# Sampling Designs for Function Recovery – Theoretical Guarantees, Comparison and Optimality

Moritz Moeller[1], Kateryna Pozharska[1,2], and Tino Ullrich[1]

[1] Faculty of Mathematics of Chemnitz University of Technology
[2] Institute of Mathematics of NAS of Ukraine

**Abstract.** In this paper we compare non-linear sampling recovery methods for multivariate function classes. In the first part of the paper we propose square root Lasso with a particular choice of the regularization parameter $\lambda > 0$ as a noise blind decoder which efficiently recovers multivariate functions from random samples. In contrast to basis pursuit denoising the algorithm does not require any additional information on the width of the function class in $L_\infty$. We then relate the findings to commonly used linear recovery methods and compare the performance in a model situation, namely periodic multivariate functions with bounded mixed derivative in $L_q$. The main observation is the fact, that square root Lasso asymptotically outperforms Smolyak's algorithm (sparse grids) in various situations. For $q = 2$ we even see that square root Lasso outperforms any linear method including recently investigated optimal least squares methods.

**Keywords:** Sampling recovery, Square root Lasso, Compressed sensing, Restricted isometry property, Best $n$-term trigonometric approximation, Periodic functions with mixed smoothness

## 1 Introduction

In this paper we extend, discuss and compare recovery methods and sampling designs for the recovery problem in classes of multivariate functions. A particular focus is put on recent developments for non-linear recovery methods, i.e., a variant of *square root Lasso* (`rLasso`) using function samples at random points, see, e.g., Adcock, Bao, Brugiapaglia [1] or H. Petersen, P. Jung [28] and the references therein. The decoder (`rLasso`) turns out to be noise blind and does not require any further information on the function class $\boldsymbol{F}$, like for instance its width in $L_\infty$, in contrast to the recently proposed variant of *basis pursuit denoising*, see Jahn, T. Ullrich, Voigtlaender [14] or Krieg [16]. Hence, the corresponding recovery operator $R_{m,\lambda}(\cdot;\mathbf{X})$ is a universal algorithm allowing for individual estimates on the respective $d$-variate periodic function $f \in \boldsymbol{F}$ of interest. To be more precise, for $m \geq \alpha \cdot d \cdot n \cdot (\log(n+1))^2 \cdot \log M$ random samples $\mathbf{X} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^m\}$ it holds for $2 \leq q \leq \infty$, as stated in Theorem 3, that

$$\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_q} \leq C n^{1/2-1/q} \Big(\sigma_n(f;\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d \cap \mathbb{Z}^d}(f;\mathcal{T}^d)_{L_\infty}\Big) . \quad (1)$$

A function is recovered from the vector $\boldsymbol{y} = f(\mathbf{X}) \in \mathbb{C}^m$ of point evaluations at random nodes where the set of nodes is fixed in advance and is used for all functions belonging to

$C(\mathbb{T}^d)$. One only has to solve the (rLasso) optimization program (12) for the coefficient vector of the approximand $R_{m,\lambda}(f;\mathbf{X})$.

We put that into perspective to other contemporary sampling recovery methods such as sparse grids (Smolyak) and linear methods based on least squares with respect to hyperbolic crosses on subsampled random points (Lsqr). Our results are collected in Figure 1 below which illustrates the regions in the $(1/p, 1/q)$ parameter domain for our model scenario on the $d$-torus, namely spaces with bounded mixed derivative $\mathbf{W}^r_p$ in $L_q$, where the different methods are known to be optimal, close to optimal or at least superior over others. As optimality measure we use the classical notion of sampling numbers introduced in (2) below. The picture is only partially complete which in turn means that there are a lot of open problems, where the reader is invited to contribute.

We consider mixed Wiener spaces $\mathcal{A}^r_{\mathrm{mix}}$ on the $d$-torus and function classes with bounded mixed derivative $\mathbf{W}^r_p$ as surveyed in Dũng, Temlyakov, T. Ullrich [8, Chapt. 2]. These spaces have a relevant history in the former Soviet Union and serve as a powerful model for multivariate approximation. Concretely, we study the situation $\mathbf{W}^r_p$ in $L_q$ where $1 < p \le 2 \le q$ and the case of small smoothness where $2 < p < \infty$ and $1/p < r \le 1/2$. We consider the worst-case setting where the error is measured in $L_q$. It turned out in [14] that for several classical smoothness spaces non-linear recovery in $L_2$ outperforms any linear method (not only sampling). The results in this paper show that this effect partially extends to $L_q$ with $2 \le q < \infty$. In fact, functions in mixed Wiener classes $\mathcal{A}^r_{\mathrm{mix}}$ provide an intrinsic sparsity with respect to the trigonometric system, such that the additional gain in the rate does not seem to be a surprise. For $r > 1/2$ Corollary 2 states that

$$\varrho_{\lceil C_{r,d} n (\log(n+1))^3 \rceil}(\mathcal{A}^r_{\mathrm{mix}})_{L_q} \lesssim n^{-(r+1/q)}(\log(n+1))^{(d-1)r+1/2}\,.$$

We determine a polynomial rate of convergence $r + 1/q$ which is at least sharp in the main rate (apart from logarithms) and outperforms any linear algorithm. The situation is not so clear when studying $\mathbf{W}^r_p$ classes in $L_q$. Surprisingly, in the case $1 < p < 2 < q$ and $1/p + 1/q > 1$ square root Lasso outperforms any sampling algorithm based upon sparse grids if $d$ is large. The acceleration only happens in the logarithmic term. From Corollary 3 and Remark 3 we obtain

$$\varrho_{\lceil C_{r,p,d} n (\log(n+1))^3 \rceil}(\mathbf{W}^r_p)_{L_q} \lesssim n^{-(r-\frac{1}{p}+\frac{1}{q})}(\log(n+1))^{(d-1)(r-2(\frac{1}{p}-\frac{1}{2}))+\frac{1}{2}}\,,$$

which shows that in the regime where $q = 2$ and $d$ is large (rLasso) has a faster asymptotic decay than any linear method and in particular (Lsqr). This effect has been observed already for basis pursuit denoising in [14]. Note that the described effects do not appear when it comes to the uniform norm. This is a consequence of a general result described in Novak, Woźniakowski [27, Chapter 4.2.2].

Other non-linear recovery methods have been considered in Dai, Temlyakov [10]. The authors complemented and partly improved the results from [14] by using greedy methods. In Krieg [16, Theorem 1] it has been shown that for certain weighted Wiener spaces the sampling recovery problem in $L_2$ using basis pursuit denoising is polynomially tractable. A similar phenomenon has been observed recently in Moeller, Stasyuk, T. Ullrich [23, Corollary 5.3] for certain Besov spaces with mixed smoothness. Indeed, if we combine (1) with Theorem 5.2 in [23] we obtain a corresponding tractability result also for (`rLasso`).

The bound in (1) has the striking advantage that one may directly insert known bounds from the literature, see [3,34] and [8, Section 7] for an overview. Other approaches, like in [16], require the embedding of the function class into the multivariate Wiener algebra $\mathcal{A}$ which is not always the case, not even for classical smoothness spaces like Sobolev spaces $\mathbf{W}_p^{1/2}$ for $p > 2$. This non-trivial fact is a corollary of the very recent Theorem 1.8 in Saucedo, Tikhonov [29].

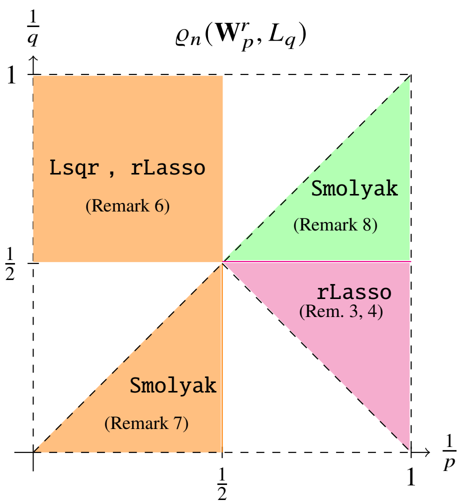


**Fig. 1.** Magenta area: Only comparison, optimality not clear. Orange area: Optimality w.r.t. Gelfand widths. Green area: Optimality w.r.t. linear widths.

Smolyak's sparse grids [30] in connection with functions providing bounded mixed derivative or difference have a significant history not only for approximation theory, see [31], [8] and the references therein, but also in scientific computing, see Bungartz, Griebel [6]. The underlying spaces do not only serve as a powerful model for multivariate approximation theory motivated from practical problems. Sparse grid algorithms allow for good (and sometimes optimal) approximation rates. It is strongly related to *hyperbolic cross approximation*. In Figure 1 we indicate the parameter regions where (`Smolyak`) is known to be optimal with respect to Gelfand/approximation numbers.

Finally, we would like to mention the recent developments in the direction of least squares methods (`Lsqr`). Beginning from the breakthrough result by Krieg, M. Ullrich [19], where it was shown that sampling recovery for reproducing kernel Hilbert spaces in $L_2$ is asymptotically equally powerful as linear approximation, the authors improved both algorithm [24,2] and error guarantee [20] until the remaining $\sqrt{\log n}$ gap has finally been sealed by Dolbeault, Krieg, M. Ullrich [11] for RKHS which are sufficiently compact in $L_2$. In Nagel, Schäfer, T. Ullrich [24] subsampled random points appeared for the first time. The final solution [11] is again heavily based on the solution of the Kadison-Singer problem [21], however it is highly non-constructive. As for the classical problem $\boldsymbol{W}_2^r$ in $L_2$ (the midpoint in Figure 1) the algorithm uses the basis functions from the hyperbolic cross (left picture in Figure 2) with $m$ frequencies and inserts $O(m)$ nodes which result from a random draw ($O(m \log m)$) together with a subsampling to $O(m)$ points (fourth picture in Figure 2). The resulting overdetermined matrix is then used to recover the coefficients from the sample vector. Apart from the Hilbert space setting, the situation $\boldsymbol{W}_p^r$ in $L_q$ has been investigated in [17,18].

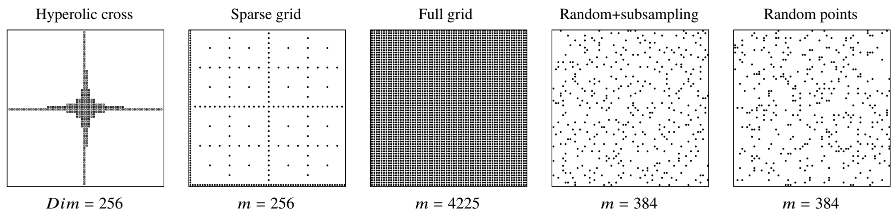


**Fig. 2.** Hyperbolic cross in the frequency domain, different sampling designs in $d = 2$

*Notation* For a number $a$, by $a_+$ we denote $\max\{a, 0\}$, and by $\log(a)$ its natural logarithm. $\mathbb{C}^n$ shall denote the complex $n$-space, where we distinguish $\|\boldsymbol{v}\|_{\ell_1} := \sum_{k=1}^n |v_k|$ and $\|\boldsymbol{v}\|_{\ell_2} := (\sum_{k=1}^n |v_k|^2)^{1/2}$. Further let $\mathbb{C}^{m\times n}$ denote the set of complex $m\times n$-matrices. Vectors and matrices are usually typesetted boldface. For a vector $\boldsymbol{v} \in \mathbb{C}^N$ and a set $S \subset \{1, ..., N\} =: [N]$ we mean by $\boldsymbol{v}_S \in \mathbb{C}^N$ the restriction of $\boldsymbol{v}$ to $S$, where all other entries are set to zero and $S^c = [N] \setminus S$. We denote by $f \in \mathcal{T}([-M, M]^d)$ that $f$ is a trigonometric polynomial with support on the frequencies in the set $[-M, M]^d$. The notation $L_q(\mathbb{T}^d)$, $1 \le q < \infty$, indicates the classical Lebesgue space of periodic functions on the $d$-torus $\mathbb{T}^d = [0, 1]^d$ with the usual modification for $q = \infty$. The notation $C(\mathbb{T}^d)$ stands for the space of continuous functions on $\mathbb{T}^d$ with the sup-norm. All other function spaces of $d$-dimensional functions will be typesetted boldface. Let $X$ and $Y$ denote two normed spaces. The norm of an element $x$ in $X$ will be denoted by $\|x\|_X$. The space of linear operators between $X$ and $Y$ will be denoted by $\mathcal{L}(X, Y)$. We write $X \hookrightarrow Y$ indicates that the identity operator from $X$ to $Y$ is continuous. For two sequences $a_n$ and $b_n$ we will write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \le c\, b_n$ for all $n \in \mathbb{N}$. We will write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. The implied constants may depend on parameters of the source and target spaces and the dimension $d$ of the domain, but not on $n$.

## 2 Best *n*-term and Linear Approximation

Let $\Omega$ denote a compact topological space and $C(\Omega)$ the set of complex-valued continuous functions on $\Omega$. The (non-linear) sampling numbers for a quasi-normed space $\boldsymbol{F}$ of functions $f\colon \Omega \to \mathbb{C}$ which is continuously embedded into $Y \cap C(\Omega)$ such that point evaluations are reasonably defined are given as follows:

$$\varrho_m(\boldsymbol{F})_Y := \inf_{\mathbf{X}=\{\mathbf{x}^1,\dots,\mathbf{x}^m\}} \ \inf_{R\colon \mathbb{C}^m \to Y} \ \sup_{\|f\|_{\boldsymbol{F}} \le 1} \|f - R\,(f(\mathbf{X}))\,\|_Y . \tag{2}$$

If one restricts oneself to linear recovery operators $R\colon \mathbb{C}^m \to Y$, then the corresponding quantity is denoted by $\varrho_m^{\text{lin}}(\boldsymbol{F})_Y$. This quantity is lower bounded by the Gelfand width $c_m(\boldsymbol{F})_Y$ defined as

$$c_m(\boldsymbol{F})_Y := \inf_{\substack{R\colon \mathbb{C}^m \to Y \\ L \in \mathcal{L}(\boldsymbol{F}, \mathbb{C}^m)}} \ \sup_{\|f\|_{\boldsymbol{F}} \le 1} \|f - R \circ L(f)\|_Y \,.$$

Note that this is not the original definition for the Gelfand numbers/widths. It is more common to use (see [8, 9.6, (9.6.1)])

$$\tilde{c}_m(\boldsymbol{F})_Y := \inf_{L \in \mathcal{L}(\boldsymbol{F}, \mathbb{C}^m)} \sup_{\substack{f \in \ker L \\ \|f\|_{\boldsymbol{F}} \le 1}} \|f\|_Y .$$

The introduced quantities only differ by a small universal constant. Namely we have

$$\tilde{c}_m(\boldsymbol{F})_Y \le c_m(\boldsymbol{F})_Y \le 2\tilde{c}_m(\boldsymbol{F})_Y .$$

If we require $R : \mathbb{C}^m \to Y$ to be a linear map we may even consider the linear widths

$$\lambda_m(\boldsymbol{F})_Y := \inf_{\substack{T \in \mathcal{L}(\boldsymbol{F}, Y) \\ \operatorname{rank}(T) \le m}} \sup_{\|f\|_{\boldsymbol{F}} \le 1} \left\| f - Tf \right\|_Y .$$

The following general relations are obvious

$$\varrho_m \ge c_m \quad \text{and} \quad \varrho_m^{\mathrm{lin}} \ge \lambda_m \ge c_m .$$

Let $I$ denote a countable index set and $\mathcal{B} = \{b_k \in C(\Omega) : k \in I\}$ a dictionary consisting of continuous functions (often the additional requirement is needed that the functions in $\mathcal{B}$ are uniformly bounded). For $n \in \mathbb{N}$, we define the set of linear combinations of $n$ elements of $\mathcal{B}$ as

$$\Sigma_n := \left\{ \sum_{j \in J} c_j b_j(\cdot) : J \subset I, |J| \le n, (c_j)_{j \in J} \in \mathbb{C}^J \right\} .$$

Furthermore, given $J \subset I$ we denote the linear span of $(b_j(\mathbf{x}))_{j \in J}$ by

$$V_J := \operatorname{span}_{\mathbb{C}} \{ b_j(\cdot) : j \in J \}.$$

Note that the set $\Sigma_n$ is "non-linear" (not a vector space), whereas the space $V_J$ is linear. For a Borel measure $\mu$ on $\Omega$ orthogonality of the $b_i(\cdot)$ with respect to $\mu$ is often of advantage for our framework. We denote by

$$\sigma_n(\boldsymbol{F}; \mathcal{B})_Y := \sup_{\|f\|_{\boldsymbol{F}} \le 1} \sigma_n(f; \mathcal{B})_Y := \sup_{\|f\|_{\boldsymbol{F}} \le 1} \inf_{g \in \Sigma_n} \|f - g\|_Y$$

the best $n$-term approximation widths and by

$$E_J(\boldsymbol{F}; \mathcal{B})_Y := \sup_{\|f\|_{\boldsymbol{F}} \le 1} E_J(f; \mathcal{B})_Y := \sup_{\|f\|_{\boldsymbol{F}} \le 1} \inf_{g \in V_J} \|f - g\|_Y$$

the error of best approximation (from $V_J$) in $Y$.

## 3 Square Root Lasso and Recovery Guarantees

Let us define a decoder known as *square root Lasso* `(rLasso)` in the literature, see H. Petersen, P. Jung [28] and the references therein. The advantage of `(rLasso)` over *basis* *pursuit denoising* as used in [14] is the "noise blindness" which results in the advantage that we do not have to incorporate additional information from the function class $\boldsymbol{F}$ of interest. This feature is also present for recent greedy methods, as observed by Dai, Temlyakov in [10], see especially the remark after their Corollary 2.1. We will tailor square root Lasso to the function recovery problem. For the general scenario described above, the decoder map $R_{m,\lambda}\colon C(\Omega) \to C(\Omega)$ is chosen in the following way.

**Definition 1.** *Let $\lambda > 0$, $J \subset I$ a finite set, $\mathbf{X} = \{x^1, \dots, x^m\} \subset \Omega$. Put*

$$\boldsymbol{A} := 1/\sqrt{m}(b_j(x^\ell))_{1\le \ell\le m, j\in J} \in \mathbb{C}^{m\times |J|}$$

*and for $\mathbf{y} = f(\mathbf{X})/\sqrt{m} \in \mathbb{C}^m$,*

$$R_{m,\lambda}(f;\mathbf{X}) := \sum_{j\in J} (\mathbf{c}^{\#}(\mathbf{y}))_j b_j(\cdot) \in V_J \subset L_\infty, \tag{3}$$

*where $\mathbf{c}^{\#}(\mathbf{y}) \in \mathbb{C}^{|J|}$ is any (fixed) solution of the square root Lasso minimization problem*

$$\inf_{z\in\mathbb{C}^{|J|}} \Big( \|z\|_{\ell_1(|J|)} + \lambda \|\boldsymbol{A}z - \mathbf{y}\|_{\ell_2(m)} \Big). \tag{4}$$

*This defines a (not necessarily linear) map $R_{m,\lambda}\colon C(\Omega) \to C(\Omega)$. The parameter $\lambda > 0$ is chosen below and may depend on other parameters.*

To show our results we rely on techniques from compressed sensing that require the RIP and one of its consequences.

**Definition 2.** *For a matrix $\boldsymbol{A} \in \mathbb{C}^{m\times N}$ we say that is satisfies the Restricted Isometry Property (RIP) of order n with RIP constant $\delta_n$ if for any n-sparse vector $\boldsymbol{v} \in \mathbb{C}^N$ that*

$$(1-\delta_n)\|\boldsymbol{v}\|_{\ell_2}^2 \le \|\mathbf{A}\cdot \boldsymbol{v}\|_{\ell_2}^2 \le (1+\delta_n)\|\boldsymbol{v}\|_{\ell_2}^2 . \tag{5}$$

**Lemma 1 ([12, Proposition 6.3]).** *Let $A \in \mathbb{C}^{m\times N}$ have RIP of order $2n$ with RIP constant $\delta_{2n}$. Then it holds for any n-sparse vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ with $\operatorname{supp}\mathbf{u} \cap \operatorname{supp}\mathbf{v} = \emptyset$ that*

$$\left|\left\langle \boldsymbol{A}\mathbf{u}, \boldsymbol{A}\mathbf{v}\right\rangle\right| \le \delta_{2n}\|\mathbf{u}\|_{\ell_2}\|\mathbf{v}\|_{\ell_2}. \tag{6}$$

**Theorem 1 (RIP implies $\ell_2$-robust NSP).** *For $\mathbf{A} \in \mathbb{C}^{m\times N}$ assume that $\delta_{2n} < \frac{1}{3}$. Then $\mathbf{A}$ satisfies the $\ell_2$-robust null space property (NSP) of order n, i.e.*

$$\|\mathbf{c}_S\|_{\ell_2} \le \frac{\rho}{\sqrt{n}}\|\mathbf{c}_{S^{\mathrm{C}}}\|_{\ell_1} + \tau\|\mathbf{A}\mathbf{c}\|_{\ell_2} \quad \forall \mathbf{c}\in\mathbb{C}^N,\ \forall S\subset[N],\ |S|\le n, \tag{7}$$

*where the constants $\rho \in (0,1)$, $\tau > 0$ depend only on $\delta_{2n}$.*

*Proof.* Let $\mathbf{c} \in \ker\mathbf{A} \setminus \{0\}$. We partition $[N]$ into the index sets

$$\begin{aligned} S_0 &:= J_n(\mathbf{c}) \\ S_1 &:= J_{2n}(\mathbf{c}) \setminus J_n(\mathbf{c}) \\ S_2 &:= J_{3n}(\mathbf{c}) \setminus J_{2n}(\mathbf{c}) \\ &\ \ \vdots \end{aligned}$$

where $J_n(\mathbf{c})$ is the index set of the largest $n$ entries of $\mathbf{c}$ in absolute value. To prove now that the `(rNSP)`, i.e. (7) holds, it is sufficient to show it for the extreme choice of $S = S_0 = J_n(\mathbf{c})$. We do this by using Lemma 1

$$
\begin{aligned}
\|\mathbf{c}_{S_0}\|_{\ell_2}^2 &\le \frac{1}{1-\delta_n}\|\mathbf{A}\mathbf{c}_{S_0}\|_{\ell_2}^2 = \frac{1}{1-\delta_n}\Big\langle \mathbf{A}\mathbf{c}_{S_0}, \mathbf{A}\mathbf{c} - \sum_{k\ge 1}\mathbf{A}\mathbf{c}_{S_k}\Big\rangle \\
&\le \frac{1}{1-\delta_n}\Big(|\langle \mathbf{A}\mathbf{c}_{S_0}, \mathbf{A}\mathbf{c}\rangle| + \sum_{k\ge 1}|\langle \mathbf{A}\mathbf{c}_{S_0}, \mathbf{A}\mathbf{c}_{S_k}\rangle|\Big) \\
&\le \frac{1}{1-\delta_n}\Big(\|\mathbf{A}\mathbf{c}_{S_0}\|_{\ell_2}\|\mathbf{A}\mathbf{c}\|_{\ell_2} + \delta_{2n}\sum_{k\ge 1}\|\mathbf{c}_{S_0}\|_{\ell_2}\|\mathbf{c}_{S_k}\|_{\ell_2}\Big) \\
&\le \frac{1}{1-\delta_n}\Big(\sqrt{\delta_n+1}\|\mathbf{c}_{S_0}\|_{\ell_2}\|\mathbf{A}\mathbf{c}\|_{\ell_2} + \delta_{2n}\|\mathbf{c}_{S_0}\|_{\ell_2}\cdot\frac{1}{\sqrt{n}}\sum_{k\ge 1}\|\mathbf{c}_{S_{k-1}}\|_{\ell_1}\Big).
\end{aligned}
$$

After division by $\|\mathbf{c}_{S_0}\|_{\ell_2}$ and Hölder's inequality, this yields

$$
\|\mathbf{c}_{S_0}\|_{\ell_2} \le \frac{\sqrt{1+\delta_{2n}}}{1-\delta_{2n}}\|\mathbf{A}\mathbf{c}\|_{\ell_2} + \frac{\delta_{2n}}{1-\delta_{2n}}\frac{1}{\sqrt{n}}\|\mathbf{c}_{S_0^C}\|_{\ell_1} + \frac{\delta_{2n}}{1-\delta_{2n}}\|\mathbf{c}_{S_0}\|_{\ell_2}.
$$

We used that fact that $\delta_{2n} \ge \delta_n$ to simplify the constants.

Now after using a bootstrapping argument, i.e. moving the last term to the left side and rearranging, we obtain

$$
\|\mathbf{c}_{S_0}\|_{\ell_2} \le \Big(1 - \frac{\delta_{2n}}{1-\delta_{2n}}\Big)^{-1}\frac{\delta_{2n}}{1-\delta_{2n}}\frac{1}{\sqrt{n}}\|\mathbf{c}_{S_0^C}\|_{\ell_1} + \Big(1 - \frac{\delta_{2n}}{1-\delta_{2n}}\Big)^{-1}\frac{\sqrt{1+\delta_{2n}}}{1-\delta_{2n}}\|\mathbf{A}\mathbf{c}\|_{\ell_2}.
$$

We set $\rho$ and $\tau$ accordingly and get the assertion. In particular we get $\rho \in (0,1)$ from the condition $\delta_{2n} < \frac{1}{3}$.

In what follows, we formulate the results only for the case of the multivariate trigonometric system

$$
\mathcal{B} = \mathcal{T}^d = \{\exp(2\pi \mathrm{i}\mathbf{k}\cdot) : \mathbf{k} \in \mathbb{Z}^d\}
$$

defined on the torus $\Omega = \mathbb{T}^d = [0,1]^d$. The improved RIP result below (Theorem 2) stays valid for other bounded orthonormal systems as shown in Brugiapaglia, Dirksen, H.C. Jung, Rauhut [5]. Here we are specifically interested in the multivariate Fourier system, which is why we rely on the result by Bourgain [4] and Haviv and Regev [13], see also [14, Theorem 2.16] for a discussion on the multivariate aspect. Note that one can consider a general system $\mathcal{B}$ in $L_2$ w.r.t. a probability measure and work then with a quasi-projection operator $P\colon L_2 \to L_2$ from [14, Definition 2.3]. For partial non-periodic setting which also works here see [14, Section 5].

In our setting of the Fourier system the operator $V_M$ takes the place of the quasi-projection $P$, where $V_M$ is the modified de la Valleé Poussin operator (see, e.g. [14, Sect. 3.1])

$$
V_M(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \hat{f}(\mathbf{k})\, v_{\mathbf{k}} \exp(2\pi \mathrm{i}\mathbf{k}\cdot\mathbf{x}), \tag{8}
$$

with weights $v_{\mathbf{k}} = \prod_{j=1}^{d} v_{k_j}$, where

$$v_{k_j} = \begin{cases} 1, & |k_j| \le M, \\ \frac{(2d+1)M - |k_j|}{2dM}, & M < |k_j| \le (2d+1)M, \\ 0, & |k_j| > (2d+1)M, \end{cases} \tag{9}$$

$\hat{f}(\mathbf{k}) = \int_{\mathbb{T}^d} f(\mathbf{x}) \exp(-2\pi \mathrm{i} \mathbf{k} \cdot \mathbf{x})\, \mathrm{d}\mathbf{x}$ are the Fourier coefficients.

This modified de la Valleé Poussin operator has a uniformly bounded operator norm. Indeed, from [14, Sect. 3.1] we obtain

$$\|V_M\|_{L_\infty \to L_\infty} \le \left(1 + \frac{1}{d}\right)^d \le e. \tag{10}$$

One may use an enumeration of $[-D, D]^d = \{\boldsymbol{k}_1, ..., \boldsymbol{k}_N\} \subset \mathbb{Z}^d$ with $N = (2D+1)^d$ and define the enumerated multivariate Fourier system as $\boldsymbol{e}_j(\cdot) := \exp(2\pi \mathrm{i} \boldsymbol{k}_j \cdot)$, $j = 1, ..., N$.

We will use points that are uniformly i.i.d. subsampled from the full grid $G(D, d) := \frac{\boldsymbol{\ell}}{2D} : \boldsymbol{\ell} \in \{0, \ldots, 2D\}^d$, see Figure 2. This is called in what follows a discrete uniform distribution $\mu_G = 1/|G| \sum_{\mathbf{x} \in G} \delta_{\mathbf{x}}$.

Let us prove the following statement which combines the robust recovery guarantee in H. Petersen and P. Jung [28, Theorem 3.1] using square root Lasso with the fact that RIP matrices of order $2n$ with sufficiently small RIP constant $\delta_{2n} < 1/3$ provide the robust $\ell_2$ null spaces property of order $n$, see Theorem 1.

**Theorem 2.** *There exist universal constants $\alpha, \beta, \gamma, \delta, \kappa > 0$ such that the following holds true. Let $D \in \mathbb{N}$, $N = (2D+1)^d$ and $n, m \in \mathbb{N}$ satisfy*

$$m \ge \alpha \cdot d \cdot n \cdot (\log(n+1))^2 \cdot \log(D+1). \tag{11}$$

*Put $\mathbf{A} = 1/\sqrt{m}(\boldsymbol{e}_j(\mathbf{x}^\ell))_{1 \le \ell \le m, 1 \le j \le N}$ for $\mathbf{x}^1, \ldots, \mathbf{x}^m \overset{iid}{\sim} \mu_G$. Then with probability at least $1 - N^{-\gamma \log(n+1)}$ with respect to the choice of $\mathbf{x}^1, \ldots, \mathbf{x}^m$ the following holds: Given any $\eta > 0$, $\mathbf{c} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^m$, and a solution $\mathbf{c}^\# \in \mathbb{C}^N$ of the* `(rLasso)` *minimization problem*

$$\inf_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\ell_1(N)} + \kappa\sqrt{n}\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_{\ell_2(m)} \tag{12}$$

*then*

$$\|\mathbf{c} - \mathbf{c}^\#\|_{\ell_1(N)} \le \beta\sigma_n(\mathbf{c})_1 + \delta\sqrt{n} \cdot \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_{\ell_2} \tag{13}$$

*and*

$$\|\mathbf{c} - \mathbf{c}^\#\|_{\ell_2(N)} \le \beta\frac{\sigma_n(\mathbf{c})_1}{\sqrt{n}} + \delta \cdot \|\mathbf{A}\mathbf{c} - \mathbf{y}\|_{\ell_2}, \tag{14}$$

*where*

$$\sigma_n(\mathbf{c})_{\ell_1} := \inf_{\mathbf{z} \in \mathbb{C}^N, \|\mathbf{z}\|_{\ell_0(N)} \le n} \|\mathbf{c} - \mathbf{z}\|_{\ell_1(N)},$$

*with $\|\mathbf{z}\|_{\ell_0(N)} := |\{1 \le j \le N : z_j \ne 0\}|$.*

Note that since $N \ge 2$, the number $1 - N^{-\gamma \log(n+1)}$ and therefore also the probability of choosing a vector of "good" sampling points $\mathbf{X} = \{\mathbf{x}^1, \ldots, \mathbf{x}^m\}$ is close to 1.

*Proof.* Choosing $\alpha$ large enough in (11) ensures that $\boldsymbol{A}$ has RIP of order $2n$ with RIP constant $\delta_{2n} < 1/3$, see [13, Theorem 3.7]. By Theorem 1 we have that $\boldsymbol{A}$ then satisfies the $\ell_2$-robust null space property (rNSP) of order $n$, which means that there are absolute constants $0 < \varrho < 1$ and $\tau > 0$ such that for all $\mathbf{c} \in \mathbb{C}^N$ and all $S \subset [N]$ with $|S| \leq n$

$$\|\mathbf{c}_S\|_{\ell_2} \leq \varrho n^{-1/2} \|\mathbf{c}_{S^c}\|_{\ell_1} + \tau \|\boldsymbol{A} \cdot \mathbf{c}\|_{(2)} ,$$

where $\|\cdot\|_{(2)} = \|\cdot\|_{\ell_2}$. This implies

$$\|\mathbf{c}_S\|_{\ell_1} \leq \varrho \|\mathbf{c}_{S^c}\|_{\ell_1} + \tau \|\boldsymbol{A} \cdot \mathbf{c}\|_{(1)} ,$$

with $\|\cdot\|_{(1)} = n^{1/2}\|\cdot\|_{\ell_2}$. It gives the $q$-robust NSP o f order $n$ for $q = 1$ and $q = 2$ with respect to norms $\|\cdot\|_{(q)}$ depending on $q$. According to Theorem 3.1 in [28] we now find a universal $\kappa > 0$ such that $\lambda = \kappa \cdot \sqrt{n}$ is a valid choice in (4) to get (13) and (14) simultaneously.

**Theorem 3.** *There exist universal constants $C, \alpha, \kappa, \gamma > 0$ such that the following holds. Let $M, n \in \mathbb{N}$ and put $D := (2d+1)M$. Setting $\lambda := \kappa\sqrt{n}$ and drawing at least*

$$m := \lceil \alpha \cdot d \cdot n \cdot (\log(n+1))^2 \cdot \log(D+1) \rceil$$

*sampling points $\mathbf{X} = \{\mathbf{x}^1, \ldots, \mathbf{x}^m\} \overset{iid}{\sim} \mu_G$, i.i.d. from the uniform measure on the grid it holds with probability at least $1 - N^{-\gamma \log(n+1)}$ for $2 \leq q \leq \infty$ that for any $f \in C(\mathbb{T}^d)$*

$$\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_q} \leq C n^{1/2-1/q} \cdot \Big( \sigma_n(f;\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d \cap \mathbb{Z}^d}(f;\mathcal{T}^d)_{L_\infty} \Big) ,$$

*where $R_{m,\lambda}$ denotes* (rLasso) *decoder from Definition 1 such that the approximant $R_{m,\lambda}(f;\mathbf{X})$ is contained in the space of trigonometric polynomials $\mathcal{T}([-(2d+1)N, (2d+1)N]^d)$.*

*Proof.* To prove Theorem 3 for $2 \leq q \leq \infty$ we first get the $L_\infty$-bound for the worst-case error and combine it via interpolation with the $L_2$-bound.

For the $L_\infty$ bound we will use the control over $\|\mathbf{c} - \mathbf{c}^{\#}\|_{\ell_1(N)}$ in Theorem 2, whereas the control on $\|\mathbf{c} - \mathbf{c}^{\#}\|_{\ell_2(N)}$ serves for the $L_2$ bound. Let $\varepsilon > 0$. Take an arbitrary $f \in \boldsymbol{F}$ and let $f^* = V_M s$, for $s$ such that $\|f - s\|_{L_\infty} \leq \sigma_n(f;\mathcal{T}^d)_{L_\infty} + \varepsilon$. The coefficient vector $\mathbf{c}$ of $f^*$ is $n$-sparse. We also set $\mathbf{y} = f(\mathbf{X})/\sqrt{m}$ and $\boldsymbol{e} = (f(\mathbf{X}) - f^*(\mathbf{X}))/\sqrt{m}$. Hence $\|\mathbf{A} \cdot \mathbf{c} - \mathbf{y}\|_{\ell_2} = \|\mathbf{e}\|_{\ell_2} \leq \|f(\mathbf{X}) - f^*(\mathbf{X})\|_{\ell_\infty}$. Then, taking into account the boundedness of the Fourier system, we have from Theorem 2

$$\begin{aligned} \|f^* - R_{m,\lambda}(f;\mathbf{X})\|_{L_\infty} &\leq \sum_{j=0}^{N} \big|(\mathbf{c}_j - \mathbf{c}_j^{\#}(\mathbf{y}))\big| \|e_j(\cdot)\|_{L_\infty} \leq \|\mathbf{c} - \mathbf{c}^{\#}\|_{\ell_1} \\ &\leq \beta \sigma_n(\mathbf{c})_1 + \delta \cdot \sqrt{n} \|\mathbf{A} \cdot \mathbf{c} - \mathbf{y}\|_{\ell_2} \\ &\leq \delta \cdot \sqrt{n} \|f(\mathbf{X}) - f^*(\mathbf{X})\|_{\ell_\infty} . \end{aligned} \tag{15}$$

Note that $\|f(\mathbf{X}) - f^*(\mathbf{X})\|_{\ell_\infty} \leq \|f - f^*\|_{L_\infty}$ therefore we get

$$\|f - f^*\|_{L_\infty} \leq \|f - V_M f\|_{L_\infty} + \|V_M f - f^*\|_{L_\infty} . \tag{16}$$

We need to bound both of these terms from above. This can be achieved by standard computations, that we decided to include for the convenience of the reader. Let $g \in \mathcal{T}\big([-M,M]^d\big)$ denote an arbitrary trigonometric polynomial. Clearly, $V_M g = g$ and therefore,

$$\begin{aligned}\|f - V_M f\|_{L_\infty} &= \|f - g + g - V_M f\|_{L_\infty} = \|f - g - V_M(f-g)\|_{L_\infty} \\ &\le \|f-g\|_{L_\infty} + \|V_M(f-g)\|_{L_\infty} \le (1+e)\|f-g\|_{L_\infty}.\end{aligned} \tag{17}$$

Taking the infimum over $g \in \mathcal{T}\big([-M,M]^d\big)$ yields

$$\|f - V_M f\|_{L_\infty} \le (1+e) E_{[-M,M]^d}(f). \tag{18}$$

Finally we can estimate the second term from (16), using $f^* = V_m s$ we get

$$\|V_M f - f^*\|_{L_\infty} = \|V_M f - V_M s\|_{L_\infty} \le e\|f - s\|_{L_\infty} = e\sigma_n(f;\mathcal{T}^d)_{L_\infty} + e\varepsilon\,. \tag{19}$$

We can combine (15), (16), (18) and (19) to obtain

$$\begin{aligned}\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_\infty} &\le \|f - f^*\|_{L_\infty} + \|f^* - R_{m,\lambda}(f;\mathbf{X})\|_{L_\infty} \\ &\le (\delta\sqrt{n}+1)\|f - f^*\|_{L_\infty} \\ &\le (\delta\sqrt{n}+1)\|f - V_M f\|_{L_\infty} + \|V_M f - f^*\|_{L_\infty} \\ &\le C\sqrt{n}\Big(\sigma_n(f;\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d\cap\mathbb{Z}^d}(f;\mathcal{T}^d)_{L_\infty}\Big)\,.\end{aligned} \tag{20}$$

We obtain the desired bound for $q = \infty$ in (20) letting $\varepsilon$ go to zero . The $L_2$-result is proven completely analogous ly. We use Parseval in (15) to step from the $L_2$-norm to the $\ell_2$-norm of the coefficients. Using the corresponding estimate in Theorem 2 we end up with

$$\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_2} \le C\left(\sigma_n(f;\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d\cap\mathbb{Z}^d}(f;\mathcal{T}^d)_{L_\infty}\right)\,.$$

By a standard interpolation argument we have that for all $f$ with $\|f\|_{\boldsymbol{F}} \le 1$ with probability at least $1 - N^{-\gamma\log(n+1)}$ it holds that

$$\begin{aligned}\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_q} &\le \|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_2}^{1-\theta}\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_\infty}^{\theta} \\ &= \|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_2}^{2/q}\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_\infty}^{1-2/q}\,,\end{aligned}$$

where the interpolation parameter $\theta$ has to be chosen in such a way that $1/q = (1-\theta)/2 + \theta/\infty$ which yields $\theta = 1 - 2/q$. This concludes the proof.

**Corollary 1.** *Under the assumptions of Theorem 3 with $M > d$ and for $2 \le q \le \infty$, it holds that*

$$\sup_{\|f\|_{\boldsymbol{F}}\le 1}\|f - R_{m,\lambda}(f;\mathbf{X})\|_{L_q} \le C\cdot n^{1/2-1/q}\cdot\left(\sigma_n(\boldsymbol{F};\mathcal{T}^d)_{L_\infty} + E_{[-M,M]^d}(\boldsymbol{F};\mathcal{T}^d)_{L_\infty}\right)\,,$$

*where $\lambda = \kappa\sqrt{n}$.*

## 4 Examples

We will now discuss examples where Corollary 1 improve s existing results in certain directions. We start in Subsection 4.1 with the mixed Wiener spaces $\mathcal{A}^r_{\mathrm{mix}}$, a generalization of the classical Wiener algebra $\mathcal{A}$. These spaces have been studied a lot due to their good embedding properties and their connection to Barron classes. Recent work on these spaces and their approximation properties was done by Jahn, T. Ullrich and Voigtlaender [14]; Kolomoitsev, Lomako, Tikhonov [15]; Krieg [16]; Moeller [22]; Moeller, Stasyuk and T. Ullrich [23]; V. K. Nguyen, V. N. Nguyen and Sickel [25] and others.

**Definition 3.** *For $r \geq 0$ we define the mixed Wiener space $\mathcal{A}^r_{\mathrm{mix}}$ of functions $f \in L_1(\mathbb{T}^d)$ with the finite norm*

$$\|f\|_{\mathcal{A}^r_{\mathrm{mix}}} := \sum_{\mathbf{k}\in\mathbb{Z}^d} \prod_{i=1}^{d} (1+|k_i|)^r |\hat{f}(\mathbf{k})|,$$

*where $\hat{f}(\mathbf{k})$ are the respective Fourier coefficients. For the univariate case we use the notation $\mathcal{A}^r$ since the n the smoothness is not mixed anymore. In the case $r = 0$ we get the Wiener algebra that will be denoted in what follows by $\mathcal{A}$.*

In Subsection 4.2 we investigate how and in which cases the (`rLasso`) can beat linear algorithms for spaces of functions with bounded mixed derivative defined in the following way. Define for $x \in \mathbb{T}$ and $r > 0$ the univariate Bernoulli kernel

$$F_r(x) := 1 + 2\sum_{k=1}^{\infty} k^{-r}\cos(2\pi k x) = \sum_{k\in\mathbb{Z}} \max\{1,|k|\}^{-r}\exp(2\pi \mathrm{i} k x)$$

and define the multivariate Bernoulli kernels as $F_r(\boldsymbol{x}) := \prod_{j=1}^d F_r(x_j), \boldsymbol{x} \in \mathbb{T}^d$.

**Definition 4.** *Let $r > 0$ and $1 < p < \infty$. Then $\mathbf{W}^r_p$ is defined as the normed space of all elements $f \in L_p(\mathbb{T}^d)$ which can be written as*

$$f = F_r * \varphi := \int_{\mathbb{T}^d} F_r(\cdot - \boldsymbol{y})\varphi(\boldsymbol{y})\, d\boldsymbol{y}$$

*for some $\varphi \in L_p(\mathbb{T}^d)$, equipped with the norm $\|f\|_{\mathbf{W}^r_p} := \|\varphi\|_{L_p(\mathbb{T}^d)}$.*

In order to prove the statements, we will use embeddings of $\mathcal{A}^r_{\mathrm{mix}}$ and $\mathbf{W}^r_p$ into the Besov spaces $\mathbf{B}^r_{p,\theta}$ of functions with bounded mixed differences.

**Definition 5.** *Let $r \geq 0$, $1 \leq \theta \leq \infty$, $1 < p < \infty$. Then the periodic Besov space $\mathbf{B}^r_{p,\theta}$ with mixed smoothness is defined as the normed space of all elements $f \in L_p(\mathbb{T}^d)$ endowed with the norm (with the usual modifications if $\theta = \infty$)*

$$\|f\|_{\mathbf{B}^r_{p,\theta}} := \Big(\sum_{\boldsymbol{s}\in\mathbb{N}_0^d} 2^{|\boldsymbol{s}|_{\ell_1} r\theta} \Big\| \sum_{\mathbf{k}\in\rho(\boldsymbol{s})} \hat{f}(\mathbf{k})\exp(2\pi\mathrm{i}\mathbf{k}\cdot\mathbf{x})\Big\|_p^{\theta}\Big)^{1/\theta},\ 1 \leq \theta < \infty,$$

*where*

$$\rho(\boldsymbol{s}) := \left\{\mathbf{k}\in\mathbb{Z}^d :\ \lfloor 2^{s_j-1}\rfloor \leq |k_j| < 2^{s_j},\ j = 1,\dots,d\right\},\ \boldsymbol{s}\in\mathbb{N}_0^d\,. \tag{21}$$

### 4.1 $L_q$-Recovery of Functions belonging to Mixed Wiener Spaces

**Corollary 2.** *Let $r > 1/2$ and $2 \le q \le \infty$. Then there is a constant $C_{r,d} > 0$ such that*

$$\varrho_{\lceil C_{r,d} n (\log(n+1))^3 \rceil}(\mathcal{A}^r_{\text{mix}})_{L_q} \lesssim n^{-(r+1/q)} (\log(n+1))^{(d-1)r+1/2}. \tag{22}$$

*Proof.* Using [14, Lemma 4.3] and choosing $M := \lfloor n^{(r+1/2)/r} \rfloor$ we obtain (22) as a direct consequence of Corollary 1.

*Remark 1.* The upper bound in Corollary 2 is sharp in the main order, which even coincides with the one for the Gelfand width. One can show this by using the good embedding properties of Wiener spaces and exact order estimates for the Gelfand numbers of the Besov spaces embeddings by Vybiral [35, Theorem 4.12]. Indeed,

$$\begin{aligned} \varrho_m(\mathcal{A}^r_{\text{mix}}(\mathbb{T}^d))_{L_q} \ge \varrho_m(\mathcal{A}^r(\mathbb{T}))_{L_q} \ge \varrho_m(B^{r+1/2}_{2,1})_{L_q} \ge \varrho_m(B^{r+1/2}_{2,1})_{B^0_{q,\infty}} \\ \ge c_m(B^{r+1/2}_{2,1})_{B^0_{q,\infty}} \asymp m^{-(r+1/q)}. \end{aligned} \tag{23}$$

In the first line of (23) we retreat to the one-dimensional setting.

*Remark 2 (Nonlinearity helps for $\mathcal{A}^r_{mix}$).*
If we compare this upper bound for non-linear approximation to lower bounds for linear approximation we can show how much better non-linear approximation is compared to linear approximation. Indeed [25, Theorem 4.7] states (in our notation, putting $r = 1$, $s = r$) that for $r > 0$ it holds that

$$\varrho^{\text{lin}}_m(\mathcal{A}^r_{\text{mix}}(\mathbb{T}^d))_{L_q} \gtrsim m^{-r} (\log m)^{(d-1)r}.$$

We have that the maximal possible difference in the rates is attained for $q = 2$ and the same main rate for $q = \infty$ when comparing linear and non-linear approximation of mixed Wiener spaces in $L_q$ spaces, since the difference between rates is always $1/q$.

The sharp upper bounds for a linear recovery from samples in a more general setting, in particular for the worst-case errors of recovery of functions from the weighted Wiener spaces by the Smolyak algorithm, were obtained in [15], see, e.g., Theorem 5.1 and Remark 6.4. In [17, Corollary 23] the upper bounds were proved for an algorithm that uses subsampled random points. They are sharp in the case $q = 2$, see also [17, Remark 24] for the comparison with the Smolyak algorithm.

### 4.2 Results for Functions with Bounded Mixed Derivative in $L_p$

In order to have access to function values we use the restriction $r > 1/p$ which implies that every equivalence class $f \in \mathbf{W}^r_p$ contains a continuous periodic function, see [8, Lemmas 3.4.1(iii) and 3.4.3]. Moreover, the embedding $\mathbf{W}^r_p \hookrightarrow C(\mathbb{T}^d)$ is then compact. The results below are partly mentioned in [14, Section 4.2]. Here we extend these results and give some further details. The overview of our findings concerning the optimality of different non-linear algorithms is presented in Figure 1. For a detailed comparison of linear recovery algorithms we refer to [18, Figure 1].

**Corollary 3 (Lower right region).** *Let $1 < p \le 2 \le q \le \infty$ and $r > 1/p$. Then there is a constant $C_{r,p,d} > 0$ such that*

$$\varrho_{\lceil C_{r,p,d} n (\log(n+1))^3 \rceil}(\mathbf{W}^r_p)_{L_q} \lesssim n^{-(r-\frac{1}{p}+\frac{1}{q})}(\log(n+1))^{(d-1)(r-2(\frac{1}{p}-\frac{1}{2}))+\frac{1}{2}}. \qquad (24)$$

*Proof.* From Corollary 1 and the argumentation from the proof of Corollary 4.14 in [14] (the class $\mathbf{W}^r_p$ is the same as $S^r_p W(\mathbb{T}^d)$ in their notation), choosing $N := 2^m$ with $n^{2r(r-1/p)^{-1}} \le N \le 2n^{2r(r-1/p)^{-1}}$ we get

$$\varrho_{\lceil C_{r,p,d}\, n \log(n+1)^3 \rceil}(\mathbf{W}^r_p)_{L_q} \le \widetilde{C} n^{1/2-1/q} \cdot \sigma_n(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}.$$

Combining this with the upper bound for the best $n$-term trigonometric approximation from [32, Thm. 2.9], we get (24).

*Remark 3 (Main rate sharp in Corollary 3).* One can show the sharpness of the main rate of convergence in Corollary 3 using the fooling argument from [26, Theorem 23] (for $d = 1$). Actually, the main rate $m^{-(r-(1/p-1/q))}$ is optimal for both linear and non-linear sampling recovery.

Note that in the region $1 < p < 2 < q < \infty$, the recovery from arbitrary linear information of functions from the class $\mathbf{W}^r_p$ in $L_q$ always outperforms (also non-linear) sampling recovery in the main rate. i.e., $\lambda_m(\mathbf{W}^r_p)_{L_q} = o(\varrho_m(\mathbf{W}^r_p)_{L_q})$ , see [7, Theorem 7.4].

Interestingly in the case $1/p + 1/q > 1$ the Gelfand widths $c_m(\mathbf{W}^r_p)_{L_q}$ decay faster in the main rate than the respective linear widths $\lambda_m(\mathbf{W}^r_p)_{L_q}$. For $1/p + 1/q \le 1$ it holds that $c_m(\mathbf{W}^r_p)_{L_q} \asymp \lambda_m(\mathbf{W}^r_p)_{L_q}$.

Let us compare the bound for `(rLasso)` from Corollary 3 with those for other recovery methods. Here we assume that $1 < p < 2 < q < \infty$, the case $q = 2$ will be discussed separately.

*Remark 4. (i) [Comparison to* `(Smolyak)`*]* In the paper [7, Cor. 7.1] an upper bound for the linear sampling numbers of $\mathbf{W}^r_p$ (the same as $S^r_{p,2}F(\mathbb{T}^d)$ with $\mu = d$ in their notation) in $L_q$ has been given for the worst-case recovery using the Smolyak algorithm, which for $r > 1/p$, $1 < p < q < \infty$ yields that

$$\varrho^{\mathrm{lin}}_m(\mathbf{W}^r_p)_{L_q} \lesssim m^{-(r-\frac{1}{p}+\frac{1}{q})}(\log m)^{(d-1)(r-\frac{1}{p}+\frac{1}{q})}. \qquad (25)$$

By the embedding $\mathbf{B}^r_{p,p} \hookrightarrow \mathbf{W}^r_p$ in case $1 < p < 2 < q < \infty$ together with [9, Thm. 5.1,(ii)] we know that we cannot do better in $L_q$ than in (25) if we restrict ourselves to sparse grid points. Hence, our non-linear approach outperforms sparse grids if $d$ is large and

$$2(1/p - 1/2) > 1/p - 1/q \Longleftrightarrow 1/p + 1/q > 1\,.$$

*(ii) [Comparison to* `(Lsqr)`*]* In [18, Cor. 21] we obtain (25) for $1 < p < 2 < q < \infty$ also with a different linear method, namely a weighted least squares estimator based on subsampled random points involving the solution of the Kadison-Singer problem [21]. We do not know if the bound given there is sharp and whether it may outperform `(rLasso)`.

*Remark 5 ($L_2$-estimates outperform any linear method).* In [14, Corollary 4.16] it was proven that for $1 < p < 2$ and $r > 1/p$ it holds that

$$\varrho_{\lceil C_{r,p,d} n(\log(n+1))^3 \rceil}(\mathbf{W}^r_p)_{L_2} \lesssim n^{-r+\frac{1}{p}-\frac{1}{2}}(\log(n+1))^{(d-1)(r-\frac{2}{p}+1)+\frac{1}{2}}.$$

As mentioned in [14, Remark 4.17], for sufficiently large $d$ the non-linear sampling numbers decay faster in this situation than the respective linear widths, which coincide in the order of decay with the linear sampling numbers.

Let us proceed with the case $p > 2$.

**Corollary 4 (Left region including small smoothness).** *Let $2 \le p < \infty$, $1 \le q < \infty$. Then there is a constant $C_{r,p,d} > 0$ such that with $m = \lceil C_{r,p,d} n(\log(n+1))^3 \rceil$*

$$\varrho_m(\mathbf{W}^r_p)_{L_q} \lesssim \begin{cases} n^{-(r-(\frac{1}{2}-\frac{1}{q})_+)}(\log(n+1))^{(d-1)(1-r)+r}, 1/p < r < 1/2, \\ n^{-(r-(\frac{1}{2}-\frac{1}{q})_+)}(\log(n+1))^{(d-1)(1-r)+r}(\log\log n)^{r+1}, r = 1/2, \\ n^{-(r-(\frac{1}{2}-\frac{1}{q})_+)}(\log n)^{(d-1)r+\frac{1}{2}}, r > 1/2. \end{cases} \tag{26}$$

*Proof.* Since $\|\cdot\|_{L_q} \le \|\cdot\|_{L_2}$ for $q \le 2$, it suffices to consider the case $2 \le q < \infty$. Further, in order to use Corollary 1, we need upper estimates for the quantities $\sigma_n(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}$ and $E_{[-M,M]^d \cap \mathbb{Z}^d}(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}$. The rate of convergence of the respective best $n$-term approximation width for $2 \le p < \infty$ is

$$\sigma_n(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty} \lesssim \begin{cases} n^{-r}(\log n)^{(d-1)(1-r)+r}, \ 1/p < r < 1/2, \\ n^{-r}(\log n)^{(d-1)(1-r)+r}(\log\log n)^{r+1}, \ r = 1/2, \\ n^{-r}(\log n)^{(d-1)r+1/2}, \ r > 1/2. \end{cases} \tag{27}$$

The case of small smoothness ($1/p < r \le 1/2$) is known from [34, Theorems 6.1, 6.2], the big smoothness case ($r > 1/2$) is taken from [32, Theorem 1.3], see also [8, Theorem 7.5.2].

In what follows we show that for an appropriately chosen $M = M(n, r, p)$, the quantity $E_{[-M,M]^d \cap \mathbb{Z}^d}(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}$ has a faster rate of convergence than the respective best $n$-term approximation, see Lemma 2 below.

Hence, Corollary 1 yields the estimate

$$\varrho_{\lceil C_{r,p,d} n(\log(n+1))^3 \rceil}(\mathbf{W}^r_p)_{L_q} \le 2n^{1/2-1/q} \cdot \sigma_n(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}.$$

To conclude the proof, we use (27).

**Lemma 2.** *Let $M \in \mathbb{N}$, $2 \le p < \infty$ and $r > 1/p$. Then it holds that*

$$E_{[-M,M]^d \cap \mathbb{Z}^d}(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty} \lesssim M^{-(r-\frac{1}{p})}.$$

*In addition, for $M$ such that $n^{2r(r-1/p)^{-1}} \le M \le 2n^{2r(r-1/p)^{-1}}$ it holds that*

$$E_{[-M,M]^d \cap \mathbb{Z}^d}(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty} \lesssim n^{-r} \lesssim \sigma_n(\mathbf{W}^r_p; \mathcal{T}^d)_{L_\infty}. \tag{28}$$

*Proof.* By the embedding $\mathbf{W}^r_p \hookrightarrow \mathbf{B}^r_{p,p}$, $p \geq 2$, and the Nikol'skii inequality, we get

$$
\begin{aligned}
E_{[-M,M]^d\cap\mathbb{Z}^d}(\mathbf{W}^r_p;\mathcal{T}^d)_{L_\infty} &\leq \sup_{\|f\|_{\mathbf{B}^r_{p,p}}\leq 1} \inf_{\mathbf{k}\in\mathbb{Z}^d\setminus[-M,M]^d} \|\hat f(\mathbf{k})\exp(2\pi i\mathbf{k}\cdot\mathbf{x})\|_{L_\infty} \\
&\leq \sup_{\|f\|_{\mathbf{B}^r_{p,p}}\leq 1} \sum_{\boldsymbol{s}\in\mathbb{N}_0^d,\exists s_j:\,2^{s_j}>M} 2^{\frac{|\boldsymbol{s}|_1}{p}} \Big\| \sum_{\mathbf{k}\in\rho(\boldsymbol{s})} \hat f(\mathbf{k})\exp(2\pi i\mathbf{k}\cdot\mathbf{x})\Big\|_p, \qquad (29)
\end{aligned}
$$

where the blocks $\rho(\boldsymbol{s})$ are defined in (21).

In what follows we use Hölder's inequality and obtain

$$
\begin{aligned}
&\sup_{\|f\|_{\mathbf{B}^r_{p,p}}\leq 1} \sum_{\boldsymbol{s}\in\mathbb{N}_0^d,\exists s_j:\,2^{s_j}>M} 2^{-|\boldsymbol{s}|_1(r-\frac{1}{p})} \Big\| \sum_{\mathbf{k}\in\rho(\boldsymbol{s})} \hat f(\mathbf{k})\exp(2\pi i\mathbf{k}\cdot\mathbf{x})\Big\|_p 2^{r|\boldsymbol{s}|_1} \\
&\leq M^{-(r-\frac{1}{p})} \left( \sum_{\boldsymbol{s}\in\mathbb{N}_0^d} 2^{-|\boldsymbol{s}|_1(r-\frac{1}{p})(1-\frac{1}{p})} \right)^{1-\frac{1}{p}} \sup_{\|f\|_{\mathbf{B}^r_{p,p}}\leq 1} \|f\|_{\mathbf{B}^r_{p,p}} \lesssim M^{-(r-\frac{1}{p})}.
\end{aligned}
$$

Choosing the parameter $M$ such that $n^{2r(r-1/p)^{-1}} \leq M \leq 2n^{2r(r-1/p)^{-1}}$ implies (28).

*Remark 6 (Left upper square ($1 < q \leq 2 \leq p < \infty$)). (i) (*`(rLasso)` *is almost sharp)* The order of Gelfand widths in this region is $c_m(\mathbf{W}^r_p)_{L_q} \asymp \lambda_m(\mathbf{W}^r_p)_{L_q} \asymp m^{-r}(\log m)^{(d-1)r}$ (see, e.g. [8, Section 9.6]). With `(rLasso)` we obtain the same main rate of convergence but additional ($d$-independent) logarithms, i.e. this non-linear method is almost optimal w.r.t. Gelfand numbers.

*(ii) (Comparison to* `(Lsqr)` *)* The sharp (w.r.t. Gelfand numbers) bound for `(Lsqr)` in the case $1 < q < 2 < p < \infty$ was obtained in [18, Cor. 21]. Note that the approach in [18] required a square summability of linear width, and cover s only the case $r > 1/2$, whereas in [17] and [33] this condition can be avoided by paying a $d$-independent logarithm.

*(iii) (Comparison to* `(Smolyak)`*)* In the considered region the right order for `(Smolyak)` is $m^{-r}(\log m)^{(d-1)(r+1/2)}$ (see [8, Thm. 5.3.1] and references therein). In fact, by the embedding $\mathbf{B}^r_{p,2} \hookrightarrow \mathbf{W}^r_p$ together with [9, Thm. 5.1,(ii)] we know that we cannot do better in $L_q$ if we restrict ourselves to sparse grid points. This estimate is worse in logarithms than those for `(rLasso)` for large $d$.

*Remark 7 (Left lower square ($2 \leq p, q < \infty$)).* We will distinguish two cases: $2 \leq p < q < \infty$ (lower triangular) and $2 < q < p < \infty$ (upper triangular).

*(i)* In the region $2 \leq p < q < \infty$ `(Smolyak)` achieves the exact order w.r.t. Gelfand numbers (see [7, Cor. 7.1]) It is better in the main rate than the bound for `(rLasso)` (which is in turn better than `(Lsqr)` from [18, Cor. 21]). Note that for $p = 2 < q \leq \infty$ `(Lsqr)` gives the same (sharp) order of decay as `(Smolyak)`.

*(ii)* For $2 < q < p < \infty$ we do not know anything about the optimality of (linear and non-linear) sampling algorithms. The known upper bounds for `(rLasso)` decay slower in the main rate than the respective Gelfand numbers. In turn the existing upper bounds for `(Smolyak)` and `(Lsqr)` are worse than those for the linear widths. Note that in this region Gelfand numbers decay faster than linear widths in the main rate.

*Remark 8 (Right upper region ($1 < p, q < 2$)). (i)* In the whole region the upper bound for `(rLasso)`, derived from Corollary 3 putting $q = 2$, does not achieve the right order of Gelfand (neither linear) numbers. We note that this fact alone, without having a matching lower bound for `(rLasso)`, does not lead to the nonoptimality of the discussed method. It is rather an open question.

*(ii)* To discuss other algorithms let us consider separately two triangular areas: $1 < p < q < 2$ (lower triangular) and $1 < q \leq p < 2$ (upper triangular). For $1 < p < q < 2$ the bound for `(Smolyak)` [7, Cor. 7.1] coincides with the order for linear widths and is better than the known upper bound for `(Lsqr)` [18] (note that the respective Gelfand numbers decay faster in this region). In the case $1 < q \leq p < 2$ we cannot say anything about the optimality of `(Smolyak)` or `(Lsqr)` w.r.t. neither Gelfand nor linear widths.

**Acknowledgement.** The first named author MM is supported by the ESF, being co-financed by the European Union and from tax revenues on the basis of the budget adopted by the Saxonian State Parliament. KP would like to acknowledge support by the Philipp Schwartz Fellowship of the Alexander von Humboldt Foundation, the German Research Foundation (DFG 403/4-1) and the budget program of the NAS of Ukraine "Support for the development of priority areas of research" (KPKVK 6541230). KP and TU would like to thank Ben Adcock for pointing out reference [5] and bringing square root Lasso as a noise blind alternative to the authors' attention during a discussion within the Session *Function recovery and discretization problems* organized by David Krieg and Kateryna Pozharska at the conference MCQMC24 in Waterloo (CA).

## References

1. Adcock, B., Bao, A., Brugiapaglia, S.: Correcting for unknown errors in sparse high-dimensional function approximation. Numer. Math. **142**(3), 667–711 (2019)
2. Bartel, F., Schäfer, M., Ullrich, T.: Constructive subsampling of finite frames with applications in optimal function recovery. Appl. Comput. Harmon. Anal. **65**, 209–248 (2023). DOI 10.1016/j.acha.2023.02.004
3. Belinskii, É.S.: Approximation of functions of several variables by trigonometric polynomials with given number of harmonics, and estimates of $\epsilon$-entropy. Analysis Mathematica **15**, 67–74 (1989)
4. Bourgain, J.: An improved estimate in the restricted isometry problem pp. 65–70 (2014). DOI 10.1007/978-3-319-09477-9_5. In: Klartag, B., Milman, E. (eds) Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics, vol. 2116. Springer, Cham.
5. Brugiapaglia, S., Dirksen, S., Jung, H.C., Rauhut, H.: Sparse recovery in bounded Riesz systems with applications to numerical methods for pdes. Applied and Computational Harmonic Analysis **53**, 231–269 (2021). DOI 10.1016/j.acha.2021.01.004
6. Bungartz, H.J., Griebel, M.: Sparse grids. Acta Numerica **13**, 147–269 (2004). DOI 10.1017/S0962492904000182
7. Byrenheid, G., Ullrich, T.: Optimal sampling recovery of mixed order Sobolev embeddings via discrete Littlewood–Paley type characterizations. Analysis Mathematica **43**, 807–820 (2017)
8. Dũng, D., Temlyakov, V.N., Ullrich, T.: Hyperbolic cross approximation. Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, Cham (2018). DOI 10.1007/978-3-319-92240-9

9. Dũng, D., Ullrich, T.: Lower bounds for the integration error for multivariate functions with mixed smoothness and optimal Fibonacci cubature for functions on the square. Math. Nachr. **288**(7), 743–762 (2015). DOI 10.1002/mana.201400048
10. Dai, F., Temlyakov, V.N.: Random points are good for universal discretization. J. Math. Anal. Appl. **529**(1), Paper No. 127,570, 28 (2024). DOI 10.1016/j.jmaa.2023.127570
11. Dolbeault, M., Krieg, D., Ullrich, M.: A sharp upper bound for sampling numbers in $L_2$. Appl. Comput. Harmon. Anal. **63**, 113–134 (2023). DOI 10.1016/j.acha.2022.12.001
12. Foucart, S., Rauhut, H.: A mathematical introduction to compressive sensing. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York (2013). DOI 10.1007/978-0-8176-4948-7
13. Haviv, I., Regev, O.: The restricted isometry property of subsampled Fourier matrices. In: Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016, pp. 163–179. Springer (2017)
14. Jahn, T., Ullrich, T., Voigtlaender, F.: Sampling numbers of smoothness classes via $\ell^1$-minimization. Journal of Complexity **79**, Paper No. 101,786, 35 (2023). DOI 10.1016/j.jco.2023.101786
15. Kolomoitsev, Y., Lomako, T., Tikhonov, S.: Sparse grid approximation in weighted Wiener spaces. Journal of Fourier Analysis and Applications **2**, 19–29 (2023). DOI https://doi.org/10.1007/s00041-023-09994-2
16. Krieg, D.: Tractability of sampling recovery on unweighted function classes. Proc. Amer. Math. Soc. Ser. B **11**, 115–125 (2024). DOI 10.1090/bproc/216
17. Krieg, D., Pozharska, K., Ullrich, M., Ullrich, T.: Sampling projections in the uniform norm. arXiv:math/2401.02220 (2024)
18. Krieg, D., Pozharska, K., Ullrich, M., Ullrich, T.: Sampling recovery in $L_2$ and other norms. arXiv:math/2305.07539 (2024)
19. Krieg, D., Ullrich, M.: Function values are enough for $L_2$-approximation. Found. Comput. Math. **21**(4), 1141–1151 (2021). DOI 10.1007/s10208-020-09481-w
20. Limonova, I., Temlyakov, V.N.: On sampling discretization in $L_2$. Journal of Mathematical Analysis and Applications (2021)
21. Marcus, A.W., Spielman, D.A., Srivastava, N.: Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. Ann. of Math. (2) **182**(1), 327–350 (2015). DOI 10.4007/annals.2015.182.1.8
22. Moeller, M.: Gelfand numbers and best $m$-term trigonometric approximation for weighted mixed Wiener classes in $L_2$. Master's thesis, TU Chemnitz, Germany (2023)
23. Moeller, M., Stasyuk, S., Ullrich, T.: High-dimensional sparse trigonometric approximation in the uniform norm and consequences for sampling recovery. arxiv:2407.15965 (2024)
24. Nagel, N., Schäfer, M., Ullrich, T.: A new upper bound for sampling numbers. Found. Comput. Math. **22**(2), 445–468 (2022). DOI 10.1007/s10208-021-09504-0
25. Nguyen, V.D., Nguyen, V.K., Sickel, W.: $s$-numbers of embeddings of weighted Wiener algebras. J. Approx. Theory **279**, 26 (2022). DOI 10.1016/j.jat.2022.105745. Id/No 105745
26. Novak, E., Triebel, H.: Function spaces in Lipschitz domains and optimal rates of convergence for sampling. Constructive Approximation **23**, 325–350 (2005)
27. Novak, E., Woźniakowski, H.: Tractability of multivariate problems. Vol. 1: Linear information, *EMS Tracts in Mathematics*, vol. 6. European Mathematical Society (EMS), Zürich (2008). DOI 10.4171/026
28. Petersen, H.B., Jung, P.: Robust instance-optimal recovery of sparse signals at unknown noise levels. Inf. Inference **11**(3), 845–887 (2022). DOI 10.1093/imaiai/iaab015
29. Saucedo, M., Tikhonov, S.: Kahane-Katznelson-de Leeuw theorem and absolute convergence of Fourier series. arXiv:2503.14133 (2025)
30. Smolyak, S.A.: Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR **148**, 1042–1045 (1963)

31. Temlyakov, V.N.: Approximation of Periodic Functions. Computational mathematics and analysis series. Nova Science Publishers (1993)
32. Temlyakov, V.N.: Constructive sparse trigonometric approximation and other problems for functions with mixed smoothness. Sbornik: Mathematics **206**(11), 1628 (2015). DOI 10.1070/SM2015v206n11ABEH004507
33. Temlyakov, V.N., Ullrich, T.: Bounds on Kolmogorov widths and sampling recovery for classes with small mixed smoothness. Journal of Complexity **67**, 101,575 (2021). DOI https://doi.org/10.1016/j.jco.2021.101575
34. Temlyakov, V.N., Ullrich, T.: Approximation of functions with small mixed smoothness in the uniform norm. Journal of Approximation Theory **277**, 105,718 (2022). DOI 10.1016/j.jat.2022.105718
35. Vybíral, J.: Widths of embeddings in function spaces. Journal of Complexity **24**(4), 545–570 (2008). DOI 10.1016/j.jco.2008.01.002