Interpolation via weighted $\ell_1$ minimization

Holger Rauhut; Rachel Ward †

November 11, 2014

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 $\ell_1$ minimization can reconstruct sparse functions from very few samples, but do not necessarily encourage smoothness. Here we show that weighted $\ell_1$ minimization effectively merges the two approaches, promoting both sparsity and smoothness in reconstruction. More precisely, we provide specific choices of weights in the $\ell_1$ objective to achieve approximation rates for functions with coefficient sequences in weighted $\ell_p$ spaces with $p \leq 1$. We consider 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, $\ell_1$ 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 values, we wish to find a suitably well-behaved function agreeing with the data up to some error. Classically, “well-behaved” 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, 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

---

*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

© 2015. This manuscript version is made available under the Elsevier user license

http://www.elsevier.com/open-access/userlicense/1.0/
are principally governed by main effects and low order interactions. In such situations, we might hope to synthesize smoothness and sparsity-based approaches.

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 structured sparse expansions are imposed by weighted $\ell_p$ coefficient spaces in the range $0 < p \leq 1$. Accordingly, we will use weighted $\ell_1$ minimization, a convex surrogate for weighted $\ell_p$ minimization with $p < 1$, as our reconstruction method of choice.

The contributions of this paper are as follows.

1. **Function approximation.** We provide the first rigorous analysis for function interpolation using weighted $\ell_1$ minimization. We show that with the appropriate choice of weights, one obtains approximation rates for functions with coefficient sequences lying in weighted $\ell_p$ 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. The number of sampling points required by weighted $\ell_1$ minimization to achieve a desired rate grows only linearly with the ambient dimension, rather than exponentially. We illustrate the improvement of weighted $\ell_1$ minimization compared to unweighted $\ell_1$ minimization on several specific examples, including spherical harmonic interpolation and tensorized Chebyshev and Legendre polynomial interpolation. We expect that the results for polynomial interpolation should have applications in uncertainty quantification [30] and, in particular, in the computation of generalized polynomial chaos expansions [16, 17].

2. **Weighted sparsity.** In order to derive recovery guarantees for weighted $\ell_1$ 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.

3. **Structured random matrices for compressive sensing.** It is by now well established [9, 40, 36] that an $s$-sparse trigonometric polynomial of maximal degree $N$ can be recovered efficiently from its values at $m \asymp s \log^4(N)$ sampling points drawn independently from the uniform measure on their domain. More generally, such near-optimal sparse recovery results hold whenever the underlying sparsity basis is uniformly bounded like the trigonometric system, so as to be incoherent with point samples [36]. However, these results become weak if the $L_\infty$ norms of the functions in the orthonormal system grow with the maximal degree $N$. If this growth is not too sharp, one may apply a preconditioning technique proposed in [38] to transform the system into a uniformly bounded one. However, often these preconditioning techniques are not enough to make the basis functions uniformly bounded. Here, we show that as long as lower-degree polynomials are preferred in the sparse expansions under consideration, one may still derive sparse recovery guarantees using weighted $\ell_1$ minimization with weights which grow at least as quickly as the $L_\infty$ norms of the functions they correspond to.

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 $\ell_p$ 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 $\ell_1$ 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 a numerical illustration.

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 isometry property. In Section 6 we finish by presenting our main results on interpolation via weighted $\ell_1$ minimization.

1.2 Weighted sparsity

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

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

Also central to our analysis will be the weighted $\ell_0$-"norm",

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

which counts the squared weights of the non-zero entries of $x$. 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 $\omega = 1$, these weighted norms reproduce the standard $\ell_p$ norms, in which case we use the standard $\ell_p$-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 instance, the exponent must scale this way in order to apply a Cauchy-Schwarz inequality to weighted $s$-sparse vectors:

$$\|x\|_{\omega,1} = \sum_{j \in \Lambda} \omega_j |x_j| \leq \sqrt{\sum_{j \in \Lambda} \omega_j^2} \sqrt{\sum_{j \in \Lambda} |x_j|^2} = \omega(A)\|x\|_2 \leq \sqrt{s}\|x\|_2, \quad \forall x : \|x\|_{\omega,0} \leq s$$

Consequently, error bounds we present will be in terms of the $\ell_{\omega,p}$ norms scaled as such; for $p = 1$ we have weighted $\ell_1$ error bounds, for $p = 2$ we have unweighted $\ell_2$ error bounds.

For $x \in \ell_{\omega,p}$ and for a subset $S$ of the index set $\Lambda$, we define $x_S \in \ell_{\omega,p}$ as the restriction of $x$ to $S$. For $s \geq 1$, the error of best weighted $s$-term approximation of the vector $x \in \ell_{\omega,p}$ is defined as

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

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 $\sigma_s(x)_{\omega,p}$ using a quantity that can easily computed from $x$ by sorting and thresholding.
1.3 Weighted $\ell_p$ spaces for smoothness and sparsity

The weighted $\ell_p$ coefficient spaces introduced in the previous section can be used to define weighted function spaces. For a bounded domain $D$, let $\psi_j : D \to \mathbb{C}, j \in \Lambda$, be a sequence of functions indexed by the set $\Lambda$ which are orthonormal with respect to the probability measure $\nu$ on $D$, that is,

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

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

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

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

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

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

$$\sigma_s(f)_{\omega,p} = \sigma_s(x)_{\omega,p}.$$ (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:

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

This estimate illustrates how a small $\ell_p$–norm for $p < 1$ supports small sparse approximation error. Conditions of the form $\| \omega \|_\infty^2 < s$ 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. replace a countably-infinite set $\Lambda$ 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 $\ell_1$ minimization

In treating the interpolation problem, we first assume that the index set $\Lambda$ is finite with $N = |\Lambda|$. Given sampling points $t_1, \ldots, t_m \in D$ and $f = \sum_{j \in \Lambda} x_j \psi_j$ we can write the vector of sample values $y = (f(t_\ell))_{\ell=1,\ldots,m}$ succinctly in matrix form as $y = Ax$, where $A$ 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 $A$. In our theorems, the sampling points are drawn independently from the orthogonalization measure $\nu$ associated to the orthonormal system $(\psi_j)$; as a consequence, the random matrix $AA^*$, 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 \( f^z(t) = \sum_{j \in \Lambda} x_j^z \psi_j(t) \) whose coefficient vector \( z \) is the solution of the weighted \( \ell_1 \) minimization program

\[
\min \| z \|_{\omega, 1} \text{ subject to } Az = y
\]

The equality constraint in the \( \ell_1 \) minimization ensures that the function \( f^z \) interpolates \( f \) at the points \( t_\ell \), that is, \( f^z(t_\ell) = f(t_\ell) \), \( \ell = 1, \ldots, m \). Let us give the following result on interpolation via weighted \( \ell_1 \) minimization with respect to \( \| . \|_{\omega, 1} \).

**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 \( s \geq 2\| \omega \|_{\infty}^2 \), fix a number of samples

\[
m \geq c_0 s \log^3(s) \log(N),
\]

and suppose that \( t_\ell \), \( \ell = 1, \ldots, m \), 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 \( z \) be the solution of

\[
\min \| z \|_{\omega, 1} \text{ subject to } Az = y
\]

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

\[
\| f - f^z \|_{L_\infty} \leq \| f - f^z \|_{\omega, 1} \leq c_1 \sigma_s(f, \omega, 1),
\]

\[
\| f - f^z \|_{L_2} \leq d_1 \sigma_s(f, \omega, 1) / \sqrt{s}.
\]

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 \( L_\infty \) norm by the weighted \( \ell_1 \) coefficient norm: for a function \( f \in S_{\omega, p} \),

\[
\| f \|_{L_\infty} = \sup_{t \in D} \left| \sum_{n=-\infty}^{\infty} x_n \psi_n(t) \right| \leq \sup_{t \in D} \sum_{n=-\infty}^{\infty} |x_n| \| \psi_n(t) \| \leq \sum_{n=-\infty}^{\infty} |x_n| \omega_n = \| f \|,
\]

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 \( L_\infty \) norm, then by the Stechkin-type estimate (5) with \( q = 1 \) we have

\[
\| f - f_0 \|_{L_\infty} \leq \| f - f_0 \|_{1} \leq (s - \| \omega \|_{\infty}^2)^{1/p} \| f \|_{\omega, p}, \quad p < 1.
\]

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 be interesting for future research.

In Section 6 we will prove a more general version of Theorem 1.1 showing robustness of weighted \( \ell_1 \) minimization to noisy samples. Using this robustness to noise, we will be able to treat the case where the index set \( \Lambda \) 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 \(t_\ell, \ell = 1, \ldots , m\), independently from the orthogonalization measure associated to \((\psi_j)_{j \in \Lambda}\). Let \(A \in \mathbb{C}^{m \times N}\) be the sampling matrix with entries \(A_{t_\ell, j} = \psi_j(t_\ell)\). 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 \(x^s\) be the solution of
\[
\min \|z\|_{\omega, 1} \text{ subject to } \|Az - y\|_2 \leq \left(\frac{m}{s}\right)^{1/2} \eta
\]
and set \(f^s(t) = \sum_{j \in \Lambda_0} x_j^s \psi_j(t)\). Then with probability exceeding \(1 - N^{-\log^3(s)}\),
\[
\|f - f^s\|_{L_{\infty}} \leq \|f - f^s\|_{\omega, 1} \leq c_1 \sigma_s(f)_{\omega, 1},
\]
\[
\|f - f^s\|_{L^2} \leq d_1 \sigma_s(f)_{\omega, 1}/\sqrt{s}.
\]
Above, \(c_0\) is an absolute constant and \(c_1, d_1\) are constants which depend only on the distortion \(\varepsilon\).

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 is possible to avoid this using greedy or iterative methods; see [23] for one such method, a weighted version of the iterative hard thresholding algorithm [3]. 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 \(\Lambda_0\) is polynomial in \(s\), then the number of samples reduces to \(m \geq C s \log^4(s)\) 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 (\(\omega\)-RIP) for a linear map \(A : \mathbb{C}^N \to \mathbb{C}^m\), which generalizes the concept of restricted isometry property in compressive sensing.

Definition 1.3 (\(\omega\)-RIP constants). For \(A \in \mathbb{C}^{m \times N}, s \geq 1,\) and weight \(\omega,\) the \(\omega\)-RIP constant \(\delta_{\omega, s}\) associated to \(A\) is the smallest number for which
\[
(1 - \delta_{\omega, s}) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_{\omega, s}) \|x\|_2^2
\]
for all \(x \in \mathbb{C}^N\) with \(\|x\|_{\omega, 0} = \sum_{j \in \text{supp}(x)} \omega_j \leq s\).

For weights \(\omega \equiv 1\), the \(\omega\)-RIP reduces to the standard RIP, as introduced in [9, 8]. For general weights \(\omega_j \geq 1\), the \(\omega\)-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. Consider weights of the general form \(\omega_j = j^{\alpha/2}\) with \(\alpha > 0\). We may then take \(N = s^{1/\alpha}\), as even single indices \(j > N\) have weighted cardinality exceeding \(s\). Observe that if \(\|x\|_{\omega, 0} \leq s\), then \(x\) is supported on an index set of (unweighted) cardinality at most \(s^{1/\alpha}s^{1/(\alpha + 1)}\). Following the approach of [1], see also [2, 4], taking a union bound and applying
covering arguments, one may argue that an $m \times N$ i.i.d. subgaussian random matrix has the $\omega$-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 $m = \mathcal{O}(s \log(N/s))$ 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 $\ell_1$ minimization

Weighted $\ell_1$ minimization has been analyzed previously in the compressive sensing literature. Weighted $\ell_1$ minimization with weights $\omega_j \in \{0, 1\}$ was introduced independently in the papers [24, 45, 44] and extended further in [22]. The paper [21] seems to be the first to provide conditions under which weighted $\ell_1$ minimization is stable and robust under weaker sufficient conditions than the analogous conditions for standard $\ell_1$ minimization for general weights. Improved sufficient conditions were recently provided for this setting in [47]. Recovery guarantees were also provided in [35], also with a focus on function interpolation with applications to polynomial chaos expansions. Still, the analysis in all of these works is based on the standard restricted isometry property and this limits the extent to which their recovery guarantees improve on those for unweighted $\ell_1$ minimization.

Weighted $\ell_1$ minimization has also been considered under probabilistic models. In [46], the vector indices are partitioned into two sets, and indices on each set have different probabilities $p_1, p_2$ of being nonzero. The papers [24, 25] 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 [33] 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 $\ell_p$.

2 Examples

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 $\ell_p$ spaces using Theorem 6.4.

2.1 Spherical harmonic interpolation

The spherical harmonics $Y^k_\ell$ form an orthonormal system for square-integrable functions on the sphere $S^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$, 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
\[
\int_0^{2\pi} \int_0^{\pi} Y^k_\ell(\varphi, \theta) Y^l_{\ell'}(\varphi, \theta) \sin(\theta) d\theta d\varphi = \delta_{\ell\ell'} \delta_{kk'}, \quad k, \ell \in \mathbb{Z}, \quad |k| \leq \ell.
\] (8)

The spherical harmonics are bounded according to \(\|Y^k_\ell\|_{L^\infty} \leq \ell^{1/2}\), and this bound is realized at the poles of the sphere, \(\theta = 0, \pi\). As shown in [39, 6], 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
\[
Z^k_\ell(\varphi, \theta) = \sin(\theta)^{1/2} Y^k_\ell(\varphi, \theta),
\]
normalized by the proper constant, is orthonormal on the sphere with respect to the measure \(d\mu = d\theta d\varphi\) by virtue of (8). The \(Z^k_\ell\) are more uniformly bounded than the spherical harmonics \(Y^k_\ell\); as noted in [27],
\[
\|Z^k_\ell\|_{L^\infty} \leq C\ell^{1/4}
\]
for a universal constant \(C\). A sharper preconditioning estimate was shown in [6] for the system
\[
\tilde{Z}^k_\ell(\varphi, \theta) := (\sin^2(\theta) \cos(\theta))^{1/6} Y^k_\ell(\varphi, \theta).
\] (9)
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 1. This system obeys the uniform bound
\[
\|\tilde{Z}^k_\ell\|_{L^\infty} \leq C\ell^{1/6},
\] (10)
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 \(\tilde{Z}^k_\ell, |k| \leq \ell\), 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 \(f(\varphi, \theta) = \sum_{\ell,k} x_{\ell,k} \tilde{Z}^k_\ell(\varphi, \theta) \in S_{\omega,1}\) and let \(\eta = \|f - f_{\Lambda_0}\|_{\omega,1}\). Draw \((\varphi_j, \theta_j), j = 1, \ldots, m, i.i.d.\) from \(d\nu\), and consider sample values \(y_j = f(\varphi_j, \theta_j), j = 1, \ldots, m\). Then with probability exceeding \(1 - N^{-\log^2(s)}\), the function \(f^* = \sum_{(\ell,k) \in \Lambda_0} x^*_{\ell,k} \tilde{Z}^k_\ell\) formed from the solution \(x^*\) of the weighted \(\ell_1\) minimization program
\[
\min_{u_{\ell,k}} \sum_{\ell,k \in \Lambda_0} \omega_{\ell,k} |u_{\ell,k}| \text{ subject to } \sum_{j=1}^m \left( \sum_{\ell,k \in \Lambda_0} u_{\ell,k} \tilde{Z}^k_\ell(\varphi_j, \theta_j) - y_j \right)^2 \leq \frac{mn^2}{s},
\]
satisfies the error bounds
\[
\|f - f^*\|_{L^\infty} \leq \|f - f^*\|_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1},
\]
\[
\|f - f^*\|_{L^2} \leq c_2 \sigma_s(f)_{\omega,1}/\sqrt{s}.
\]
It is informative to compare these results with previously available bounds for unweighted \( \ell_1 \) minimization. Using the same sampling distribution \( dv = |\tan(\theta)|^{1/3}d\theta d\varphi \) and number of basis elements \( N = s^6 \), existing bounds for unweighted \( \ell_1 \) minimization (see [6]) 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 (see [39] for more details). That is, significantly more measurements \( m \) are required to achieve the . (A rate for \( L_{\infty} \) is not available for unweighted \( \ell_1 \)-minimization). However, stronger assumptions on \( f \) are required in the sense that the result above requires the \textit{weighted} best \( s \)-term approximation error to be small while the bound from [39] 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 \( D = \mathbb{T}^d \) are given by

\[
\psi_k(t) = \psi_{k_1}(t_1)\psi_{k_2}(t_2) \ldots \psi_{k_d}(t_d), \quad k \in \mathbb{Z}^d,
\]

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 \( \ell_1 \) minimization with properly chosen weights.

2.2.1 Chebyshev polynomials

Consider the tensorized Chebyshev polynomials on \( D = [-1,1]^d \):

\[
C_k(t) = C_{k_1}(t_1)C_{k_2}(t_2) \ldots C_{k_d}(t_d), \quad k \in \mathbb{N}^d,
\]

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

\[
d\mu = \frac{dt}{(2\pi)^d \prod_{j=1}^d (1 - t_j^2)^{1/2}}.
\]

This motivates us to apply Theorem 1.2 with weights

\[
\omega_k = \prod_{\ell=1}^d (k_\ell + 1)^{1/2},
\]

noting that \( \|C_k\|_\infty \leq \omega_k \). (More generally, one could also work with weights of the form \( \omega_k = 2^{|k|_1} v_k \), where \( v_k \) tends to infinity as \( |k|_2 \to \infty \).) Such weights encourage both sparse and low order tensor products of Chebyshev polynomials. The subset of indices

\[
H^d_s = \{ k \in \mathbb{N}_0^d : \omega_k^2 \leq s \} = \left\{ k \in \mathbb{N}_0^d : \prod_{\ell=1}^d (k_\ell + 1) \leq s \right\}
\]

forms a hyperbolic cross. As argued in [28, Proof of Theorem 4.9], see also [15], the size of a hyperbolic cross can be bounded according to

\[
|H^d_s| \leq e^{2s^{2+\log_2(d)}}.
\]
Corollary 2.2. Consider the tensorized Chebyshev polynomial basis \((C_k^s)\) for \([-1,1]^d\), and weights \(\omega_k^s\) as in (13). Let \(\Lambda_0 = \{k \in \mathbb{N}_0^d : \omega_k^2 \leq s/2\}\), and let \(N = |\Lambda_0| \leq c^2(s/2)^{2+\log_2(d)}\). Fix a number of samples

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

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

\[
\min \|z\|_{\omega,1} \text{ subject to } \| Az - y \|_2 \leq \sqrt{m/s} \| f - f_{ \Lambda_0 } \|_{\omega,1}
\]

and set \(f^d(t) = \sum_{k \in \Lambda_0} x_k^d C_k(t)\). Then with probability exceeding \(1 - N^{-\log^3(s)}\),

\[
\| f - f^d \|_\infty \leq \| f - f^d \|_{\omega,1} \leq c_1 \sigma_3(f)_{\omega,1},
\]

\[
\| f - f^d \|_2 \leq d_1 \sigma_3(f)_{\omega,1}/\sqrt{s}.
\]

Above, \(c_0, c_1\), and \(d_1\) are universal constants.

Note that with the stated estimate of \(N, m\) satisfies (14) 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 \(D = [-1,1]^d\):

\[
L_k(t) = L_{k_1}(t_1) L_{k_2}(t_2) \ldots L_{k_d}(t_d), \quad k \in \mathbb{N}^d,
\]

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

\[
\| L_k \|_\infty \leq \prod_{\ell=1}^d (k_\ell + 1)^{1/2},
\]

and we may apply Theorem 1.2 with hyperbolic cross weights \(\omega_k = \prod_{\ell=1}^d (k_\ell + 1)^{1/2}\) 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 \(\omega_k\) as in (13) and with \(\Lambda_0, N,\) and \(m\) as in Corollary 2.2. Consider a function \(f = \sum_{k \in \Lambda} x_k L_k\), and suppose that \(t_\ell, \ell = 1, \ldots, m,\) are drawn i.i.d. from the tensorized uniform measure on \([-1,1]^d\). Let \(A \in \mathbb{C}^{m \times N}\) be the associated sampling matrix with entries \(A_{\ell,j} = \psi_j(t_\ell)\). From samples \(y_\ell = f(t_\ell), \ell = 1, \ldots, m,\) let \(x^d\) be the solution of

\[
\min \|z\|_{\omega,1} \text{ subject to } \| Az - y \|_2 \leq \sqrt{m/s} \| f - f_{ \Lambda_0 } \|_{\omega,1}
\]

and set \(f^d(t) = \sum_{k \in \Lambda_0} x_k^d L_k(t)\). Then with probability exceeding \(1 - N^{-\log^3(s)}\),

\[
\| f - f^d \|_\infty \leq \| f - f^d \|_{\omega,1} \leq c_1 \sigma_3(f)_{\omega,1},
\]

\[
\| f - f^d \|_2 \leq d_1 \sigma_3(f)_{\omega,1}/\sqrt{s}.
\]

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
\[ v(t) = \left(\frac{\pi}{2}\right)^{1/2}(1-t^2)^{1/4}, \quad t \in [-1, 1], \]
and recalling Theorem 7.3.3 from [42] which states that, for all $j \geq 1$,
\[
\sup_{t \in [-1,1]} v(t) |L_j(t)| \leq \sqrt{2 + 1/j} \leq \sqrt{3}. \tag{18}
\]
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 [38].

### 2.3 Numerical illustrations

Although the interplay between sparsity and smoothness in function approximation becomes more pronounced in high-dimensional problems, we still observe benefits of weighted $\ell_1$ minimization in univariate polynomial interpolation problems. In this section we provide an illustration of this effect.

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 [19]). One method for minimizing the effect of Runge’s phenomenon is to carefully choose interpolation nodes – e.g., Chebyshev nodes for algebraic polynomial interpolation or equispaced nodes for trigonometric interpolation. Other 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 and which has been shown to be stable with respect to random sampling [14], or by doing weighted $\ell_2$ regularization [29], e.g. use for interpolation the function $f^2(t) = \sum_{j \in A} x_j^2 \psi_j(t)$, where the coefficient vector $x^2$ solves the minimization problem
\[
\min \sum_{j \in A} \alpha_j^2 z_j^2 \text{ subject to } Az = y \tag{19}
\]
as in (1.4).

In this section we provide evidence that weighted $\ell_1$ minimization can be significantly more robust to perturbations in the choice of the sampling points than these other methods, specifically unweighted $\ell_1$ minimization, least squares projections, weighted $\ell_2$ regularization, and exact polynomial interpolation.

For our numerical experiments, we follow the examples in [14] and consider on $[-1, 1]$ the smooth function
\[ f(t) = \frac{1}{1 + 25t^2}, \]
which was originally considered by Runge [41] to illustrate the instability of polynomial interpolation at equispaced points. We repeat the following experiment 100 times: draw $m = 30$ sampling points $x_1, x_2, \ldots, x_m$, i.i.d. from a measure $\mu$ on $D = [-1, 1]$ 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 $\ell_1$ approximation, weighted $\ell_2$ approximation
with weights $\alpha_j = j$ in (19), and weighted $\ell_1$ approximations with weights $\omega_j = j$ and weights $\omega_j = j^{1/2}$. We also compare to exact inversion, where we fit the sampling points with a polynomial whose maximal degree -1 (degrees of freedom) matches the number of sampling points. In Figures 2-3 we display the interpolations resulting from all 100 experiments overlaid so as to illustrate the variance of each interpolation method with respect to the choice of sampling points. In all experiments, we fix in the $\ell_1$ and $\ell_2$ minimization a maximal polynomial degree $N = 100$. For the least squares solution to be stable [14], we project onto the span of the first $d = 15$ basis elements.

In this example, we observe that the weighted $\ell_1$ solutions are more consistently accurate than the other methods, including exact polynomial interpolation. There is mild sparsity present in this example; Runge’s function is an even function and so all odd coefficients in its Legendre / trigonometric expansion are zero. Indeed, the weighted $\ell_1$ solutions tend to pick up this sparsity and have zero odd coefficients, unlike the other interpolation methods do not pick up on this sparsity.

We emphasize once more that this is just an illustration. For univariate polynomial interpolation in the setting where one can design carefully-chosen sampling points, e.g., Chebyshev nodes, weighted $\ell_1$ minimization should not outperform exact polynomial interpolation.

![Interpolations](image.png)

**Figure 2:** Overlaid interpolations of the function $f(t) = \frac{1}{1 + 25t^2}$, $t \in [-1, 1]$, by real trigonometric polynomials using various reconstruction methods as described in Section 2.3. Different interpolations correspond to different random draws of $m = 30$ sampling points from the uniform measure on $[-1, 1]$.

### 3 Weighted sparsity and quasi-best $s$-term approximations

In this section we revisit some important technical results pertaining to weighted $\ell_p$ spaces that were touched upon in the introduction. First, unlike unweighted $s$-term approximations
to different random draws of various reconstruction methods as described in Section 2.3. Different interpolations correspond to different random draws of $m = 30$ sampling points according the Chebyshev measure on $[-1, 1]$.

for finite vectors, the weighted $s$-term approximations $\sigma_s(x)_{\omega, p} = \inf_{z : \|z\|_{\omega, 0} \leq s} \|x - z\|_{\omega, p}$ are not straightforward to compute in general. Nevertheless, we can approximate $\sigma_s(x)_{\omega, p}$ using a quantity that can easily be computed from $x$ by sorting and thresholding, which we will call the quasi-best $s$-term approximation.

Let $v$ 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 $\pi$ 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 $x_S$ a weighted quasi-best $s$-term approximation to $x$ and define the corresponding error of weighted quasi-best $s$-term approximation as

$$\tilde{\sigma}_s(x)_{\omega, p} = \|x - x_S\|_{\omega, p} = \|x_{\tilde{S}}\|_{\omega, p}.$$  

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

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

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

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

$$\|x_S\|_{\omega, p} \leq \|x_\tilde{S}\|_{\omega, p}.$$

Assume without loss of generality that the terms in $x$ 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

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

Figure 3: Overlaid interpolations of the function $f(t) = \frac{1}{1+t^2}$ by Legendre polynomials using various reconstruction methods as described in Section 2.3. Different interpolations correspond to different random draws of $m = 30$ sampling points according the Chebyshev measure on $[-1, 1]$.  

(a) Least squares  
(b) Weighted $\ell_2$, $\omega_j = j^{1/2}$  
(c) Exact inversion  

(d) Unweighted $\ell_1$  
(e) Weighted $\ell_1$, $\omega_j = j^{1/2}$  
(f) Weighted $\ell_1$, $\omega_j = j$
Now, let

\[
(\sum_{n_1} |x_1|^p \omega_1^{-p}, \ldots, |x_1|^p \omega_1^{-p}, (1 - r_1) |x_1|^p \omega_1^{-p}, \sum_{n_2} |x_2|^p \omega_2^{-p}, \ldots, |x_2|^p \omega_2^{-p}, (1 - r_2) |x_2|^p \omega_2^{-p}, \ldots).
\]

We constructed \( z \) so that the first \( n_1 \) terms in \( z \) sum to \( |x_1|^p \omega_1^{-2p} \), the next \( n_2 \) terms sum to \( |x_2|^p \omega_2^{-2p} \), and so on. Then This completes the proof.

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) an appropriate weighted vector norm.

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

\[
\sigma_s(x)_{\omega,q} \leq \tilde{\sigma}_s(x)_{\omega,q} \leq \left( s - \|\omega\|_\infty^2 \right)^{1/p - 1/q} \|x\|_{\omega,p}.
\]

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

\[
s - \|\omega\|_\infty^2 \leq s - \omega^2_{\pi(k)} < \omega(S) \leq s.
\]

Then

\[
\tilde{\sigma}_s(x)_{\omega,p} \leq \sum_{j \in S} |x_j|^p \omega_j^{-p} \leq \max_{j \in S} \{ |x_j|^p \omega_j^{-p} \} \sum_{j \notin S} |x_j|^q \omega_j^{-q} \leq \left( \max_{j \notin S} |x_j| \omega_j^{-1} \right)^{p-q} \|x\|_{q,q}.
\]

Now let \( \alpha_k := (\sum_{j \in S} \omega_j^2)^{-1} \omega_k^2 \leq (s - \|\omega\|_\infty^2)^{-1} \omega_k^2 \). 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

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

Combining the above estimates yields

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

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 for unweighted \( \ell_1 \) minimization, one can derive reconstruction guarantees for weighted \( \ell_1 \) minimization via appropriate weighted versions of the null space property and restricted isometry property \([13, 9]\). 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 $\omega$, a matrix $A \in \mathbb{C}^{m \times N}$ is said to satisfy the weighted robust null space property of order $s$ with constants $\rho \in (0, 1)$ and $\tau > 0$ if

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

The inequalities stated in the next theorem are crucial for deriving error bounds for recovery via weighted $\ell_1$ minimization.

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

$$\|z - x\|_{\omega, 1} \leq \frac{1 + \rho}{1 - \rho} (\|z\|_{\omega, 1} - \|x\|_{\omega, 1} + 2\sigma_s(x)_{\omega, 1}) + \frac{2\tau\sqrt{s}}{1 - \rho} \|Av(z - x)\|_2$$

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

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

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

$$\|x\|_{\omega, 1} + \|(x - z)_{S^c}\|_{\omega, 1} \leq \|x_{S^c}\|_{\omega, 1} + \|x_S\|_{\omega, 1} + \|x_{S^c}\|_{\omega, 1} + \|z_{S^c}\|_{\omega, 1}$$

$$\leq 2\|x_{S^c}\|_{\omega, 1} + \|(x - z)_S\|_{\omega, 1} + \|z_S\|_{\omega, 1} + \|z_{S^c}\|_{\omega, 1} = 2\sigma_s(x)_{\omega, 1} + \|(x - z)_S\|_{\omega, 1}. \quad (26)$$

Rearranging and setting $\nu := z - x$ leads to

$$\|v_{S^c}\|_{\omega, 1} \leq \|z\|_{\omega, 1} - \|x\|_{\omega, 1} + \|v_S\|_{\omega, 1} + 2\sigma_s(x)_{\omega, 1}. \quad (25)$$

The Cauchy-Schwarz inequality implies

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

and therefore by (22)

$$\|v_S\|_{\omega, 1} \leq \sqrt{s}\|v_S\|_2 \leq \rho \|v_{S^c}\|_{\omega, 1} + \tau \sqrt{s}\|Av\|_2. \quad (26)$$

We combine with (25) to arrive at

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

Using (26) once more finally gives

$$\|x - z\|_{\omega, 1} = \|v_S\|_{\omega, 1} + \|v_{S^c}\|_{\omega, 1} \leq (1 + \rho)\|v_{S^c}\|_{\omega, 1} + \tau \sqrt{s}\|Av\|_2$$

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

We pass to the proof of (24). Let \( S \) with \( \omega(S) \leq s \) be such that \( \|v - v_S\|_2 = \|v_{S^c}\|_2 = \tilde{\sigma}_s(v)_{\omega,2} \) (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

\[
\|x - z\|_2 \leq \|(x - z)_{S^c}\|_2 + \|(x - z)_S\|_2
\]

\[
\leq \frac{1}{\sqrt{s - \|\omega\|_\infty^2}} \|x - z\|_{\omega,1} + \frac{\rho}{\sqrt{s}} \|(x - z)_{S^c}\|_{\omega,1} + \tau \|A(x - z)\|_2
\]

\[
\leq \frac{1 + \rho}{\sqrt{s - \|\omega\|_\infty^2}} \|x - z\|_{\omega,1} + \tau \|A(x - z)\|_2
\]

\[
\leq \frac{2(1 + \rho)^2}{(1 - \rho)\sqrt{s - \|\omega\|_\infty^2}} (\|z\|_{\omega,1} - \|x\|_{\omega,1} + \sigma_s(x)_{\omega,1})
\]

\[
+ \left( \tau + \frac{2\tau(1 + \rho)\sqrt{s}}{(1 - \rho)\sqrt{s - \|\omega\|_\infty^2}} \right) \|A(x - z)\|_2.
\]

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

\[ \square \]

As a consequence of the previous result we obtain error bounds for sparse recovery via weighted \( \ell_1 \) minimization.

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

\[ \min_{z \in \mathbb{C}^N} \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \leq \eta. \]

Then the reconstruction error satisfies

\[
\|x - x^z\|_{\omega,1} \leq c_1 \sigma_s(x)_{\omega,1} + d_1 \sqrt{s}\eta \quad (27)
\]

\[
\|x - x^z\|_2 \leq \frac{\sigma_s(x)_{\omega,1}}{\sqrt{s}} + d_2 \eta, \quad (28)
\]

where the second bound additionally assumes \( s \geq 2\|\omega\|_\infty^2 \). The constants \( c_1, c_2, d_1, d_2 > 0 \) depend only on \( \rho \) and \( \tau \).

***Proof.*** The reconstruction errors follow from the error bounds in Theorem 4.2 with \( z = x^\# \), noting that and \( \|A(x - x^\#)\|_2 \leq \|Ax - y\|_2 + \|Ax^\# - y\|_2 \leq 2\eta \).

\[ \square \]

**Remark 4.4.** In the case of noiseless measurements, the previous result gives error bounds for equality-constrained weighted \( \ell_1 \) 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 [20] 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 (\( \omega \)-RIP) constant \( \delta_{\omega,s} \) associated to a matrix \( A \) as the smallest number such that

\[
(1 - \delta_{\omega,s})\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_{\omega,s})\|x\|_2^2 \quad \text{for all} \ x \ \text{with} \ \|x\|_{\omega,1} \leq s.
\]
We say that \( A \) satisfies a weighted restricted isometry property (\( \omega \)-RIP) if \( \delta_{\omega,s} \) is small for \( s \) relatively large compared to \( m \). The \( \omega \)-RIP implies the weighted robust null space property and therefore the error bounds (27) and (28) for recovery via weighted \( \ell_1 \) minimization as shown in the following result.

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

\[
\delta_{\omega,3s} < 1/3
\]

for \( s \geq 2\|\omega\|_\infty^2 \). Then \( A \) 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. [20, 36]) the \( \omega \)-RIP constants can be rewritten as

\[
\delta_{\omega,s} = \max_{S \subseteq \{1, \ldots, N\}, \omega(S) \leq s} \| A_S^* A_S - \text{Id} \|_{2 \to 2},
\]

where \( A_S \) denotes the submatrix of \( A \) restricted to the columns indexed by \( S \).

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

\[
\langle (Au, Av) \rangle \leq \delta_{\omega,s+t} \|u\|_2 \|v\|_2.
\]

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

\[
\langle (Au, Av) \rangle = \langle (A_S u_S, A_S v_S) - (u_S, v_S) \rangle = \langle (A_S^* A_S - \text{Id}) u_S, v_S \rangle \leq \|A_S^* A_S - \text{Id}\|_{2 \to 2} \|u_S\|_2 \|v_S\|_2 \leq \delta_{\omega,s+t} \|u\|_2 \|v\|_2.
\]

This completes the proof. \( \square \)

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

**Proof of Theorem 4.5.** Let \( v \in \mathbb{C}^N \) and \( S \subseteq \{1, \ldots, N\} \) with \( \omega(S) \leq s \). We partition \( S^c \) into blocks \( S_1, S_2, \ldots \) with \( s - \|\omega\|_\infty^2 \leq \omega(S_i) \leq s \) according to the nonincreasing rearrangement of \( v_{S^c} \cdot \omega_{S^c}^{-1} \), that is, \( |v_j| \omega_j^{-1} \leq |v_k| \omega_k^{-1} \) for all \( j \in S_i \) and all \( k \in S_{i-1}, \ell \geq 2 \). Then we estimate

\[
\|v_S + v_{S_1}\|_2^2 \leq \frac{1}{1 - \delta_{\omega,2s}} \|A(v_S + v_{S_1})\|_2^2 = \frac{1}{1 - \delta_{\omega,2s}} \left\langle A(v_S + v_{S_1}), Av - \sum_{\ell \geq 2} Av_{S_\ell} \right\rangle
\]

\[
= \frac{1}{1 - \delta_{\omega,2s}} \left\langle A(v_S + v_{S_1}), Av \right\rangle - \sum_{\ell \geq 2} \left\langle A(v_S + v_{S_1}), Av_{S_\ell} \right\rangle
\]

\[
\leq \frac{1}{1 - \delta_{\omega,2s}} \left( \sqrt{1 + \delta_{\omega,2s}} \|v_S + v_{S_1}\|_2 \|Av\|_2 + \delta_{\omega,3s} \|v_S + v_{S_1}\|_2 \sum_{\ell \geq 2} \|v_{S_\ell}\|_2 \right),
\]

where we have used Lemma 4.6 in the third line. Dividing by \( \|v_S + v_{S_1}\|_2 \) and using the fact that \( \delta_{\omega,2s} \leq \delta_{\omega,3s} \) we arrive at

\[
\|v_S\|_2 \leq \|v_S + v_{S_1}\|_2 \leq \frac{\delta_{\omega,3s}}{1 - \delta_{\omega,3s}} \sum_{\ell \geq 2} \|v_{S_\ell}\|_2 + \frac{\sqrt{1 + \delta_{\omega,3s}}}{1 - \delta_{\omega,3s}} \|Av\|_2.
\]

Now for \( k \in S_\ell \), set \( \alpha_k = (\sum_{j \in S_\ell} \omega_j^2)^{-1} \omega_k^2 = (s - \|\omega\|_\infty^2) \omega_k^{-1} \). Then \( \sum_{k \in S_\ell} \alpha_k = 1 \) and \( |v_j| \omega_j^{-1} \leq \sum_{k \in S_{\ell-1}} \alpha_k |v_k| \omega_k^{-1} \leq (s - \|\omega\|_\infty^2)^{-1} \sum_{k \in S_{\ell-1}} |v_k| \omega_k \) for all \( j \in S_\ell, \ell \geq 2 \), by our
construction of the partitioning. By the Cauchy-Schwarz inequality and since \( s \geq 2\|\omega\|_\infty^2 \), this gives
\[
\|v_{S_t}\|_2 \leq \frac{\sqrt{s}}{s - \|\omega\|_2^2} \|v_{S_{t-1}}\|_{\omega,1} \leq \frac{2}{\sqrt{s}} \|v_{S_{t-1}}\|_{\omega,1}.
\]
Therefore,
\[
\|v_S\|_2 \leq \frac{2\delta_{\omega,3s}}{(1 - \delta_{\omega,3s})\sqrt{s}} \sum_{t \geq 1} \|v_{S_t}\|_{\omega,1} + \frac{\sqrt{1 + \delta_{\omega,3s}}}{1 - \delta_{\omega,3s}} \|Av\|_2
\]
\[
\leq \frac{2\delta_{\omega,3s}}{(1 - \delta_{\omega,3s})\sqrt{s}} \|v_{S^c}\|_{\omega,1} + \frac{\sqrt{1 + \delta_{\omega,3s}}}{1 - \delta_{\omega,3s}} \|Av\|_2.
\]
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. [20]. 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 \( \omega \)-RIP. Precisely, the main theorem of this section is that matrices arising from orthonormal systems satisfy the \( \omega \)-RIP as long as the weights grow at least as quickly as the \( L_\infty \) norms of the functions they correspond to. This extends existing unweighted RIP results for finite \( \omega \)-RIP orthonormal systems such that \( \sup_{j \in \Lambda} \|\psi_j\|_\infty \leq K \) for some constant \( K \geq 1 \). For bounded orthonormal systems, the following unweighted RIP estimates have been shown.

**Proposition 5.1** (Theorems 4.4 and 8.4, [36]). Fix parameters \( \delta, \gamma \in (0, 1) \). Let \( (\psi_j)_{j \in \Lambda} \) be a bounded orthonormal system with uniform bound \( K \). Suppose
\[
m \geq CK^2\delta^{-2} s \log^2(s) \log(m) \log(N),
\]
\[
m \geq DK^2\delta^{-2} s \log(1/\gamma),
\]
where \( N = |\Lambda| \). Assume that \( t_1, t_2, \ldots, t_m \) are drawn independently from the orthogonalization measure \( \nu \) associated to the orthonormal system. Then with probability exceeding \( 1 - \gamma \), the normalized sampling matrix \( \tilde{A} \in \mathbb{C}^{m \times N} \) 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, \( \delta_s \leq \delta \).

In fact, we can allow \( \|\psi_j\|_\infty \) to depend on \( j \) if we ask only for \( \omega \)-RIP with weights \( \omega_j = \|\psi_j\|_\infty \), or more generally, \( \omega_j \geq \|\psi_j\|_\infty \). This is the content of the following theorem.

**Theorem 5.2** (\( \omega \)-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 \( (\psi_j) \). Then with probability exceeding \( 1 - \gamma \), the normalized sampling matrix \( \tilde{A} \in \mathbb{C}^{m \times N} \) 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, \( \delta_{s,\omega} \leq \delta \).
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. 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 [11]. The remainder of this section is reserved for the proof of Theorem 5.2.

**Proof of Theorem 5.2.** The proof proceeds in a similar manner to those in [40, 36, 20], with some adaptations to account for the weights – see the application of Maurey’s lemma (Lemma 5.3) – and with a twist from [11] leading to the slight improvement in the logarithmic factor. Note that our analysis improves the result of [11] in terms of the probability estimate.

Introducing the set
\[ T_{\omega, s, N} = \{ x \in \mathbb{C}^N, \| x \|_2 \leq 1, \| x \|_{0, \omega} \leq s \}, \tag{31} \]
we can rephrase the weighted isometry constant of $A$ as
\[ \delta_{\omega, s} = \sup_{x \in T_{\omega, s, N}} |\langle (A^* A - \text{Id}) x, x \rangle|. \]
The quantity
\[ \| B \|_s := \sup_{z \in T_{\omega, s, N}} |\langle B z, z \rangle| \tag{32} \]
defines a semi-norm on matrices $B \in \mathbb{C}^{N \times N}$, and we can write
\[ \delta_{\omega, s} = \| A^* A - \text{Id} \|_s. \]
Consider the random variable associated to a column of the adjoint matrix,
\[ X_\ell = \left( \overline{\psi_j (t_\ell)} \right)_{j \in \Lambda}. \tag{33} \]
By orthonormality of the system $(\psi_j)$ we have $\mathbb{E} X_\ell X_\ell^* = \text{Id}$, and the restricted isometry constant equals
\[ \delta_{\omega, s} = \left\| \frac{1}{m} \sum_{\ell=1}^m X_\ell X_\ell^* - \text{Id} \right\|_s = \frac{1}{m} \left\| \sum_{\ell=1}^m (X_\ell X_\ell^* - \mathbb{E} X_\ell X_\ell^*) \right\|_s. \]
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 $\epsilon = (\epsilon_1, \ldots, \epsilon_m)$, i.e., a sequence of independent Rademacher variables $\epsilon_\ell$ taking the values $+1$ and $-1$ with equal probability, also independent of the variables $X_\ell$. Symmetrization, see e.g. [31, Lemma 6.3] or [36, Lemma 6.7], yields
\[ \mathbb{E} \delta_{\omega, s} \leq \frac{2}{m} \mathbb{E} \left\| \sum_{\ell=1}^m \epsilon_\ell X_\ell X_\ell^* \right\|_s = \frac{2}{m} \mathbb{E} X \mathbb{E} \epsilon \sup_{x \in T_{\omega, s, N}} |\langle \sum_{\ell=1}^m \epsilon_\ell X_\ell X_\ell^* x, x \rangle| \]
\[ = \frac{2}{m} \mathbb{E} X \mathbb{E} \epsilon \sup_{x \in T_{\omega, s, N}} |\sum_{\ell=1}^m \epsilon_\ell |\langle X_\ell, x \rangle|^2|. \tag{34} \]
Conditional on $(X_\ell)$, 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 $N(T, d, u)$ are defined as the smallest number of balls with respect to $d$ and centered at points of $T$ necessary to cover $T$. For fixed $(X_\ell)$, we work with the (pseudo-)metric
\[ d(x, z) = \left( \sum_{\ell=1}^m (|\langle X_\ell, x \rangle|^2 - |\langle X_\ell, z \rangle|^2)^2 \right)^{1/2}. \]
Then Dudley’s inequality \([32, 31, 36, 20]\) implies that

\[
E_x \sup_{x \in T^s_{\omega, N}} |\left(\sum_{\ell=1}^m \epsilon_\ell X_\ell X_\ell^* x, x\right)| \leq 4\sqrt{2} \int_0^\infty \sqrt{\log(N(T^s_{\omega, N}, d, u))} du.
\]

In order to continue we estimate the metric \(d\) using Hölder’s inequality with exponents \(p \geq 1\) and \(q \geq 1\) satisfying \(1/p + 1/q = 1\) to be specified later on. For \(x, z \in T^s_{\omega, N}\), this gives

\[
d(x, z) = \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle| + |\langle X_\ell, z \rangle|\right)^2 \left(|\langle X_\ell, x \rangle| - |\langle X_\ell, z \rangle|\right)^2\right)^{1/2}
\]

\[
\leq \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle| + |\langle X_\ell, z \rangle|\right)^2\right)^{1/(2p)} \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle - |\langle X_\ell, z \rangle|\right)^2\right)^{1/(2q)}
\]

\[
\leq 2 \sup_{x \in T^s_{\omega, N}} \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle|\right)^2\right)^{1/(2p)} \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle - |\langle X_\ell, z \rangle|\right)^2\right)^{1/(2q)}.
\]

(35)

In the standard analysis \([40, 36, 20]\), this bound is applied for \(p = 1, q = \infty\). Following \([11]\), we will achieve a slightly better factor by working with a different value of \(p\) to be determined later.

For any realization of \((X_\ell)\), we have \(|\langle X_\ell, x \rangle|\leq \|\psi_j\|_\infty \leq \omega_j\) by assumption. For \(x \in T^s_{\omega, N}\) with \(S = \text{supp} x\) we have \(\sum_{j \in S} \omega_j^2 \leq s\), resulting in

\[
|\langle X_\ell, x \rangle| \leq \sum_{j \in S} \omega_j |x_j| \leq \left(\sum_{j \in S} \omega_j^2\right)^{1/2} \|x\|_2 \leq \sqrt{s}.
\]

(36)

This gives

\[
\sup_{x \in T^s_{\omega, N}} \left(\sum_{\ell=1}^m \left(|\langle X_\ell, x \rangle|\right)^2\right)^{1/(2p)} = \sup_{x \in T^s_{\omega, N}} \left(\sum_{\ell=1}^m \left|\langle X_\ell, x \rangle\right|^2 \left|\langle X_\ell, x \rangle\right|^{2(p-1)}\right)^{1/(2p)}
\]

\[
\leq s^{(p-1)/(2p)} \left(\sup_{x \in T^s_{\omega, N}} \sum_{\ell=1}^m \left|\langle X_\ell, x \rangle\right|^2\right)^{1/(2p)}.
\]

Introducing the (semi-)norm

\[
\|x\|_{x, q} = \left(\sum_{\ell=1}^m \left|\langle X_\ell, x \rangle\right|^{2q}\right)^{1/(2q)}
\]

and using basic properties of covering numbers, we obtain

\[
E_x \sup_{x \in T^s_{\omega, N}} |\sum_{\ell=1}^m \epsilon_\ell X_\ell X_\ell^* x, x\rangle|
\]

\[
\leq C_1 s^{(p-1)/(2p)} \left(\sup_{x \in T^s_{\omega, N}} \sum_{\ell=1}^m \left|\langle X_\ell, x \rangle\right|^2\right)^{1/(2p)} \int_0^\infty \sqrt{\log(N(T^s_{\omega, N}, \|x\|_{x, q}, u))} du,
\]

(37)

where \(C_1\) is a suitable constant. Next, we estimate the covering numbers appearing above in two different ways.
Lemma 5.3. The convex hull of a set $A$ to deduce a covering number bound which is good for larger values of $u$. It follows from (36) that, for $x \in T^s_{\omega,N}$,

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

(38)

Denoting $B_S$ to be the $\ell_2$ unit ball of vectors with support in $S$ and applying the volumetric covering number bound (see e.g. [36, Proposition 10.1]) gives

$$
\mathcal{N}(T^s_{\omega,N}, \| \cdot \|_{X,q}, u) \leq \sum_{S \subseteq \lambda \omega(S) \leq s} \mathcal{N}(B_S, \sqrt{sm^{1/(2q)}} \| \cdot \|_2, u)
$$

$$
\leq \left( \frac{N}{s} \right)^{1 - 2s} \left( 1 + 2s \right) \leq (eN/s)^2 \left( 1 + 2s \right),
$$

where we have applied [20, Proposition C.3] (see also [36, p. 72]) in the last step.

We use Maurey’s lemma [10] – see also [26, 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, conv($U$) denotes the convex hull of a set $U$.

**Lemma 5.3.** For a normed space $X$, consider a finite set $U \subset X$ of cardinality $N$, and assume that for every $L \in \mathbb{N}$ and $(u_1, \ldots, u_L) \in U^L$, $\mathbb{E}_\epsilon \| \sum_{j=1}^{L} \epsilon_j u_j \|_X \leq A\sqrt{L}$, where $\epsilon$ denotes a Rademacher vector. Then for every $u > 0$,

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

The constant $c > 0$ is universal.

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

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

Here, $e_j$ denotes the $j$-th canonical unit vector. For a Rademacher vector $\epsilon = (\epsilon_1, \ldots, \epsilon_L)$, and $u_1, \ldots, u_L \in U$ we have

$$
\mathbb{E}_\epsilon \left\| \sum_{j=1}^{L} \epsilon_j u_j \right\|_{X,q} \leq \left( \mathbb{E} \left\| \sum_{j=1}^{L} \epsilon_j u_j \right\|_{X,q}^{2q} \right)^{1/(2q)} = \left( \mathbb{E} \left\| \sum_{\ell=1}^{m} \sum_{j=1}^{L} \epsilon_j \langle X_\ell, u_j \rangle \right\|_{X,q}^{2q} \right)^{1/(2q)}
$$

$$
= \left( \sum_{\ell=1}^{m} \mathbb{E} \left\| \sum_{j=1}^{L} \epsilon_j \langle X_\ell, u_j \rangle \right\|_{X,q}^{2q} \right)^{1/(2q)} \leq 2e^{-1/2} \sqrt{2q} \left( \sum_{\ell=1}^{m} \left\| \langle X_\ell, u_j \rangle \right\|_{X,q}^{2q} \right)^{1/(2q)}.
$$

In the last step, we have applied Khintchine’s inequality, see e.g. [36, Corollary 6.9]. Using that $|\langle X_\ell, u_j \rangle| \leq \|u_j\|_X \leq \omega_k$, we have, for any vector $u_j \in U$, say $u_j = \omega_k^{-1} e_k$, that

$$
|\langle X_\ell, u_j \rangle| = |\omega_k^{-1} (X_\ell)_{ik}| \leq 1.
$$

Therefore, $\left\| \langle X_\ell, u_j \rangle \right\|_{X,q}^{2q} \leq \sqrt{L}$ for any $L$ and

$$
\mathbb{E}_\epsilon \left\| \sum_{j=1}^{L} \epsilon_j u_j \right\|_{X,q} \leq 2e^{-1/2} \sqrt{2qm^{1/(2q)}} \sqrt{L}.
$$

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

$$
\sqrt{\log \mathcal{N}(T^s_{\omega,N}, \| \cdot \|_{X,q}, u)} \leq \sqrt{\log \mathcal{N}(\text{conv}(U), \| \cdot \|_{X,q}, \sqrt{2q}u)} \leq C_2 \sqrt{qm^{1/q} s \log(4N)} u^{-1}
$$
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 \( \sqrt{sm^{1/(2q)}} \) because for \( u > \sqrt{sm^{1/(2q)}} \) we have \( \mathcal{N}(T^N_{\omega}, \| \cdot \|_{\mathbf{x},q}, u) = 1 \) by (38). Splitting then the Dudley integral into two parts and using the appropriate bounds for the covering numbers, we obtain, for \( \kappa \in (0, \sqrt{sm^{1/(2q)}}) \),

\[
\int_0^\infty \sqrt{\log(\mathcal{N}(T^N_{\omega}, \| \cdot \|_{\mathbf{x},q}, u))} \, du \\
\leq \int_0^\kappa \sqrt{s \log(eN/s) + 2s \log(1 + 2\sqrt{sm^{1/(2q)}/u})} \, du \\
+ C_2 \sqrt{qm^{1/q} \log(4N)} \int_\kappa ^{\sqrt{sm^{1/(2q)}}} u^{-1} \, du \\
\leq \kappa \sqrt{s \log(eN/s) + 2s \kappa} \sqrt{\log(e(1 + \sqrt{sm^{1/(2q)}/\kappa}))} \\
+ C' \sqrt{qm^{1/q} \log(4N) \log(\sqrt{sm^{1/(2q)}/\kappa})}.
\]

In the last step, we have applied [36, Lemma 10.3]. Choosing \( \kappa = m^{1/(2q)} \) yields

\[
\int_0^\infty \sqrt{\log(\mathcal{N}(T^N_{\omega}, \| \cdot \|_{\mathbf{x},q}, u))} \, du \leq C_3 \sqrt{qsm^{1/q} \log(N) \log^2(s)}.
\]

A combination with (37) and (34) gives

\[
\mathbb{E} \delta_{\omega, s} \leq \frac{C_3 s^{(p-1)/(2p)}}{m^{1-1/(2q)}} \sqrt{qm^{1/q} \log(N) \log^2(s)} \mathbb{E} \sup_{x \in T^N_{\omega}} \left( \sum_{i=1}^m |\langle X_i, x \rangle| \right)^{1/(2p)} \\
\leq \frac{C_3 s^{1/2+(p-1)/(2p)} \sqrt{q \log(N) \log^2(s)}}{m^{1-1/(2q)}} \mathbb{E} \left( \frac{1}{m} \left\| \sum_{i=1}^m X_i X_i^* - \text{Id} \right\|_s + \|\text{Id}\|_s \right)^{1/(2p)} \\
\leq \frac{C_3 s^{1/2+(p-1)/(2p)} \sqrt{q \log(N) \log^2(s)}}{m^{1/2}} \sqrt{\mathbb{E} \delta_{\omega, s} + 1}.
\]

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 \( s^{(p-1)/(2p)} \leq s^{1/(2 \log(s))} = e^{1/2} \)
and

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

Completing squares finally shows that

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

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

\[
m \geq C_6 \delta^{-2} s \log^3(s) \log(N)
\]

with \( C_6 = 4C_5^2 \). Similarly to [36, Section 8.6] we write

\[
\delta_{\omega, s} = \frac{1}{m} \sup_{(z, \omega) \in Q^{N, \omega}_s} \left( \sum_{i=1}^m (X_i X_i^* - \text{Id}) z, \omega \right)
\]
where $Q_{s,N}^ω$ denotes a dense countable subset of
\[
Q_{s,N}^ω = \bigcup_{S \subseteq \mathcal{A}, \omega(S) \leq s} Q_S, \quad Q_S = \{(x, w) : \|x\|_2 = \|w\|_2 = 1, \text{supp } x, \text{supp } w \subseteq S\}.
\]

With the functions $f_{z,w}(X) = \langle (XX^* - \text{Id})x, w \rangle$ we can write $δ_{ω,s}$ as the supremum of an empirical process
\[
δ_{ω,s} = \frac{1}{m} \sup_{(z,w) \in Q_{s,N}^ω} \sum_{\ell=1}^m f_{z,w}(X_\ell).
\]

Since $\mathbb{E}X_\ell^tX_\ell^* = \text{Id}$ we have for all $z, w$. Further, for $(x, w) \in Q_S$ with $\omega(S) \leq s$ and for any realization of $X_\ell$, we have
\[
|f_{z,w}(X_\ell)| \leq \max\{1, \max_{x : \text{supp } x \subseteq S} |\langle X_\ell X_\ell^* x, x \rangle|\} = \max\{1, \max_{x : \text{supp } x \subseteq S} |\langle X_\ell, x \rangle|^2\} \leq s.
\]

Moreover,
\[
\mathbb{E}|f_{z,w}(X_\ell)|^2 = \mathbb{E}|\langle (X_\ell X_\ell^* - \text{Id})z, w \rangle|^2
= \mathbb{E}|\langle X_\ell X_\ell^* z, w \rangle|^2 - 2 \Re(\mathbb{E}[\langle X_\ell X_\ell^* z, w \rangle \langle w, z \rangle]) + |\langle z, w \rangle|^2
= \mathbb{E}|\langle X_\ell, z \rangle|^2|\langle X_\ell, w \rangle|^2 - |\langle z, w \rangle|^2 \leq s \mathbb{E}|\langle X_\ell, z \rangle|^2 = s.
\]

With these bounds for $f_{z,w}(X_\ell)$ together with (40), the Bernstein inequality for the supremum of an empirical process, see e.g. [36, Theorem 6.25], [5] or [20, Theorem 8.42], yields, for $δ \in (0, 1)$,
\[
P(δ_{ω,s} \geq δ) \leq P(δ_{ω,s} \geq \mathbb{E}δ_{ω,s} + δ/2)
= P\left(\sup_{(z,w) \in Q_{s,N}^ω} \sum_{\ell=1}^m f_{z,w}(X_\ell) \geq \mathbb{E}\sup_{(z,w) \in Q_{s,N}^ω} \sum_{\ell=1}^m f_{z,w}(X_\ell) + δm/2\right)
\leq \exp\left(-\frac{(δm/2)^2/2}{ms + 2s(δm/2) + (δm/2)s/3}\right) \leq \exp\left(-\frac{δ^2m}{C_7s}\right),
\]
where $C_7 = 8(1 + 2 + 1/6) \leq 26$. The last term is bounded by $γ \in (0, 1)$ if $m \geq C_7δ^{-2}s \log(1/γ)$. Altogether we have $δ_{ω,s} \leq δ$ with probability at least $1 - γ$ if
\[
m \geq C_8δ^{-2}s \max\{\log^3(s) \log(N), \log(1/γ)\},
\]
where $C_8 = \max\{C_6, C_7\}$. This completes the proof. □

6 Main interpolation estimates

Using the concepts of weighted null space and weighted restricted isometry property, and together with Theorem 5.2 the $ω$-RIP for orthonormal systems, we now prove Theorems 1.1 and 1.2 concerning interpolation via weighted $\ell_1$ minimization, and a more general result for functions with coefficients in weighted $\ell_p$ 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 $ω_j ≥ \|\psi_j\|_\infty$. For $s ≥ 2 \max_j |ω_j|^2$ and $γ \in (0, 1)$, fix a number of samples
\[
m \geq cs \max\{\log^3(s) \log(N), \log(1/γ)\}.
\]

Suppose that $t_\ell, \ell = 1, \ldots, m$, are drawn independently from the orthogonalization measure associated to the $(\psi_j)$. Let $A \in \mathbb{C}^{m \times N}$ 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 \(y_\ell = f(t_\ell) + \xi_\ell, \ell = 1, \ldots, m\), with \(\|\xi\|_2 \leq \eta\), let \(x^*\) be the solution of

\[
\min \|z\|_{\omega,1} \text{ subject to } \|Az - y\|_2 \leq \eta
\]

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

\[
\|f - f^*\|_{L_\infty} \leq \|f - f^*\|_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1} + d_1 \eta \sqrt{\frac{s}{m}},
\]

\[
\|f - f^*\|_{L_2} \leq c_2 \frac{\sigma_s(f)_{\omega,1}}{\sqrt{s}} + d_2 \eta \sqrt{\frac{1}{m}}.
\]

Above, \(c_0, c_1, d_1, c_2\), and \(d_2\) are universal constants.

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

\[
\|f - f^*\|_{\infty} \leq \sum_{j=1}^{N} |x_j - x_j^*| \|\psi_j\|_{\infty} \leq \|x - x^*\|_{\omega,1} = \|f - f^*\|_{\omega,1}.
\]

\[\square\]

**Remark 6.2.** Theorem 1.1 in the introduction corresponds to the special case of Theorem 6.1 when 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 \(\Lambda\) is countably infinite then we first have to restrict to a suitable finite subset \(\Lambda_0\) before applying weighted \(\ell_1\) 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 \(f_0 = \sum_{j \in \Lambda_0} x_j \psi_j\), \(\Lambda\) into \(\Lambda_0\) and \(\Lambda_R\), decomposing \(f\) 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 samples in hopes of applying Theorem 6.1.

The remainder of the proof 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 \((\psi_j)\), the random variables \(|f_R(t_\ell)|^2\) are independent and identically distributed, with

\[
\mathbb{E}\left(|f_R(t_\ell)|^2\right) = \sum_{j \in \Lambda_R} |x_j|^2. \tag{43}
\]

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

\[
\sum_{j \in \Lambda_R} |x_j|^2 \leq \frac{2}{s} \sum_{j \in \Lambda} |x_j|^2 \omega_j^2 \leq \frac{2}{s} \left( \sum_{j \in \Lambda_R} |x_j| \omega_j \right)^2 = \frac{2}{s} \|f_R\|_{\omega,1}^2.
\]
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 \( |f_{R}(t_\ell)|^2 - \mathbb{E}(\{f_{R}(t_\ell)|^2) \) is bounded by

\[
\mathbb{E}(\{f_{R}(t_\ell)|^2 - \mathbb{E}(\{f_{R}(t_\ell)|^2) \)^2 \leq \mathbb{E}(\{f_{R}(t_\ell)|^4) \leq \|f_{R}\|_{\omega,1}^2 \mathbb{E}(\{f_{R}(t_\ell)|^2) \leq \frac{2}{s} \|f_{R}\|_{\omega,1}^4.
\]

We now apply Bernstein’s inequality to obtain 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)

\[
\mathbb{P} \left\{ \frac{1}{m} \sum_{\ell=1}^{m} |f_{R}(t_\ell)|^2 \geq \frac{4}{s} \|f_{R}\|_{\omega,1}^2 \right\} \leq N^{- \log^3(s)}.
\]

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 \( \ell_p \) spaces \( p \leq 1 \)

A weakness of Theorem 1.2 is 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 \( \omega \) in a suitable way. In order to illustrate the idea we start by recalling the error bound

\[
\|f - f^2\|_{L_\infty} \leq c_1 \sigma_s(f)_{\omega,1} + d_1 \eta \sqrt{\frac{s}{m}},
\]

valid in the finite-dimensional setting: \( f = \sum_{j \in \Lambda} x_j \psi_j, |\Lambda| = N \in \mathbb{N} \), and the samples are perturbed, \( \sum_{\ell=1}^{m} y_\ell - f(t_\ell) \leq \eta^2 \). In the infinite-dimensional setting, where \( f = \sum_{j \in \Lambda} x_j \psi_j \) but \( \Lambda \) possibly countably infinite, we will treat the samples \( y_\ell = f(t_\ell) \) as perturbed samples of a finite-dimensional approximation \( f_0 = \sum_{j \in \Lambda_0} x_j \psi_j \) as in the probabilistic error analysis in the proof of Theorem 1.2, for some suitable \( \Lambda_0 \subset \Lambda \). For a parameter \( \alpha > 0 \), the approximation error can then be bounded using

\[
\|f - f_0\|_{L_\infty} \leq \sum_{j \notin \Lambda_0} |x_j| \|\psi_j\|_{\infty} \leq \max_{j \notin \Lambda_0} \{ \|\psi_j\|_{\infty} v_j^{-\alpha} \} \sum_{j \notin \Lambda_0} |x_j| v_j^\alpha \leq \max_{j \notin \Lambda_0} \{ w_j v_j^{-\alpha} \} \|f\|_{\tilde{v},1,1}.
\]

On the right hand side, we obtain the norm \( \|f\|_{\tilde{v},1,1} \). 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 \( \omega \) and \( 0 < p < 1 \), set \( \alpha = \frac{2}{p} - 1 \). Then \( \|x\|_{\tilde{\omega},1,1} \leq \|x\|_{\omega,p} \).
Proof. First observe that
\[
\left( \max_{j \in \Lambda_0} |x_j| \omega_j^{2/p-1} \right)^p = \max |x_j|^p \omega_j^{2-p} \leq \sum_{j \in \Lambda_0} |x_j|^p \omega_j^{2-p} = \|x\|^p_{\omega, p}.
\]
The claimed inequality follows then from Theorem 3.2 by error bound (44) with
\[
\left( \max_{j \in \Lambda_0} |x_j| \omega_j^\alpha \right)^{1-p} \|x\|_{\omega, p} \leq \|x\|_{\omega_0}^{1-p} \|x\|_{\omega, p} = \|x\|_{\omega_0}.
\]
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
\[
\sigma_s(f_0)_{\omega_0} \leq cs^{1-1/p} \|f_0\|_{\omega, p} \leq cs^{1-1/p} \|f\|_{v, p}.
\]
We aim to provide a bound of the second term on the right-hand side of (44) of the same order. Now \( \eta := \sqrt{\sum_{\ell=1}^\infty |f_0(t_\ell) - f(t_\ell)|^2} \leq m \|f - f_0\|_{L_\infty} \), so by (45) and Lemma 6.3,
\[
\sqrt{\frac{s}{m}} \eta \leq \sqrt{s} \max_{j \notin \Lambda_0} \{w_j \nu_j^{-\alpha}\} \|f\|_{\omega_0, 1} \leq \sqrt{s} \max_{j \notin \Lambda_0} \{w_j \nu_j^{-\alpha}\} \|f\|_{v, p},
\]
where we have applied Lemma 6.3 with \( \alpha = \frac{2}{p} - 1 \) in the last step. The choice \( \Lambda_0 = \Lambda_0^{(s,p)} \) with
\[
\Lambda_0^{(s,p)} := \{j \in \Lambda : \omega_j \nu_j^{1-2/p} \geq s^{1/2-1/p}\}
\]
therefore gives
\[
\eta \leq \sqrt{ms^{1/2-1/p}} \|f\|_{v, p} \quad \text{so that} \quad \sqrt{\frac{s}{m}} \eta \leq s^{1-1/p} \|f\|_{v, p},
\]
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 \( (\omega_j \nu_j^{1-2/p})_{j \in \Lambda} \) is a sequence which converges to 0 as \( |j| \to \infty \). Moreover, if \( v \) satisfies
\[
\nu_j^{2/p-1} \geq 2^{1/p-1/2} \nu_j^{2/p} = 2^{1/p-1/2} \omega_j \cdot \omega_j^{2(1/p-1/2)},
\]
then we have for all \( j \) satisfying \( \omega_j^2 \geq s/2 \) that
\[
\nu_j^{2/p-1} \geq 2^{1/p-1/2} \omega_j(s/2)^{1/p-1/2} = \omega_j s^{1/p-1/2}.
\]
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 \( (\omega_j \nu_j^{1-2/p})_{j \in \Lambda} \) 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 \( \omega, v \) 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
\[
m \geq c_0 s \max\{ \log^3(s) \log(N^{(s,p)}), \log(1/\gamma) \}.
\]
Suppose that the sampling points \( t_{\ell}, \ell = 1, \ldots, m \), are drawn independently at random according to the orthogonalization measure for \( (\psi_j) \). 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 \( A \) 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 \( x^* \) be the solution to
\[
\min \|z\|_{\omega,1} \text{ subject to } \|Az - y\|_2 \leq \tau s^{1/2-1/p} \sqrt{m} \|f\|_{v,p}
\]
and set \( f^* = \sum_{j \in \Lambda_0} x^*_j \psi_j \). Then \( \|f - f^*\|_{\infty} \leq C_{\tau} s^{1-1/p} \|f\|_{v,p} \).

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
\[
\|f - f^*\|_{L_\infty} \leq \|f - f_0\|_{L_\infty} + \|f_0 - f^*\|_{L_\infty}.
\]
Since with probability at least \( 1 - \gamma \) under the stated assumption on \( m \), the matrix \( A \) has the \( \omega \)-RIP and thereby the weighted null space property of order \( s \), we have
\[
\|f_0 - f^*\|_{L_\infty} \leq C_{\tau} s^{1-1/p} \|f\|_{v,p}
\]
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. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[4] T. Blumensath and M. E. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Info. Theory*, 55(4):1872–1882, 2009.

[5] 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.

[6] N. Burq, S. Dyatlov, R. Ward, and M. Zworski. Weighted eigenfunction estimates with applications to compressed sensing. *SIAM J. Math. Anal.*, 44(5):3481–3501, 2012.

[7] G. Byrenheid, D. Dung, W. Sickel, and T. Ullrich. Sampling on energy-norm based sparse grids for the optimal recovery of sobolev type functions in $H^s$. *Preprint*, 2014.

[8] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.

[9] 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.

[10] 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.

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

[12] F. Cobos, T. Kühn, and W. Sickel. Optimal approximation of Sobolev functions in the sup-norm. *preprint*, 2014.

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

[14] A. Cohen, M. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. of Comput. Math.*, 13(5):819–834, 2013.

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

[16] 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.

[17] 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.

[18] L. Fagiano and M. Khammash. Simulation of stochastic systems via polynomial chaos expansions and convex optimization. *Phys. Rev. E.*, 86, 2012.

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

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

[21] 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.
[22] L. Jacques. A short note on compressed sensing with partially known signal support. *Signal Processing*, 90:3308–3312, 2010.

[23] J. Jo. Iterative hard thresholding for weighted sparse approximation. *arXiv preprint arXiv:1312.3582*, 2013.

[24] A. Khajehnejad, W. Xu, A. Salman, and B. Hassibi. Weighted $\ell_1$ minimization for sparse recovery with prior information. In *ISIT 2009*, pages 483–487. IEEE, 2009.

[25] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Analyzing weighted $\ell_1$ minimization for sparse recovery with nonuniform sparse models. *Signal Processing, IEEE Transactions on*, 59:1985–2001, 2011.

[26] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67:1877–1904, 2014.

[27] I. Krasikov. An upper bound on Jacobi polynomials. *J. Approx. Theory*, 149:116–130, 2007.

[28] T. Kühn, W. Sickel, and T. Ullrich. Approximation of mixed order Sobolev functions on the d-torus – Asymptotics, preasymptotics and d-dependence. *preprint*, 2014. arXiv:1312.6386.

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

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

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

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

[33] S. Misra and P. Parrilo. Analysis of weighted $\ell_1$-minimization for model-based compressed sensing. *Arxiv*, 2013.

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

[35] J. Peng, J. Hampton, and A. Doostan. A weighted $\ell_1$-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267:92–111, 2014.

[36] 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.

[37] H. Rauhut and C. Schwab. Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations. *Preprint*, 2014. arXiv:1410.4929.

[38] H. Rauhut and R. Ward. Sparse Legendre expansions via $\ell_1$-minimization. *J. Approx. Theory*, 164(5):517–533, 2010.

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

[40] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
[41] C. Runge. Über empirische Funktionen und die interpolation zwischen äquidistanten Ordinaten. *Z Math Physik*, 46:224–243, 1901.

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

[43] V. Temlyakov. On approximate recovery of functions with bounded mixed derivative. *J. Complexity*, 9(1):41–59, 1993.

[44] 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.

[45] 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.

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

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