(t)$ where the coefficient vector $\boldsymbol{x}^\sharp$ solves the minimization problem

$$\min \sum_{j\in\Lambda} \omega_j^2 z_j^2 \text{ subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$$

where $\boldsymbol{A}$ is the sampling matrix as in (1.4).

In this section we provide numerical evidence that weighted ℓ_1 regularization can significantly outperform unweighted ℓ_1 minimization, least squares projections, and weighted ℓ_2 regularization in reducing the effect of oscillatory artifacts in polynomial interpolation, and more generally provides more accurate reconstructions which are less sensitive to perturbations in the choice of sampling points.

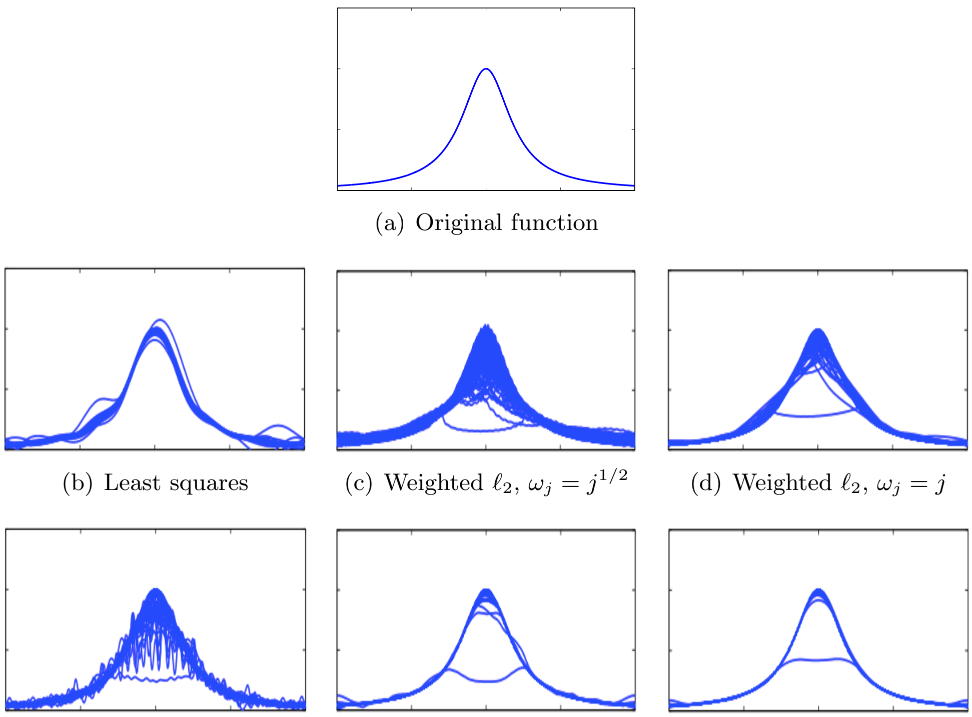
For our numerical experiments, we follow the examples in [11] and consider on $[-1,1]$ the smooth function

$$f_1(x) = \frac{1}{1+25x^2},$$

which was originally considered by Runge [34] to illustrate the instability of polynomial interpolation at equispaced points. We also consider the non-smooth function

$$f_2(x) = |x|.$$

For f_1 and f_2, we repeat the following experiment 100 times: draw $m = 25$ sampling points $x_1, x_2, \ldots, x_m$, i.i.d. from a measure μ on $\mathcal{D} = [-1,1]$ and compute the noise-free observations $y_k = f(x_k)$: we will use the uniform measure for real trigonometric polynomial interpolation and the Chebyshev measure for Legendre polynomial interpolation. We then compare the least squares approximation, unweighted ℓ_1 approximation, weighted ℓ_2 approximations with weights $\omega_j = j$ and $\omega_j = j^{1/2}$, and weighted ℓ_1 approximations with weights $\omega_j = j$ and weights $\omega_j = j^{1/2}$. In figures 3-6 we display the interpolations resulting from all 100 experiments, overlaid so as to illustrate the sensitivity of each interpolation method to the choice of sampling points. In all experiments, we fix in the ℓ_1 and ℓ_2 minimization a maximal polynomial degree $N = 100$. For the least squares solution to be stable [11], we project onto the span of the first $d = 50$ basis elements. In all experiments, we observe that the weighted ℓ_1 interpolants are more accurate and more robust with respect to the choice of sampling points.



(a) Original function

(b) Least squares

(c) Weighted ℓ_2, $\omega_j = j^{1/2}$

(d) Weighted ℓ_2, $\omega_j = j$

(e) Unweighted ℓ_1

(f) Weighted ℓ_1, $\omega_j = j^{1/2}$

(g) Weighted ℓ_1, $\omega_j = j$

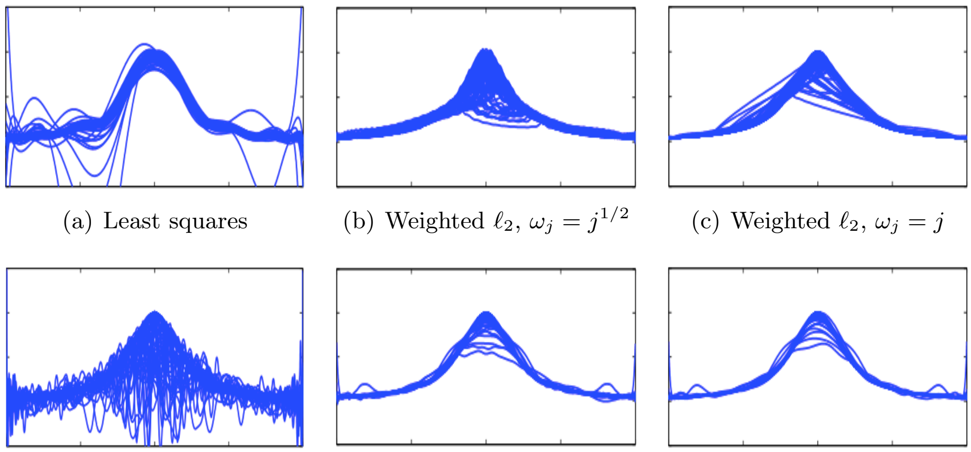
Figure 3: Overlaid interpolations of the function $f_1(x) = \frac{1}{1+25x^2}$ by real trigonometric polynomials using various reconstruction methods.

2.3.2 Spherical harmonic interpolation

We now numerically compare the performance of weighted ℓ_1 minimization with various weights in reconstructing functions on the sphere using spherical harmonic interpolations (see also the theoretical results in Section 2.1).

We consider the smooth function

$$f_1(\theta, \varphi) = \frac{1}{|\theta|^2 + .1}, \quad 0 \le \theta \le \pi, \quad 0 \le \varphi < 2\pi$$



(a) Least squares (b) Weighted ℓ_2, $\omega_j = j^{1/2}$ (c) Weighted ℓ_2, $\omega_j = j$

(d) Unweighted ℓ_1 (e) Weighted ℓ_1, $\omega_j = j^{1/2}$ (f) Weighted ℓ_1, $\omega_j = j$

Figure 4: Overlaid interpolations of the function $f_1(x) = \frac{1}{1+25x^2}$ by Legendre polynomials using various reconstruction methods.

which has a localized maximum at the north pole, and the function

$$f_2(\theta, \varphi) = \frac{1}{|\theta - \pi/2| + \pi/6},$$

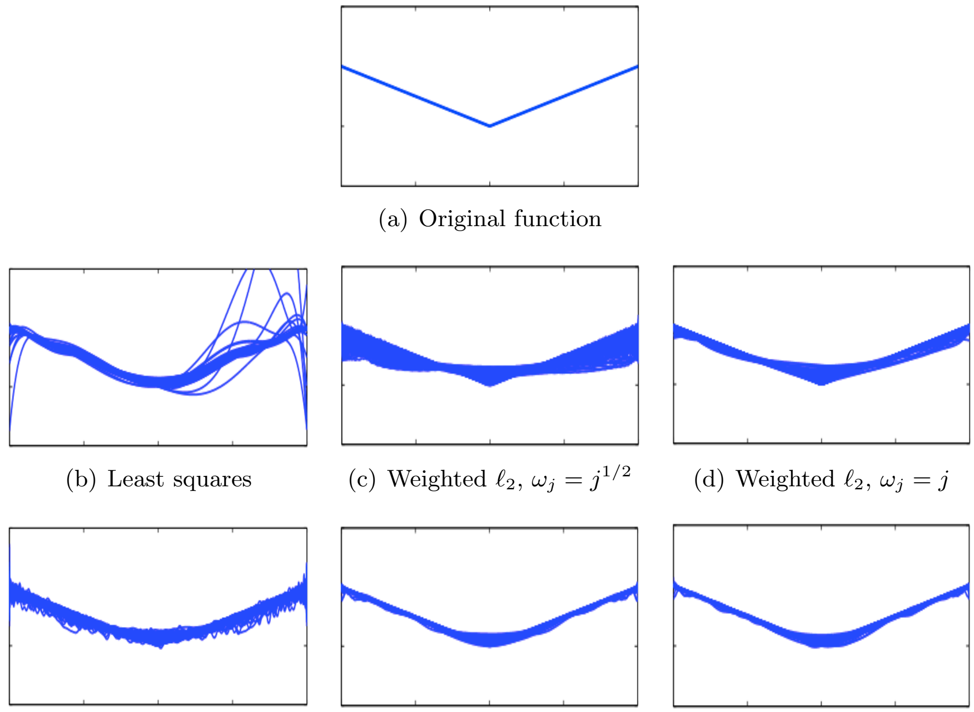
which has a ring of maxima around the equator.

In the following experiments we use for interpolation $m = 30$ sampling points (θ_j, φ_j) i.i.d. with respect to the spherical tangent measure $d\nu = |\tan(\theta)|^{1/3} d\theta d\varphi$. We compare the performance of unweighted ℓ_1 minimization with weighted ℓ_1 minimization using weights $\omega_{k,\ell} = \ell^{1/6}$ as described in Theorem 2.1, and weights $\omega_{k,\ell} = \ell^{1/2}$. In all experiments, we restrict in the ℓ_1 minimization program to the first $N = 15^2 = 225$ spherical harmonics per the usual ordering.

Figure 7 illustrates the approximations resulting from a representative randomized sampling of points. We find that weighted ℓ_1 minimization using larger weights tends to give the most accurate reconstructions, while using smaller weights as supported by Theorem 2.1 nevertheless outperforms unweighted ℓ_1 minimization.

3 Weighted sparsity and quasi-best s-term approximations

In this section we revisit some important technical results pertaining to weighted ℓ_p spaces that were touched upon in the introduction. First, unlike unweighted s-term approximations for finite vectors, the weighted s-term approximations $\sigma_s(\boldsymbol{x})_{\omega,p} = \inf_{\boldsymbol{z}:\|\boldsymbol{z}\|_{\omega,0} \leq s} \|\boldsymbol{x} -$



(a) Original function

(b) Least squares (c) Weighted ℓ_2, $\omega_j = j^{1/2}$ (d) Weighted ℓ_2, $\omega_j = j$

(e) Unweighted ℓ_1 (f) Weighted ℓ_1, $\omega_j = j^{1/2}$ (g) Weighted ℓ_1, $\omega_j = j$

Figure 5: Overlaid interpolations of the function $f_2(x) = |x|$ by real trigonometric polynomials using various reconstruction methods.

$\boldsymbol{z}\|_{\omega,p}$ are not straightforward to compute in general. Nevertheless, we can approximate $\sigma_s(\boldsymbol{x})_{\omega,p}$ using a quantity that can easily computed from $\boldsymbol{x}$ by sorting and thresholding, which we will call the quasi-best s-term approximation.

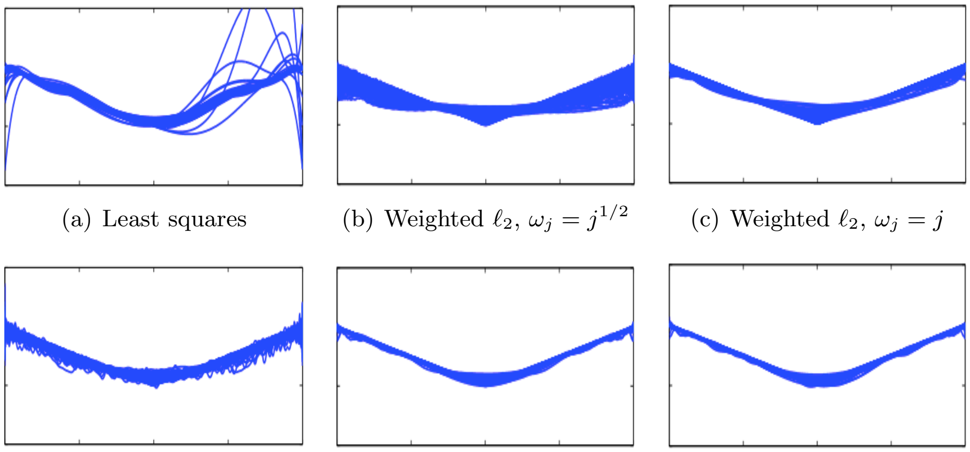
Let $\boldsymbol{v}$ denote the non-increasing rearrangement of the sequence $(|x_j|^p \omega_j^{-p})$, that is, $v_j = |x_{\pi(j)}|^p \omega_{\pi(j)}^{-p}$ for some permutation π such that $v_1 \geq v_2 \geq \cdots \geq 0$. Let k be the maximal number such that $\sum_{j=1}^{k} \omega_{\pi(j)}^2 \leq s$ and set $S = \{\pi(1), \pi(2), \ldots, \pi(k)\}$ so that $\omega(S) \leq s$. Then we call $\boldsymbol{x}_S$ a weighted quasi-best s-term approximation to $\boldsymbol{x}$ and define the corresponding error of weighted quasi-best s-term approximation as

$$\widetilde{\sigma}_s(\boldsymbol{x})_{\omega,p} = \|\boldsymbol{x} - \boldsymbol{x}_S\|_{\omega,p} = \|\boldsymbol{x}_{S^c}\|_{\omega,p}.$$

By definition, $\sigma_s(\boldsymbol{x})_{\omega,p} \leq \widetilde{\sigma}_s(\boldsymbol{x})_{\omega,p}$. We also have a converse inequality relating the two s-term approximations in the case of bounded weights.

Lemma 3.1. *Suppose that $s \geq \|\boldsymbol{\omega}\|_\infty^2$. Then*

$$\widetilde{\sigma}_{3s}(\boldsymbol{x})_{\omega,p} \leq \sigma_s(\boldsymbol{x})_{\omega,p}$$



(a) Least squares (b) Weighted ℓ_2, $\omega_j = j^{1/2}$ (c) Weighted ℓ_2, $\omega_j = j$

(d) Unweighted ℓ_1 (e) Weighted ℓ_1, $\omega_j = j^{1/2}$ (f) Weighted ℓ_1, $\omega_j = j$

Figure 6: Overlaid interpolations of the function $f_2(x) = |x|$ by Legendre polynomials using various reconstruction methods.

Proof. Let $\boldsymbol{x}_S$ be the weighted best s-term approximation to $\boldsymbol{x}$, and let $\boldsymbol{x}_{\tilde{S}}$ be the weighted quasi-best $3s$-term approximation to $\boldsymbol{x}$. Because the supports of $\boldsymbol{x}_S$ and $\boldsymbol{x} - \boldsymbol{x}_S$, and also $\boldsymbol{x}_{\tilde{S}}$ and $\boldsymbol{x} - \boldsymbol{x}_{\tilde{S}}$, do not overlap, it suffices to show that

$$\|\boldsymbol{x}_S\|_{\omega,p} \leq \|\boldsymbol{x}_{\tilde{S}}\|_{\omega,p}.$$

Assume without loss of generality that the terms in $\boldsymbol{x}$ are ordered so that $|x_j|^p \omega_j^{-p} \geq |x_{j+1}|^p \omega_{j+1}^{-p}$ for all j; let J be the largest integer such that $\sum_{j=1}^{J} \omega_j^2 \leq 3s$, and $\tilde{S} = \{1, 2, \ldots, J\}$. Because $|\omega_j|^2 \leq s$, we know also that $\sum_{j=1}^{J} \omega_j^2 \geq 2s$.

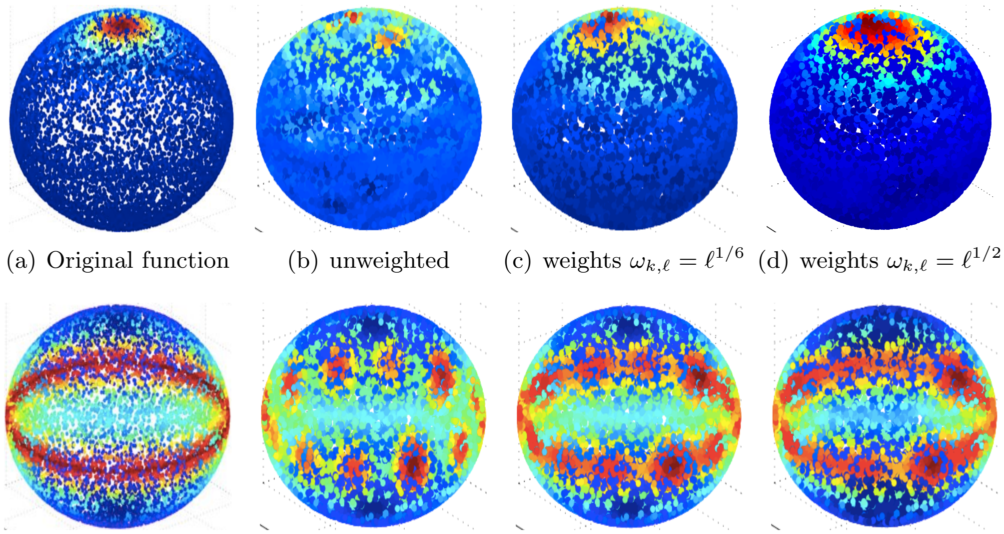
Let $n_j = \lfloor \omega_j^2 + 1 \rfloor$ be the largest integer less than or equal to $\omega_j^2 + 1$, and let $r_j = n_j - \omega_j^2$. Then $\sum_{j \in S} \omega_j^2 \leq s$ implies that

$$\sum_{j \in S} n_j \leq \sum_{j \in S} \omega_j^2 + |S| \leq s + s \leq \sum_{j=1}^{J} \omega_j^2 \leq \sum_{j=1}^{J} n_j \tag{20}$$

Now, let

$$\boldsymbol{z} = \Big(\underbrace{|x_1|^p \omega_1^{-p}, \ldots, |x_1|^p \omega_1^{-p}, (1 - r_1)|x_1|^p \omega_1^{-p}}_{n_1 \text{ coefficients}}, \underbrace{|x_2|^p \omega_2^{-p}, \ldots, |x_2|^p \omega_2^{-p}, (1 - r_2)|x_2|^p \omega_2^{-p}}_{n_2 \text{ coefficients}}, \ldots \Big).$$

We constructed $\boldsymbol{z}$ so that the first n_1 terms in $\boldsymbol{z}$ sum to $|x_1|^p \omega_1^{2-p}$, the next n_2 terms sum



(a) Original function (b) unweighted (c) weights $\omega_{k,\ell} = \ell^{1/6}$ (d) weights $\omega_{k,\ell} = \ell^{1/2}$

(e) Original function (f) unweighted (g) weights $\omega_{k,\ell} = \ell^{1/6}$ (h) weights $\omega_{k,\ell} = \ell^{1/2}$

Figure 7: Comparing unweighted and weighted ℓ_1 minimization for spherical harmonic interpolation. For the top function, the relative L_∞ errors between the original and reconstructed functions are (b) .87, (c) .67, and (d) .49, respectively. For the bottom function, the errors are (f) .42, (g) .25, and (h) .22, respectively.

to $|x_2|^p \omega_2^{2-p}$, and so on. Then

$$
\begin{aligned}
\|\boldsymbol{x}_S\|_{\omega,p}^p &:= \max_S \sum_{j\in S} \omega_j^2 |x_j|^p \omega_j^{-p} \text{ subject to } \sum_{j\in S} \omega_j^2 \leq s \\
&\leq \max_S \sum_{j\in S} \omega_j^2 |x_j|^p \omega_j^{-p} \text{ subject to } \sum_{j\in S} n_j \leq \sum_{j=1}^{J} n_j \\
&\leq \max_\Lambda \sum_{k\in\Lambda} z_k \text{ subject to } \|\Lambda\|_0 \leq \sum_{j=1}^{J} n_j \\
&\leq \sum_{j=1}^{J} \omega_j^{2-p} |x_j|^p = \|\boldsymbol{x}_{\tilde{S}}\|_{\omega,p}^p
\end{aligned}
$$

This completes the proof. $\square$

In the remainder of this section, we prove the Stechkin-type estimate (5) which bounds the quasi-best s-term approximation of a vector (and hence also the best s-term approximation) by an appropriate weighted vector norm.

Theorem 3.2. *For $p < q \leq 2$, let $\boldsymbol{x} \in \ell_{\omega,p}$. Then, for $s > \|\boldsymbol{\omega}\|_\infty^2$,*

$$\sigma_s(\boldsymbol{x})_{\omega,q} \leq \widetilde{\sigma}_s(\boldsymbol{x})_{\omega,q} \leq \left(s - \|\boldsymbol{\omega}\|_\infty^2\right)^{1/q-1/p} \|\boldsymbol{x}\|_{\omega,p}. \tag{21}$$

Proof. Let S be the support of the weighted quasi-best s-term approximation, so that $\widetilde{\sigma}_s(\boldsymbol{x})_{\omega,p} = \|\boldsymbol{x} - \boldsymbol{x}_S\|_{\omega,p}$. Since the number k in the construction of S is maximal, we have with π denoting the corresponding permutation,

$$s - \|\boldsymbol{\omega}\|_\infty^2 \leq s - \omega_{\pi(k)}^2 < \omega(S) \leq s.$$

Then

$$\widetilde{\sigma}_s(\boldsymbol{x})_{\omega,p}^p \leq \sum_{j \notin S} |x_j|^p \omega_j^{2-p} \leq \max_{j \notin S}\{|x_j|^{p-q}\omega_j^{q-p}\} \sum_{j \notin S} |x_j|^q \omega_j^{2-q} \leq \left(\max_{j \notin S} |x_j| \omega_j^{-1}\right)^{p-q} \|\boldsymbol{x}\|_{\omega,q}^q.$$

Now let $\alpha_k := (\sum_{j \in S} \omega_j^2)^{-1} \omega_k^2 \leq (s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \omega_k^2$. Then $\sum_{j \in S} \alpha_k = 1$. Moreover, by definition of S we have $|x_j|\omega_j^{-1} \leq |x_k|\omega_k^{-1}$ for all $k \in S$ and $j \notin S$. This implies

$$\left(\max_{j \notin S} |x_j| \omega_j^{-1}\right)^q \leq \sum_{k \in S} \alpha_k |x_k|^q \omega_k^{-q} \leq (s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \sum_{k \in S} \omega_k^{2-q} |x_k|^q \leq (s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \|\boldsymbol{x}\|_{\omega,q}^q.$$

Combining the above estimates yields

$$\widetilde{\sigma}_s(\boldsymbol{x})_{\omega,p}^p \leq \left((s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \|\boldsymbol{x}\|_{\omega,q}^q\right)^{(p-q)/q} \|\boldsymbol{x}\|_{\omega,q}^q$$

which is equivalent to the claim. □

Theorem 3.2 will be used in deriving weighted null space properties and weighted restricted isometry properties in the following sections.

4 Weighted null space and restricted isometry property

As is the case for unweighted ℓ_1 minimization, one can derive reconstruction guarantees for weighted ℓ_1 minimization via appropriate weighted versions of the null space property and restricted isometry property [10, 7]. Below we work out these approaches.

4.1 Weighted null space property

We start directly with a robust version of the null space property in the weighted case.

Definition 4.1 (Weighted robust null space property)**.** *Given a weight $\boldsymbol{\omega}$, a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ is said to satisfy the weighted robust null space property of order s with constants $\rho \in (0,1)$ and $\tau > 0$ if*

$$\|\boldsymbol{v}_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|\boldsymbol{v}_{S^c}\|_{\omega,1} + \tau \|\boldsymbol{A}\boldsymbol{v}\|_2 \quad \text{for all } \boldsymbol{v} \in \mathbb{C}^N \text{ and all } S \subset [N] \text{ with } \omega(S) \leq s. \tag{22}$$

The inequalities stated in the next theorem are crucial for deriving error bounds for recovery via weighted ℓ_1 minimization.

Theorem 4.2. *Suppose that* $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ *is such that* (22) *holds for* $\rho \in (0,1)$ *and* $\tau > 0$. *Then, for all* $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$, *we have*

$$\|\boldsymbol{z}-\boldsymbol{x}\|_{\omega,1} \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + 2\sigma_s(\boldsymbol{x})_{\omega,1}\right) + \frac{2\tau\sqrt{s}}{1-\rho}\|\boldsymbol{A}(\boldsymbol{z}-\boldsymbol{x})\|_2 \tag{23}$$

and, additionally assuming $s \geq 2\|\boldsymbol{\omega}\|_\infty^2$,

$$\|\boldsymbol{x}-\boldsymbol{z}\|_2 \leq \frac{C_1}{\sqrt{s}}\left(\|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + 2\sigma_s(\boldsymbol{x})_{\omega,1}\right) + C_2\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{z})\|_2. \tag{24}$$

Proof. We start with the proof of (23). Let S with $\omega(S) \leq s$ be such that $\sigma_s(\boldsymbol{x})_{\omega,1} = \|\boldsymbol{x}-\boldsymbol{x}_S\|_{\omega,1} = \|\boldsymbol{x}_{S^c}\|_{\omega,1}$. The triangle inequality gives

$$\begin{aligned}&\|\boldsymbol{x}\|_{\omega,1} + \|(\boldsymbol{x}-\boldsymbol{z})_{S^c}\|_{\omega,1} \leq \|\boldsymbol{x}_{S^c}\|_{\omega,1} + \|\boldsymbol{x}_S\|_{\omega,1} + \|\boldsymbol{x}_{S^c}\|_{\omega,1} + \|\boldsymbol{z}_{S^c}\|_{\omega,1}\\ &\leq 2\|\boldsymbol{x}_{S^c}\|_{\omega,1} + \|(\boldsymbol{x}-\boldsymbol{z})_S\|_{\omega,1} + \|\boldsymbol{z}_S\|_{\omega,1} + \|\boldsymbol{z}_{S^c}\|_{\omega,1} = 2\sigma_s(\boldsymbol{x})_{\omega,1} + \|(\boldsymbol{x}-\boldsymbol{z})_S\|_{\omega,1} + \|\boldsymbol{z}\|_1.\end{aligned}$$

Rearranging and setting $\boldsymbol{v} := \boldsymbol{z}-\boldsymbol{x}$ leads to

$$\|\boldsymbol{v}_{S^c}\|_{\omega,1} \leq \|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + \|\boldsymbol{v}_S\|_{\omega,1} + 2\sigma_s(\boldsymbol{x})_{\omega,1}. \tag{25}$$

The Cauchy-Schwarz inequality implies

$$\|\boldsymbol{v}_S\|_{\omega,1} = \sum_{j\in S}|v_j|\omega_j \leq \sqrt{\sum_{j\in S}|v_j|^2}\sqrt{\sum_{j\in S}\omega_j^2} = \sqrt{\omega(S)}\|\boldsymbol{v}_S\|_2 \leq \sqrt{s}\|\boldsymbol{v}_S\|_2,$$

and therefore by (22)

$$\|\boldsymbol{v}_S\|_{\omega,1} \leq \sqrt{s}\|\boldsymbol{v}_S\|_2 \leq \rho\|\boldsymbol{v}_{S^c}\|_{\omega,1} + \tau\sqrt{s}\|\boldsymbol{A}\boldsymbol{v}\|_2. \tag{26}$$

We combine with (25) to arrive at

$$\|\boldsymbol{v}_{S^c}\|_{\omega,1} \leq \frac{1}{1-\rho}\left(\|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + \tau\sqrt{s}\|\boldsymbol{A}\boldsymbol{v}\|_2 + 2\sigma_s(\boldsymbol{x})_{\omega,1}\right).$$

Using (26) once more finally gives

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{z}\|_{\omega,1} &= \|\boldsymbol{v}_S\|_{\omega,1} + \|\boldsymbol{v}_{S^c}\|_{\omega,1} \leq (1+\rho)\|\boldsymbol{v}_{S^c}\|_{\omega,1} + \tau\sqrt{s}\|\boldsymbol{A}\boldsymbol{v}\|_2\\ &\leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + 2\sigma_s(\boldsymbol{x})_{\omega,1}\right) + \frac{2\tau\sqrt{s}}{1-\rho}\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{z})\|_2.\end{aligned}$$

We pass to the proof of (24). Let S with $\omega(S) \leq s$ be such that $\|\boldsymbol{v} - \boldsymbol{v}_S\|_2 = \|\boldsymbol{v}_{S^c}\|_2 = \widetilde{\sigma}_s(\boldsymbol{v})_{\omega,2}$ (recalling that $\|\cdot\|_2 = \|\cdot\|_{\omega,2}$). Using the weighted Stechkin estimate (21) and the robust null space property (22) as well as the error bound (23) we obtain

$$\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{z}\|_2 &\leq \|(\boldsymbol{x} - \boldsymbol{z})_{S^c}\|_2 + \|(\boldsymbol{x} - \boldsymbol{z})_S\|_2 \\
&\leq \frac{1}{\sqrt{s - \|\boldsymbol{\omega}\|_\infty^2}} \|\boldsymbol{x} - \boldsymbol{z}\|_{\omega,1} + \frac{\rho}{\sqrt{s}} \|(\boldsymbol{x} - \boldsymbol{z})_{S^c}\|_{\omega,1} + \tau \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{z})\|_2 \\
&\leq \frac{1 + \rho}{\sqrt{s - \|\boldsymbol{\omega}\|_\infty^2}} \|\boldsymbol{x} - \boldsymbol{z}\|_{\omega,1} + \tau \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{z})\|_2 \\
&\leq \frac{2(1 + \rho)^2}{(1 - \rho)\sqrt{s - \|\boldsymbol{\omega}\|_\infty^2}} \left(\|\boldsymbol{z}\|_{\omega,1} - \|\boldsymbol{x}\|_{\omega,1} + \sigma_s(\boldsymbol{x})_{\omega,1}\right) \\
&+ \left(\tau + \frac{2\tau(1 + \rho)\sqrt{s}}{(1 - \rho)\sqrt{s - \|\boldsymbol{\omega}\|_\infty^2}}\right) \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{z})\|_2.
\end{aligned}$$

Since $s \geq 2\|\boldsymbol{\omega}\|_\infty^2$ the statement follows with $C_1 = 2\sqrt{2}(1 + \rho)^2/(1 - \rho)$ and $C_2 = \tau + 2\sqrt{2}\tau(1 + \rho)/(1 - \rho)$. □

As an easy consequence of the previous result we obtain error bounds for sparse recovery via weighted ℓ_1 minimization.

Corollary 4.3. *Let $\boldsymbol{A} \in \mathbb{C}^{m \times N}$ satisfy the weighted robust null space property of order s and constants $\rho \in (0, 1)$ and $\tau > 0$. For $\boldsymbol{x} \in \mathbb{C}^N$ and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$ with $\|\boldsymbol{e}\|_2 \leq \eta$, let $\boldsymbol{x}^\sharp$ be the solution of*

$$\min_{\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}}} \|\boldsymbol{z}\|_1 \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta.$$

Then the reconstruction error satisfies

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_{\omega,1} \leq c_1 \sigma_s(\boldsymbol{x})_{\omega,1} + d_1 \sqrt{s}\eta \tag{27}$$

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_2 \leq c_2 \frac{\sigma_s(\boldsymbol{x})_{\omega,1}}{\sqrt{s}} + d_2 \eta, \tag{28}$$

where the second bound additionally assumes $s \geq 2\|\boldsymbol{\omega}\|_\infty^2$. The constants $c_1, c_2, d_1, d_2 > 0$ depend only on ρ and τ.

Proof. The reconstruction errors follow from the error bounds in Theorem 4.2 with $\boldsymbol{z} = \boldsymbol{x}^\#$, noting that $\|\boldsymbol{x}^\#\|_1 - \|\boldsymbol{x}\|_1 \leq 0$ and $\|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^\#)\|_2 \leq \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 + \|\boldsymbol{A}\boldsymbol{x}^\# - \boldsymbol{y}\|_2 \leq 2\eta$. □

Remark 4.4. In the case of noiseless measurements, the previous result gives error bounds for equality-constrained weighted ℓ_1 minimization by setting $\eta = 0$.
Moreover, with a similar technique as used for the previous result, one can generalize (27) and (28) to error bounds in weighted $\ell_{\omega,p}$ for $1 \leq p \leq 2$, see [17] for the unweighted case.

4.2 Weighted restricted isometry property

It is often unclear how to show the weighted null space property directly for a given matrix. In the unweighted case, it therefore has become useful to work instead with the restricted isometry property, which implies the null space property. As introduced in Definition 1.3, we define the weighted restricted isometry ($\boldsymbol{\omega}$-RIP) constant $\delta_{\omega,s}$ associated to a matrix $\boldsymbol{A}$ as the smallest number such that

$$(1-\delta_{\omega,s})\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1+\delta_{\omega,s})\|\boldsymbol{x}\|_2^2 \quad \text{for all } \boldsymbol{x} \text{ with } \|\boldsymbol{x}\|_{\omega,0} \leq s.$$

We say that $\boldsymbol{A}$ satisfies a weighted restricted isometry property ($\boldsymbol{\omega}$-RIP) if $\delta_{\omega,s}$ is small for s relatively large compared to m. The $\boldsymbol{\omega}$-RIP implies the weighted robust null space property and therefore the error bounds (27) and (28) for recovery via weighted ℓ_1 minimization as shown in the following result.

Theorem 4.5. *Let $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ with $\boldsymbol{\omega}$-RIP constant*

$$\delta_{\omega,3s} < 1/3 \tag{29}$$

for $s \geq 2\|\boldsymbol{\omega}\|_\infty^2$. Then $\boldsymbol{A}$ satisfies the weighted robust null space property of order s with constants $\rho = 2\delta_{\omega,3s}/(1-\delta_{\omega,3s}) < 1$ and $\tau = \sqrt{1+\delta_{\omega,3s}}/(1-\delta_{\omega,3s})$.

Before proving Theorem 4.5, we make the following observations. As in the unweighted case (see e.g. [17, 30]) the $\boldsymbol{\omega}$-RIP constants can be rewritten as

$$\delta_{\omega,s} = \max_{S\subset[N],\omega(S)\leq s} \|\boldsymbol{A}_S^*\boldsymbol{A}_S - \mathrm{Id}\|_{2\to 2},$$

where $\boldsymbol{A}_S$ denotes the submatrix of $\boldsymbol{A}$ restricted to the columns indexed by S.

Lemma 4.6. *If $\boldsymbol{u},\boldsymbol{v} \in \mathbb{C}^N$ are such that $\|\boldsymbol{u}\|_{\omega,0} \leq s$, $\|\boldsymbol{v}\|_{\omega,0} \leq t$ and* $\operatorname{supp}\boldsymbol{u} \cap \operatorname{supp}\boldsymbol{v} = \emptyset$ *then*

$$|\langle \boldsymbol{A}\boldsymbol{u}, \boldsymbol{A}\boldsymbol{v}\rangle| \leq \delta_{\omega,s+t}\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2.$$

Proof. Let $S = \operatorname{supp}\boldsymbol{u} \cup \operatorname{supp}\boldsymbol{v}$ so that $\omega(S) \leq s+t$. Since $\langle \boldsymbol{u},\boldsymbol{v}\rangle = 0$ we have

$$\begin{aligned}|\langle \boldsymbol{A}\boldsymbol{u}, \boldsymbol{A}\boldsymbol{v}\rangle| &= |\langle \boldsymbol{A}_S\boldsymbol{u}_S, \boldsymbol{A}_S\boldsymbol{v}_S\rangle - \langle \boldsymbol{u}_S,\boldsymbol{v}_S\rangle| = |\langle (\boldsymbol{A}_S^*\boldsymbol{A}_S - \mathrm{Id})\boldsymbol{u}_S, \boldsymbol{v}_S\rangle| \\ &\leq \|\boldsymbol{A}_S^*\boldsymbol{A}_S - \mathrm{Id}\,\|_{2\to 2}\|\boldsymbol{u}_S\|_2\|\boldsymbol{v}_S\|_2 \leq \delta_{\omega,s+t}\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2.\end{aligned}$$

This completes the proof. □

Now we are prepared for the proof of the main result of this section.

Proof of Theorem 4.5. Let $\boldsymbol{v} \in \mathbb{C}^N$ and $S \subset [N]$ with $\omega(S) \le s$. We partition S^c into blocks $S_1, S_2, \dots$ with $s - \|\boldsymbol{\omega}\|_\infty^2 \le \omega(S_\ell) \le s$ according to the nonincreasing rearrangement of $v_{S^c} \cdot \omega_{S^c}^{-1}$, that is, $|v_j|\omega_j^{-1} \le |v_k|\omega_k^{-1}$ for all $j \in S_\ell$ and all $k \in S_{\ell-1}$, $\ell \ge 2$. Then we estimate

$$
\begin{aligned}
\|\boldsymbol{v}_S + \boldsymbol{v}_{S_1}\|_2^2 &\le \frac{1}{1-\delta_{\omega,2s}} \|\boldsymbol{A}(\boldsymbol{v}_S + \boldsymbol{v}_{S_1})\|_2^2 = \frac{1}{1-\delta_{\omega,2s}} \left\langle \boldsymbol{A}(\boldsymbol{v}_S + \boldsymbol{v}_{S_1}), \boldsymbol{A}\boldsymbol{v} - \sum_{\ell \ge 2} \boldsymbol{A}\boldsymbol{v}_{S_\ell} \right\rangle \\
&= \frac{1}{1-\delta_{\omega,2s}} \left(\langle \boldsymbol{A}(\boldsymbol{v}_S + \boldsymbol{v}_{S_1}), \boldsymbol{A}\boldsymbol{v} \rangle - \sum_{\ell \ge 2} \langle \boldsymbol{A}(\boldsymbol{v}_S + \boldsymbol{v}_{S_1}), \boldsymbol{A}\boldsymbol{v}_{S_\ell} \rangle \right) \\
&\le \frac{1}{1-\delta_{\omega,2s}} \left(\sqrt{1+\delta_{\omega,2s}} \|\boldsymbol{v}_S + \boldsymbol{v}_{S_1}\|_2 \|\boldsymbol{A}\boldsymbol{v}\|_2 + \delta_{\omega,3s} \|\boldsymbol{v}_S + \boldsymbol{v}_{S_1}\|_2 \sum_{\ell \ge 2} \|\boldsymbol{v}_{S_\ell}\|_2 \right),
\end{aligned}
$$

where we have used Lemma 4.6 in the third line. Dividing by $\|\boldsymbol{v}_S + \boldsymbol{v}_{S_1}\|_2$ and using the fact that $\delta_{\omega,2s} \le \delta_{\omega,3s}$ we arrive at

$$
\|\boldsymbol{v}_S\|_2 \le \|\boldsymbol{v}_S + \boldsymbol{v}_{S_1}\|_2 \le \frac{\delta_{\omega,3s}}{1-\delta_{\omega,3s}} \sum_{\ell \ge 2} \|\boldsymbol{v}_{S_\ell}\|_2 + \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,3s}} \|\boldsymbol{A}\boldsymbol{v}\|_2.
$$

Now for $k \in S_\ell$, set $\alpha_k = (\sum_{j \in S_\ell} \omega_j^2)^{-1} \omega_k^2 \le (s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \omega_k^2$. Then $\sum_{k \in S_\ell} \alpha_k = 1$ and $|v_j|\omega_j^{-1} \le \sum_{k \in S_{\ell-1}} \alpha_k |v_k| \omega_k^{-1} \le (s - \|\boldsymbol{\omega}\|_\infty^2)^{-1} \sum_{k \in S_{\ell-1}} |v_k| \omega_k$ for all $j \in S_\ell$, $\ell \ge 2$, by our construction of the partitioning. By the Cauchy-Schwarz inequality and since $s \ge 2\|\boldsymbol{\omega}\|_\infty^2$ this gives

$$
\|\boldsymbol{v}_{S_\ell}\|_2 \le \frac{\sqrt{s}}{s - \|\boldsymbol{\omega}\|_2^2} \|\boldsymbol{v}_{S_{\ell-1}}\|_{\omega,1} \le \frac{2}{\sqrt{s}} \|\boldsymbol{v}_{S_{\ell-1}}\|_{\omega,1}.
$$

Therefore,

$$
\begin{aligned}
\|\boldsymbol{v}_S\|_2 &\le \frac{2\delta_{\omega,3s}}{(1-\delta_{\omega,3s})\sqrt{s}} \sum_{\ell \ge 1} \|\boldsymbol{v}_{S_\ell}\|_{\omega,1} + \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,3s}} \|\boldsymbol{A}\boldsymbol{v}\|_2 \\
&\le \frac{2\delta_{\omega,3s}}{(1-\delta_{\omega,3s})\sqrt{s}} \|\boldsymbol{v}_{S^c}\|_{\omega,1} + \frac{\sqrt{1+\delta_{\omega,3s}}}{1-\delta_{\omega,3s}} \|\boldsymbol{A}\boldsymbol{v}\|_2.
\end{aligned}
$$

This yields the desired estimate with $\tau = \sqrt{1+\delta_{\omega,3s}}/(1-\delta_{\omega,3s})$ and $\rho = 2\delta_{\omega,3s}/(1-\delta_{\omega,3s})$ which is strictly smaller than 1 if $\delta_{\omega,3s} < 1/3$. □

We remark that we did not attempt to provide the optimal constant in (29). Improvements can be achieved by pursuing more complicated arguments, see e.g. [17]. Also, conditions involving $\delta_{\omega,2s}$ instead of $\delta_{\omega,3s}$ are possible.

5 Weighted RIP estimates for orthonormal systems

In this section, we provide a quite general class of structured random matrices which satisfy the $\boldsymbol{\omega}$-RIP. For finite orthonormal systems $(\phi_j)_{j\in\Lambda}$ which are *bounded*, i.e., $\sup_{j\in\Lambda}\|\psi_j\|_\infty \leq K$ for some constant $K \geq 1$, the following unweighted RIP estimates have been shown.

Proposition 5.1 (Theorems 4.4 and 8.4, [30])**.** *Fix parameters $\delta, \gamma \in (0,1)$. Let $(\psi_j)_{j\in\Lambda}$ be a bounded orthonormal system with uniform bound K. Suppose*

$$\begin{aligned} m &\geq CK^2\delta^{-2}s\log^2(s)\log(m)\log(N), \\ m &\geq DK^2\delta^{-2}s\log(1/\gamma), \end{aligned} \tag{30}$$

where $N = |\Lambda|$. Assume that $t_1, t_2, \ldots, t_m$ are drawn independently from the orthogonalization measure ν associated to the orthonormal system. Then with probability exceeding $1-\gamma$, the normalized sampling matrix $\tilde{\boldsymbol{A}} \in \mathbb{C}^{m\times N}$ with entries $\tilde{A}_{\ell,k} = \frac{1}{\sqrt{m}}\psi_k(t_\ell)$, $\quad \ell \in [m], k \in [N]$, satisfies the restricted isometry property of order s, that is, $\delta_s \leq \delta$.

We could not expect to get standard RIP if $\|\psi_j\|_\infty$ grows with j, no matter what normalization we impose. However, we can allow $\|\psi_j\|_\infty$ to depend on j if we ask only for $\boldsymbol{\omega}$-RIP with weights $\omega_j = \|\psi_j\|_\infty$, or more generally, $\omega_j \geq \|\psi_j\|_\infty$. The main theorem of this section is that matrices arising from orthonormal systems satisfy the $\boldsymbol{\omega}$-RIP as long as the weights grow at least as quickly as the L_∞ norms of the functions they correspond to.

Theorem 5.2 ($\boldsymbol{\omega}$-RIP for orthonormal systems)**.** *Fix parameters $\delta, \gamma \in (0,1)$. Let $(\psi_j)_{j\in\Lambda}$ be an orthonormal system of finite size $N = |\Lambda|$. Consider weights satisfying $\omega_j \geq \|\psi_j\|_\infty$. Fix*

$$m \geq C\delta^{-2}s\max\{\log^3(s)\log(N), \log(1/\gamma)\}$$

and suppose that $t_1, t_2, \ldots, t_m$ are drawn independently from the orthogonalization measure associated to the (ψ_j). Then with probability exceeding $1-\gamma$, the normalized sampling matrix $\tilde{\boldsymbol{A}} \in \mathbb{C}^{m\times N}$ with entries $\tilde{A}_{\ell,k} = \frac{1}{\sqrt{m}}\psi_k(t_\ell)$ satisfies the weighted restricted isometry property of order s, that is, $\delta_{\omega,s} \leq \delta$.

We remark that if $K = \max_j \|\psi_j\|_\infty$ is a constant independent or only mildly dependent on N, then Theorem 5.2 essentially reduces to Proposition 5.1 concerning the unweighted RIP for bounded orthonormal systems. Note, however, that in the restricted parameter regime of $s \lesssim \log(N)$, the above result gives a slight improvement over the classical result stated in Proposition 5.1 – generalizing the main result of [9]. The remainder of this section is reserved for the proof of Theorem 5.2.

Proof of Theorem 5.2. The proof proceeds similar to those in [33, 30, 17], with some adaptations to account for the weights – see the application of Maurey's lemma (Lemma 5.3) –

and with a twist from [9] leading to the slight improvement in the logarithmic factor. Note that our analysis improves the result of [9] in terms of the probability estimate.

Introducing the set

$$T_\omega^{s,N} = \{\boldsymbol{x} \in \mathbb{C}^N, \|\boldsymbol{x}\|_2 \leq 1, \|\boldsymbol{x}\|_{0,\omega} \leq s\}, \tag{31}$$

we can rephrase the weighted isometry constant of $\boldsymbol{A}$ as

$$\delta_{\omega,s} = \sup_{\boldsymbol{x} \in T_\omega^{s,N}} |\langle (\boldsymbol{A}^*\boldsymbol{A} - \mathrm{Id})\boldsymbol{x}, \boldsymbol{x}\rangle|.$$

The quantity

$$|||\boldsymbol{B}|||_s := \sup_{\boldsymbol{z} \in T_\omega^{s,N}} |\langle \boldsymbol{B}\boldsymbol{z}, \boldsymbol{z}\rangle| \tag{32}$$

defines a semi-norm on matrices $\boldsymbol{B} \in \mathbb{C}^{N \times N}$, and we can write

$$\delta_{\omega,s} = |||\boldsymbol{A}^*\boldsymbol{A} - \mathrm{Id}|||_s.$$

Consider the random variable associated to a column of the adjoint matrix,

$$\boldsymbol{X}_\ell = \left(\overline{\psi_j(t_\ell)}\right)_{j \in \Lambda} \tag{33}$$

By orthonormality of the system (ψ_j) we have $\mathbb{E}\boldsymbol{X}_\ell \boldsymbol{X}_\ell^* = \mathrm{Id}$, and the restricted isometry constant equals

$$\delta_{\omega,s} = \left|\left|\left| \frac{1}{m}\sum_{\ell=1}^m \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* - \mathrm{Id} \right|\right|\right|_s = \frac{1}{m}\left|\left|\left| \sum_{\ell=1}^m (\boldsymbol{X}_\ell \boldsymbol{X}_\ell^* - \mathbb{E}\boldsymbol{X}_\ell \boldsymbol{X}_\ell^*) \right|\right|\right|_s.$$

As a first step we estimate the expectation of $\delta_{\omega,s}$ and later use a concentration result to deduce the probability estimate. We introduce a Rademacher sequence $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$, i.e., a sequence of independent Rademacher variables ϵ_ℓ taking the values $+1$ and -1 with equal probability, also independent of the variables $\boldsymbol{X}_\ell$. Symmetrization, see e.g. [26, Lemma 6.3] or [30, Lemma 6.7], yields

$$\begin{aligned}
\mathbb{E}\delta_{\omega,s} &\leq \frac{2}{m}\mathbb{E}\left|\left|\left| \sum_{\ell=1}^m \epsilon_\ell \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \right|\right|\right|_s = \frac{2}{m}\mathbb{E}_{\boldsymbol{X}}\mathbb{E}_{\boldsymbol{\epsilon}} \sup_{\boldsymbol{x} \in T_\omega^{s,N}} |\langle \sum_{\ell=1}^m \boldsymbol{\epsilon}_\ell \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \boldsymbol{x}, \boldsymbol{x}\rangle| \\
&= \frac{2}{m}\mathbb{E}_{\boldsymbol{X}}\mathbb{E}_\epsilon \sup_{\boldsymbol{x} \in T_\omega^{s,N}} |\sum_{\ell=1}^m \epsilon_\ell |\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^2|. \qquad (34)
\end{aligned}$$

Conditional on $(\boldsymbol{X}_\ell)$, we arrive at a Rademacher (in particular, subgaussian) process indexed by $T_\omega^{s,N}$. For a set T, a metric d and given $u > 0$, the covering numbers $\mathcal{N}(T, d, u)$

are defined as the smallest number of balls with respect to d and centered at points of T necessary to cover T. For fixed $(\boldsymbol{X}_\ell)$, we work with the (pseudo-)metric

$$d(\boldsymbol{x},\boldsymbol{z}) = \left(\sum_{\ell=1}^{m}(|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^2 - |\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|^2)^2\right)^{1/2}.$$

Then Dudley's inequality [27, 26, 30, 17] implies that

$$\mathbb{E}_{\boldsymbol{\epsilon}} \sup_{\boldsymbol{x}\in T_\omega^{s,N}} |\langle \sum_{\ell=1}^{m} \epsilon_\ell X_\ell X_\ell^* \boldsymbol{x}, \boldsymbol{x}\rangle| \le 4\sqrt{2}\int_0^\infty \sqrt{\log(\mathcal{N}(T_\omega^{s,N}, d, u))}du.$$

In order to continue we estimate the metric d using Hölder's inequality with exponents $p \ge 1$ and $q \ge 1$ satisfying $1/p + 1/q = 1$ to be specified later on. For $\boldsymbol{x},\boldsymbol{z} \in T_\omega^{s,N}$, this gives

$$\begin{aligned} d(\boldsymbol{x},\boldsymbol{z}) &= \left(\sum_{\ell=1}^{m}(|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle| + |\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|)^2(|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle| - |\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|)^2\right)^{1/2} \\ &\le \left(\sum_{\ell=1}^{m}(|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle| + |\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|)^{2p}\right)^{1/(2p)} \left(\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}-\boldsymbol{z}\rangle|^{2q}\right)^{1/(2q)} \\ &\le 2\sup_{\boldsymbol{x}\in T_\omega^{s,N}} \left(\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^{2p}\right)^{1/(2p)} \left(\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}-\boldsymbol{z}\rangle|^{2q}\right)^{1/(2q)}. \end{aligned} \tag{35}$$

In the standard analysis [33, 30, 17], this bound is applied for $p = 1$, $q = \infty$. Following [9], we will achieve a slightly better log-factor by working with a different value of p to be determined later.

For any realization of $(\boldsymbol{X}_\ell)$, we have $|(\boldsymbol{X}_\ell)_j| \le \|\psi_j\|_\infty \le \omega_j$. For $\boldsymbol{x} \in T_\omega^{s,N}$ with $S = \operatorname{supp}\boldsymbol{x}$ we have $\sum_{j\in S}\omega_j^2 \le s$, resulting in

$$|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle| \le \sum_{j\in S}\omega_j|x_j| \le (\sum_{j\in S}\omega_j^2)^{1/2}\|\boldsymbol{x}\|_2 \le \sqrt{s}. \tag{36}$$

This gives

$$\begin{aligned} \sup_{\boldsymbol{x}\in T_\omega^{s,N}} \left(\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^{2p}\right)^{1/(2p)} &= \sup_{\boldsymbol{x}\in T_\omega^{s,N}} \left(\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^{2}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^{2(p-1)}\right)^{1/(2p)} \\ &\le s^{(p-1)/(2p)}\left(\sup_{\boldsymbol{x}\in T_\omega^{s,N}}\sum_{\ell=1}^{m}|\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^2\right)^{1/(2p)}. \end{aligned}$$

Introducing the (semi-)norm

$$\|\boldsymbol{x}\|_{X,q} = \left(\sum_{\ell=1}^{m} |\langle \boldsymbol{X}_\ell, \boldsymbol{x} \rangle|^{2q} \right)^{1/(2q)}$$

and using basic properties of covering numbers, we obtain

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{\epsilon}} \sup_{\boldsymbol{x} \in T_\omega^{s,N}} |\langle \sum_{\ell=1}^{m} \epsilon_\ell \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \boldsymbol{x}, \boldsymbol{x} \rangle| \\ &\leq C_1 s^{(p-1)/(2p)} \left(\sup_{\boldsymbol{x} \in T_\omega^{s,N}} \sum_{\ell=1}^{m} |\langle \boldsymbol{X}_\ell, \boldsymbol{x} \rangle|^2 \right)^{1/(2p)} \int_0^\infty \sqrt{\log(\mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{\boldsymbol{X},q}, u))} du, \end{aligned} \tag{37}$$

where C_1 is a suitable constant. Next, we estimate the covering numbers appearing above in two different ways.

Let us first derive a bound which is good for small values of u. It follows from (36) that, for $\boldsymbol{x} \in T_\omega^{s,N}$,

$$\|\boldsymbol{x}\|_{\boldsymbol{X},q} \leq \left(\sum_{\ell=1}^{m} (\sqrt{s} \|\boldsymbol{x}\|_2)^{2q} \right)^{1/2q} = \sqrt{s} m^{1/(2q)} \|\boldsymbol{x}\|_2. \tag{38}$$

Denoting B_S to be the ℓ_2 unit ball of vectors with support in S and applying the volumetric covering number bound (see e.g. [30, Proposition 10.1]) gives

$$\begin{aligned} \mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{\boldsymbol{X},q}, u) &\leq \sum_{S \subset \Lambda : \omega(S) \leq s} \mathcal{N}(B_S, \sqrt{s} m^{1/(2q)} \|\cdot\|_2, u) \\ &\leq \binom{N}{s} \left(1 + \frac{2\sqrt{s} m^{1/(2q)}}{u} \right)^{2s} \leq (eN/s)^s \left(1 + \frac{2\sqrt{s} m^{1/(2q)}}{u} \right)^{2s}, \end{aligned}$$

where we have applied [17, Proposition C.3] (see also [30, p. 72]) in the last step.

We use Maurey's lemma [8], see also [22, Lemma 4.2] for the precise form below, in order to deduce a covering number bound which is good for larger values of u. Below, $\operatorname{conv}(U)$ denotes the convex hull of a set U.

Lemma 5.3. *For a normed space X, consider a finite set $\mathcal{U} \subset X$ of cardinality N, and assume that for every $L \in \mathbb{N}$ and $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_L) \in \mathcal{U}^L$, $\mathbb{E}_{\boldsymbol{\epsilon}} \| \sum_{j=1}^{L} \epsilon_j \boldsymbol{u}_j \|_X \leq A\sqrt{L}$, where $\boldsymbol{\epsilon}$ denotes a Rademacher vector. Then for every $u > 0$,*

$$\log \mathcal{N}(\operatorname{conv}(\mathcal{U}), \|\cdot\|_X, u) \leq c(A/u)^2 \log N.$$

The constant $c > 0$ is universal.

To apply this lemma, we first observe that $T_\omega^{s,N} \subset \sqrt{2s}\,\mathrm{conv}(U)$, where

$$U = \{\pm \omega_j^{-1}\boldsymbol{e}_j, \pm i\omega_j^{-1}\boldsymbol{e}_j, j \in \Lambda\}.$$

Here, $\boldsymbol{e}_j$ denotes the j-th canonical unit vector. For a Rademacher vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_L)$, and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_L \in U$ we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\epsilon}} \| \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j \|_{X,q} &\leq \left(\mathbb{E} \| \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j \|_{X,q}^{2q} \right)^{1/(2q)} = \left(\mathbb{E} \sum_{\ell=1}^m |\langle \boldsymbol{X}_\ell, \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j \rangle|^{2q} \right)^{1/(2q)} \\
&= \left(\sum_{\ell=1}^m \mathbb{E} \left| \sum_{j=1}^L \epsilon_j \langle \boldsymbol{X}_\ell, \boldsymbol{u}_j \rangle \right|^{2q} \right)^{1/(2q)} \leq 2e^{-1/2}\sqrt{2q} \left(\sum_{\ell=1}^m \| (\langle \boldsymbol{X}_\ell, \boldsymbol{u}_j \rangle)_{j=1}^L \|_2^{2q} \right)^{1/(2q)}.
\end{aligned}$$

In the last step, we have applied Khintchine's inequality, see e.g. [30, Corollary 6.9]. Using that $|(\boldsymbol{X}_\ell)_k| \leq \|\psi_k\|_\infty \leq \omega_k$, we have, for any vector $\boldsymbol{u}_j \in U$, say $\boldsymbol{u}_j = \omega_k^{-1}\boldsymbol{e}_k$, that

$$|\langle \boldsymbol{X}_\ell, \boldsymbol{u}_j \rangle| = |\omega_k^{-1} (\boldsymbol{X}_\ell)_k| \leq 1.$$

Therefore, $\|(\langle \boldsymbol{X}_\ell, \boldsymbol{u}_j \rangle)_{j=1}^L\|_2 \leq \sqrt{L}$ for any L and

$$\mathbb{E}_{\boldsymbol{\epsilon}} \| \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j \|_{\boldsymbol{X},q} \leq 2e^{-1/2}\sqrt{2q} m^{1/(2q)} \sqrt{L}.$$

An application of Lemma 5.3 with $A = 2e^{-1/2}\sqrt{2q}m^{1/(2q)}$ yields

$$\sqrt{\log \mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{\boldsymbol{X},q}, u)} \leq \sqrt{\log \mathcal{N}(\mathrm{conv}(U), \|\cdot\|_{\boldsymbol{X},q}, u/\sqrt{2s})} \leq C_2 \sqrt{q m^{1/q} s \log(4N)}\, u^{-1}$$

with $C_2 = 4e^{-1/2}\sqrt{c}$.

Observe that it is enough to choose the upper integration bound in the Dudley type integral as $\sqrt{s}m^{1/(2q)}$ because for $u > \sqrt{s}m^{1/(2q)}$ we have $\mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{X,q}, u) = 1$ by (38). Splitting then the Dudley integral into two parts and using the appropriate bounds for the covering numbers, we obtain, for $\kappa \in \left(0, \sqrt{s}m^{1/(2q)}\right)$,

$$\begin{aligned}
&\int_0^\infty \sqrt{\log(\mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{\boldsymbol{X},q}, u))} du \\
&\leq \int_0^\kappa \sqrt{s \log(eN/s) + 2s \log(1 + 2\sqrt{s}m^{1/(2q)}/u)} du \\
&+ C_2 \sqrt{q m^{1/q} s \log(4N)} \int_\kappa^{\sqrt{s}m^{1/(2q)}} u^{-1} du \\
&\leq \kappa \sqrt{s \log(eN/s)} + \sqrt{2s}\kappa \sqrt{\log(e(1 + \sqrt{s}m^{1/(2q)}))} + C' \sqrt{q m^{1/q} s \log(4N)} \log(\sqrt{s}m^{1/(2q)}/\kappa).
\end{aligned}$$

In the last step, we have applied [30, Lemma 10.3]. Choosing $\kappa = m^{1/(2q)}$ yields

$$\int_0^\infty \sqrt{\log(\mathcal{N}(T_\omega^{s,N}, \|\cdot\|_{\boldsymbol{X},q}, u))} du \leq C_3 \sqrt{qsm^{1/q}\log(N)\log^2(s)}.$$

A combination with (37) and (34) gives

$$\begin{aligned}
\mathbb{E}\delta_{\omega,s} &\leq \frac{C_3 s^{(p-1)/(2p)}\sqrt{qm^{1/q}s\log(N)\log^2(s)}}{m}\mathbb{E}\sup_{\boldsymbol{x}\in T_\omega^{s,N}}\left(\sum_{\ell=1}^m |\langle \boldsymbol{X}_\ell, x\rangle|^2\right)^{1/(2p)} \\
&\leq \frac{C_3 s^{1/2+(p-1)/(2p)}\sqrt{q\log(N)\log^2(s)}}{m^{1-1/(2q)}m^{-1/(2p)}}\mathbb{E}\left(\frac{1}{m}\left\|\left\|\sum_{\ell=1}^m X_\ell X_\ell^* - \mathrm{Id}\right\|\right\|_s + \|\|\mathrm{Id}\|\|_s\right)^{1/(2p)} \\
&\leq \frac{C_3 s^{1/2+(p-1)/(2p)}\sqrt{q\log(N)\log^2(s)}}{m^{1/2}}\sqrt{\mathbb{E}\delta_{\omega,s}+1}.
\end{aligned}$$

Hereby, we have applied Hölder's inequality and used that $1/q + 1/p = 1$ as well as $p \geq 1$. Choosing $p = 1+1/\log(s)$ and $q = 1+\log(s)$ gives $s^{(p-1)/(2p)} \leq s^{(p-1)/2} = s^{1/(2\log(s))} = e^{1/2}$ and

$$\mathbb{E}\delta_{\omega,s} \leq C_4\sqrt{\frac{s\log(N)\log^3(s)}{m}}\sqrt{\mathbb{E}\delta_{\omega,s}+1}.$$

Completing squares finally shows that

$$\mathbb{E}\delta_{\omega,s} \leq C_5\sqrt{\frac{s\log(N)\log^3(s)}{m}} \tag{39}$$

provided the term under the square root is bounded by 1.

For the **probability bound**, we show that $\delta_{\omega,s}$ does not deviate much from its expectation. By (39), $\mathbb{E}\delta_{\omega,s} \leq \delta/2$ for some $\delta \in (0,1)$ if

$$m \geq C_6\delta^{-2}s\log^3(s)\log(N) \tag{40}$$

with $C_6 = 4C_5^2$.

Similarly to [30, Section 8.6] we write

$$\delta_{\omega,s} = \frac{1}{m}\sup_{(\boldsymbol{z},\boldsymbol{w})\in Q_{\omega,*}^{s,N}} \Re\left\langle \sum_{\ell=1}^m (\boldsymbol{X}_\ell\boldsymbol{X}_\ell^* - \mathrm{Id})\boldsymbol{z}, \boldsymbol{\omega}\right\rangle$$

where $Q_{\omega,*}^{s,N}$ denotes a dense countable subset of

$$Q_\omega^{s,N} = \bigcup_{S\subset\Lambda,\omega(S)\leq s} Q_S, \qquad Q_S = \{(\boldsymbol{z},\boldsymbol{w}) : \|\boldsymbol{z}\|_2 = \|\boldsymbol{w}\|_2 = 1, \operatorname{supp}\boldsymbol{z}, \operatorname{supp}\boldsymbol{w} \subset S\}$$

With the functions $f_{\boldsymbol{z},\boldsymbol{w}}(X) = \Re\langle (\boldsymbol{X}\boldsymbol{X}^* - \mathrm{Id})\boldsymbol{x}, \boldsymbol{w}\rangle$ we can write $\delta_{\omega,s}$ as the supremum of an empirical process

$$\delta_{\omega,s} = \frac{1}{m} \sup_{(\boldsymbol{z},\boldsymbol{w}) \in Q_{\omega,*}^{s,N}} \sum_{\ell=1}^{m} f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell).$$

Since $\mathbb{E}\boldsymbol{X}_\ell \boldsymbol{X}_\ell^* = \mathrm{Id}$ we have $f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell) = 0$ for all $\boldsymbol{z}, \boldsymbol{w}$. Further, for $(\boldsymbol{x}, \boldsymbol{w}) \in Q_S$ with $\omega(S) \leq s$ and for any realization of $\boldsymbol{X}_\ell$, we have

$$|f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell)| \leq \max\{1, \max_{\substack{\boldsymbol{x}:\operatorname{supp}\boldsymbol{x} \subset S \\ \|\boldsymbol{x}\|_2 = 1}} |\langle \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \boldsymbol{x}, \boldsymbol{x}\rangle|\} = \max\{1, \max_{\substack{\boldsymbol{x}:\operatorname{supp}\boldsymbol{x} \subset S \\ \|\boldsymbol{x}\|_2 = 1}} |\langle \boldsymbol{X}_\ell, \boldsymbol{x}\rangle|^2\} \leq s.$$

Moreover,

$$\begin{aligned}
\mathbb{E}|f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell)|^2 &= \mathbb{E}|\langle (\boldsymbol{X}_\ell \boldsymbol{X}_\ell^* - \mathrm{Id})\boldsymbol{z}, \boldsymbol{w}\rangle|^2 \\
&= \mathbb{E}|\langle \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \boldsymbol{z}, \boldsymbol{w}\rangle|^2 - 2\Re(\mathbb{E}[\langle \boldsymbol{X}_\ell \boldsymbol{X}_\ell^* \boldsymbol{z}, \boldsymbol{w}\rangle]\langle \boldsymbol{w}, \boldsymbol{z}\rangle) + |\langle \boldsymbol{z}, \boldsymbol{w}\rangle|^2 \\
&= \mathbb{E}[|\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|^2 |\langle \boldsymbol{X}_\ell, \boldsymbol{w}\rangle|^2] - |\langle \boldsymbol{z}, \boldsymbol{w}\rangle|^2 \leq s\mathbb{E}|\langle \boldsymbol{X}_\ell, \boldsymbol{z}\rangle|^2 = s.
\end{aligned}$$

With these bounds for $f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell)$ together with (40), the Bernstein inequality for the supremum of an empirical process, see e.g. [30, Theorem 6.25], [4] or [17, Theorem 8.42], yields, for $\delta \in (0, 1)$,

$$\begin{aligned}
\mathbb{P}(\delta_{\omega,s} \geq \delta) &\leq \mathbb{P}(\delta_{\omega,s} \geq \mathbb{E}\delta_{\omega,s} + \delta/2) \\
&= \mathbb{P}(\sup_{(\boldsymbol{z},\boldsymbol{w}) \in Q_{\omega,*}^{s,N}} \sum_{\ell=1}^{m} f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell) \geq \mathbb{E} \sup_{(\boldsymbol{z},\boldsymbol{w}) \in Q_{\omega,*}^{s,N}} \sum_{\ell=1}^{m} f_{\boldsymbol{z},\boldsymbol{w}}(\boldsymbol{X}_\ell) + \delta m/2) \\
&\leq \exp\left(-\frac{(\delta m/2)^2/2}{ms + 2s(\delta m/2) + (\delta m/2)s/3} \right) \leq \exp\left(-\frac{\delta^2 m}{C_7 s} \right),
\end{aligned} \tag{41}$$

where $C_7 = 8(1 + 2 + 1/6) \leq 26$. The last term is bounded by $\gamma \in (0, 1)$ if $m \geq C_7 \delta^{-2} s \log(1/\gamma)$. Altogether we have $\delta_{\omega,s} \leq \delta$ with probability at least $1 - \gamma$ if

$$m \geq C_8 \delta^{-2} s \max\{\log^3(s)\log(N), \log(1/\gamma)\},$$

where $C_8 = \max\{C_6, C_7\}$. This completes the proof. $\square$

6 Putting it all together: main results

Using the concepts of weighted null space and restricted isometry properties, and together with Theorem 5.2 concerning the ω-RIP for orthonormal systems, we now prove Theorems 1.1 and 1.2 concerning interpolation via weighted ℓ_1 minimization, and a more general result for functions with coefficients in weighted ℓ_p spaces. We first state a finite-dimensional result which allows for noisy measurements.

Theorem 6.1. *Suppose $(\psi_j)_{j\in\Lambda}$ is an orthonormal system with $|\Lambda| = N$ finite. Consider weights $\omega_j \geq \|\psi_j\|_\infty$. For $s \geq 2\max_j |\omega_j|^2$ and $\gamma \in (0,1)$, fix a number of samples*

$$m \geq c_0 s \max\{\log^3(s)\log(N), \log(1/\gamma)\}, \tag{42}$$

suppose that t_ℓ, $\ell = 1,\dots,m$, are drawn independently from the orthogonalization measure associated to the (ψ_j). Let $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ be the sampling matrix with entries $A_{\ell,k} = \psi_k(t_\ell)$. Then with probability exceeding $1-\gamma$, the following holds for all functions $f = \sum_{j\in\Lambda} x_j\psi_j$. Given noisy samples $y_\ell = f(t_\ell) + \xi_\ell$, $\ell = 1,\dots,m$, with $\|\xi\|_2 \leq \eta$, let $\boldsymbol{x}^\sharp$ be the solution of

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta$$

and set $f^\sharp(t) = \sum_{j\in\Lambda} x_j^\sharp \psi_j(t)$. Then

$$\|f - f^\sharp\|_{L_\infty} \leq |||\, f - f^\sharp \,|||_{\omega,1} \leq c_1\sigma_s(f)_{\omega,1} + d_1\sqrt{s/m}\,\eta,$$
$$\|f - f^\sharp\|_{L_2} \leq c_2\frac{\sigma_s(f)_{\omega,1}}{\sqrt{s}} + d_2\eta/\sqrt{m}.$$

Above, $c_0, c_1, d_1, c_2,$ and d_2 are universal constants.

Proof. By Theorem 5.2, the normalized sampling matrix $\tilde{\boldsymbol{A}} = \frac{1}{\sqrt{m}}\boldsymbol{A} \in \mathbb{C}^{m\times N}$ satisfies the $\boldsymbol{\omega}$-RIP of order s with $\delta_{\omega,s} \leq 1/3$ with probability exceeding $1-\gamma$, at the stated number of measurements in (42). Given that $\tilde{\boldsymbol{A}}$ satisfies the $\boldsymbol{\omega}$-RIP, Theorem 4.5 implies that $\tilde{\boldsymbol{A}}$ satisfies the weighted null space property with constants $\rho < 1$ and $\tau > 0$. The bounds for $|||\, f - f^\sharp \,|||_{\omega,1}$ and $\|f - f^\sharp\|_{L_2}$ follow by applying Corollary 4.3. To get the bound on $\|f - f^\sharp\|_{L_\infty}$, recall that because $\|\psi_j\|_\infty \leq \omega_j$ by assumption, the reconstruction error in $\ell_{\omega,1}$ implies

$$\begin{aligned} \|f - f^\sharp\|_\infty &\leq \sum_{j=1}^N |x_j - x_j^\sharp|\|\psi_j\|_\infty \leq \|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_{\omega,1} \\ &= |||\, f - f^\sharp \,|||_{\omega,1}. \end{aligned}$$

□

Remark 6.2. Theorem 1.1 in the introduction corresponds to the special case of Theorem 6.1 where there is no noise, $\eta = 0$, and with $\gamma = N^{-\log^3(s)}$ chosen to balance both terms in the maximum in (42) so that $m \geq c_0 s\log^3(s)\log(N)$ implies the stated error bounds with probability at least $1 - N^{-\log^3(s)}$.

6.1 Proof of Theorem 1.2

If the index set Λ is countably infinite then we first have to restrict to a suitable finite subset Λ_0 before applying weighted ℓ_1 minimization for reconstruction. The suitable finite subset we consider is $\Lambda_0 = \{j : \omega_j^2 \leq s/2\}$. The basic idea is to treat the samples of $f = \sum_{j \in \Lambda} x_j \psi_j$ as perturbed or noisy samples of the finite-dimensional approximation $f_0 = \sum_{j \in \Lambda_0} x_j \psi_j$, decomposing Λ into Λ_0 and Λ_R, and f as $f = f_0 + f_R$, and treating $f_R(t_j) = \sum_{j \in \Lambda_R} x_j \psi_j(t_j)$ as noise on the observed sampling in hopes of applying Theorem 6.1. The remainder of the proof is to show that the error $\sum_{\ell=1}^m |f_R(t_\ell)|^2 = \eta^2$ is small with high probability. Since the sampling points $t_1, t_2, \ldots, t_m$ are drawn i.i.d. from the orthogonalization measure associated to (ψ_j), the random variables $|f_R(t_\ell)|^2$ are independent and identically distributed, with

$$\mathbb{E}\Big(|f_R(t_\ell)|^2\Big) = \sum_{j \in \Lambda_R} |x_j|^2. \tag{43}$$

Since $\omega_j^2 \geq s/2$ for $j \in \Lambda_R$ by construction, we have

$$\sum_{j \in \Lambda_R} x_j^2 \leq \frac{2}{s} \sum_{j \in \Lambda_R} x_j^2 \omega_j^2 \leq \frac{2}{s}\Big(\sum_{j \in \Lambda_R} |x_j| \omega_j\Big)^2 = \frac{2}{s}\|f_R\|_{\omega,1}^2.$$

We further have the sup-norm bound $|f_R(t_\ell)| \leq \|f_R\|_{L_\infty} \leq \sum_{j \in \Lambda_R} |x_j| \omega_j = \|f_R\|_{\omega,1}$. Therefore, the variance of the mean-zero variable $|f_R(t_\ell)|^2 - \mathbb{E}\Big(|f_R(t_\ell)|^2\Big)$ is bounded by

$$\mathbb{E}\Big(|f_R(t_\ell)|^2 - \mathbb{E}\Big(|f_R(t_\ell)|^2\Big)\Big)^2 \leq \mathbb{E}\Big(|f_R(t_\ell)|^4\Big) \leq \|f_R\|_{\omega,1}^2 \mathbb{E}\Big(|f_R(t_\ell)|^2\Big) \leq \frac{2}{s}\|f_R\|_{\omega,1}^4.$$

We now apply Bernstein's inequality to certify the probability bound

$$\mathbb{P}\left\{\left|\frac{1}{m}\sum_{\ell=1}^m |f_R(t_\ell)|^2 - \sum_{j \in \Lambda_R} x_j^2\right| \geq \kappa\right\} \leq \exp\left\{-\frac{m\kappa^2/2}{2\|f_R\|_{\omega,1}^4/s + \kappa\|f_R\|_{\omega,1}^2/3}\right\}.$$

Setting $\kappa = \frac{3}{s}\|f_R\|_{\omega,1}^2$ in Bernstein's inequality gives

$$\mathbb{P}\left\{\left|\frac{1}{m}\sum_{\ell=1}^m |f_R(t_\ell)|^2 - \sum_{j \in \Lambda_R} x_j^2\right| \geq \frac{3}{s}\|f_R\|_{\omega,1}^2\right\} \leq \exp\left\{-\frac{3m}{2s}\right\}.$$

For the number of measurements $m = c_0 s \log^3(s) \log(N)$ stated in Theorem 1.2, we therefore have by (43)

$$\mathbb{P}\left\{\frac{1}{m}\sum_{\ell=1}^m |f_R(t_\ell)|^2 \geq \frac{1}{s}\|f_R\|_{\omega,1}^2\right\} \leq N^{-\log^3(s)}.$$

Note that $\sigma_s(f)_{\omega,1} = \sigma_s(f_0)_{\omega,1} + \|f_R\|_{\omega,1}$ since the best weighted s-term approximations to f and f_0 are the same. Theorem 1.2 results then by application of Theorem 6.1 with $\gamma = N^{-\log^3(s)}$.

6.2 Interpolation estimates in weighted ℓ_p spaces, $p \leq 1$

Theorem 1.2 is somewhat weak in the sense that for a random draw of the sampling points, it gives guarantees with high probability only for a fixed function. In order to derive uniform recovery bounds, or guarantees *for all functions* in a given class for a single set of measurements, as opposed to guarantees for a particular function, we need to introduce a positive weight vector v which dominates the weight vector ω in a suitable way. In order to illustrate the idea we start by recalling the error bound

$$\|f - f^\sharp\|_{L_\infty} \leq c_1 \sigma_s(f)_{\omega,1} + d_1\sqrt{s}\eta/\sqrt{m}, \tag{44}$$

valid in the finite-dimensional setting, where $f = \sum_{j\in\Lambda} x_j\psi_j$ and the samples are perturbed, $\sum_{\ell=1}^m |y_\ell - f(t_\ell)|^2 \leq \eta^2$. As in the probabilistic error analysis, we treat the samples $y_\ell = f(t_\ell)$ as perturbed samples of a finite-dimensional approximation $f_0 = \sum_{j\in\Lambda_0} x_j\psi_j$ for some suitable $\Lambda_0 \subset \Lambda$. For a parameter $\alpha > 0$, the approximation error can then be bounded using

$$\|f - f_0\|_{L_\infty} \leq \sum_{j\notin\Lambda_0} |x_j|\|\psi_j\|_\infty \leq \max_{j\notin\Lambda_0}\{\|\psi_j\|_\infty v_j^{-\alpha}\} \sum_{j\notin\Lambda_0} |x_j| v_j^\alpha \leq \max_{j\notin\Lambda_0}\{w_j v_j^{-\alpha}\} \,|\!|\!|\, f \,|\!|\!|_{v^\alpha,1}. \tag{45}$$

On the right hand side, we obtain the norm $|\!|\!|\, f \,|\!|\!|_{v^\alpha,1}$. Recall, however, that for our compressive sensing approximation we can impose $|\!|\!|\, f \,|\!|\!|_{v,p}$ to be small for a small value of $p < 1$. The following estimate will be useful for comparing weighted p and 1-norms.

Lemma 6.3. *For a weight $\boldsymbol{\omega}$ and $0 < p < 1$, set $\alpha = 2/p - 1$. Then $\|\boldsymbol{x}\|_{\omega^\alpha,1} \leq \|\boldsymbol{x}\|_{\omega,p}$.*

Proof. First observe that

$$\left(\max_{j\in\Lambda_0} |x_j|\omega_j^{2/p-1}\right)^p = \max_j |x_j|^p \omega_j^{2-p} \leq \sum_{j\in\Lambda_0} |x_j|^p\omega_j^{2-p} = \|\boldsymbol{x}\|_{\omega,p}^p.$$

The claimed inequality follows then from

$$\begin{aligned}\|\boldsymbol{x}\|_{\omega^\alpha,1} &= \sum_{j\in\Lambda_0} |x_j|\omega_j^\alpha \leq \left(\max_j |x_j|^{1-p}\omega_j^{\alpha-2+p}\right)\sum_j |x_j|^p\omega_j^{2-p} \\ &= \left(\max_j |x_j|\omega_j^{2/p-1}\right)^{1-p} \|\boldsymbol{x}\|_{\omega,p}^p \leq \|\boldsymbol{x}\|_{\omega,p}^{1-p}\|\boldsymbol{x}\|_{\omega,p}^p = \|\boldsymbol{x}\|_{\omega,p}.\end{aligned}$$

□

Assuming that $v \geq \omega$ and $s \geq 2\max_{j\in\Lambda_0} w_j^2$, say, the first term in the finite-dimensional error bound (44) with f replaced by f_0 can be estimated using the Stechkin-type estimate of Theorem 3.2 by

$$\sigma_s(f_0)_{\omega,1} \leq cs^{1-1/p} \,|\!|\!|\, f_0 \,|\!|\!|_{\omega,p} \leq cs^{1-1/p} \,|\!|\!|\, f \,|\!|\!|_{v,p}. \tag{46}$$

We aim to provide a bound of the second term on the right-hand side of (44) of the same order. Now $\eta := \sqrt{\sum_{\ell=1}^{m} |f_0(t_\ell) - f(t_\ell)|^2} \leq \sqrt{m}\|f - f_0\|_{L_\infty}$, so by (45) and Lemma 6.3,

$$\frac{\sqrt{s}\eta}{\sqrt{m}} \leq \sqrt{s} \max_{j \notin \Lambda_0}\{w_j v_j^{-\alpha}\} \left|\!\left|\!\left| f \right|\!\right|\!\right|_{v^\alpha,1} \leq \sqrt{s} \max_{j \notin \Lambda_0}\{w_j v_j^{-\alpha}\} \left|\!\left|\!\left| f \right|\!\right|\!\right|_{v,p},$$

where we have applied Lemma 6.3 with $\alpha = 2/p - 1$ in the last step. The choice $\Lambda_0 = \Lambda_0^{(s,p)}$ with

$$\Lambda_0^{(s,p)} := \{j \in \Lambda : \omega_j v_j^{1-2/p} \geq s^{1/2-1/p}\} \tag{47}$$

therefore gives

$$\eta \leq \sqrt{m} s^{1/2-1/p} \left|\!\left|\!\left| f \right|\!\right|\!\right|_{v,p} \quad \text{so that} \quad \frac{\sqrt{s}\eta}{\sqrt{m}} \leq s^{1-1/p} \left|\!\left|\!\left| f \right|\!\right|\!\right|_{v,p},$$

and we have balanced the two error terms in (44) after applying (46). We still need to choose the weight v so that $\Lambda_0^{s,p}$ is a finite set (ideally with size polynomial in s) and such that the technical assumption $\max_{j \in \Lambda_0^{(s,p)}} w_j^2 \leq s/2$ is satisfied. The finiteness of $\Lambda_0^{(s,p)}$ is ensured when $(\omega_j v_j^{1-2/p})_{j \in \Lambda}$ is a sequence which converges to 0 as $|j| \to \infty$. Moreover, if v satisfies

$$v_j^{2/p-1} \geq 2^{1/p-1/2} \omega_j^{2/p} = 2^{1/p-1/2} \omega_j \cdot \omega_j^{2(1/p-1/2)}, \tag{48}$$

then we have for all j satisfying $\omega_j^2 \geq s/2$ that

$$v_j^{2/p-1} \geq 2^{1/p-1/2} \omega_j (s/2)^{1/p-1/2} = \omega_j s^{1/p-1/2}.$$

Then, in light of (47), all $j \in \Lambda_0^{(s,p)}$ satisfy $\max_{j \in \Lambda_0^{(s,p)}} \omega_j^2 \leq s/2$. Inequality (48) is satisfied if $v_j \geq 2\omega_j^{1/(1-p/2)}$. In particular, the choice

$$v_j = 2\omega_j^2$$

is valid for all values of $p \in (0,1]$. In this case $(\omega_j v_j^{1-2/p})_{j \in \Lambda}$ converges to 0 as $|j| \to \infty$ if and only if $(\omega_j^{-1})_{j \in \Lambda}$ converges to 0 as $|j| \to \infty$. We can now state the main result.

Theorem 6.4. *Let $p \in (0,1]$ and let ω, v be weights satisfying $\omega_j \geq \|\psi_j\|_\infty$ and $v_j \geq 2\omega_j^{1/(1-p/2)}$. For given $s \in \mathbb{N}$, let $\Lambda_0^{(s,p)} = \Lambda_0$ of size $N^{(s,p)}$ be the set of indices*

$$\Lambda_0 := \{j \in \Lambda : \omega_j v_j^{1-2/p} \geq s^{1/2-1/p}\}$$

Fix a number m of samples

$$m \geq c_0 s \max\{\log^3(s) \log(N^{(s,p)}), \log(1/\gamma)\}. \tag{49}$$

Suppose that the sampling points t_ℓ, $\ell = 1, \ldots, m$, are drawn independently at random according to the orthogonalization measure for (ψ_j). Then with probability exceeding $1-\gamma$ the following holds for all f with $\|\!| f |\!\|_{v,p} < \infty$.

Let $y_\ell = f(t_\ell)$, $\ell = 1, \ldots, m$, and $\boldsymbol{A}$ be the $m \times N^{s,p}$ sampling matrix with entries $A_{j,\ell} = \psi_j(t_\ell)$, $j \in \Lambda_0$. For $\tau \geq 1$, let $\boldsymbol{x}^\sharp$ be the solution to

$$\min \|\boldsymbol{z}\|_{\omega,1} \text{ subject to } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \tau s^{1/2-1/p}\sqrt{m}\, \|\!| f |\!\|_{v,p} \tag{50}$$

and set $f^\sharp = \sum_{j\in\Lambda_0} x_j^\sharp \psi_j$. Then $\|f - f^\sharp\|_\infty \leq C_\tau s^{1-1/p} \|\!| f |\!\|_{v,p}$.

Proof. Consider $f = \sum_{j\in\Lambda} x_j\psi_j$, and associated $f_0 = \sum_{j\in\Lambda_0^{(s,p)}} x_j\psi_j$. We have

$$\|f - f^\sharp\|_{L_\infty} \leq \|f - f_0\|_{L_\infty} + \|f_0 - f^\sharp\|_{L_\infty}.$$

Since with probability at least $1-\gamma$ under the stated assumption on m, the matrix $\boldsymbol{A}$ has the $\boldsymbol{\omega}$-RIP and thereby the weighted null space property of order s, we have

$$\|f_0 - f^\sharp\|_{L_\infty} \leq C\tau s^{1-1/p} \|\!| f |\!\|_{v,p}$$

by the observations preceding the statement of the theorem. Furthermore,

$$\|f - f_0\|_\infty \leq s^{1/2-1/p} \|\!| f |\!\|_{v,p} \leq s^{1-1/p} \|\!| f |\!\|_{v,p}$$

by (45), Lemma 6.3, and the definition of $\Lambda_0^{(s,p)}$. This concludes the proof with $C_\tau = C\tau + 1$. □

Acknowledgments

Holger Rauhut acknowledges funding by the European Research Council through the grant StG 258926 and support by the Hausdorff Center for Mathematics at the University of Bonn. Rachel Ward was supported in part by an Alfred P. Sloan Research Fellowship, a Donald D. Harrington Faculty Fellowship, ONR Grant N00014-12-1-0743, NSF CAREER Award, and an AFOSR Young Investigator Program Award. Both Holger Rauhut and Rachel Ward would like to thank the Institute for Mathematics and Its Applications for its hospitality during a stay where this work was initiated.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr Approx*, 28:253–263, 2008.

[2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inf. Theor.*, 56(4):1982–2001, Apr. 2010.

[3] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Info. Theory*, 55(4):18721882, 2009.

[4] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R., Math., Acad. Sci. Paris*, 334(6):495–500, 2002.

[5] N. Burq, S. Dyatlov, R. Ward, and M. Zworski. Weighted eigenfunction estimates with applications to compressed sensing. *SIAM J. Math. Anal.*, 44(5):34813501, 2012.

[6] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:42034215, 2005.

[7] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[8] B. Carl. Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. *Ann. Inst. Fourier (Grenoble)*, 35(3):79–118, 1985.

[9] M. Cheraghchi, V. Guruswami, and A. Velingker. Restricted isometry of Fourier matrices and list decodability of random linear codes. *SODA*, pages 432–442, 2013.

[10] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *JAMS*, 22(1):211–231, 2009.

[11] A. Cohen, M. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Preprint*, 2011.

[12] A. Cohen, R. DeVore, S. Foucart, and H. Rauhut. Recovery of functions of many variables via compressive sensing. *In Proc. SampTA 2011, Singapore*, 2011.

[13] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best n-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.

[14] A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl.*, 9(1):11–47, 2011.

[15] L. Fagiano and M. Khammash. Simulation of stochastic systems via polynomial chaos expansions and convex optimization. *Arxiv*, 2012.

[16] G. Folland. *Fourier Analysis and its Applications.* American Mathematical Soc., 1992.

[17] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[18] M. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inf. Theor.*, 58(2):1122–1134, 2012.

[19] L. Jacques. A short note on compressed sensing with partially known signal support. *Signal Processing*, 90:3308–3312, 2010.

[20] A. Khajehnejad, W. Xu, A. Salman, and B. Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior information. In *ISIT 2009*, pages 483–487. IEEE, 2009.

[21] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models. *preprint*, 2010.

[22] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, to appear.

[23] I. Krasikov. An upper bound on Jacobi polynomials. *JAT*, 149:116–130, 2007.

[24] S. Kunis and D. Potts. Stability results for scattered data interpolation by trigonometric polynomials. *SIAM Journal on Scientific Computing*, 29(4):1403–1419, 2007.

[25] O. Le Maitre and O. Knio. *Spectral Methods for Uncertainty Quantification.* Springer, 2010.

[26] M. Ledoux and M. Talagrand. *Probability in Banach Spaces.* Springer-Verlag, 1991.

[27] M. Lifshits. *Lectures on Gaussian Processes.* Springer Briefs in Mathematics. Springer, 2012.

[28] S. Misra and P. Parrilo. Analysis of weighted ℓ_1-minimization for model-based compressed sensing. *Arxiv*, 2013.

[29] F. J. Narcowich, R. Schaback, and J. D. Ward. Approximation in Sobolev spaces by kernel expansions. *J. Approx. Theory*, 114:70–83, 2002.

[30] H. Rauhut. Compressive Sensing and Structured Random Matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.

[31] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *J. Approx. Theory*, 164(5):517–533, 2010.

[32] H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. *In Proc. SampTA*, 2011.

[33] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[34] C. Runge. Über empirische Funktionen und die interpolation zwischen äquidistanten Ordinaten. *Z Math Physik*, 46:224–243, 1901.

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

[36] N. Vaswani and W. Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE International Symposium on Information Theory, ISIT*, pages 488–492, 2009.

[37] R. von Borries, C. Miosso, and C. Potes. Compressed sensing using prior information. *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP*, 12-14:121 – 124, 2007.

[38] W. Xu. Compressive sensing for sparse approximations: Construction, algorithms and analysis. *Ph.D. dissertation, California Institute of Technology, Department of Electrical Engineering*, 2010.

[39] X. Yu and S. Baek. Sufficient conditions on stable recovery of sparse signals with partial support information. *IEEE Signal Processing Letters*, 20(5), 2013.