

Design Techniques for Multi-Core Neural Network Accelerators on Radiation-Hardened FPGAs

*Original*

Design Techniques for Multi-Core Neural Network Accelerators on Radiation-Hardened FPGAs / Portaluri, A., Azimi, S., Sterpone, L. - ELETTRONICO. - (2023), pp. 16-22. (IEEE International Symposium on Parallel and Distributed Computing Bucharest (Romania) 10-12 July 2023) [10.1109/ISPDC59212.2023.00023].

*Availability:*

This version is available at: 11583/2979316 since: 2023-06-12T09:09:28Z

*Publisher:*

IEEE

*Published*

DOI:10.1109/ISPDC59212.2023.00023

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

IEEE postprint/Author's Accepted Manuscript

©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

# COMPRESSED SENSING FOR INVERSE PROBLEMS AND THE SAMPLE COMPLEXITY OF THE SPARSE RADON TRANSFORM

GIOVANNI S. ALBERTI, ALESSANDRO FELISI, MATTEO SANTACESARIA,  
AND S. IVAN TRAPASSO

**ABSTRACT.** Compressed sensing allows for the recovery of sparse signals from few measurements, whose number is proportional to the sparsity of the unknown signal, up to logarithmic factors. The classical theory typically considers either random linear measurements or subsampled isometries and has found many applications, including accelerated magnetic resonance imaging, which is modeled by the subsampled Fourier transform. In this work, we develop a general theory of infinite-dimensional compressed sensing for abstract inverse problems, possibly ill-posed, involving an arbitrary forward operator. This is achieved by considering a generalized restricted isometry property, and a quasi-diagonalization property of the forward map.

As a notable application, for the first time, we obtain rigorous recovery estimates for the sparse Radon transform (i.e., with a finite number of angles  $\theta_1, \dots, \theta_m$ ), which models computed tomography, in both the parallel-beam and the fan-beam settings. In the case when the unknown signal is  $s$ -sparse with respect to an orthonormal basis of compactly supported wavelets, we prove stable recovery under the condition

$$m \gtrsim s,$$

up to logarithmic factors.

---

*Date:* May 6, 2024.

*2020 Mathematics Subject Classification.* 42C40; 44A12; 60B20; 94A20.

*Key words and phrases.* Radon transform, infinite-dimensional compressed sensing, sparse recovery, sample complexity, wavelets, quasi-diagonalization, coherence, inverse problems.

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA8655-20-1-7027. We acknowledge the support of Fondazione Compagnia di San Paolo. Co-funded by the European Union (ERC, SAMPDE, 101041040). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The authors are members of the “Gruppo Nazionale per l’Analisi Matematica, la Probabilità e le loro Applicazioni”, of the “Istituto Nazionale di Alta Matematica”.

## CONTENTS

1. Introduction	2
1.1. State of the art	2
1.2. Main contributions	3
1.3. Related research	8
1.4. Discussion and open questions	10
1.5. Structure of the paper	11
2. Setup	12
2.1. Weighted setup and notation	12
2.2. Dictionaries, measurements and the minimization problem	13
3. A general result on sample complexity	16
3.1. Main result	16
3.2. Quantitative bounds on $\ G^{-1}\ $ and quasi-diagonalization	18
3.3. Analysis of the recovery error	23
4. The sparse Radon transform	24
5. Proofs of the main results	29
5.1. Preliminary results	29
5.2. The robust null space property	30
5.3. The generalized restricted isometry property	31
5.4. Sufficient conditions for the g-RIP via subsampling	33
5.5. Proof of Theorem 3.5	34
5.6. Quantitative bounds on $\ G^{-1}\ $	38
5.7. The sparse Radon transform	42
References	45
Appendix A. Proof of Theorem 5.7	50
A.1. The case $f_\nu = 1$	50
A.2. The general case	57

## 1. INTRODUCTION

1.1. **State of the art.** Compressed sensing (CS) [38, 48, 51, 54, 52] has been a very active research area within applied mathematics over the last two decades, characterized by a fruitful exchange between theoretical analysis and technological advances. The focus of CS is the recovery of sparse signals, namely those that can be expressed as a linear combination of a small number of vectors from a reference dictionary. Leveraging sparsity allows for obtaining stable recovery results with a smaller number of measurements than in standard sampling theory.

Over the years, in CS theory there has been a transition from the finite-dimensional setting to the infinite-dimensional one [8, 5, 1, 14, 9]. This shift was largely stimulated

by the need to encompass increasingly challenging sensing tasks, which happen to be more faithfully framed in an analog world in terms of suitable operators and function spaces. In fact, the analysis of problems where one aims at recovering a physical quantity given some indirect measurements is the main focus of the theory of inverse problems (IP), with a long tradition in mathematical and numerical analysis [73, 82, 92, 65]. Most IP are described by integral or differential equations, and are therefore naturally modeled in an infinite-dimensional setting.

The reconstruction guarantees currently available in the mathematical literature on IP rely on the assumption that an observer has access to an arbitrarily large number of measurements, except for a limited number of special cases (see [13, 63, 15, 11] and references therein, in which, however, no sparsity information is exploited, but only the finite-dimensionality of the unknown). It is clear that in real-life scenarios one can only take a finite number of measurements, usually corrupted by noise and also suffering from limited control on the procedure due to experimental constraints. This discussion calls for the development of a *sample complexity* theory of ill-posed IP, namely, a study of the minimum (finite) number of measurements needed to achieve stable and accurate reconstruction.

It is intuitively clear that the principles of CS could play a role here. Indeed, if the unknown quantity is sparse in some dictionary, then one could approximately locate the signal in some finite-dimensional subspace of the domain of the measurement operator. In other words, a sparsity condition comes along with an inherently finite-dimensional setup. Not surprisingly, such a reduction step is quite delicate and may eventually destroy important information on the genuine infinite-dimensional structure. The framework of *generalized sampling* [6, 8, 7, 5] addresses this issue, as it relies on finding appropriate sampling and sparsity truncation bandwidths in order to perform the finite-dimensional reduction in a stable way.

Although ideas from CS have influenced the analysis of IP (e.g., sparse regularization [17, 59, 58]), the state of the art is still far from being completely satisfactory, mostly due to technical limitations and a case-by-case approach aimed at leveraging the peculiar features of each model. Despite a large body of literature on the numerical analysis of sparsity-promoting approaches for inverse problems (see, e.g., [66, 70, 55]), rigorous theoretical guarantees are missing (see [64, 14, 12] for some results in this direction).

## 1.2. Main contributions.

1.2.1. *Summary.* The purpose of the present article is to move the first steps towards a unified framework for the study of the sample complexity of ill-posed IP under realistic sparsity constraints, by extending the current theory of CS. The main goal is to find the connection between the number of samples needed to achieve recovery of a signal and its sparsity with respect to a suitable dictionary.

In addition to the formulation of the general framework for studying the sample complexity, we present an application to the sparse angle tomography problem. Basically, the goal is to obtain information on the inner structure of an object from the knowledge of a limited number of projections obtained via the Radon transform. Tomographic techniques appear in a variety of real-life situations, including non-destructive testing, biomedical and geophysical imaging, and analysis of astronomical signals. The analysis of tomography IP is a well-established topic and still a very active research area [83, 84, 93, 92, 62].

While the sparse Radon transform has been thoroughly studied from the numerical point of view and ideas of CS have found applications to tomographic imaging shortly after its appearance [94, 71, 61, 70, 69], rigorous results on the sample complexity for the Radon transform and the related technical challenges have been beyond the reach of the CS theory so far. Our main contribution in this respect is to show that signals supported in the unit ball of the plane, assumed to be compressible in a suitable wavelet dictionary, can be reconstructed with high probability if a sufficiently large number of samples of the corresponding Radon transform at different angles is available – see Theorems 4.4, 4.5 and 4.6 below. We consider both the parallel-beam and the fan-beam settings. In simplified terms, we obtain the estimate

$$\text{number of angles} \gtrsim \text{sparsity},$$

up to logarithmic factors.

Let us briefly illustrate the key aspects of both the abstract sample complexity framework and its application to the sparse Radon inversion.

*1.2.2. Sample complexity of sparse ill-posed IP.* Let  $\mathcal{H}_1, \mathcal{H}_2$  be separable complex Hilbert spaces. We assume that the measurements are represented by a family of uniformly bounded linear maps  $(F_t)_{t \in \mathcal{D}}: \mathcal{H}_1 \rightarrow \mathcal{H}_2$ , indexed over a measure space  $(\mathcal{D}, \mu)$  endowed with a probability distribution  $\nu$  such that  $d\nu = f_\nu d\mu$  for a density function  $f_\nu \in L^1(\mu)$ ,  $\|f_\nu\|_{L^1} = 1$ , satisfying  $f_\nu \geq c_\nu$  for some  $0 < c_\nu \leq 1$ . The forward map  $F: \mathcal{H}_1 \rightarrow L^2_\mu(\mathcal{D}; \mathcal{H}_2)$  of the model is thus naturally defined by the relationship  $F_t u = F u(t)$ , for almost every  $t \in \mathcal{D}$ . This setting encompasses both scalar measurements given by inner products with respect to a certain dictionary and by pointwise evaluations. Moreover, we consider measurements taking values in a Hilbert space; this is related to the use of block-sampling strategies in CS [86, 20].

We fix once for all an orthonormal basis  $(\phi_i)_{i \in \Gamma}$  of  $\mathcal{H}_1$  as a reference dictionary, and let  $\Phi: \mathcal{H}_1 \rightarrow \ell^2(\Gamma)$  be the corresponding analysis operator – here  $\Gamma$  is an (at most) countable index set, e.g.,  $\Gamma = \mathbb{N}$  or  $\Gamma \subset \mathbb{N} \times \mathbb{Z}$ . The signals we consider will be sparse with respect to  $(\phi_i)_i$ .

The dictionary, the measurement operators and the probability density are assumed to be intertwined by a *coherence bound* of the form

$$\|F_t \phi_i\|_{\mathcal{H}_2} \leq B \sqrt{f_\nu(t)}, \quad t \in \mathcal{D}, i \in \Gamma,$$

for some  $B \geq 1$ .

Consider now an unknown signal  $u^\dagger \in \mathcal{H}_1$ . We assume that there exists a finite subset of indices  $\Lambda \subset \Gamma$ , with  $|\Lambda| = M$ , such that  $\|P_\Lambda^\perp x^\dagger\|$  is small, where we set  $x^\dagger := \Phi u^\dagger$  and  $P_\Lambda$  denotes the orthogonal projection on  $\text{span}\{e_i : i \in \Lambda\}$ ,  $(e_i)_{i \in \Gamma}$  being the canonical basis of  $\ell^2(\Gamma)$ . Therefore,  $\Lambda$  should be thought of as a rough estimate of the support of  $x^\dagger$ . One could consider more general forms of tail decay rates for  $\|P_\Lambda^\perp x^\dagger\|_2$ , which are usually linked to the regularity of  $u^\dagger$  and  $\Phi$ . We will exploit this connection in Sections 3 and 4 in order to discuss applications to relevant signal classes. For simplicity, here we assume  $P_\Lambda^\perp x^\dagger = 0$ .

Given  $m$  i.i.d. samples  $t_1, \dots, t_m \in \mathcal{D}$  drawn from  $\nu$ , the corresponding measurements are

$$y_k = F_{t_k} u^\dagger + \varepsilon_k, \quad k = 1, \dots, m,$$

where the noise vector  $\varepsilon \in \mathcal{H}_2^m$  is such that  $\|\varepsilon_k\|_{\mathcal{H}_2} \leq \beta$  for some  $\beta > 0$ , for every  $k$  – in other words, we have that the measurement indexed by  $t_k$  comes with a noise  $\varepsilon_k$  that is uniformly bounded over the samples.

We attempt to reconstruct  $x^\dagger$  (and, consequently,  $u^\dagger = \Phi^* x^\dagger$ ) via the  $\ell^1$ -minimization problem

$$(1) \quad \min_{x \in \ell^2(\Lambda)} \|x\|_1 \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* x - y_k\|_{\mathcal{H}_2}^2 \leq \beta^2,$$

The transition to a finite-dimensional setup is performed via a stable truncation of the forward map  $F$  of the model. To be more precise, we consider the  $M \times M$  matrix given by

$$G = \sqrt{P_\Lambda \Phi F^* F \Phi^* P_\Lambda^*}.$$

Stability ultimately reduces to the invertibility of  $G$ , which can be ensured if  $F$  satisfies a restricted injectivity assumption known as the FBI property, which is usually met in many IP of interest – for example whenever the operator  $F$  is injective, see Definition 3.6 below for further details. More generally, this condition can be relaxed if a form of elastic net regularization is introduced in the minimization problem (1), so that  $G$  can be replaced by  $G_\alpha = \sqrt{P_\Lambda \Phi F^* F \Phi^* P_\Lambda^* + \alpha^2 I_\Lambda}$  for  $\alpha > 0$ , so that the invertibility of  $G_\alpha$  is no longer a concern. For simplicity, this generalization will not be treated in this work.

A largely simplified version of our abstract result for the sample complexity of abstract inverse problems (Theorem 3.5) reads as follows.

**Theorem.** *Under the assumptions outlined so far, there exist universal constants  $C_0, C_1, C_2 > 0$  such that, for any integer  $2 \leq s \leq |\Lambda|$ , if  $\widehat{x} \in \ell^2(\Lambda)$  is a solution of the minimization problem (1) and the number of samples satisfies*

$$m \geq C_0 B^2 \|G^{-1}\|^4 s,$$

*up to logarithmic factors, then the following recovery estimate holds with overwhelming probability:*

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 \frac{\sigma_s(x^\dagger)_1}{\sqrt{s}} + C_2 c_\nu^{-1/2} \|G^{-1}\| \beta,$$

*the term  $\sigma_s(x^\dagger)_1 = \inf\{\|x^\dagger - z\|_1 : z \in \ell^2(\Gamma), |\text{supp}(z)| \leq s\}$  representing the error of best  $s$ -sparse approximation to  $x^\dagger$ .*

Most of the subsequent efforts are directed towards keeping the quantity  $\|G^{-1}\|$  under control, in light of its role in the previous result. In Section 3.2 we discuss several strategies to provide explicit bounds for  $\|G^{-1}\|$ , ultimately in connection with the singular values of  $F\Phi^*P_\Lambda^*$ . In the case where the singular values are not accessible, we introduce a workaround relying on a compatibility condition that is typically met in practice, namely that the measurement operator of the IP approximately acts as a diagonal operator on  $(\phi_i)_{i \in \Gamma}$ , similarly to the wavelet-vaguelette decomposition [46]. Roughly speaking, the *quasi-diagonalization* condition introduced in Definition 3.7 captures how the forward map affects the sparsity of the signal, hence it allows us to obtain bounds for  $\|G^{-1}\|$  by mimicking a singular value decomposition.

Moreover, in the spirit of statistical inverse learning [21, 27, 29] (see Section 1.3 below for a brief account), it is possible to further optimize the recovery estimates to make them dependent only on the parameters that can be tuned by the observer, namely the noise level  $\beta$  and the number of samples  $m$ . More precisely, under suitable conditions on the linear and nonlinear approximation errors of  $u^\dagger$  with respect to the dictionary  $(\Phi_i)_i$  (typically met if  $u^\dagger$  belongs to certain classes of signals, such as cartoon-like images – see the end of §1.2.3 for more details), provided that  $m$  is suitably chosen as a function of the noise, we obtain estimates of the form

$$\|x^\dagger - \widehat{x}\|_2 \leq c\beta^\alpha, \quad \|x^\dagger - \widehat{x}\|_2 \leq c \frac{1}{m^{\alpha'}},$$

up to logarithmic factors, for a suitable constant  $c > 0$ . Here,  $\alpha, \alpha' > 0$  depend explicitly on the linear and nonlinear approximation error rates and on the degree of ill-posedness of the forward map  $F$ . The above estimates hold with high probability, but it would be possible to derive similar estimates for the error in expectation.

We address the reader to Section 3 for further details and comments on these results, as well as to Sections 5.2, 5.3 and 5.4 for a detailed exposition of how the standard techniques of CS (in particular, the restricted isometry property and its

consequences) must be generalized in order to accommodate the difficulties arising in ill-posed IP.

1.2.3. *The sparse Radon transform.* In Section 4 we discuss how recovery results from finite samples can be obtained for the inversion of the sparse angle Radon transform. Precisely, we consider the reconstruction of a signal supported in the unit ball in  $\mathbb{R}^2$  from samples of the Radon transform along angles  $\theta \in [0, 2\pi)$ .

Recall that the Radon transform along the angle  $\theta \in [0, 2\pi)$  of a square integrable signal  $u$  with compact support in the plane is defined by

$$\mathcal{R}_\theta u(s) := \int_{e_\theta^\perp} u(y + se_\theta) dy, \quad s \in \mathbb{R},$$

$dy$  being the 1D Lebesgue measure on the slice  $e_\theta^\perp$ . The full Radon transform  $\mathcal{R}u \in L^2([0, 2\pi) \times \mathbb{R})$  is defined by  $\mathcal{R}u(s, \theta) = \mathcal{R}_\theta u(s)$ .

Let us describe the dictionary with respect to which the unknown signal  $u^\dagger \in L^2(\mathcal{B}_1)$  is assumed to be sparse/compressible. We consider an orthonormal basis of suitable 2D wavelets  $(\phi_{j,n})_{(j,n) \in \Gamma}$  with compact support. We assume that there exists  $\Lambda \subset \Gamma$  of size  $|\Lambda| = M$  such that  $\text{supp } x^\dagger \subset \Lambda$  (recall that  $x^\dagger = \Phi u^\dagger$ ) – hence  $\|P_\Lambda^\perp x^\dagger\|_2 = 0$ . In particular, it is natural to assume  $\Lambda = \Lambda_{j_0} = \{(j, n) \in \Gamma : j \leq j_0\}$  for some  $j_0 \in \mathbb{N}$ , namely  $u^\dagger$  belongs to a subspace of signals with finite, finest scale  $j_0$  in the multiscale decomposition associated with the wavelet basis.

Our goal is to provide recovery estimates for  $u^\dagger$  given  $m$  samples of the corresponding Radon transform at different angles  $\theta_1, \dots, \theta_m \in \mathbb{S}^1$ , drawn i.i.d. uniformly at random from  $\mathbb{S}^1$ . We consider here only the parallel-beam setting, the fan-beam sampling pattern is discussed in Section 4.

Setting  $W = \text{diag}(2^{j/2})_{j,n}$ , the nonlinear recovery of  $u$  comes through the solution of the minimization problem

$$(2) \quad \min_u \|W^{-1}\Phi u\|_1 \quad : \quad \frac{1}{m} \sum_{k=1}^m \|\mathcal{R}_{\theta_k} u - y_k\|_{L^2(\mathbb{R})}^2 \leq \beta^2,$$

where the minimum is taken over  $\text{span}(\phi_{j,n})_{(j,n) \in \Lambda_{j_0}}$ , the sparse-angle measurements are given by  $y_k = \mathcal{R}_{\theta_k} u^\dagger + \varepsilon_k$ , and the noise  $\varepsilon \in (L^2(\mathbb{R}))^m$  satisfies the uniform bound  $\|\varepsilon_k\|_{L^2(\mathbb{R})} \leq \beta$  for every  $k$ .

A simplified form of our main result (Theorem 4.5) on the sparse Radon reconstruction reads as follows.

**Theorem.** *Under the assumptions outlined above, there exist constants  $C_0, C_1, C_2 > 0$  depending only on the wavelet basis such that, for any  $2 \leq s \leq |\Lambda_{j_0}|$ , if  $\hat{u} \in \text{span}(\phi_{j,n})_{(j,n) \in \Lambda_{j_0}}$  is a solution of the minimization problem (2) and the number of samples satisfies*

$$m \geq C_0 s,$$

up to logarithmic factors, then the following recovery estimate holds with overwhelming probability:

$$\|u^\dagger - \widehat{u}\|_{L^2} \leq C_1 2^{j_0/2} \frac{\sigma_s(W^{-1}\Phi u^\dagger)_1}{\sqrt{s}} + C_2 2^{j_0/2} \beta.$$

Let us emphasize that the use of weights allows us to obtain a recovery estimate with a smaller number of samples than the ones required when the abstract result above is applied to the Radon setting – see Theorem 4.4.

Finally, we wish to provide a brief exposition of the result proved in Theorem 4.6. If we relax the assumption  $\|P_{\Lambda_{j_0}}^\perp x^\dagger\|_2 = 0$  to a tail bound  $\|P_{\Lambda_{j_0}}^\perp x^\dagger\|_2 \leq C 2^{-aj_0}$  for some  $a, C > 0$ , and the signal  $u^\dagger$  satisfies the compressibility condition  $\sigma_s(\Phi u^\dagger)_1 \leq C s^{1/2-p}$  for some  $p > 0$ , then the recovery estimates reduce to

$$\|\widehat{u} - u^\dagger\|_{L^2} \leq c \beta^{\frac{2a}{2a+1}}, \quad \|u^\dagger - \widehat{u}\|_{L^2} \leq c \left(\frac{1}{m}\right)^{\frac{ap}{a+p}},$$

up to logarithmic factors. These estimates hold under the assumption that the noise is deterministic. Assuming the noise to be a random variable, under natural assumptions (e.g., see [21]) it would be possible to obtain more refined estimates in expectation.

This result is particularly useful for signals from classes with known regularity with respect to a multiresolution dictionary. For instance, it is known that such sparsity constraints in wavelet space with  $a = p = 1/2$  corresponds to the class of *cartoon-like images*, consisting of 2D piecewise smooth functions apart from mild discontinuities along curves [47, 75, 80]. As a result, the reconstruction of a cartoon-like image  $u^\dagger$  from  $m \asymp \beta^{-2}$  (see Theorem 4.6) noisy samples of the Radon transform at different angles via sparse minimization as above is successful with an error  $\|\widehat{u} - u^\dagger\|_{L^2} \leq C \beta^{1/2}$ .

**1.3. Related research.** Let us briefly discuss other works in the literature that share some common features with the present contribution.

Several approaches to obtain recovery guarantees for sampling and reconstruction in infinite-dimension can be found in the literature. The framework of *generalized sampling* was developed in [4, 10] to obtain guarantees in separable Hilbert spaces. This setting allows one to recover an element of a finite-dimensional subspace given finitely many inner products with the elements of an arbitrary frame. The issues related to the stability of the reconstruction were addressed by introducing the *balancing property* [8], which is well suited in the case of measurements modelled by a unitary operator. Several works have investigated this issue for a variety of nonlinear problems in different scenarios – we mention [13, 63, 15, 11] in this regard.

One of the main results in this paper (Theorem 3.5) generalizes several previously-known coherence-based CS results in finite [32] and infinite dimension [64, 8, 5, 14, 1] in many respects. Indeed, our theory is able to deal with linear ill-posed problems, generally involving compact operators with anisotropic and possibly vector-valued

measurements. Since our framework is based on a very weak definition of the sampling/measurement map, which is able to encompass at once both scalar products and pointwise evaluations, this result also intersects the scope of approximation theory, especially in connection with sparse approximation with polynomials [91] and sparse high-dimensional function approximation [3, 26, 45, 2].

The main notion that allows us to handle the ill-posedness of the forward map and the geometry of the measurements is the *generalized restricted isometry property* (g-RIP) (see Section 5.3), an extension of the classical *restricted isometry property* (RIP), which has been widely studied in CS due to its fundamental relevance – see for instance [39, 38, 33]. A number of variations on the theme can be found in the literature. In particular, we mention the D-RIP [36, 74], which takes into account possible redundancy in the dictionaries; the RIPL [16], which is suitable for signals with sparsity in levels; and the G-RIPL [1], which is suited to deal with infinite-dimensional signals. The g-RIP considered in this work corresponds to the G-RIPL without levels, and with a weighted sparsity. The role of the matrix  $G$  is slightly different in this work, where the main objective is to take into account the ill-posedness of the inverse problem. Alternative approaches to the RIP in CS have been proposed in the literature – see for instance [19, 37]. In [59] an interesting comparison is carried out between the RIP and the weaker notion of *source conditions*; however, it might be difficult to verify if the latter are satisfied in practical scenarios.

The g-RIP is employed here to exploit an approximate diagonalizability condition that is met in many cases of interest in applications – see Definition 3.7 of *quasi-diagonalizability*. In fact, similar approaches have already appeared in the literature, a prominent example being the wavelet-vaguelette decomposition (WVD) introduced in [46] and further developed in [35, 41] for more general dictionaries. Let us also point out that in [64] the authors exploit diagonalization properties related to a WVD of the forward map in order to obtain recovery estimates, although the setting and the results differ in several respects from our approach. Moreover, the exact diagonalizability assumption of [64] is too strict for most cases of interest, especially for the sparse Radon transform discussed in this work.

We were able to derive refined estimates in Theorem 3.11 by leveraging a notion of coherence across scales that is tailored to the multiscale structure of the sparsifying dictionary  $(\phi_{j,n})_{(j,n)\in\Gamma}$  – see Assumption 3.10. Alternative approaches that use multilevel information on coherence are available in the literature. For instance, nonuniform coherence patterns were already observed in some works on MRI [76, 78, 79], and have been considered together with multilevel sampling schemes [96, 1, 5, 77]. Our approach considers a modified notion of coherence – see Remark 3.16 – that extends these approaches by taking into account the presence of ill-posedness quantified by a certain decay of the singular values of the forward map.

The main application of our abstract framework – the inversion of the sparse Radon transform – is a well-known challenge for the IP community. A sparsity-promoting method with respect to wavelets based on a Besov-norm penalty has been considered in [61], where numerical results are also presented, showing substantial improvement with respect to more classical approaches such as filtered back-projection (FBP) and Tikhonov regularization. Other wavelet-like dictionaries (e.g., shearlets) have been used in the literature to tackle the problem of reconstructing a signal in tomography from sparse data – see [30, 88, 28]. The reconstruction in [49] is performed via sampling in the frequency domain along lines of a signal that is given by a linear combination of Dirac deltas. Under suitable assumptions, it is proved that the number of samples required for exact reconstruction is proportional to the number of deltas. In [70, 68], a numerical study is pursued to investigate how many samples are needed for accurate reconstruction in terms of the sparsity of tomographic images, where sparsity is meant here as the number of nonzero pixels. The numerical experiments presented show that the number of scalar samples required for accurate reconstruction is proportional to the sparsity of the signal.

Finally, we mention an alternative approach to derive recovery guarantees for IP that are similar in spirit to those presented in this paper, which is given by statistical inverse learning theory, a field that lies at the interface between IP and statistical learning. In a nutshell, the formulation of the problem goes as follows. Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces,  $\mathcal{H}_2$  consisting of functions defined on some input space  $(\mathcal{X}, \nu)$ , where  $\nu$  is a probability distribution on  $\mathcal{X}$ . Let  $F: \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be a linear operator. Given  $u \in \mathcal{H}_1$ , if  $X_1, \dots, X_m \stackrel{\text{i.i.d.}}{\sim} \nu$ , consider

$$Y_k = Fu(X_k) + \varepsilon_k, \quad k = 1, \dots, m,$$

where  $\varepsilon_k$  are independent noise random variables. The goal is to recover  $u$  from the observed values  $Y_1, \dots, Y_m$ . The case  $F = \text{Id}$  reduces to the classical nonparametric regression problem. Reconstruction estimates in this scenario are obtained, for instance, in [21, 89, 90, 57].

**1.4. Discussion and open questions.** Let us discuss here some limitations of our approach and potential directions for future research.

- The methods presented in this paper, as well as most of the literature on CS (for some exceptions on specific models, see [85, 22, 23, 56, 18]), are suited to treat only linear problems, and so are applicable only to linear inverse problems. Nonlinear inverse problems appear in many domains, especially in parameter identification problems for partial differential equations [73, 65]. Notable examples are the Calderón problem for electrical impedance tomography [24, 97] and the inverse scattering problem [42, 31]. It would be interesting to develop a theory of nonlinear CS to handle these nonlinear problems.

- We consider a fixed deterministic noise  $\varepsilon \in \mathcal{H}_2^m$ , representing a worst case scenario. A common model used in statistical inverse problems [72, 95] and in statistical inverse learning theory [21, 57] considers statistical noise, namely, each  $\varepsilon_k$  is sampled i.i.d. following a probability distribution on  $\mathcal{H}_2$ . Understanding how this different setting would affect our method and our reconstruction estimates is an intriguing issue. In particular, since our approach allows us to use variable-density sampling strategies, it would be interesting to study optimality rates for a class of probability distributions.
- In this work, we assume that the unknown signal  $u^\dagger$  is sparse with respect to an orthonormal basis  $(\phi_i)_{i \in \Gamma}$  of  $\mathcal{H}_1$ . We expect that the whole machinery would work also under relaxed assumptions, for instance, if  $(\phi_i)_{i \in \Gamma}$  is a frame and/or a Riesz basis, as in [87, 14, 26]. Such generalizations would allow us to consider dictionaries that are more general than wavelet orthonormal bases, such as curvelets or shearlets, which are known to provide better rates for the nonlinear approximation error [34, 60, 75]. As a consequence, this is expected to give better sample complexity estimates. We leave this extension for future research.
- In the model of the sparse Radon transform we considered a finite number of angles  $\theta_k$  but, for each fixed angle, we sample the Radon transform for every translation  $s \in [-1, 1]$ . As such, this is a semi-discrete model for the measurements, and it is natural to wonder whether the same estimates, or similar ones, hold true with a sampling also in the variable  $s$ . The work [49] contains some results in this direction, but with sampling in frequency and in the setting of Dirac delta spikes.
- We have opted to present only one application (the sparse Radon transform) of the abstract sample complexity result for inverse problems. However, the proposed approach is flexible enough to encompass a number of classical problems, including the recovery of wavelet coefficients of a  $L^2$  signal from its samples in the Fourier domain, as well as deconvolution problems with finite samples. To the best of our knowledge, this is the first time that IP so different in nature (namely, involving “discrete” and “continuous” samples) can be handled at one time by specializing the same abstract framework. We decided to postpone this analysis to a separate contribution.

**1.5. Structure of the paper.** Let us briefly illustrate the organization of the contents. In Section 2 we offer a detailed treatment of the preliminary notions that are necessary to frame and state the main results. After such a preparation, in Section 3 we formulate our general result for the sample complexity of ill-posed IP under sparsity constraints. The main application is the sparse angle tomography problem, which is thoroughly discussed in Section 4. All the proofs of both main and ancillary

results can be found in Section 5, with the exception of the proof of a key technical result (Theorem 5.7), which can be found in Appendix A.

## 2. SETUP

**2.1. Weighted setup and notation.** Let  $\mathbb{N}$  denote the set of positive integer numbers. For  $n \in \mathbb{N}$ , we denote the set of the positive integers up to  $n$  by  $[n]$ , namely  $[n] = \{1, \dots, n\}$ .

Let  $\Gamma$  be a finite or countable index set, e.g.,  $\Gamma = \mathbb{N}$  or  $\Gamma \subseteq \mathbb{N} \times \mathbb{Z}$ . In our setting, it is the index set of the dictionary representing the unknown signals. We will consider also a finite subset of indices  $\Lambda \subseteq \Gamma$ , which stands for the index set where the signal reconstruction is carried out. We denote by  $P_\Lambda$  the orthogonal projection on  $\ell^2(\Gamma)$  defined by

$$(P_\Lambda x)_i = \begin{cases} x_i & (i \in \Lambda) \\ 0 & (i \notin \Lambda) \end{cases}.$$

The image of  $P_\Lambda$  is thus  $\ell_\Lambda^2(\Gamma) := \text{span}\{e_i : i \in \Lambda\}$ ,  $(e_i)_{i \in \Gamma}$  being the canonical basis of  $\ell^2(\Gamma)$ . With an abuse of notation, we often identify  $\ell_\Lambda^2(\Gamma)$  with  $\ell^2(\Lambda)$  or with  $\mathbb{C}^M$ , where  $M = |\Lambda|$ . We denote the corresponding adjoint map by  $\iota_\Lambda$ , that is the canonical embedding  $\ell_\Lambda^2(\Gamma) \rightarrow \ell^2(\Gamma)$ . We also set  $P_\Lambda^\perp := I - P_\Lambda$ , where  $I$  is the identity operator.

We say that a sequence  $\omega \in \mathbb{R}^\Gamma$  is a vector of *weights* if  $\omega_i \geq 1$  for every  $i \in \Gamma$ . The *weighted size* of a set  $S \subseteq \Gamma$  is defined by

$$\omega(S) := \sum_{i \in S} \omega_i^2,$$

whenever  $S$  is finite. Denoting by  $|S|$  the cardinality of a set  $S$ , the following inequalities are immediately verified:

$$|S| \leq \omega(S) \leq |S| \sup_{i \in S} \omega_i^2.$$

For  $0 < p \leq 2$  we introduce the set

$$\ell_\omega^p(\Gamma) := \{x \in \mathbb{C}^\Gamma : \|x\|_{p,\omega} < \infty\},$$

where the  $\omega$ -weighted  $\ell^p$  norm is defined by

$$\|x\|_{p,\omega} := \left( \sum_{i \in \Gamma} |x_i|^p \omega_i^{2-p} \right)^{1/p}.$$

The limit case  $p \rightarrow 0$  leads to the notion of *weighted sparsity*. To be precise, denoting the support of  $x \in \mathbb{C}^\Gamma$  by  $\text{supp}(x) = \{i \in \Gamma : x_i \neq 0\}$ , we say that  $x$  is  $s$ - $\omega$ -sparse if

$$\|x\|_{0,\omega} := \omega(\text{supp } x) \leq s.$$

Note that in the unweighted case (i.e.,  $\omega_i = 1$  for all  $i \in \Gamma$ ) we obtain the usual  $\ell^p$  norm  $\|x\|_p$  and the cardinality of the support  $\|x\|_0 = |\text{supp } x|$ .

We also introduce the *error of best  $\omega$ -weighted  $s$ -sparse approximation* of  $x \in \ell^2(\Gamma)$  with respect to the weighted  $\ell^p$  norm:

$$\sigma_s(x)_{p,\omega} := \inf\{\|x - y\|_{p,\omega} : y \in \mathbb{C}^\Gamma, \|y\|_{0,\omega} \leq s\}.$$

Equivalently, we have that

$$\sigma_s(x)_{p,\omega} = \inf\{\|x_{S^c}\|_{p,\omega} : S \subset \Gamma, \omega(S) \leq s\},$$

where  $x_S := P_S x$  and  $S^c = \Gamma \setminus S$ . If  $\tilde{S} \subset \Gamma$  is a set such that  $\sigma_s(x)_{p,\omega} = \|x_{\tilde{S}^c}\|_{p,\omega}$  and  $\omega(\tilde{S}) \leq s$ , we say that  $x_{\tilde{S}}$  is a *best  $s$ - $\omega$ -sparse approximation to  $x$*  with respect to  $\ell^p_\omega$ . Note that  $x_{\tilde{S}}$  is not unique in general.

**2.2. Dictionaries, measurements and the minimization problem.** The Hilbert spaces appearing below are assumed to be complex and separable.

*Hilbert spaces.* Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces. We will refer to  $\mathcal{H}_1$  as the space of signals and to  $\mathcal{H}_2$  as the space of measurements.

*Dictionary.* Let  $(\phi_i)_{i \in \Gamma}$  be an orthonormal basis of  $\mathcal{H}_1$ . Let  $\Phi: \mathcal{H}_1 \rightarrow \ell^2(\Gamma)$  be the corresponding analysis operator, namely  $\Phi u := (\langle u, \phi_i \rangle_{\mathcal{H}_1})_{i \in \Gamma}$ . The corresponding synthesis operator  $\Phi^*: \ell^2(\Gamma) \rightarrow \mathcal{H}_1$  is given by  $\Phi^* x = \sum_{i \in \Gamma} x_i \phi_i$ .

*Measurement space.* Let  $(\mathcal{D}, \mu)$  be a measure space. We think of  $\mathcal{D}$  as a space of parameters associated with the measurements. Indeed, we perform random sampling with respect to a probability measure  $\nu$  on  $\mathcal{D}$  that is absolutely continuous with respect to  $\mu$  – namely, there exists a positive function  $f_\nu \in L^1(\mu)$  such that  $\|f_\nu\|_{L^1} = 1$ , and  $d\nu := f_\nu d\mu$ . The choice of the probability distribution turns out to be crucial in order to obtain an optimal sample complexity; however, it will not be very relevant for the purpose of the present work. We will investigate this matter further in future work.

*Measurement operators and forward map.* Let  $(F_t)_{t \in \mathcal{D}}: \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be a family of uniformly bounded linear maps, with  $\|F_t\| \leq C_F$  for every  $t \in \mathcal{D}$ , modelling the measurement operators in our problem. We suppose that

$$\mathcal{D} \ni t \mapsto F_t u \in \mathcal{H}_2$$

belongs to the Bochner space  $L^2_\mu(\mathcal{D}; \mathcal{H}_2)$  for every  $u \in \mathcal{H}_1$ . We define the forward map

$$F: \mathcal{H}_1 \rightarrow L^2_\mu(\mathcal{D}; \mathcal{H}_2), \quad (Fu)(t) = F_t u,$$

for almost every  $t \in \mathcal{D}$ . We assume that  $F$  is bounded with  $\|F\|_{\mathcal{H}_1 \rightarrow L^2_\mu(\mathcal{D}; \mathcal{H}_2)} \leq C_F$ .

The attentive reader will notice that assuming (non-uniform) boundedness of the measurement operators would affect only the estimates involving the truncation error – see Proposition 5.8. Nevertheless, as the uniform boundedness assumption is satisfied in the context of the applications discussed below, we prefer to trade off some generality for consistency.

*Noise.* Letting  $u^\dagger \in \mathcal{H}_1$  denote the unknown signal and if  $t_1, \dots, t_m$  are i.i.d. samples from the distribution  $\nu$  on  $\mathcal{D}$ , the noisy measurements may be written as

$$y_k := F_{t_k} u^\dagger + \varepsilon_k, \quad k = 1, \dots, m,$$

with  $\varepsilon_k \in \mathcal{H}_2$ . We assume the following uniform bound on the noise:

$$(3) \quad \max_k \|\varepsilon_k\|_{\mathcal{H}_2} \leq \beta,$$

for some  $\beta \geq 0$ .

*Truncation error.* Let  $\Lambda \subseteq \Gamma$  be finite. We assume that there exists  $r \geq 0$  such that  $\|P_\Lambda^\perp x^\dagger\|_2 \leq r$ , where  $x^\dagger = \Phi u^\dagger$ . The truncation error  $\|P_\Lambda^\perp x^\dagger\|_2$  is often referred to as linear approximation error [44]. We show below (see Example 4.7) that, if  $u^\dagger$  and  $\Phi$  are sufficiently regular in a sense to be specified, then the decay rate of  $\|P_\Lambda^\perp x^\dagger\|_2$  can be explicitly given in terms of  $M = |\Lambda|$ .

*Minimization problem.* We consider the following minimization problem:

$$(4) \quad \min_{x \in \ell^2(\Lambda)} \|x\|_{1,\omega} \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_\Lambda x - y_k\|_{\mathcal{H}_2}^2 \leq \eta^2,$$

for some  $\eta \geq 0$ , where we recall that  $\iota_\Lambda = (P_\Lambda)^*$ . Since the problem involves the minimization of a convex functional under a quadratic constraint, there exists at least one solution – unless the constraint condition is empty.

*Remark 2.1.* The  $\ell^1$  minimization problem (4) is known in the literature as the *synthesis formulation*. An alternative approach is the so-called *analysis formulation*

$$\min_u \|\Phi u\|_1 \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} u - y_k\|_{\mathcal{H}_2}^2 \leq \eta^2,$$

where the minimum is taken over  $\text{span}(\phi_i)_{i \in \Lambda}$ .

For general dictionaries, the two approaches are not equivalent – see for instance [50]. However, equivalence holds for orthonormal bases, which is the case in the present work. Indeed,  $\Phi$  and its inverse  $\Phi^*$  are isometric isomorphisms of  $\mathcal{H}_1$  and  $\ell^2(\Gamma)$ . We adopt the synthesis approach in the statement of theorems of the current section as this formulation is more natural in the context of compressed sensing and the restricted isometry property – see Section 5.3. Instead, in Section 4 we opt for the analysis approach, which best fits the spirit of inverse problems.

*Two examples.* We now briefly discuss two examples of problems that fit into our framework: the reconstruction of the wavelet coefficients of a sparse signal from Fourier measurements and the recovery of a sparse function in a basis of polynomials from a finite number of pointwise evaluations. Even though they have been thoroughly studied – see, for instance, [54, Section 12.1] – our framework generalizes compressed sensing in bounded orthonormal systems in many respects, including the introduction of a possibly ill-posed forward map and the presence of more general vector-valued measurement operators.

**Example 2.2 (Reconstruction of wavelet coefficients from Fourier samples).**

We consider the problem of reconstructing the wavelet coefficients of a signal supported in  $\mathbb{T} = [0, 1]$  by sampling the Fourier coefficients in an appropriate bandwidth  $[N]_{\pm} := \{-N, -N+1, \dots, N-1, N\}$ ; we refer the interested reader to [38, 79, 78].

In this context, we have that  $\mathcal{H}_1 = L^2(\mathbb{T})$  is the space of signals and  $\mathcal{H}_2 = \mathbb{C}$  is the space of scalar measurements. We consider an orthonormal basis  $(\phi_i)_{i \in \Gamma} = (\phi_{j,n})_{(j,n) \in \Gamma}$  of boundary ([40]) or periodic ([81, Section 3.11]) Daubechies wavelets. The index set  $\Gamma \subset \mathbb{N} \times \mathbb{Z}$  here includes all the possible scales, indexed by  $j \in \mathbb{N}$ , and all the possible values of the translation parameter, indexed by  $n \in \mathbb{Z}$ . In general, the range of  $n$  depends on the scale  $j$ .

As measurement space for this model we consider  $(\mathcal{D}, \mu) = ([N]_{\pm}, \mathbf{c})$ , where  $\mathbf{c}$  is the counting measure. The forward map  $F: L^2(\mathbb{T}) \rightarrow \mathbb{C}^{2N+1}$  is given by  $F = P_N \mathcal{F}$ , where  $\mathcal{F}: L^2(\mathbb{T}) \rightarrow \ell^2(\mathbb{Z})$  is the Fourier transform

$$\mathcal{F}u(t) = \int_{\mathbb{T}} u(x) e^{-2\pi i x t} dx, \quad t \in \mathbb{Z},$$

and  $P_N: \ell^2(\mathbb{Z}) \rightarrow \mathbb{C}^{2N+1}$  is the projection on the components indexed by  $[N]_{\pm}$ . For every  $t \in [N]_{\pm}$ , the measurement operators  $F_t: L^2(\mathbb{T}) \rightarrow \mathbb{C}$  are given by  $F_t u = \mathcal{F}u(t) = F u(t)$ . We choose  $f_{\nu}(0) := 1/C_{\nu}$  and  $f_{\nu}(t) := 1/(|t|C_{\nu})$  for  $t \neq 0$ , where  $C_{\nu} := 1 + \sum_{t=1}^N 2/t$  is a normalizing constant; see (i) in Remark 3.2 for further insights on this choice.

The truncation bandwidth  $\Lambda \subset \Gamma$  is the subset of indices  $(j, n)$  such that  $j \leq j_0$ . In other words, we approximate the original signal with one having a finite finest scale  $j_0$ .

**Example 2.3 (Approximation with polynomials).** We consider the classical problem of reconstructing a polynomial of degree at most  $M$  that is sparse with respect to an orthonormal basis of polynomials, given a finite number of pointwise samples [91].

Let  $L_w^2(-1, 1)$  be the space of square-integrable functions with respect to a weight  $w: [-1, 1] \rightarrow \mathbb{R}_+$ , where  $w dx$  is a probability measure. Let  $(p_i)_{i \in \mathbb{N}}$  be an orthonormal basis of polynomials of  $L_w^2(-1, 1)$ . We fix the dictionary  $(\phi_i)_{i \in \Gamma} = (p_i)_{i \in \Gamma}$ ,  $\Gamma = [M+1]$ ,

and define the space of polynomials of degree at most  $M$  by  $\mathcal{P}_M := \text{span}(p_i)_{i \in \Gamma} \subset L_w^2(-1, 1)$ . The model under consideration is represented by  $\mathcal{H}_1 = \mathcal{P}_M$  and  $\mathcal{H}_2 = \mathbb{C}$ .

The measurement space is given by  $(\mathcal{D}, \mu) = ([-1, 1], w dx)$ . The forward map coincides with the identity ( $F = I$ ) and, for every  $t \in [-1, 1]$ , the measurement operators  $F_t: \mathcal{P}_M \rightarrow \mathbb{C}$  are given by  $F_t u = u(t)$ . We choose  $\Lambda = \Gamma$  and  $f_\nu \equiv 1$ , so that  $\nu = \mu$ .

### 3. A GENERAL RESULT ON SAMPLE COMPLEXITY

**3.1. Main result.** In order to state our main result we need to introduce a few assumptions.

**Assumption 3.1.** The following coherence bound is satisfied:

$$\|F_t \phi_i\|_{\mathcal{H}_2} \leq B \sqrt{f_\nu(t)} \omega_i, \quad t \in \mathcal{D}, \quad i \in \Gamma,$$

for some  $B \geq 1$ , where  $f_\nu$  was defined in Section 2.2 and  $\omega_i \geq 1$ .

*Remark 3.2.* Let us discuss the occurrence of this bound in the two examples already discussed.

- (i) Consider the Fourier-wavelet sensing problem discussed in Example 2.2. It is known from [67, Theorem 2.1] that the following estimates hold under suitable regularity assumptions on the scaling functions of the wavelet dictionary:

$$\begin{aligned} |\mathcal{F} \phi_{j,n}(0)| &\leq C, \quad (j, n) \in \Gamma, \\ |\mathcal{F} \phi_{j,n}(t)| &\leq \frac{C}{\sqrt{|t|}}, \quad t \in \mathbb{Z} \setminus \{0\}, \quad (j, n) \in \Gamma \end{aligned}$$

for some  $C \geq 1$ . Letting  $C_\nu := 1 + \sum_{t=1}^N 2/t$ ,  $f_\nu(0) = 1/C_\nu$  and  $f_\nu(t) = 1/(|t|C_\nu)$  for  $t \neq 0$  and  $B := C\sqrt{C_\nu}$ , then the above estimate can be recast in the form of Assumption 3.1 with  $\omega_i \equiv 1$ .

- (ii) Consider the problem in Example 2.3. In the case of approximation with Legendre polynomials  $(p_i)_{i \in \mathbb{N}}$  on  $[-1, 1]$  (see [91] or [98, pag. 504]), we have

$$\|p_i\|_\infty \leq \sqrt{2i-1}, \quad i \in \mathbb{N}.$$

Assumption 3.1 can thus be satisfied provided that the weights are chosen in accordance with the constraint  $\omega_i \geq \sqrt{2i-1}$ .

**Assumption 3.3.** The following estimate is satisfied for some  $c_\nu > 0$ :

$$c_\nu \leq f_\nu \leq 1.$$

*Remark 3.4.* The inequality is a rather technical assumption and, as will be clear from the proof of Theorem 3.5, it is related to the fact that we suppose to have only a uniform bound on the noise level for the samples, see (3). By strengthening the

requirement on the noise, this assumption can be dropped, as will be shown in future work. We do not include this extension here as it is not relevant for the sparse Radon problem treated in this work.

If  $A$  is a matrix or an operator, we let  $\|A\|$  denote its operator norm.

**Theorem 3.5.** *Consider the setting of Section 2.2. Let Assumptions 3.1 and 3.3 be satisfied. Set  $M = |\Lambda|$  and define the following  $M \times M$  matrix:*

$$G := \sqrt{P_\Lambda \Phi F^* F \Phi^* \iota_\Lambda}.$$

We assume that  $G$  is invertible.

Let  $x^\dagger \in \ell^2(\Gamma)$  be such that  $\|P_\Lambda^\perp x^\dagger\|_2 \leq r$ . Consider  $m$  i.i.d. samples  $t_1, \dots, t_m \in \mathcal{D}$  drawn from the probability distribution  $\nu$ . Let  $y_k := F_{t_k} \Phi^* x^\dagger + \varepsilon_k$ , for  $k = 1, \dots, m$ , with  $\max_k \|\varepsilon_k\|_{\mathcal{H}_2} \leq \beta$ .

There exist absolute constants  $C_0, C_1, C_2 > 0$  such that the following holds. Let  $\hat{x} \in \ell^2(\Lambda)$  be a solution of the minimization problem

$$(5) \quad \min_{x \in \ell^2(\Lambda)} \|x\|_{1,\omega} \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_\Lambda x - y_k\|_{\mathcal{H}_2}^2 \leq (\beta + C_3 \|G^{-1}\|^{-1} r)^2.$$

Fix  $s \in [M]$  with  $s \geq \|\omega\|_\infty^2/4$  and set  $\tau := B^2 \|G^{-1}\|^4 \|G\|^2 s$ . For any  $\gamma \in (0, 1)$ , if

$$m \geq C_0 \tau \max\{\log^3 \tau \log M, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|x^\dagger - \hat{x}\|_{1,\omega} \leq C_1 \sigma_s(P_\Lambda x^\dagger)_{1,\omega} + C_2 \sqrt{s} c_\nu^{-1/2} (\|G^{-1}\| \beta + C_3 r) + \|P_\Lambda^\perp x^\dagger\|_{1,\omega},$$

$$\|x^\dagger - \hat{x}\|_2 \leq C_1 \frac{\sigma_s(P_\Lambda x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 c_\nu^{-1/2} (\|G^{-1}\| \beta + C_3 r),$$

where  $C_3 := c_\nu^{-1/2} C_4 (\|F \Phi^* \iota_{\Gamma \setminus \Lambda}\| \|G^{-1}\| + 1)$ , and  $C_4$  depends only on  $C_F$ .

Theorem 3.5 presents a recovery guarantee of nonuniform type, i.e., valid with overwhelming probability for a fixed unknown signal. Nonetheless, all the tools used in the proof are inherently suited for uniform recovery except for Proposition 5.8 only, which is necessary to obtain refined estimates on the truncation error. It is possible to obtain a uniform recovery result (that is, valid for a whole family of suitably sparse signals) by replacing the term  $\|G^{-1}\| \beta + C_3 r$  in the estimates with the possibly worse bound  $\|G^{-1}\|(\beta + r)$ . In particular, the two terms coincide if  $x^\dagger$  is supported in  $\Lambda$  and therefore  $r = 0$ , leading to uniform recovery estimates for this class of signals.

The factor  $\log^3 \tau$  appearing in the estimates is known to be suboptimal – see for instance [26]. Optimizing the exponent of the logarithm would lead to a substantial increase in the length of the proofs and falls outside the scope of the present work.

In order to apply this result to specific cases of interest, several quantities must be explicitly computed, the most crucial one being  $\|G^{-1}\|$ . This will require new tools and will be the focus of §3.2 below; we anticipate here a few general observations.

A sufficient condition for the invertibility of  $G$  that is typically satisfied in several inverse problems of interest is a restricted injectivity property for  $F$ , as detailed below.

**Definition 3.6.** Let  $\mathcal{H}$  be a complex and separable Hilbert space. A linear operator  $V: \ell^2(\Gamma) \rightarrow \mathcal{H}$  is said to possess the *finite basis injectivity (FBI) property* if, for every finite family of indices  $\Lambda \subset \Gamma$ , the restriction of  $V$  to  $\ell^2(\Lambda)$  is injective.

Given an orthonormal basis  $(\phi_i)_{i \in \Gamma}$  of a Hilbert space  $\mathcal{H}_1$  with analysis operator  $\Phi$ , an operator  $F: \mathcal{H}_1 \rightarrow \mathcal{H}$  is said to possess the *FBI property with respect to  $\Phi$*  (in short:  $\Phi$ -FBI) if  $F\Phi^*$  has the FBI property.

Suppose that  $F$  is  $\Phi$ -FBI. Then we have that  $F\Phi^*\iota_\Lambda$  is injective for every finite  $\Lambda \subset \Gamma$ , which implies that  $G \in \mathbb{C}^{M \times M}$  is positive definite and thus invertible.

As already observed in [25], the FBI condition is quite natural in the context of inverse problems with sparsity constraints – in fact, it is less restrictive than full injectivity, even if the latter is more common in the IP literature. An elementary example of a non-injective FBI operator on  $\ell^2(\mathbb{N})$  is the following one:

$$V e_i = \begin{cases} \sum_j c_j e_j & (i = 1) \\ e_{i-1} & (i \geq 2), \end{cases}$$

where  $(e_i)_{i \in \mathbb{N}}$  is the standard basis of  $\ell^2(\mathbb{N})$  and  $(c_j)_{j \in \mathbb{N}}$  is any complex sequence such that  $\sum_j |c_j|^2 < \infty$  and  $c_j \neq 0$  for every  $j$ .

Another quantity that needs to be explicitly computed in order to apply Theorem 3.5 is  $C_3 = c_\nu^{-1/2} C_4 (\|F\Phi^*\iota_{\Gamma \setminus \Lambda}\| \|G^{-1}\| + 1)$ . We find that under reasonable conditions (see Theorem 3.11), the quantity  $C_3$  can be bounded by a constant depending only on  $C_F$ , on  $c_\nu$  and on the quasi-diagonalization bounds.

**3.2. Quantitative bounds on  $\|G^{-1}\|$  and quasi-diagonalization.** We have already shown that if  $F$  is  $\Phi$ -FBI then the matrix  $G$  is non-degenerate. In order to make the estimates of Theorem 3.5 effective, we now derive explicit bounds for  $\|G^{-1}\|$ . In the literature on infinite-dimensional compressed sensing, such bounds were obtained with the so-called *balancing property* [8], which is well suited in the case of measurements modeled by a unitary operator. Here we somewhat exploit the (possible) ill-posedness of our forward map as well as a *quasi-diagonalization* property, introduced below, that is satisfied by many forward maps of interest, including the Radon transform. We then derive ad hoc bounds that are of a different flavour compared to the ones obtained with the balancing property. Potential combinations of the balancing property with the quasi-diagonalization property are left for future work.

We note that  $\|G^{-1}\|$  can be explicitly bounded using the singular values of  $F\Phi^*$ . Indeed, we have that

$$\|G^{-1}\| \leq \sigma_M(F\Phi^* \iota_\Lambda)^{-1},$$

where  $\sigma_M(F\Phi^* \iota_\Lambda)$  is the  $M$ -th singular value of  $F\Phi^* \iota_\Lambda$ , with  $M = |\Lambda|$ .

*Quasi-diagonalization.* The previous estimate depends on the singular values of  $F\Phi^* \iota_\Lambda$ . In practical scenarios, estimating  $\sigma_M(F\Phi^* \iota_\Lambda)$  is non-trivial. For this reason, we now discuss an alternative approach for estimating  $\|G^{-1}\|$ , which is a consequence of a condition that is typically met in several contexts – more precisely, a condition that holds when it is possible to identify a dictionary that approximately diagonalizes the action of the forward map of the model.

We consider now (orthonormal) multiresolution dictionaries  $(\phi_{j,n})_{(j,n) \in \Gamma}$  with two indices (e.g., orthonormal wavelets). The parameter  $j$  represents the scale, and we use the convention that higher values of  $j$  correspond to finer scales.

**Definition 3.7.** We say that  $F$  satisfies the *quasi-diagonalization property* with respect to  $(\Phi, b)$ ,  $b \geq 0$ , if there exist  $c, C > 0$  such that

$$(6) \quad c \sum_{(j,n) \in \Gamma} 2^{-2bj} |x_{j,n}|^2 \leq \|F\Phi^* x\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 \leq C \sum_{(j,n) \in \Gamma} 2^{-2bj} |x_{j,n}|^2, \quad x \in \ell^2(\Gamma).$$

Roughly speaking, the previous property entails the fact that the action of the forward map on the elements of the dictionary  $(\phi_{j,n})_{(j,n) \in \Gamma}$  approximately coincides with that of a diagonal operator with dyadic coefficients depending only on the parameter  $b$  and the “scale” index  $j$ : as a rule of thumb, the larger  $b$  is, the more smoothing  $F$  is – hence, the ill-posedness of the corresponding inverse problem gets worse.

We now provide a set of conditions ensuring that the quasi-diagonalization property is satisfied. Recall that  $H^b(\mathbb{R}^d)$ ,  $b \in \mathbb{R}$ , is the Sobolev space of distributions  $u \in \mathcal{S}'(\mathbb{R}^d)$  such that  $(I - \Delta)^{b/2} u \in L^2(\mathbb{R}^d)$ , while  $H^b(\mathbb{T}^d)$  is the Sobolev space of distributions  $u \in \mathcal{S}'(\mathbb{T}^d)$  such that  $(I - \Delta)^{b/2} u \in L^2(\mathbb{T}^d)$ , where  $(I - \Delta)^{b/2}$  corresponds to the Fourier multiplier  $(1 + |\cdot|^2)^{b/2}$  on  $\mathbb{R}^d$  or on  $\mathbb{Z}^d$ , respectively.

**Proposition 3.8.** *Let  $\mathcal{H}_1$  be either  $L^2(\mathbb{T}^d)$  or  $L^2(\mathbb{R}^d)$ . Assume that there exist  $c, C > 0$ ,  $b \geq 0$  such that:*

- *The forward map  $F$  satisfies*

$$(7) \quad c \|u\|_{H^{-b}}^2 \leq \|Fu\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 \leq C \|u\|_{H^{-b}}^2, \quad u \in \mathcal{H}_1.$$

- *The dictionary  $(\phi_{j,n})_{(j,n) \in \Gamma}$  is an orthonormal basis of  $\mathcal{H}_1$  such that*

$$(8) \quad c \|u\|_{H^{-b}}^2 \leq \sum_{(j,n) \in \Gamma} 2^{-2bj} |\langle u, \phi_{j,n} \rangle_{L^2}|^2 \leq C \|u\|_{H^{-b}}^2, \quad u \in \mathcal{H}_1,$$

*which we will refer to as the Littlewood-Paley property of the dictionary.*

Then  $F$  satisfies the quasi-diagonalization property (6) with respect to  $(\Phi, b)$ .

*Remark 3.9.* If  $(\phi_{j,n})_{(j,n) \in \Gamma}$  is a  $q$ -regular wavelet system (in the sense of [81, Section 2.2]), then the Littlewood-Paley condition (8) is satisfied for every  $-q < b < q$  – see for instance [81, Theorem 8 in Section 2.8 and Section 3.11] and [80, Theorem 9.2]. Compactly supported  $q$ -regular wavelet system exist for every fixed value of  $q \in \mathbb{N}$  – see [81, Theorem 3 in Section 3.8].

*Coherence bounds across scales.* The results of Theorem 3.5 can be improved in the quasi-diagonalization regime with respect to  $(\Phi, b)$  as soon as additional coherence bounds are available, tailored to the multiscale structure of the dictionary  $(\phi_{j,n})_{(j,n) \in \Gamma}$ .

In particular, we consider the following tighter version of Assumption 3.1.

**Assumption 3.10.** The following coherence bound is satisfied:

$$\|F_t \phi_{j,n}\|_{\mathcal{H}_2} \leq B \frac{\sqrt{f_\nu(t)}}{d_{j,n}} \omega_{j,n}, \quad t \in \mathcal{D}, (j,n) \in \Gamma.$$

for some  $B \geq 1$ ,  $1 \leq d_{j,n} \leq 2^{bj}$ , where  $b$  is the constant appearing in the quasi-diagonalization (6),  $f_\nu$  was defined in Section 2.2 and  $\omega_{j,n} \geq 1$ .

Remark 3.16 below gives further insights on the role of the parameters  $d_{j,n}$ .

We can take advantage of such refined coherence information to substantially reduce the number of samples that are needed to achieve recovery in Theorem 3.5, retaining the same recovery rates.

For  $j \in \mathbb{N}$ , we introduce the index subsets with finest scale  $j$

$$(9) \quad \Lambda_j := \{(j', n) \in \Gamma : j' \leq j\}.$$

Moreover, we always assume that  $|\Lambda_j| < +\infty$ .

**Theorem 3.11.** Consider the setting of Section 2.2. Let  $(\phi_{j,n})_{(j,n) \in \Gamma}$  be a multiresolution (orthonormal) dictionary that satisfies (6). Let Assumptions 3.10 and 3.3 be satisfied. Fix  $j_0 \in \mathbb{N}$  and set  $M := |\Lambda_{j_0}|$ .

Let  $x^\dagger \in \ell^2(\Gamma)$  be such that  $\|P_{\Lambda_{j_0}}^\perp x^\dagger\| \leq r$ . Consider  $m$  i.i.d. samples  $t_1, \dots, t_m \in \mathcal{D}$  drawn from the probability distribution  $\nu$ . Let  $y_k := F_{t_k} \Phi^* x^\dagger + \varepsilon_k$ , for  $k = 1, \dots, m$ , with  $\max_k \|\varepsilon_k\|_{\mathcal{H}_2} \leq \beta$ .

There exist constants  $C_0, C_1, C_2, C_4 > 0$ , which depend only on the quasi-diagonalization bounds in (6) and on  $C_F$ , such that the following holds. Let  $\zeta \in [0, 1]$  and  $W := \text{diag}(2^{bj})_{(j,n) \in \Lambda_{j_0}}$ . Let  $\hat{x}$  be a solution of the following minimization problem:

$$(10) \quad \min_{x \in \ell^2(\Lambda_{j_0})} \|W^{-\zeta} x\|_{1,\omega} : \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_{\Lambda_{j_0}} x - y_k\|_{\mathcal{H}_2}^2 \leq (\beta + C_3 2^{-bj_0} r)^2,$$

where  $C_3 = c_\nu^{-1/2} C_4$ .

Fix  $s \in [M]$  with  $s \geq \|\omega\|_\infty^2/4$  and set

$$\tau := B^2 \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}) 2^{2(1-\zeta)bj_0} s.$$

If

$$(11) \quad m \geq C_0 \tau \max\{\log^3 \tau \log M, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|W^{-\zeta} x^\dagger - W^{-\zeta} \hat{x}\|_2 \leq C_1 \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-\zeta} x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 2^{-\zeta bj_0} c_\nu^{-1/2} (2^{bj_0} \beta + C_3 r).$$

In particular, we have that

$$\|x^\dagger - \hat{x}\|_2 \leq C_1 2^{\zeta bj_0} \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-\zeta} x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 c_\nu^{-1/2} (2^{bj_0} \beta + C_3 r).$$

*Remark 3.12.* When  $(\phi_{j,n})_{j,n}$  is a wavelet-like dictionary, the parameter  $M = |\Lambda_{j_0}|$  in Theorem 3.11 usually satisfies  $M \asymp 2^{cj_0}$  for some positive integer  $c$  – see, for instance, equation (15) below. In these cases, the quantity  $\log M$  in the sample complexity can be therefore replaced with  $cj_0$ .

*Remark 3.13.* Thanks to the quasi-diagonalization property (6), the quantities related to the matrix  $G$  in Theorem 3.5 assume an explicit form in Theorem 3.11. Moreover, thanks to refined information about the coherence given by Assumption 3.10 and the flexibility given by the parameter  $\zeta$ , the sample complexity improves in Theorem 3.11 from Theorem 3.5; indeed, the sample complexity in the previous scenario was given by  $m = m(\tau)$ , where it is possible to show that

$$\tau := B^2 \|G^{-1}\|_4^4 s \asymp B^2 2^{4bj_0} s,$$

which, recalling that  $d_{j,n} \geq 1$  and  $\zeta \in [0, 1]$ , is in general larger than

$$\tau := B^2 \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}) 2^{2(1-\zeta)bj_0} s.$$

In practice, this will lead to significant improvements in the estimates for the Radon transform, as will be shown in Section 4.

The tunable parameter  $\zeta$  in Theorem 3.11 shows that there is a trade-off between the sample complexity of the problem and the reconstruction error: larger values of  $\zeta$  correspond to a lower number of measurements, but to a possibly larger reconstruction error. We now discuss two extremal cases of special interest of Theorem 3.11 in the choice of the parameter  $\zeta$ . First, we consider the case  $\zeta = 0$ , which yields the largest number of measurements, but possibly the best bound for the reconstruction error.

**Corollary 3.14.** *Consider the setting of Theorem 3.11, and suppose that the same hypotheses are satisfied. Let  $\widehat{x}$  be a solution of the following minimization problem:*

$$(12) \quad \min_{x \in \ell^2(\Lambda_{j_0})} \|x\|_{1,\omega} \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_{\Lambda_{j_0}} x - y_k\|_{\mathcal{H}_2}^2 \leq (\beta + C_3 2^{-bj_0} r)^2.$$

If

$$\tau := B^2 \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}) 2^{2bj_0} s.$$

and

$$m \geq C_0 \tau \max\{\log^3 \tau \log M, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 \frac{\sigma_s(P_{\Lambda_{j_0}} x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 c_\nu^{-1/2} (2^{bj_0} \beta + C_3 r).$$

Next, we consider the case  $\zeta = 1$  in Theorem 3.11, which yields the smallest number of measurements, but a possibly larger bound for the reconstruction error.

**Corollary 3.15.** *Consider the setting of Theorem 3.11, and suppose that the same hypotheses are satisfied. Let  $\widehat{x}$  be a solution of the following minimization problem:*

$$\min_{x \in \ell^2(\Lambda_{j_0})} \|W^{-1}x\|_{1,\omega} \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_{\Lambda_{j_0}} x - y_k\|_{\mathcal{H}_2}^2 \leq (\beta + C_3 2^{-bj_0} r)^2,$$

If

$$(13) \quad \tau := B^2 \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}) s$$

and

$$(14) \quad m \geq C_0 \tau \max\{\log^3 \tau \log M, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|F\Phi^* x^\dagger - F\Phi^* \widehat{x}\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)} \leq C_1 \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-1} x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 c_\nu^{-1/2} (\beta + 2^{-bj_0} C_3 r).$$

In particular, we have that

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 2^{bj_0} \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-1} x^\dagger)_{1,\omega}}{\sqrt{s}} + C_2 c_\nu^{-1/2} (2^{bj_0} \beta + C_3 r).$$

As a direct consequence of Corollary 3.15, we obtain the following exact recovery result. If there is no truncation error ( $x^\dagger \in \Lambda_{j_0}$ , i.e.  $r = 0$ ), if  $x^\dagger$  is  $s$ -sparse ( $\sigma_s(P_{\Lambda_{j_0}} W^{-1} x^\dagger)_1 = \sigma_s(x^\dagger)_1 = 0$ ) and if there is no noise ( $\beta = 0$ ), then

$$\widehat{x} = x^\dagger,$$

provided that  $m$  satisfies (14).

*Remark 3.16.* Let us comment on the sample complexity appearing in (13) and (14). Consider for simplicity the case where  $f_\nu = 1$ ,  $\omega_{j,n} = 1$  and  $\zeta = 1$ . In classical CS, where  $(F_t)_{t \in \mathcal{D}}$  are the measurement operators and  $F$  is an isometry, a relevant quantity appearing in the sample complexity is the *coherence*  $B > 0$  of the system, such that

$$\max_{(j,n) \in \Lambda_{j_0}} \|F_t \phi_{j,n}\|_{\mathcal{H}_2} \leq B.$$

When a source of ill-posedness is introduced (e.g., if  $F$  is a compact operator), this quantity must be suitably normalized with the  $L^2$  norm of  $F \phi_{j,n}$ . Note that in this setting Assumption 3.10 reads  $\|F_t \phi_{j,n}\|_{\mathcal{H}_2} \leq B d_{j,n}^{-1}$ . On the other hand, the quasi-diagonalization property implies that  $\|F \phi_{j,n}\|_{L^2_\mu(\mathcal{D}; \mathcal{H}_2)} \asymp 2^{-bj}$ . We are led to consider the notion of relative coherence of the system, that is

$$\max_{(j,n) \in \Lambda_{j_0}} (B d_{j,n}^{-1} 2^{bj}).$$

It is evident from (13) and (14) that the relative coherence of the system plays the same role of the usual coherence in classical CS, where the recovery guarantees are given for  $m \gtrsim B^2 s$ , up to logarithmic terms.

**3.3. Analysis of the recovery error.** We now derive explicit estimates in terms of the noise  $\beta$  and the number of measurements  $m$ . To accomplish our goal, we need to impose additional assumptions on the growth of the parameters  $d_{j,n}$  in Assumption 3.10 and on the regularity of the signal  $x^\dagger$ . The latter includes a certain decay of the truncation error  $\|P_{\Lambda_{j_0}}^\perp x^\dagger\|_2$  in terms of  $|\Lambda_{j_0}|$  – namely,  $\|P_{\Lambda_{j_0}}^\perp x^\dagger\|_2 \leq C 2^{-aj_0}$ , where  $|\Lambda_{j_0}| \asymp 2^{cj_0}$  – and of  $p$ -compressibility for some  $p \geq 1/2$ , which means that  $\sigma_s(x^\dagger)_1 \leq C s^{1/2-p}$ .

Those estimates are natural in many problems where the signal  $x^\dagger$  is assumed to belong to a specific class of signals. For instance, in the case of cartoon-like images, we have  $a = 1/2$  and approximate  $p$ -compressibility with  $p = 1/2$ , in a sense to be specified – see Example 4.7.

**Theorem 3.17.** *Consider the setting of Theorem 3.14 and suppose that the same hypotheses are satisfied. Suppose that the following estimates hold:*

$$|\Lambda_j| \leq C 2^{cj}, \quad d_{j,n} \geq 2^{dj},$$

$$\sigma_s(x^\dagger)_1 \leq Cs^{1/2-p}, \quad \|P_{\Lambda_j}^\perp x^\dagger\|_2 \leq C2^{-aj},$$

for some  $a, c, p, C > 0$  and  $0 \leq d \leq b$ .

Fix  $\gamma \in (0, 1)$  and let  $\beta \in (0, 1)$  be sufficiently small. Let  $j_0 := \lfloor 1/(a+b) \log(1/\beta) \rfloor$  and let  $m$  be sufficiently large.

There exist constants  $C_0, C_1 > 0$  depending only on  $a, b, c, p, C$ , on  $B$  appearing in Assumption 3.10, on the quasi-diagonalization bounds in (6), on  $C_F$  and logarithmically on  $\gamma$ , such that the following holds.

Let  $\hat{x}$  be a solution of the minimization problem (12). With probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|x^\dagger - \hat{x}\|_2 \leq C_1 \left( \frac{\log^{4/p}(1/\beta)}{\beta^{2\frac{2b-d}{a+b}p} m^p} + \beta^{\frac{a}{a+b}} \right).$$

In particular, if

$$m = \lfloor C_0 \beta^{-2\frac{(2b-d)}{a+b} - \frac{a}{p(a+b)}} \log^4(1/\beta) \rfloor,$$

then

$$\|x^\dagger - \hat{x}\|_2 \leq C_1 \beta^{\frac{a}{a+b}},$$

or, equivalently,

$$\|x^\dagger - \hat{x}\|_2 \leq C_1 \left( \frac{\log^4 m}{m} \right)^{\frac{ap}{a+4bp-2dp}}.$$

*Remark 3.18.* Following a favored approach in the statistical inverse learning literature – see [21], for instance – and for simplicity of the exposition, we adopted an asymptotic formulation in Theorem 3.17, where  $\beta$  is sufficiently small and  $m$  is sufficiently large. Analyzing the proof of the Theorem, it is possible to make this statement quantitatively precise.

#### 4. THE SPARSE RADON TRANSFORM

We now consider the problem of reconstructing a function supported in the unit ball  $\mathcal{B}_1 \subset \mathbb{R}^2$  by sampling either the corresponding Radon or the fan beam transform along angles  $\theta \in \mathbb{S}^1$ .

*The Radon and the fan beam transform.* Let us recall the definition of the two-dimensional Radon transform.

**Definition 4.1.** Let  $K \subset \mathbb{R}^2$  be a compact set. Let  $u \in L^2(K)$  and  $\theta \in [0, 2\pi)$ . The Radon transform along the angle  $\theta$ ,  $\mathcal{R}_\theta: L^2(K) \rightarrow L^2(\mathbb{R})$ , is defined by

$$\mathcal{R}_\theta u(s) := \int_{e_\theta^\perp} u(y + se_\theta) dy,$$

where  $dy$  is the 1D Lebesgue measure on  $e_\theta^\perp$  and  $e_\theta := (\cos \theta, \sin \theta)$ .

The (full) Radon transform  $\mathcal{R}: L^2(K) \rightarrow L^2([0, 2\pi) \times \mathbb{R})$  is defined by  $\mathcal{R}u(\theta, s) := \mathcal{R}_\theta u(s)$ .

With an abuse of notation, we will write  $\mathcal{R}_\theta$  even when  $\theta \in \mathbb{S}^1$ , with the natural identification between  $[0, 2\pi)$  and  $\mathbb{S}^1$ .

To be precise,  $\mathcal{R}_\theta u$  is defined for functions in a dense subspace of  $L^2(K)$  at first, in such a way that the integral expression makes sense, for instance  $C^0(K)$ . The continuous extension to  $L^2(K)$  and the well-posedness of both  $\mathcal{R}_\theta$  and  $\mathcal{R}$  are ensured by the following result.

**Lemma 4.2** ([83, Theorem 1.6]).  *$\mathcal{R}_\theta$  and  $\mathcal{R}$  are well-defined continuous operators. Moreover, the norms of  $\mathcal{R}$  and  $\{\mathcal{R}_\theta\}_{\theta \in \mathbb{S}^1}$  can be uniformly bounded by a constant depending only on  $|K|$ .*

We also consider the fan beam transform, which is more relevant in the applications.

**Definition 4.3.** Let  $0 < d < \rho$ . Let  $u \in L^2(\mathcal{B}_d)$  and  $\theta \in [0, 2\pi)$ . The fan beam transform from the angle  $\theta \in [0, 2\pi)$  at distance  $\rho$ ,  $\mathcal{D}_\theta: L^2(\mathcal{B}_d) \rightarrow L^2(-\frac{\pi}{2}, \frac{\pi}{2})$ , is defined by

$$\mathcal{D}_\theta u(\alpha) := \int_{\mathbb{R}} u(\rho e_\theta + t e_{\theta+\alpha}) dt.$$

The (full) fan beam transform  $\mathcal{D}: L^2(\mathcal{B}_d) \rightarrow L^2([0, 2\pi) \times (-\frac{\pi}{2}, \frac{\pi}{2}))$  is defined by  $\mathcal{D}u(\theta, \alpha) := \mathcal{D}_\theta u(\alpha)$ .

This is nothing but a different parametrization of the Radon transform. We consider functions whose support is contained in the ball of radius  $d$ , and take radial lines with centers on the sphere of radius  $\rho$ .

*Compactly supported wavelets.* Let us briefly review a standard construction of an orthonormal basis  $(\phi_{j,n})_{j,n}$  of  $L^2(\mathbb{R}^2)$  consisting of compactly supported wavelets – see [43] for a comprehensive discussion.

We start with a preliminary construction in  $L^2(\mathbb{R})$ . For  $j \in \mathbb{N}$  and  $n \in \mathbb{Z}$ , given a compactly supported scaling function  $\chi \in L^2(\mathbb{R})$  and a mother wavelet  $\psi \in L^2(\mathbb{R})$ , we define

$$\chi_n := \tau_n \chi, \quad \psi_{j,n} := D_{j-1}^1 \tau_n \psi,$$

where  $D_j^1 f := 2^{j/2} f(2^j \cdot)$  is a dyadic dilation of  $f$  in 1D and  $\tau_n f := f(\cdot - n)$  denotes a translation by  $n$ .

For an appropriate choice of  $\chi$  and  $\psi$ , we have that  $\{\chi_n\}_n \cup \{\psi_{j,n}\}_{j,n}$  is an orthonormal basis of  $L^2(\mathbb{R})$  – see [43, Chapter 6]. Moreover, as shown in [81, Theorem 3 in Section 3.8],  $\chi$  and  $\psi$  can be taken in  $C^2(\mathbb{R})$  – see Remark 3.9 for further details.

We now build an orthonormal basis using the standard construction of orthonormal separable wavelets – see for instance [80, Section 7.7]. Recall that the tensor product of functions  $f, g$  of one variable is defined by  $(f \otimes g)(x, y) := f(x)g(y)$ . For  $j \in \mathbb{N}$  and  $n_1, n_2 \in \mathbb{Z}$ , we set

$$\begin{aligned} \chi_{n_1, n_2} &:= \chi_{n_1} \otimes \chi_{n_2}, & \psi_{0, n_1, n_2}^{(1)} &:= \psi_{0, n_1} \otimes \chi_{n_2}, \\ \psi_{0, n_1, n_2}^{(2)} &:= \chi_{n_1} \otimes \psi_{0, n_2}, & \psi_{0, n_1, n_2}^{(3)} &:= \psi_{0, n_1} \otimes \psi_{0, n_2} \end{aligned}$$

and  $\psi_{j, n_1, n_2}^\varepsilon := D_j^2 \psi_{0, n_1, n_2}^\varepsilon$ , where  $D_j^2 f := 2^j f(2^j \cdot)$  is the dyadic dilation of  $f$  in 2D. It can be proved that

$$\{\chi_{n_1, n_2}\}_{n_1, n_2} \cup \{\psi_{j, n_1, n_2}^{(\varepsilon)}\}_{j, n_1, n_2, \varepsilon}, \quad \varepsilon \in \{1, 2, 3\},$$

is an orthonormal basis of  $L^2(\mathbb{R}^2)$ .

Finally, we obtain the orthonormal basis  $(\phi_{j, n})_{j, n}$  by rearranging the wavelets defined above, according to the following rule: for  $j = 0$ ,  $n = (n_1, n_2) \in \mathbb{Z}^2$  and, for  $j \geq 1$ ,  $n = (n_1, n_2, \varepsilon) \in \mathbb{Z}^2 \times \{1, 2, 3\}$ , set

$$\phi_{0, n} := \chi_{n_1, n_2}; \quad \phi_{j, n} := \psi_{j, n_1, n_2}^{(\varepsilon)}.$$

Let  $\Gamma$  be the set of indices  $(j, n)$  such that  $\text{supp } \phi_{j, n} \cap \mathcal{B}_1 \neq \emptyset$ , namely,

$$\Gamma = \{(j, n) \in (\{0\} \times \mathbb{Z}^2) \cup (\mathbb{N} \times (\mathbb{Z}^2 \times \{1, 2, 3\})) : \text{supp } \phi_{j, n} \cap \mathcal{B}_1 \neq \emptyset\}.$$

In what follows, we will consider the space  $\mathcal{H}_1 := \overline{\text{span}(\phi_{j, n})_{(j, n) \in \Gamma}}$ . Notice that  $L^2(\mathcal{B}_1) \subseteq \mathcal{H}_1$ . It is easy to realize that

$$(15) \quad |\{(j', n) \in \Gamma : j' = j\}| \asymp 2^{2j}, \quad j \in \mathbb{N}.$$

Indeed, by construction, the number of wavelets at scale  $j + 1$  is approximately given by the corresponding number at scale  $j$  multiplied by  $2^2$ . In light of (9), this implies that  $|\Lambda_j| \asymp 2^{2j}$ .

Finally, we choose  $0 < d < \rho$  such that  $\text{supp } u \subseteq \mathcal{B}_d$  for every  $u \in \mathcal{H}_1$ , so that  $\mathcal{D}u$  is well-defined for every  $u \in \mathcal{H}_1$ .

*Main results.* Our main recovery results for the sparse Radon transform read as follows.

**Theorem 4.4.** *Consider a wavelet dictionary  $(\phi_{j, n})_{j, n \in \Gamma}$  defined as above and the corresponding analysis operator  $\Phi: \mathcal{H}_1 \rightarrow \ell^2(\Gamma)$ . Let  $F$  be either  $\mathcal{R}$  or  $\mathcal{D}$ . There exist constants  $C_0, C_1, C_2, C_3 > 0$ , which depend only on the chosen wavelet basis in the case of the Radon transform and also on  $d$  and  $\rho$  in the case of the fan beam transform, such that the following holds.*

*Fix  $j_0 \in \mathbb{N}$ . Let  $u^\dagger \in L^2(\mathcal{B}_1)$  be such that  $\|P_{\Lambda_{j_0}}^\perp \Phi u^\dagger\|_2 \leq r$ . Consider  $m$  i.i.d. samples  $\theta_1, \dots, \theta_m \in [0, 2\pi)$  drawn from the uniform distribution on  $[0, 2\pi)$ .*

Let  $y_k := F_{\theta_k} u^\dagger + \varepsilon_k$  for  $k = 1, \dots, m$ , with  $\max_k \|\varepsilon_k\|_{L^2} \leq \beta$ . Let  $\hat{u}$  be a solution of the minimization problem

$$(16) \quad \min_u \|\Phi u\|_1 \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{\theta_k} u - y_k\|_{L^2}^2 \leq (\beta + C_3 2^{-j_0/2} r)^2,$$

where the minimum is taken over  $\text{span}(\phi_{j,n})_{(j,n) \in \Lambda_{j_0}}$ .

Let  $s \in [\Lambda_{j_0}]$ ,  $s \geq 2$ , and set  $\tau := 2^{j_0} s$ . For any  $\gamma \in (0, 1)$ , if

$$m \geq C_0 \tau \max\{j_0 \log^3 \tau, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$(17) \quad \|u^\dagger - \hat{u}\|_{L^2} \leq C_1 \frac{\sigma_s(P_{\Lambda_{j_0}} \Phi u^\dagger)_1}{\sqrt{s}} + C_2(2^{j_0/2} \beta + r).$$

The above result corresponds to the case  $\zeta = 0$  in Theorem 3.11 (see Corollary 3.14), in which a smaller reconstruction error is obtained by making more measurements. In the next theorem, we consider the case with a minimum number of measurements (i.e., the case where  $\zeta = 1$ , considered in Corollary 3.15), proportional to the sparsity  $s$  up to logarithmic factors. This is best suited for the recovery of  $Fu^\dagger$  and for the exact recovery of  $u^\dagger$  in absence of noise, but yields worse error bounds in general.

**Theorem 4.5.** *Consider the setting of Theorem 4.4 and suppose that the same assumptions are satisfied.*

*There exist constants  $C_0, C_1, C_2 > 0$ , which depend only on the chosen wavelet basis in the case of the Radon transform and also on  $d$  and  $\rho$  in the case of the fan beam transform, such that the following holds. Let  $W := \text{diag}(2^{j/2})_{j,n}$ . Let  $\hat{u}$  be a solution of the minimization problem*

$$\min_u \|W^{-1} \Phi u\|_1 \quad : \quad \frac{1}{m} \sum_{k=1}^m \|F_{\theta_k} u - y_k\|_{L^2}^2 \leq (\beta + C_3 2^{-j_0/2} r)^2,$$

where the minimum is taken over  $\text{span}(\phi_{j,n})_{(j,n) \in \Lambda_{j_0}}$ .

Let  $s \in [\Lambda_{j_0}]$ ,  $s \geq 2$ . If

$$(18) \quad m \geq C_0 s \max\{j_0 \log^3 s, \log(1/\gamma)\},$$

then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|Fu^\dagger - F\hat{u}\|_{L^2(\mathbb{S}^1 \times \mathbb{R})} \leq C_1 \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-1} \Phi u^\dagger)_1}{\sqrt{s}} + C_2(\beta + 2^{-j_0/2} r).$$

In particular, we have that

$$(19) \quad \|u^\dagger - \hat{u}\|_{L^2} \leq C_1 2^{j_0/2} \frac{\sigma_s(P_{\Lambda_{j_0}} W^{-1} \Phi u^\dagger)_1}{\sqrt{s}} + C_2(2^{j_0/2} \beta + r).$$

As anticipated, in Theorem 4.5 we have a lower number of measurements than in Theorem 4.4. On the other hand, the error estimate (19) is possibly worse than the estimate (17), because of the quantity  $2^{j_0/2}$  multiplying the first term of the right hand side. However, it should be observed that  $\sigma_s(P_{\Lambda_{j_0}} W^{-1} \Phi u^\dagger)_1 \leq \sigma_s(P_{\Lambda_{j_0}} \Phi u^\dagger)_1$ , and so the comparison is not straightforward.

As a direct consequence of Theorem 4.5, we obtain the following exact recovery result. If there is no truncation error ( $u^\dagger \in \text{span}(\phi_{j,n})_{(j,n) \in \Lambda_{j_0}}$ , i.e.,  $r = 0$ ), if  $\Phi u^\dagger$  is  $s$ -sparse ( $\sigma_s(P_{\Lambda_{j_0}} W^{-1} \Phi u^\dagger)_1 = \sigma_s(\Phi u^\dagger)_1 = 0$ ) and if there is no noise ( $\beta = 0$ ), then

$$\hat{u} = u^\dagger,$$

provided that  $m$  satisfies (18).

As in Section 3.3, we now derive explicit estimates in terms of the noise  $\beta$  and the number of measurements  $m$ .

**Theorem 4.6.** *Consider the setting of Theorem 4.4 and suppose that the same hypotheses are satisfied.*

*Suppose that the following estimates hold:*

$$\sigma_s(\Phi u^\dagger)_1 \leq C s^{1/2-p}, \quad \|P_{\Lambda_j}^\perp \Phi u^\dagger\|_2 \leq C 2^{-aj},$$

for some  $a, p, C > 0$ .

Fix  $\gamma \in (0, 1)$  and let  $\beta \in (0, 1)$  be sufficiently small. Let  $j_0 := \lfloor 2/(2a+1) \log(1/\beta) \rfloor$  and let  $m$  be sufficiently large.

There exist constants  $C_0, C_1 > 0$  depending only on the chosen wavelet basis and on  $a, p, C$ , logarithmically on  $\gamma$  (and on  $d$  and  $\rho$  in the case of the fan beam transform) such that the following holds.

Let  $\hat{u}$  be a solution of the minimization problem (16). Then, with probability exceeding  $1 - \gamma$ , the following recovery estimate holds:

$$\|u^\dagger - \hat{u}\|_{L^2} \leq C_1 \left( \frac{\log^{4/p}(1/\beta)}{\beta^{\frac{2p}{2a+1}} m^p} + \beta^{\frac{2a}{2a+1}} \right).$$

In particular, if

$$m = \lfloor C_0 \beta^{-\frac{2}{2a+1} - \frac{2a}{p(2a+1)}} \log^4(1/\beta) \rfloor.$$

then

$$\|u^\dagger - \hat{u}\|_{L^2} \leq C_1 \beta^{\frac{2a}{2a+1}},$$

or, equivalently,

$$\|u^\dagger - \hat{u}\|_{L^2} \leq C_1 \left( \frac{\log^4 m}{m} \right)^{\frac{ap}{a+p}}.$$

Analogous estimates have been obtained for the sparse Radon transform in the context of statistical inverse learning [27, 29]. These are similar in spirit to the ones presented above, but comparison is non-trivial since the setting is substantially different as sparsity is not considered there, while here it plays a crucial role.

**Example 4.7** (Cartoon-like images). In the case of cartoon-like images supported in  $\mathcal{B}_1$  (see [80, Section 9.2.4] – in short,  $C^2$  signals apart from  $C^2$  edges), with reference to the notation of Theorem 4.6, we have that  $a = 1/2$  and  $p$ -compressibility holds with  $p = 1/2$  up to logarithmic terms – see the end of Section 5.7. There exist constants  $C_0, C_1 > 0$ , which only depend on the chosen wavelet basis in the case of the Radon transform and also on  $d$  and  $\rho$  in the case of the fan beam transform, such that the following bound on the  $L^2$  error holds, up to logarithmic terms:

$$(20) \quad \|u^\dagger - \widehat{u}\|_{L^2} \leq C_1 \left( \frac{1}{\beta^{1/2} m^{1/2}} + \beta^{1/2} \right).$$

Choosing  $m = \lfloor C_0 \beta^{-2} \log^4(1/\beta) \rfloor$ , we get that, up to logarithmic terms,

$$\|u^\dagger - \widehat{u}\|_{L^2} \leq C_1 \beta^{1/2}, \quad \|u^\dagger - \widehat{u}\|_{L^2} \leq C_1 \left( \frac{1}{m} \right)^{1/4}.$$

We refer the reader to Section 5.7 for a proof of (20).

## 5. PROOFS OF THE MAIN RESULTS

In this section, we will usually write  $X \lesssim Y$  if the underlying inequality holds up to a positive constant  $C > 0$ , that is  $X \leq CY$ . Similarly, we write  $X \gtrsim Y$  if  $X \geq CY$  for some  $C > 0$ . In particular,  $X \asymp Y$  means that both  $X \lesssim Y$  and  $X \gtrsim Y$  hold. These constants may depend in general on other quantities related to the results from which the inequalities are inferred, the dependence being made explicit in the corresponding statements.

**5.1. Preliminary results.** Determining the best  $s$ - $\omega$ -sparse approximation of a vector with respect to  $\ell_\omega^p$ , introduced in Section 2.1, is a quite challenging problem in practice. A surrogate notion of quasi-best sparse approximation has been introduced in [91, Section 3]. In fact, the latter construction happens to be powerful enough for the purposes of the proofs given below, hence we briefly review the basic facts in this connection for the benefit of the reader.

Given  $x \in \ell^2(\Gamma)$ ,  $\omega \in [1, +\infty)^\Gamma$  and a positive integer  $s$ , consider the permutation  $\pi: \Gamma \rightarrow \Gamma$  associated with the non-increasing rearrangement of  $(|x_i| \omega_i^{-1})_{i \in \Gamma}$ . Let  $i_0 \in \Gamma$  be the largest index such that

$$\omega(S_{i_0}) \leq s, \quad S_{i_0} := \{\pi(1), \dots, \pi(i_0)\}.$$

We thus define  $x_{S_{i_0}}$  to be the *quasi-best  $s$ - $\omega$ -sparse approximation to  $x$* . Notice that this notion does not depend on a  $\ell_\omega^p$  norm, whereas the companion *error of quasi-best  $s$ - $\omega$ -sparse approximation to  $x \in \ell^2(\Gamma)$  with respect to  $\ell_\omega^p$*  is defined by  $\tilde{\sigma}_s(x)_{p,\omega} := \|x_{S_{i_0}^c}\|_{p,\omega}$ . From the very definition we have that

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

For later use we recall the following *Stechkin-type inequality* – see [3, Lemma 3.12] for a proof. If  $s > 0$  and  $0 < p < q \leq 2$ , then

$$(21) \quad \sigma_s(x)_{q,\omega} \leq \tilde{\sigma}_s(x)_{q,\omega} \leq s^{1/q-1/p} \|x\|_{p,\omega}, \quad x \in \ell^2(\Gamma).$$

To simplify the notation in the proofs, in what follows we will also introduce the following sampling operator  $A: \ell^2(\Gamma) \rightarrow \mathcal{H}_2^m$  associated with a sample  $(t_1, \dots, t_m) \in \mathcal{D}^m$ :

$$Ax := \frac{1}{\sqrt{m}} (F_{t_1} \Phi^* x, \dots, F_{t_m} \Phi^* x).$$

With an abuse of notation, we will use  $A$  to denote also the restriction  $A: \ell^2(\Lambda) \rightarrow \mathcal{H}_2^m$  to some finite  $\Lambda \subset \Gamma$ .

Moreover, with reference to problem (4), we define  $\varepsilon := \frac{1}{\sqrt{m}} (\varepsilon_1, \dots, \varepsilon_m)$  and  $y := \frac{1}{\sqrt{m}} (y_1, \dots, y_m)$ . Notice that the condition  $\max_k \|\varepsilon_k\|_{\mathcal{H}_2} \leq \beta$  implies that

$$\|\varepsilon\|_{\mathcal{H}_2^m}^2 = \frac{1}{m} \sum_{k=1}^m \|\varepsilon_k\|_{\mathcal{H}_2}^2 \leq \beta^2.$$

Moreover, the following identity holds for  $x \in \ell^2(\Lambda)$ :

$$\|Ax - y\|_{\mathcal{H}_2^m}^2 = \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_\Lambda x - y_k\|^2,$$

where  $y_k = F_{t_k} \Phi^* x^\dagger + \varepsilon_k$ , for  $k = 1, \dots, m$ , as defined in Section 2.2.

**5.2. The robust null space property.** The typical strategy to obtain recovery estimates in a RIP-based setting requires to initially introduce a suitably designed version of the *robust null space property* (RNSP), the latter being crucial to ensure stable distance bounds and thus uniform recovery results. We recall that all the Hilbert spaces are assumed to be complex and separable, unless otherwise noted.

**Definition 5.1.** Let  $\mathcal{H}$  be a Hilbert space and let  $\Lambda$  be a finite or countable set. Consider a vector of weights  $\omega \in [1, +\infty)^\Lambda$ , an integer  $s \geq 1$ , and real parameters  $0 < \rho < 1$ ,  $\kappa > 0$ . An operator  $A \in \mathcal{L}(\ell^2(\Lambda), \mathcal{H}^m)$  is said to satisfy the robust null space property (RNSP) with respect to  $(\omega, \rho, \kappa, s)$  if

$$\|x_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|x_{S^c}\|_{1,\omega} + \kappa \|Ax\|_{\mathcal{H}^m}, \quad \forall x \in \ell^2(\Lambda), \quad S \subset \Lambda : \omega(S) \leq s.$$

*Remark 5.2.* A sufficient condition in order for  $A$  to satisfy the RNSP with respect to  $(\omega, \rho, \kappa, s)$  can be stated as follows: there exists  $0 < \rho' < 1/2$  and  $\kappa' > 0$  such that

$$(22) \quad \|x_S\|_2 \leq \frac{\rho'}{\sqrt{s}} \|x\|_{1,\omega} + \kappa' \|Ax\|_{\mathcal{H}^m}, \quad \forall x \in \ell^2(\Lambda), \quad S \subset \Lambda : \omega(S) \leq s.$$

This is indeed an easy consequence of the estimate

$$\|x\|_{1,\omega} = \|x_S\|_{1,\omega} + \|x_{S^c}\|_{1,\omega} \leq \sqrt{s} \|x_S\|_2 + \|x_{S^c}\|_{1,\omega},$$

which follows in turn by the Cauchy-Schwarz inequality. The relationships between the involved constants are given by

$$\rho = \frac{\rho'}{1 - \rho'}, \quad \kappa = \frac{\kappa'}{1 - \rho'}.$$

We now show that the RNSP implies suitable distance bounds. Indeed, the proof of the following results is largely inspired by standard arguments in the theory of compressed sensing – see for instance [53, Theorem 5] and [91, Theorem 4.2]. The proof is essentially identical to [2, Lemma 6.24]; adapting it to the case of Hilbert-valued measurement requires only some minor modifications.

**Theorem 5.3 (RNSP  $\Rightarrow$  distance bounds).** *Let  $\mathcal{H}$  be a Hilbert space and assume that  $A \in \mathcal{L}(\ell^2(\Lambda), \mathcal{H}^m)$  satisfies the RNSP with respect to  $(\omega, \rho, \kappa, s)$  as in Definition 5.1. Then, for all  $x, z \in \ell^1_\omega(\Lambda)$ ,*

$$\|x - z\|_{1,\omega} \leq c'_1 (\|z\|_{1,\omega} - \|x\|_{1,\omega} + 2\sigma_s(x)_{1,\omega}) + c'_2 \sqrt{s} \|A(x - z)\|_{\mathcal{H}^m},$$

where we set

$$c'_1 = \frac{1 + \rho}{1 - \rho}, \quad c'_2 = \frac{2\kappa}{1 - \rho}.$$

Moreover

$$\|x - z\|_2 \leq \frac{c_1}{\sqrt{s}} (\|z\|_{1,\omega} - \|x\|_{1,\omega} + 2\sigma_s(x)_{1,\omega}) + c_2 \|A(x - z)\|_{\mathcal{H}^m},$$

where

$$c_1 = (\rho + 1)c'_1, \quad c_2 = (\rho + 1)c'_2 + \kappa.$$

**5.3. The generalized restricted isometry property.** We now define a generalized version of the standard restricted isometry property (RIP), which corresponds to the G-RIPL without levels, and with a weighted sparsity, introduced in [1, Definition 3.5].

**Definition 5.4 (g-RIP).** Let  $\mathcal{H}$  be a Hilbert space. Given  $G \in \mathcal{L}(\ell^2(\Lambda))$ ,  $0 \leq \delta < 1$ ,  $\lambda > 0$  and a vector of weights  $\omega \in [1, +\infty)^\Lambda$ , we say that  $A \in \mathcal{L}(\ell^2(\Lambda), \mathcal{H}^m)$ ,  $m \in \mathbb{N}$ , satisfies the *generalized restricted isometry property (g-RIP)* with respect to  $(G, \omega, \delta, \lambda)$  if

$$(1 - \delta) \|Gx\|_2^2 \leq \|Ax\|_{\mathcal{H}^m}^2 \leq (1 + \delta) \|Gx\|_2^2, \quad x \in \ell^2(\Lambda), \quad \|x\|_{0,\omega} \leq \lambda.$$

*Remark 5.5 (Diagonal invariance of the g-RIP).* Consider the case where  $\Lambda$  is finite. Let  $Z := (z_i)_{i \in \Lambda}$  be a diagonal matrix with  $z_i \neq 0$ . Since  $\text{supp}(x') = \text{supp}(Zx')$ , we have  $\|x'\|_{0,\omega} = \|Zx'\|_{0,\omega}$ . As a result, after the change of variable  $x = Zx'$ , it is clear that  $A$  satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$  if and only if  $AZ$  satisfies the g-RIP with respect to  $(GZ, \omega, \delta, \lambda)$ .

We now provide sufficient conditions under which the g-RIP implies the RNSP. We also refer the reader to [1, Theorem 5.5] for a version of this result adapted to sparsity in levels.

**Theorem 5.6 (g-RIP  $\Rightarrow$  RNSP).** Fix  $\rho' \in (0, 1/2)$  and  $\delta \in (0, 1)$ . Assume that  $A \in \mathcal{L}(\ell^2(\Lambda), \mathcal{H}^m)$  satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$  for some invertible  $G \in \mathcal{L}(L^2(\Lambda))$ , weights  $\omega \in [1, +\infty)^\Lambda$  and  $\lambda \in \mathbb{R}^+$ . For any  $s \geq \|\omega\|_\infty^2/4$ , if

$$\lambda \geq \frac{5}{\rho'^2} \frac{1+\delta}{1-\delta} \|G^{-1}\|^2 \|G\|^2 s,$$

then  $A$  satisfies the RNSP with respect to  $(\omega, \rho, \kappa, s)$ , where

$$\rho = \frac{\rho'}{1-\rho'}, \quad \kappa = \frac{\|G^{-1}\|}{(1-\rho')\sqrt{1-\delta}}.$$

*Proof.* In view of Remark 5.2, it suffices to prove that (22) is satisfied.

Fix  $x \in \ell^2(\Lambda)$  and  $S \subset \Lambda$  with  $\omega(S) \leq s$ . Consider  $x_{S^c}$ . Having in mind the construction of the quasi-best  $\lambda$ - $\omega$ -sparse approximation to  $x$  in Section 5.1, we now proceed as follows.

Let  $\pi: \mathbb{N} \rightarrow S^c$  be the enumeration yielding the non-increasing rearrangement of  $(|x_k| \omega_k^{-1})_{k \in S^c}$ , so that  $|x_{\pi(i)}| \omega_{\pi(i)}^{-1} \geq |x_{\pi(i+1)}| \omega_{\pi(i+1)}^{-1}$ . Set  $i_0 = 0$  and for  $j \geq 1$  iteratively define  $i_j \in \mathbb{N}$  to be the largest index such that  $i_j > i_{j-1}$  and

$$\omega(\Omega_j) \leq \tilde{\lambda}, \quad \Omega_j := \{\pi(i_{j-1} + 1), \dots, \pi(i_j)\},$$

where

$$(23) \quad \tilde{\lambda} = \frac{4}{\rho'^2} \frac{1+\delta}{1-\delta} \|G^{-1}\|^2 \|G\|^2 s.$$

Note that  $\lambda \geq \tilde{\lambda} + \frac{1}{\rho'^2} \frac{1+\delta}{1-\delta} \|G^{-1}\|^2 \|G\|^2 s \geq \tilde{\lambda} + s$  since  $\rho' \in (0, 1/2)$ . We stress that each  $\Omega_j$  is non-empty, thanks to the fact that  $\tilde{\lambda} \geq 4s \geq \|\omega\|_\infty^2$  by assumption, and it is finite since  $\omega_i \geq 1$ .

It is straightforward to realize that  $\cup_{j \geq 1} \Omega_j = S^c$ . Moreover, for  $j \geq 1$ , it follows from the definition that  $x_{\Omega_j}$  is the quasi-best  $\tilde{\lambda}$ - $\omega$ -sparse approximation to  $x_{(\Omega_j \cup \Omega_{j+1})}$ . The Stechkin-type inequality (21) gives

$$(24) \quad \|x_{\Omega_j}\|_2 = \tilde{\sigma}_{\tilde{\lambda}}(x_{(\Omega_{j-1} \cup \Omega_j)})_{2,\omega} \leq \frac{1}{\sqrt{\tilde{\lambda}}} \|x_{(\Omega_{j-1} \cup \Omega_j)}\|_{1,\omega}, \quad j \geq 2.$$

Since  $x_S + x_{\Omega_1} = x - \sum_{j \geq 2} x_{\Omega_j}$  by construction, the g-RIP for  $A$  yields

$$\begin{aligned}
\|x_S\|_2 &\leq \|x_S + x_{\Omega_1}\|_2 \leq \|G^{-1}\| \|G(x_S + x_{\Omega_1})\|_2 \\
(25) \qquad &\leq \frac{\|G^{-1}\|}{\sqrt{1-\delta}} \|A(x_S + x_{\Omega_1})\|_{\mathcal{H}^m} \\
&\leq \frac{\|G^{-1}\|}{\sqrt{1-\delta}} \left( \sum_{j \geq 2} \|Ax_{\Omega_j}\|_{\mathcal{H}^m} + \|Ax\|_{\mathcal{H}^m} \right).
\end{aligned}$$

Now, by (24), we have

$$(26) \qquad \|Ax_{\Omega_j}\|_{\mathcal{H}^m} \leq \sqrt{1+\delta} \|Gx_{\Omega_j}\|_2 \leq \sqrt{1+\delta} \|G\| \|x_{\Omega_j}\|_2 \leq \sqrt{1+\delta} \|G\| \frac{1}{\sqrt{\lambda}} \|x_{(\Omega_{j-1} \cup \Omega_j)}\|_{1,\omega}.$$

We have

$$(27) \qquad \sum_{j \geq 2} \|x_{(\Omega_{j-1} \cup \Omega_j)}\|_{1,\omega} \leq \sum_{j \geq 2} \|x_{\Omega_{j-1}}\|_{1,\omega} + \|x_{\Omega_j}\|_{1,\omega} \leq 2\|x\|_{1,\omega}.$$

Combining (25), (26) and (27) we obtain

$$\|x_S\|_2 \leq \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \|G^{-1}\| \|G\| \frac{2\sqrt{s}}{\sqrt{\lambda}} \frac{1}{\sqrt{s}} \|x\|_{1,\omega} + \frac{\|G^{-1}\|}{\sqrt{1-\delta}} \|Ax\|_{\mathcal{H}^m}.$$

It is now enough to realize that (23) implies

$$\frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \|G^{-1}\| \|G\| \frac{2\sqrt{s}}{\sqrt{\lambda}} = \rho'.$$

The proof of (22), hence of the claim, is concluded.  $\square$

**5.4. Sufficient conditions for the g-RIP via subsampling.** We now provide sufficient conditions for the g-RIP to be satisfied by the sampling operator of the model.

**Theorem 5.7 (Sampling  $\Rightarrow$  g-RIP).** *Consider the setting of Section 2.2. Suppose that Assumptions 3.1 and 3.3 are satisfied. Define*

$$G := \sqrt{P_\Lambda \Phi F^* F \Phi^* \iota_\Lambda}.$$

*Suppose that  $G$  is invertible.*

*Let  $(t_1, \dots, t_m) \in \mathcal{D}^m$ ,  $m \in \mathbb{N}$ , be independent and identically distributed samples from  $\nu$  (with possible repetitions to be kept). Let  $A$  be the sampling operator associated with  $(t_1, \dots, t_m)$ , namely*

$$A := \left( \frac{1}{\sqrt{m}} F_{t_k} \Phi^* \iota_\Lambda \right)_{k=1}^m,$$

and introduce the operator

$$(28) \quad Q \in \mathcal{L}(\mathcal{H}^m), \quad Q(h_1, \dots, h_m) = (f_\nu(t_1)^{-1/2}h_1, \dots, f_\nu(t_m)^{-1/2}h_m).$$

Set  $M = |\Lambda|$  and  $\lambda \in [M]$ . Assume that  $\tau := B^2 \|G^{-1}\|^2 \lambda \geq 3$ . There exists a universal constant  $C > 0$  such that, for any  $\gamma, \delta \in (0, 1)$ , if

$$(29) \quad m \geq C\delta^{-2}\tau \max\{\log^3 \tau \log M, \log(1/\gamma)\},$$

then  $QA$  satisfies the  $g$ -RIP with respect to  $(G, \omega, \delta, \lambda)$  with probability exceeding  $1 - \gamma$ .

The proof of this result follows a well-established pathway in the theory of compressed sensing, aimed at obtaining probabilistic bounds for  $\delta$  – see for instance the arguments and techniques used in the proofs of [1, Theorem 3.6], [74, Theorem 3.1], [77, Theorem 3.2] and [91, Theorem 5.2]. We refer the reader to Appendix A for a proof of Theorem 5.7.

**5.5. Proof of Theorem 3.5.** The last ingredient needed to obtain the proof of Theorem 3.5 is the following estimate on the error arising from the truncation of the infinite dimensional unknown  $x^\dagger$  to the finite-dimensional set  $\Lambda$ .

**Proposition 5.8 (Truncation error).** *Consider the setting of Theorem 3.5. Suppose that  $\|P_\Lambda^\perp x^\dagger\|_2 \leq r$  and*

$$(30) \quad m \geq \|G_1^{-1}\|^2 \|G_2^{-1}\|^2 \|G_2\|^2 \log(1/\gamma)$$

for some invertible  $M \times M$  matrices  $G_1, G_2$  such that one of these conditions is satisfied:

- (1)  $G_1 = G_2$ ,
- (2)  $\|G_1^{-1}\| \geq c$  for some constant  $c$ .

Let  $Q$  be defined as in (28). Then, with probability exceeding  $1 - \gamma$ , we have that

$$\|G_2^{-1}\| \|QA P_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m} \leq C_1 c_\nu^{-1/2} (\|F\Phi^* \iota_{\Gamma \setminus \Lambda}\| \|G_2^{-1}\| + 1) r,$$

where  $C_1 > 0$  is a constant depending only on  $C_F$  (and on  $c$  in the second case).

*Proof.* Let  $u_R := \Phi^* P_\Lambda^\perp x^\dagger$  and  $c_F := \|F\Phi^* \iota_{\Gamma \setminus \Lambda}\|$ . We apply Lemma A.1 to the random variables  $Y_k(t) := \|QF_{t_k} u_R\|_{\mathcal{H}_2}^2 - \|F u_R\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2$  in the case where the family of functions  $f_x$  consists just of the identity mapping; as a consequence, we obtain as a special case the classical version of Bernstein's inequality. Notice that, in this case,

$$\|QA P_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m}^2 = \frac{1}{m} \sum_{k=1}^m f_\nu(t_k)^{-1} \|F_{t_k} u_R\|_{\mathcal{H}_2}^2.$$

We claim that  $\mathbb{E}(f_\nu(t_k)^{-1}\|F_{t_k}u_R\|_{\mathcal{H}_2}^2 - \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2) = 0$ . Indeed, by definition of  $F$ , we have that

$$\int_{\mathcal{D}} \|F_t u\|_{\mathcal{H}_2}^2 d\mu(t) = \|Fu\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2, \quad u \in \mathcal{H}_1,$$

which is equivalent to the claim. Moreover, we have that

$$\begin{aligned} |f_\nu(t_k)^{-1}\|F_{t_k}u_R\|_{\mathcal{H}_2}^2 - \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2| &\leq c_\nu^{-1}\|F_{t_k}u_R\|_{\mathcal{H}_2}^2 + \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2 \\ &\leq 2C_F^2 c_\nu^{-1} r^2 =: K, \end{aligned}$$

where we have used the fact that  $\|F_t\|, \|F\| \leq C_F$ . Finally, we have that

$$\begin{aligned} &\mathbb{E}|f_\nu(t_k)^{-1}\|F_{t_k}u_R\|_{\mathcal{H}_2}^2 - \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2|^2 \\ &= \mathbb{E}|f_\nu(t_k)^{-1/2}\|F_{t_k}u_R\|_{\mathcal{H}_2}|^4 - \left(\mathbb{E}\|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2\right)^2 \\ &\leq C_F^2 c_\nu^{-1} r^2 \mathbb{E}|f_\nu(t_k)^{-1/2}\|F_{t_k}u_R\|_{\mathcal{H}_2}|^2 \\ &= C_F^2 c_\nu^{-1} r^2 \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2 \leq C_F^2 c_\nu^{-1} c_F^2 r^4 =: \Sigma^2. \end{aligned}$$

By the Bernstein inequality (Lemma A.1) applied to  $\varepsilon m$  in place of  $\varepsilon$ , we obtain

$$\begin{aligned} &\mathbb{P}\left(\|QAu_R\|^2 \geq \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2 + \varepsilon\right) \\ &= \mathbb{P}\left(\sum_{k=1}^m f_\nu(t_k)^{-1}\|F_{t_k}u_R\|_{\mathcal{H}_2}^2 \geq m\|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2 + m\varepsilon\right) \\ &\leq \exp\left(-\frac{1}{C} \frac{m^2 \varepsilon^2}{m C_F^2 c_\nu^{-1} c_F^2 r^4 + m C_F^2 c_\nu^{-1} r^4 c_F^2 + C_F^2 c_\nu^{-1} r^2 m \varepsilon}\right) \\ &\leq \exp\left(-\frac{1}{C C_F^2 c_\nu^{-1} c_F^2 r^4 + r^2 \varepsilon} m \varepsilon^2\right), \end{aligned}$$

where  $C > 0$  is an absolute constant.

Therefore, with probability exceeding  $1 - \gamma$ , the following estimate holds:

$$\|QAu_R\|_{\mathcal{H}_2^m}^2 \leq \|Fu_R\|_{L_\mu^2(\mathcal{D};\mathcal{H}_2)}^2 + \varepsilon,$$

provided that

$$m \geq \tilde{C} \frac{c_F^2 r^4 + r^2 \varepsilon}{\varepsilon^2} \log(1/\gamma),$$

where  $\tilde{C} := C C_F^2 c_\nu^{-1}$ .

We now invert the relation between  $m = m(\varepsilon)$  and  $\varepsilon$ , obtaining  $\varepsilon = \varepsilon(m)$ . Then, the inequality above will be trivially satisfied for  $m$  as in (30) and  $\varepsilon = \varepsilon(m)$ . We have

that  $\varepsilon = \varepsilon(m)$  is given by

$$\varepsilon = \frac{\tilde{C} \log(1/\gamma)r^2 + \sqrt{\tilde{C}^2 \log^2(1/\gamma)r^4 + 4\tilde{C} \log(1/\gamma)c_F^2 r^4 m}}{2m}.$$

Therefore, we obtain that, with probability exceeding  $1 - \gamma$ ,

$$\begin{aligned} \|QAu_R\|_{\mathcal{H}_2^m}^2 &\leq \|Fu_R\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 + \varepsilon \\ &\leq c_F^2 r^2 + \frac{\tilde{C} \log(1/\gamma)}{m} r^2 + \sqrt{\frac{\tilde{C} \log(1/\gamma)}{m}} c_F r^2 \\ &\leq \left( c_F^2 + \frac{CC_F^2}{\|G_1^{-1}\|^2 \|G_2^{-1}\|^2 \|G_2\|^2} + \frac{\sqrt{CC_F^2 c_F}}{\|G_1^{-1}\| \|G_2^{-1}\| \|G_2\|} \right) c_\nu^{-1} r^2. \end{aligned}$$

We now distinguish between the two cases in the statement of the Proposition. If  $G_1 = G_2 =: G$ , then

$$\begin{aligned} \|G^{-1}\|^2 \|QAu_R\|_{\mathcal{H}_2^m}^2 &\leq \left( c_F^2 \|G^{-1}\|^2 + \frac{CC_F^2}{\|G^{-1}\|^2 \|G\|^2} + \frac{\sqrt{CC_F^2 c_F}}{\|G\|} \right) c_\nu^{-1} r^2 \\ &\leq \left( c_F^2 \|G^{-1}\|^2 + CC_F^2 + c_F \|G^{-1}\| \sqrt{CC_F^2} \right) c_\nu^{-1} r^2 \\ &\leq C_1^2 (c_F^2 \|G^{-1}\|^2 + 1) c_\nu^{-1} r^2, \end{aligned}$$

with  $C_1$  depending only on  $C_F$ .

On the other hand, if  $\|G_1^{-1}\| \geq c$ , we get that

$$\begin{aligned} \|G_2^{-1}\|^2 \|QAu_R\|_{\mathcal{H}_2^m}^2 &\leq \left( c_F^2 \|G_2^{-1}\|^2 + \frac{CC_F^2}{\|G_1^{-1}\|^2 \|G_2\|^2} + \frac{\sqrt{CC_F^2 c_F}}{\|G_1^{-1}\| \|G_2\|} \|G_2^{-1}\| \right) c_\nu^{-1} r^2 \\ &\leq \left( c_F^2 \|G_2^{-1}\|^2 + CC_F^2 c^{-2} \|G_2^{-1}\|^2 + c_F \sqrt{CC_F^2} c^{-1} \|G_2^{-1}\|^2 \right) c_\nu^{-1} r^2 \\ &\leq C_1^2 (c_F^2 \|G_2^{-1}\|^2 + 1) c_\nu^{-1} r^2, \end{aligned}$$

with  $C_1$  depending only on  $C_F$  and  $c$ .  $\square$

The proof of Theorem 3.5 follows by a concatenation of the results already proved in this section.

*Proof of Theorem 3.5.* We have that  $m$  satisfies (29) in Theorem 5.7 for

$$\lambda = C_0 \|G^{-1}\|^2 \|G\| s.$$

As a consequence of Theorem 5.7 we have that  $QA$ , where  $Q$  is defined as in (28), satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$ , with  $\delta = 1/2$ . Using Theorem 5.6 (with  $\delta = 1/2$  and  $\rho' = 1/3$ ), we have that  $QA$  satisfies the RNSP with respect to

$(\omega, \rho, \kappa, s)$  with  $\rho = 1/2$  and  $\kappa = \frac{3}{2\sqrt{1-\delta^*}}\|G^{-1}\| = \frac{3}{\sqrt{2}}\|G^{-1}\|$ . We have that  $C_0 = \frac{5}{\rho^2}(1+\delta)/(1-\delta) = 135$ .

Notice that  $x^\dagger \in \ell^2(\Gamma)$ , while the g-RIP property of  $QA$  holds on the finite dimensional subspace  $\ell^2(\Lambda) \subset \ell^2(\Gamma)$ . We can split the error term as follows:

$$\|x^\dagger - \hat{x}\|_2 \leq \|P_\Lambda x^\dagger - \hat{x}\|_2 + \|P_\Lambda^\perp x^\dagger\|_2 \leq \|P_\Lambda x^\dagger - \hat{x}\|_2 + r.$$

We now estimate the first term. With reference to the  $\ell^2$  bounds in Theorem 5.3 applied to  $P_\Lambda x^\dagger$  in place of  $x^\dagger$  we get the following estimate:

$$\|P_\Lambda x^\dagger - \hat{x}\|_2 \leq \frac{c_1}{\sqrt{s}} (\|\hat{x}\|_{1,\omega} - \|P_\Lambda x^\dagger\|_{1,\omega} + 2\sigma_s(P_\Lambda x^\dagger)_{1,\omega}) + c_2 \|G^{-1}\| \|QA(P_\Lambda x^\dagger - \hat{x})\|_{\mathcal{H}_2^m}$$

for absolute constants  $c_1, c_2 > 0$ . Notice that  $P_\Lambda x^\dagger$  satisfies the constraint of problem (5); indeed, we can exploit Assumption 3.3 and Proposition 5.8 applied in the case  $G_1 = G_2 = G$  to conclude that the following bound holds with probability exceeding  $1 - \gamma$ :

$$\begin{aligned} \left( \frac{1}{m} \sum_{k=1}^m \|F_{t_k} \Phi^* \iota_\Lambda x - y_k\|_{\mathcal{H}_2}^2 \right)^{\frac{1}{2}} &= \|AP_\Lambda x^\dagger - y\|_{\mathcal{H}_2^m} \\ &\leq \|Ax^\dagger - y\|_{\mathcal{H}_2^m} + \|AP_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m} \\ &\leq \beta + \|Q^{-1}\| \|QAP_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m} \\ &\leq \beta + C_3 \|G^{-1}\|^{-1} r. \end{aligned}$$

Using the fact that  $\hat{x}$  is a minimizer of (5), we get that

$$(31) \quad \|P_\Lambda x^\dagger - \hat{x}\|_2 \leq \frac{2c_1}{\sqrt{s}} \sigma_s(P_\Lambda x^\dagger)_{1,\omega} + c_2 \|G^{-1}\| \|QA(P_\Lambda x^\dagger - \hat{x})\|_{\mathcal{H}_2^m}.$$

We can split the last term on the RHS of (31) as follows:

$$\|QA(P_\Lambda x^\dagger - \hat{x})\|_{\mathcal{H}_2^m} \leq \|QAx^\dagger - Qy\|_{\mathcal{H}_2^m} + \|QA\hat{x} - Qy\|_{\mathcal{H}_2^m} + \|QAP_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m}.$$

The first term satisfies

$$\|QAx^\dagger - Qy\|_{\mathcal{H}_2^m} \leq c_\nu^{-1/2} \beta$$

by the assumption on the noise level and by Assumption 3.3. The second term satisfies

$$\|QA\hat{x} - Qy\|_{\mathcal{H}_2^m} \leq c_\nu^{-1/2} (\beta + C_3 \|G^{-1}\|^{-1} r)$$

because  $\hat{x}$  satisfies the constraint of problem (5). Finally, the third term satisfies

$$\|QAP_\Lambda^\perp x^\dagger\|_{\mathcal{H}_2^m} \leq C_3 \|G^{-1}\|^{-1} r \leq c_\nu^{-1/2} C_3 \|G^{-1}\|^{-1} r$$

with probability exceeding  $1 - \gamma$  by Proposition 5.8 applied in the case  $G_1 = G_2 = G$ .

These bounds together imply the following bound on the last term in the RHS of (31):

$$\|G^{-1}\| \|QA(P_\Lambda x^\dagger - \widehat{x})\|_{\mathcal{H}_2^m} \leq 2c_\nu^{-1/2} (\|G^{-1}\| \beta + C_3 r).$$

The bound given by Proposition 5.8 holds together with  $QA$  satisfying the g-RIP with probability exceeding  $1 - 2\gamma$ . This concludes the proof in the  $\ell^2$  case. The  $\ell_\omega^1$  case is dealt analogously.  $\square$

**5.6. Quantitative bounds on  $\|G^{-1}\|$ .** We prove the results stated in Section 3.2.

*Proof of Proposition 3.8.* The assumptions imply that

$$\|Fu\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 \asymp \|u\|_{H^{-b}}^2 \asymp \sum_{(j,n) \in \Gamma} 2^{-2bj} |\langle u, \phi_{j,n} \rangle|^2.$$

Consider now  $x \in \ell^2(\Gamma)$  and set  $u := \Phi^* x$ . Since  $(\phi_{j,n})_{(j,n) \in \Gamma}$  is an orthonormal basis by assumption, we have  $x_{j,n} = \langle u, \phi_{j,n} \rangle_{L^2}$ . Then

$$\|F\Phi^* x\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 = \|Fu\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 \asymp \sum_{(j,n) \in \Gamma} 2^{-2bj} |\langle u, \phi_{j,n} \rangle|^2 = \sum_{(j,n) \in \Gamma} 2^{-2bj} |x_{j,n}|^2,$$

hence  $F$  satisfies the quasi-diagonalization property as claimed.  $\square$

**Proposition 5.9.** *Suppose that an operator  $F$  satisfies the quasi-diagonalization property (6) with respect to  $(\Phi, b)$   $b \geq 0$ . Let  $G := \sqrt{P_\Lambda \Phi F^* F \Phi^* \iota_\Lambda}$  and  $Z := \text{diag}(z_{j,n})_{(j,n) \in \Lambda}$  be a diagonal matrix, with  $z_{j,n} > 0$  for every  $(j,n) \in \Lambda$ . Then the following estimate holds:*

$$\|(GZ)^{-1}\|^2 \asymp \max_{(j,n) \in \Lambda} (z_{j,n}^{-2} 2^{2bj}),$$

where the implicit constants depend on the quasi-diagonalization constants in (6). In particular, if  $z_{j,n} \equiv 1$  and  $\Lambda = \Lambda_{j_0}$ , then

$$\|G^{-1}\|^2 \leq C 2^{2bj_0}.$$

*Proof.* Set  $W := \text{diag}(2^{bj})_{(j,n) \in \Lambda}$ . By (6) we have that

$$\|GZx\|_2^2 = \|F\Phi^* \iota_\Lambda Zx\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2 \geq c \|W^{-1} Zx\|_2^2 \geq c \min_{(j,n) \in \Lambda} (z_{j,n}^2 2^{-2bj}) \|x\|_2^2.$$

Combining these two estimates yields

$$\|x\|_2^2 \leq \frac{1}{c} \max_{(j,n) \in \Lambda} (z_{j,n}^{-2} 2^{2bj}) \|GZx\|_2^2.$$

This proves the upper bound on  $\|(GZ)^{-1}\|^2$ .

On the other hand, let  $(j', n') \in \Lambda$  be such that

$$z_{j',n'}^2 2^{-2bj'} = \min_{(j,n) \in \Lambda} (z_{j,n}^2 2^{-2bj}).$$

Then we have that

$$\|GZe_{j',n'}\|_2^2 \leq C\|W^{-1}Ze_{j',n'}\|_2^2 = C \min_{(j,n) \in \Lambda} (z_{j,n}^2 2^{-2bj}).$$

Let  $f_{j',n'} := \frac{GZe_{j',n'}}{\sqrt{C} \min_{(j,n) \in \Lambda} (z_{j,n}^2 2^{-2bj})}$ . The previous bound implies that  $\|f_{j',n'}\|_2 \leq 1$ . Therefore, we have that

$$\|(GZ)^{-1}\|^2 = \sup_{\|y\|_2 \leq 1} \|(GZ)^{-1}y\|_2^2 \geq \|(GZ)^{-1}f_{j',n'}\|_2^2 = \frac{1}{C \min_{(j,n) \in \Lambda} (z_{j,n}^2 2^{-2bj})}.$$

This proves the lower bound on  $\|(GZ)^{-1}\|^2$ .  $\square$

*Proof of Theorem 3.11.* Let  $D := \text{diag}(d_{j,n})_{(j,n) \in \Lambda_{j_0}}$  and let  $\tilde{D}$  be the operator acting on  $\text{span}(\phi_{j,n})_{(j,n) \in \Gamma}$  by  $\tilde{D}\phi_{j,n} := d_{j,n}\phi_{j,n}$  for  $(j,n) \in \Lambda_{j_0}$ ,  $\tilde{D}\phi_{j,n} = \phi_{j,n}$  otherwise, so that  $\Phi^*D = \tilde{D}\Phi^*$ . Moreover, we extend  $W$ , and therefore  $W^\zeta$ , to an operator on  $\ell^2(\Gamma)$ , setting  $W = \text{diag}(w_{j,n})_{j,n}$ , where

$$w_{j,n} = \begin{cases} 2^{bj} & \text{if } (j,n) \in \Lambda_{j_0}, \\ 1 & \text{otherwise.} \end{cases}$$

Similarly, set  $\tilde{W}^\zeta = \Phi^*W^\zeta\Phi$ .

Consider the measurement operators given by  $(F_t\tilde{D})$ . The companion forward map is given by  $F\tilde{D}$ , while the truncated matrix (i.e., playing the role of  $G$ ) is given by  $\tilde{G} := \sqrt{D^*G^*GD}$  and the corresponding sampling matrix is  $AD$ . Furthermore,  $\|(\tilde{G})^{-1}\| = \|(GD)^{-1}\|$ .

The following bounds hold by Proposition 5.9 and by the quasi-diagonalization property:

$$(32) \quad \|(GD)^{-1}\|^2 \asymp \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}), \quad \|(GW^\zeta)^{-1}\|^2 \asymp 2^{2(1-\zeta)bj_0}, \quad \|GW^\zeta\| \lesssim 1,$$

where the implicit constants depend on the quasi-diagonalization constants. Let  $\lambda$  be defined by

$$\lambda := \tilde{C}_0 \|GW^\zeta\|^2 \|(GW^\zeta)^{-1}\|^2 s,$$

where  $\tilde{C}_0 = 135$  is the constant appearing in the expression for  $\lambda$  in Theorem 5.6 with  $\delta = 1/2$  and  $\rho' = 1/3$ .

Using these bounds, we deduce that, if  $C_0$  is large enough, the sample complexity  $m$  in (11) satisfies (29), namely,

$$m \geq \frac{C}{4} \tau \max\{\log^3 \tau \log M, \log(1/\gamma)\}$$

with  $\tau = B^2 \|(GD)^{-1}\|^2 \lambda \geq 3$ . By Theorem 5.7, we conclude that  $QAD$  satisfies the g-RIP with respect to  $(GD, \omega, 1/2, \lambda)$  with overwhelming probability. The diagonal

invariance of the g-RIP proved in Remark 5.5 shows that the latter result implies that  $QA$  satisfies the g-RIP with respect to  $(G, \omega, 1/2, \lambda)$ .

By virtue of the diagonal invariance of the g-RIP illustrated in Remark 5.5,  $QAW^\zeta$  equivalently satisfies the g-RIP with respect to  $(GW^\zeta, \omega, 1/2, \lambda)$ . Theorem 5.6 now implies that  $QAW^\zeta$  satisfies the RNSP with respect to  $(\omega, \rho, \kappa, s)$  with  $\rho = 1/2$  and  $\kappa = \frac{3}{\sqrt{2}} \|(GW^\zeta)^{-1}\| \lesssim 2^{(1-\zeta)bj_0}$ .

The following identity holds:

$$y = AW^\zeta z^\dagger + \varepsilon,$$

where  $z^\dagger := W^{-\zeta} x^\dagger$ . Notice that  $\|P_{\Lambda_{j_0}}^\perp z^\dagger\|_2 = \|W^{-\zeta} P_{\Lambda_{j_0}}^\perp x^\dagger\|_2 \leq 2^{-\zeta bj_0} r =: r'$ . Let  $\widehat{z}$  be a solution of the problem

$$\min_{z \in \ell^2(\Lambda_{j_0})} \|z\|_{1, \omega} : \|AW^\zeta z - y\|_{\mathcal{H}_2^m}^2 \leq (\beta + C_3 2^{(\zeta-1)bj_0} r')^2 = (\beta + C_3 2^{-bj_0} r)^2,$$

where  $C_3$  is a constant to be assigned later. We also introduce the notation  $\widehat{x} := W^\zeta \widehat{z}$ . The change of variable  $x = W^\zeta z$  shows that this problem is equivalent to (10).

We now proceed as in the proof of Theorem 3.5. Since

$$\|W^{-\zeta} x^\dagger - W^{-\zeta} \widehat{x}\|_2 = \|z^\dagger - \widehat{z}\|_2 \leq \|P_{\Lambda_{j_0}} z^\dagger - \widehat{z}\|_2 + r',$$

it is enough to obtain a bound for  $\|P_{\Lambda_{j_0}} z^\dagger - \widehat{z}\|_2$ .

The companion forward map associated with the measurement operators  $(F_t \widetilde{W}^\zeta)$  is given by  $F \widetilde{W}^\zeta$ . Using the quasi-diagonalization property (6), we deduce that  $\|F \widetilde{W}^\zeta\| \leq C$ , where  $C$  is the constant in the quasi-diagonalization upper bound; therefore, an upper bound  $C_{F \widetilde{W}^\zeta}$  associated with this forward map is given by  $C$ . Using Proposition 5.8 applied in the case  $G_1 = GD$  and  $G_2 = GW^\zeta$  (indeed notice that, by (32),  $\|(GD)^{-1}\|$  is bounded from below by a constant depending on the quasi-diagonalization bounds, since  $d_{j,n} \leq 2^{bj}$ ), we get that, with probability exceeding  $1 - \gamma$ , the following bound holds:

$$\begin{aligned} \|QAW^\zeta P_{\Lambda_{j_0}}^\perp z^\dagger\|_{\mathcal{H}_2^m} &\leq C c_\nu^{-1/2} \|(GW^\zeta)^{-1}\|^{-1} (\|F \Phi^* W^\zeta_{\ell_{\Gamma \setminus \Lambda_{j_0}}}\| \|(GW^\zeta)^{-1}\| + 1) r' \\ &= C c_\nu^{-1/2} (\|F \Phi^* W^\zeta_{\ell_{\Gamma \setminus \Lambda_{j_0}}}\| + \|(GW^\zeta)^{-1}\|^{-1}) r' \\ &\leq C c_\nu^{-1/2} 2^{(\zeta-1)bj_0} r' = C c_\nu^{-1/2} 2^{-bj_0} r, \end{aligned}$$

where  $C > 0$  depends only on the quasi-diagonalization constants; we have also used (32) to bound  $\|(GW^\zeta)^{-1}\|^{-1} \lesssim 2^{(\zeta-1)bj_0}$  and the quasi-diagonalization property to bound  $\|F \Phi^* W^\zeta_{\ell_{\Gamma \setminus \Lambda_{j_0}}}\| \lesssim 2^{(\zeta-1)bj_0}$ . We set  $C_3 := C c_\nu^{-1/2}$ . Arguing as in the proof of Theorem 3.5, we can infer that  $z^\dagger$  satisfies the constraints of the minimization problem; using the distance bounds from Theorem 5.3, we conclude.  $\square$

*Proof of Corollary 3.14.* It is enough to apply Theorem 3.11 in the case  $\zeta = 0$ .  $\square$

*Proof of Corollary 3.15.* It suffices to apply Theorem 3.11 in the case  $\zeta = 1$  and to notice that, by the quasi-diagonalization property (6),

$$\|F\Phi^*x^\dagger - F\Phi^*\widehat{x}\|_{L^2_\mu(\mathcal{D};\mathcal{H}_2)} \asymp \|W^{-1}x^\dagger - W^{-1}\widehat{x}\|_2. \quad \square$$

*Proof of Theorem 3.17.* We first introduce the convenient notation

$$\tilde{m} := m/(\tilde{C}_0 B \log^3 \tau \log M) \gtrsim m/(\tilde{C}_0 B c_{j_0} \log^3 \tau),$$

where  $\tilde{C}_0 = C_0 \log(1/\gamma)$  and  $C_0, B$  are the constants appearing in Theorem 3.14 and  $\tau := 2^{bj_0} s$  for some  $s \geq 3$ . Notice that, by the choice of  $j_0$ , we have  $j_0 \asymp \log(1/\beta)$  and

$$\log \tau \leq 2bj_0 + \log s \leq bj_0 + \log M \asymp j_0 \asymp \log(1/\beta).$$

We conclude that

$$(33) \quad \tilde{m} \gtrsim m/\log^4(1/\beta).$$

Suppose that  $\beta$  is sufficiently small so that  $\log^3 \tau \log M \geq \log(1/\gamma)$  – recall that  $j_0 = j_0(\beta)$ . By Theorem 3.14, if

$$\tilde{m} \geq 2^{2(b-d)j_0} 2^{2bj_0} s \geq \max_{(j,n) \in \Lambda_{j_0}} (d_{j,n}^{-2} 2^{2bj}) 2^{2bj_0} s$$

then, with probability exceeding  $1 - \gamma$ , the following error estimate holds:

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 \left( \frac{\sigma_s(x^\dagger)_1}{\sqrt{s}} + 2^{bj_0} \beta + r \right).$$

Suppose that  $2^{-bj_0} \asymp \beta^\zeta$  for some  $\zeta > 0$ . Therefore  $\tilde{m} = \beta^{-2\frac{(2b-d)}{b}\zeta} s$  and

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 \left( s^{-p} + \beta^{1-\zeta} + \beta^{\frac{a}{b}\zeta} \right).$$

Imposing  $1 - \zeta = a\zeta/b$ , we get that  $\zeta = b/(a+b)$ , which yields  $2^{-bj_0} \asymp \beta^{\frac{b}{a+b}}$ .

Combining the expressions for  $s$  and  $\zeta$  with the estimates, we get

$$\|x^\dagger - \widehat{x}\|_2 \leq C_1 \left( \frac{\beta^{-2\frac{(2b-d)}{a+b}p}}{\tilde{m}^p} + \beta^{\frac{a}{a+b}} \right).$$

Using (33), the first estimate in the claim follows.

Moreover, if  $\tilde{m} = \beta^{-2\frac{(2b-d)}{a+b} - \frac{a}{p(a+b)}}$ , then  $\tilde{m}^p = \beta^{-2\frac{(2b-d)}{a+b}p - \frac{a}{a+b}}$  and therefore

$$\frac{\beta^{-2\frac{(2b-d)}{a+b}p}}{\tilde{m}^p} = \beta^{\frac{a}{a+b}}.$$

The second estimate follows.

The last estimate is obtained by inverting the relation  $\tilde{m} = \tilde{m}(\beta)$ , using again (33) and observing that  $\log m \asymp \log 1/\beta$ .  $\square$

**5.7. The sparse Radon transform.** We recall some notation from Section 4.

The space of signals for both problems is given by  $\mathcal{H}_1 = \overline{\text{span}(\phi_{j,n})_{(j,n) \in \Gamma}}$ , while  $\mathcal{H}_2 = L^2(\mathbb{R})$  is the codomain for the Radon transform at a fixed angle  $\theta \in [0, 2\pi)$  and  $\mathcal{H}_2 = L^2(-\pi/2, \pi/2)$  is the codomain for the fan beam transform from the angle  $\theta$ . Let  $K$  be the compact set defined by

$$K := \overline{\bigcup_{(j,n) \in \Gamma} \text{supp } \phi_{j,n}},$$

so that  $\mathcal{H}_1 \subset L^2(K)$ .

A suitable measurement space for this model is  $(\mathcal{D}, \mu) = ([0, 2\pi), d\theta)$ , where  $d\theta$  is the uniform probability measure on  $[0, 2\pi)$ . For every  $\theta \in \mathbb{S}^1$ ,  $F_\theta: \mathcal{H}_1 \subset L^2(K) \rightarrow L^2(\mathbb{R})$  is given by either the Radon transform at a fixed angle  $F_\theta = \mathcal{R}_\theta$  or the fan beam transform from a fixed angle  $F_\theta = \mathcal{D}_\theta$ . The forward map  $F: \mathcal{H}_1 \subset L^2(K) \rightarrow L^2([0, 2\pi) \times \mathbb{R})$  is represented by either the Radon transform  $F = \mathcal{R}$  or the fan beam transform  $F = \mathcal{D}$ . Notice that the codomain of the Radon transform  $L^2([0, 2\pi) \times \mathbb{R})$  fits into our framework via the canonical identification  $L^2([0, 2\pi) \times \mathbb{R}) \cong L^2([0, 2\pi), L^2(\mathbb{R}))$ , and the same applies to the fan beam transform.

**Lemma 5.10.** *The following inequalities hold for some constants  $c_1, C_1 > 0$  depending on  $K$ :*

$$c_1 \|u\|_{H^{-1/2}} \leq \|\mathcal{R}u\|_{L^2} \leq C_1 \|u\|_{H^{-1/2}}, \quad u \in L^2(K).$$

*Proof.* The proof is analogous to [83, Theorem 5.1], where the inequalities are proved for  $u \in L^2(\mathcal{B}_1)$ .  $\square$

*Quasi-diagonalization.* Theorem 5.10 implies that the Radon transform satisfies condition (7) with  $b = 1/2$ . On the other hand, if the wavelets are sufficiently regular as detailed in Remark 3.9, the Littlewood-Paley property (8) holds with  $b = 1/2$ . Then, Proposition 3.8 yields the following quasi-diagonalization property:

$$(34) \quad c_1 \sum_{(j,n) \in \Gamma} 2^{-j} |x_{j,n}|^2 \leq \|\mathcal{R}\Phi^*x\|_{L^2}^2 \leq C_1 \sum_{(j,n) \in \Gamma} 2^{-j} |x_{j,n}|^2, \quad x \in \ell^2(\Gamma).$$

We now prove a quasi-diagonalization property for the fan beam transform using the following identity (see, for instance, [83, Section III.3] or [92, Section 16.3]):

$$\mathcal{D}u(\theta, \alpha) = \mathcal{R}u(\theta + \alpha - \pi/2, \rho \sin \alpha).$$

We have that

$$\begin{aligned} \|\mathcal{D}u\|_{L^2}^2 &= \int_{\mathbb{S}^1} \int_{-\pi/2}^{\pi/2} |\mathcal{R}u(\theta + \alpha - \pi/2, \rho \sin \alpha)|^2 d\alpha d\theta \\ &= \int_{\mathbb{S}^1} \int_{-\pi/2}^{\pi/2} |\mathcal{R}u(\theta, \rho \sin \alpha)|^2 d\alpha d\theta \end{aligned}$$

$$= \int_{\mathbb{S}^1} \int_{-r}^r |\mathcal{R}(\theta, s)|^2 \frac{1}{\sqrt{\rho^2 - s^2}} ds d\theta,$$

which implies that

$$(35) \quad \rho^{-1/2} \|\mathcal{R}u\|_{L^2} \leq \|\mathcal{D}u\|_{L^2} \leq (\rho^2 - d^2)^{-1/4} \|\mathcal{R}u\|_{L^2}.$$

This, together with Lemma 5.10 and (34), implies that  $\mathcal{D}$  satisfies the quasi-diagonalization property with  $b = 1/2$ , with constants depending on the chosen wavelet basis and on  $d$  and  $\rho$ .

*Coherence bounds.* We notice that, by construction, for every line  $\ell \subset \mathbb{R}^2$ ,  $\text{supp } \phi_{j,n} \cap \ell$  is contained in a segment of length  $C2^{-j}$ , where  $C$  depends only on the chosen wavelet basis. We then have that

$$\begin{aligned} \|\mathcal{R}_\theta \phi_{j,n}\|_{L^2(\mathbb{R})}^2 &= \int_{\mathbb{R}} \left| \int_{\theta^\perp} \phi_{j,n}(y + se_\theta) dy \right|^2 ds \\ &\leq C2^{-j} \int_{\mathbb{R}} \int_{\theta^\perp} |\phi_{j,n}(y + se_\theta)|^2 dy ds \\ &= C2^{-j} \|\phi_{j,n}\|_{L^2(\mathbb{R}^2)}^2 = C2^{-j}. \end{aligned}$$

In the same way, we have that

$$\begin{aligned} \|\mathcal{D}_\theta \phi_{j,n}\|_{L^2}^2 &= \int_{-\pi/2}^{\pi/2} \left| \int_{\mathbb{R}} \phi_{j,n}(\rho e_\theta + te_{\theta+\alpha}) dt \right|^2 d\alpha \\ &\leq C2^{-j} \int_{-\pi/2}^{\pi/2} \int_{\mathbb{R}} |\phi_{j,n}(\rho e_\theta + te_{\theta+\alpha})|^2 dt d\alpha \\ &= C2^{-j} \int_{-\pi/2}^{\pi/2} \int_{\mathbb{R}} |\phi_{j,n}(\rho e_\theta + te_{\theta+\alpha})|^2 t \frac{1}{t} dt d\alpha \\ &\leq \frac{C2^{-j}}{\rho - d} \int_{-\pi/2}^{\pi/2} \int_{\mathbb{R}} |\phi_{j,n}(\rho e_\theta + te_{\theta+\alpha})|^2 t dt d\alpha = \frac{C2^{-j}}{\rho - d}. \end{aligned}$$

Assumption 3.10 is then satisfied in both cases by  $d_{j,n} = 2^{j/2}$  and a constant  $B$ , which, in the case of the fan beam transform, depends on  $d$ ,  $\rho$  and the chosen wavelet basis:

$$(36) \quad \|F_\theta \phi_{j,n}\|_{L^2} \leq \frac{B}{2^{j/2}}, \quad \theta \in [0, 2\pi), (j, n) \in \Gamma.$$

In both cases, the quasi-diagonalization property, together with Proposition 5.9, implies the following estimates:

$$\|(GD)^{-1}\|^2 \leq C \max_{(j,n) \in \Lambda_{j_0}} (2^j 2^{-j}) = C,$$

and

$$\|G^{-1}\|^2 \leq C2^{j_0}.$$

Moreover, by (15), we have that  $\log M \asymp 2j_0$ .

*Proof of Theorem 4.4.* It suffices to apply Corollary 3.14 to the Radon setting, using the quasi-diagonalization properties (34) and (35) and the coherence bounds (36).  $\square$

*Proof of Theorem 4.5.* It suffices to apply Corollary 3.15 to the Radon setting, using the quasi-diagonalization properties (34) and (35) and the coherence bounds (36).  $\square$

*Proof of Theorem 4.6.* It suffices to apply Theorem 3.17 to the Radon setting, using the quasi-diagonalization properties (34) and (35) and the coherence bounds (36).  $\square$

*Proof of (20).* We mainly follow the arguments used in [80, Section 9.3.1]. From the construction of the basis of compactly supported wavelets, we have that  $\|\phi_{j,n}\|_{L^2} = 1$  and  $|\text{supp}(\phi_{j,n})| = C2^{-2j}$ . Therefore, we get that

$$|\langle u^\dagger, \phi_{j,n} \rangle| \leq \|u^\dagger\|_{L^\infty} \|\phi_{j,n}\|_{L^1} \leq C2^{-j}.$$

On the other hand, in view of Remark 3.9 with  $b = -2$ , we have that  $|\langle u^\dagger, \phi_{j,n} \rangle| \lesssim 2^{-2j}$  if  $\text{supp}(\phi_{j,n})$  does not intersect the discontinuities of  $u^\dagger$ .

Let  $\Sigma_j^1$  be the set of indices of wavelets at scale  $j$  whose support intersects the discontinuities of  $u^\dagger$ , and  $\Sigma_j^2$  be the set of indices of wavelets at scale  $j$  whose support does not intersect the discontinuities of  $u^\dagger$ . It is possible to see, arguing as in the part on *NonLinear Approximation of Piecewise Regular Images* of [80, Section 9.3.1], that  $|\Sigma_j^1| \asymp 2^j$  and  $|\Sigma_j^2| \asymp 2^{2j}$ . Therefore we have that

$$\begin{aligned} \|P_{\Lambda_{j_0}}^\perp \Phi u^\dagger\|_2^2 &= \sum_{j>j_0} \sum_n |\langle u^\dagger, \phi_{j,n} \rangle|^2 \\ &= \sum_{j>j_0} \left( \sum_{(j,n) \in \Sigma_j^1} |\langle u^\dagger, \phi_{j,n} \rangle|^2 + \sum_{(j,n) \in \Sigma_j^2} |\langle u^\dagger, \phi_{j,n} \rangle|^2 \right) \\ &\lesssim \sum_{j>j_0} (2^j 2^{-2j} + 2^{2j} 2^{-4j}) \lesssim 2^{-j_0}. \end{aligned}$$

This implies that  $\|P_{\Lambda_j}^\perp \Phi u^\dagger\|_2 \leq C2^{-aj}$  is satisfied with  $a = 1/2$ .

From a similar argument (see formula (9.63) in [80]), it follows that  $|(\Phi u^\dagger)^{(i)}| \leq i^{-1}$ , where  $|(\Phi u^\dagger)^{(\cdot)}|$  is the non-increasing rearrangement of  $|\Phi u^\dagger|$ . It is not possible to infer a  $p$ -compressibility estimate since  $\Phi u^\dagger$  might not even be summable and therefore we cannot resort to Theorem 4.6. Instead, we will exploit the fact that the sparsity error

in the recovery estimates of Theorem 4.4 depends on the truncated signal  $P_{\Lambda_{j_0}} \Phi u^\dagger$ . More precisely, for a certain fixed sparsity  $s \geq 1$ , we have that

$$\sigma_s(P_{\Lambda_{j_0}} \Phi u^\dagger)_1 \leq \sum_{i=s}^M |(\Phi u^\dagger)^{(i)}| \leq \sum_{i=s}^M \frac{1}{i} \lesssim \log(M/s) \leq \log(M),$$

where  $M = |\Lambda_{j_0}|$ . Recalling that  $M \asymp 2^{2j_0}$  and choosing  $j_0$  as a function of  $\beta$  as in Theorem 4.6, we get that  $\log(M) \lesssim \log(1/\beta)$ . Therefore we conclude that

$$\sigma_s(P_{\Lambda_{j_0}} \Phi u^\dagger)_1 \lesssim \log(1/\beta) s^{1/2-p}$$

with  $p = 1/2$ . The rest of the estimates are deduced in the exact same way as in the proof of Theorem 4.6, which imply (20).  $\square$

## REFERENCES

- [1] B. Adcock, V. Antun, and A. C. Hansen. Uniform recovery in infinite-dimensional compressed sensing and applications to structured binary sampling. *Appl. Comput. Harmon. Anal.*, 55:1–40, 2021.
- [2] B. Adcock, S. Brugiapaglia, N. Dexter, and S. Moraga. On efficient algorithms for computing near-best polynomial approximations to high-dimensional, Hilbert-valued functions from limited samples. *arXiv preprint arXiv:2203.13908*, 2022.
- [3] B. Adcock, S. Brugiapaglia, and C. G. Webster. *Sparse Polynomial Approximation of High-Dimensional Functions*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2022.
- [4] B. Adcock, A. Hansen, and C. Poon. Beyond consistent reconstructions: Optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem. *SIAM J. Math. Anal.*, 45, 01 2013.
- [5] B. Adcock, A. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *Forum Math. Sigma*, 5, 02 2017.
- [6] B. Adcock, A. Hansen, B. Roman, and G. Teschke. Generalized sampling: Stable reconstructions, inverse problems and compressed sensing over the continuum. volume 182 of *Advances in Imaging and Electron Physics*, pages 187–279. Elsevier, 2014.
- [7] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *J. Fourier Anal. Appl.*, 18(4):685–716, 2012.
- [8] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Found. Comput. Math.*, 16(5):1263–1323, 2016.
- [9] B. Adcock and A. C. Hansen. *Compressive imaging: structure, sampling, learning*. Cambridge University Press, Cambridge, 2021. With contributions by Vegard Antun.
- [10] B. Adcock, A. C. Hansen, and C. Poon. On optimal wavelet reconstructions from Fourier samples: Linearity and universality of the stable sampling rate. *Appl. Comput. Harmon. Anal.*, 36(3):387–415, 2014.
- [11] G. S. Alberti, A. Arroyo, and M. Santacesaria. Inverse problems on low-dimensional manifolds. *Nonlinearity*, 36(1):734–808, 2023.
- [12] G. S. Alberti, P. Campodonico, and M. Santacesaria. Compressed sensing photoacoustic tomography reduces to compressed sensing for undersampled Fourier measurements. *SIAM J. Imaging Sci.*, 14(3):1039–1077, 2021.

- [13] G. S. Alberti and M. Santacesaria. Calderón’s inverse problem with a finite number of measurements. *Forum Math. Sigma*, 7:e35, 2019.
- [14] G. S. Alberti and M. Santacesaria. Infinite dimensional compressed sensing from anisotropic measurements and applications to inverse problems in PDE. *Appl. Comput. Harmon. Anal.*, 50:105–146, 2021.
- [15] G. S. Alberti and M. Santacesaria. Infinite-dimensional inverse problems with finite measurements. *Arch. Ration. Mech. Anal.*, 243(1):1–31, 2022.
- [16] A. Bastounis and A. C. Hansen. On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels. *SIAM J. Imaging Sci.*, 10(1):335–371, 2017.
- [17] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numer.*, 27:1–111, 2018.
- [18] A. Berk, S. Brugiapaglia, B. Joshi, Y. Plan, M. Scott, and Ö. Yilmaz. A coherence parameter characterizing generative compressed sensing with fourier measurements. *IEEE Journal on Selected Areas in Information Theory*, 3(3):502–512, 2022.
- [19] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, 37(4):1705 – 1732, 2009.
- [20] J. Bigot, C. Boyer, and P. Weiss. An analysis of block sampling strategies in compressed sensing. *IEEE Transactions on Information Theory*, 62(4):2125–2139, 2016.
- [21] G. Blanchard and N. Mücke. Optimal rates for regularization of statistical inverse learning problems. *Found. Comput. Math.*, 18(4):971–1013, 2018.
- [22] T. Blumensath. Compressed sensing with nonlinear observations and related nonlinear optimization problems. *IEEE Trans. Inform. Theory*, 59(6):3466–3474, 2013.
- [23] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 537–546, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.
- [24] L. Borcea. Electrical impedance tomography. *Inverse Problems*, 18(6):R99–R136, 2002.
- [25] K. Bredies and D. A. Lorenz. Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.*, 14(5-6):813–837, 2008.
- [26] S. Brugiapaglia, S. Dirksen, H. C. Jung, and H. Rauhut. Sparse recovery in bounded Riesz systems with applications to numerical methods for PDEs. *Appl. Comput. Harmon. Anal.*, 53:231–269, 2021.
- [27] T. A. Bubba, M. Burger, T. Helin, and L. Ratti. Convex regularization in statistical inverse learning problems. *Inverse Probl. Imaging*, 17(6):1193–1225, 2023.
- [28] T. A. Bubba, T. Heikkilä, H. Help, S. Huotari, Y. Salmon, and S. Siltanen. Sparse dynamic tomography: a shearlet-based approach for iodine perfusion in plant stems. *Inverse Problems*, 36(9):094002, sep 2020.
- [29] T. A. Bubba and L. Ratti. Shearlet-based regularization in statistical inverse learning with an application to x-ray tomography. *Inverse Problems*, 38(5):Paper No. 054001, 43, 2022.
- [30] T. A. Bubba, S. Siltanen, M. Lassas, M. März, and Z. Purisha. Shearlet-based regularization in sparse dynamic tomography. page 31, 08 2017.
- [31] F. Cakoni, D. Colton, and H. Haddar. *Inverse Scattering Theory and Transmission Eigenvalues*, volume 88 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016.

- [32] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.
- [33] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9):589–592, 2008.
- [34] E. J. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with  $c^2$  singularities. *Communications on Pure and Applied Mathematics*, 57:219 – 266, 02 2004.
- [35] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *The Annals of Statistics*, 30(3):784–842, 2002.
- [36] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31(1):59–73, 2011.
- [37] E. J. Candès and Y. Plan. A probabilistic and ripples theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.
- [38] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [39] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [40] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1(1):54–81, 1993.
- [41] F. Colonna, G. Easley, K. Guo, and D. Labate. Radon transform inversion using the shearlet representation. *Appl. Comput. Harmon. Anal.*, 29:232–250, 09 2010.
- [42] D. Colton and R. Kress. *Inverse acoustic and electromagnetic scattering theory*, volume 93 of *Applied Mathematical Sciences*. Springer, New York, third edition, 2013.
- [43] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [44] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.
- [45] N. Dexter, H. Tran, and C. Webster. A mixed  $\ell_1$  regularization approach for sparse simultaneous approximation of parameterized PDEs. *ESAIM Math. Model. Numer. Anal.*, 53(6):2025–2045, 2019.
- [46] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126, 1995.
- [47] D. L. Donoho. Sparse components of images and optimal atomic decompositions. *Constr. Approx.*, 17(3):353–382, 2001.
- [48] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [49] C. Dossal, V. Duval, and C. Poon. Sampling the Fourier transform along radial lines. *SIAM J. Numer. Anal.*, 55(6):2540–2564, 2017.
- [50] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. In *2006 14th European Signal Processing Conference*, pages 1–5, 2006.
- [51] Y. C. Eldar and G. Kutyniok, editors. *Compressed sensing*. Cambridge University Press, Cambridge, 2012. Theory and applications.
- [52] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 205–256. Springer, New York, 2015.
- [53] S. Foucart. Stability and robustness of  $\ell^1$ -minimizations with Weibull matrices and redundant dictionaries. *Linear Algebra and its Applications*, 441:4–21, 2014.

- [54] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*, volume Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.
- [55] H. Garde and K. Knudsen. Sparsity prior for electrical impedance tomography with partial data. *Inverse Probl. Sci. Eng.*, 24(3):524–541, 2016.
- [56] A. C. Gilbert, H. W. Levinson, and J. C. Schotland. Nonlinear Iterative Hard Thresholding for Inverse Scattering. *SIAM Journal on Imaging Sciences*, 13(1):108–140, Jan. 2020.
- [57] M. Giordano and R. Nickl. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems*, 36(8):085001, 35, 2020.
- [58] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with  $l^q$  penalty term. *Inverse Problems*, 24(5):055020, 13, 2008.
- [59] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of  $\ell^1$ -regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.
- [60] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, 39(1):298–318, 2007.
- [61] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparse tomography. *SIAM J. Sci. Comput.*, 35(3):B644–B665, 2013.
- [62] P. C. Hansen, J. Jørgensen, and W. R. B. Lionheart. *Computed Tomography: Algorithms, Insight, and Just Enough Theory*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2021.
- [63] B. Harrach. Uniqueness and Lipschitz stability in electrical impedance tomography with finitely many electrodes. *Inverse Problems*, 35(2):024005, 2019.
- [64] E. Herrholz and G. Teschke. Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. *Inverse Problems*, 26(12):125012, 24, 2010.
- [65] V. Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, Cham, third edition, 2017.
- [66] B. Jin and P. Maass. Sparsity regularization for parameter identification problems. *Inverse Problems*, 28(12):123001, 70, 2012.
- [67] A. D. Jones, B. Adcock, and A. C. Hansen. Analyzing the structure of multidimensional compressed sensing problems through coherence. *CoRR*, abs/1610.07497, 2016.
- [68] J. Jørgensen, S. Coban, W. Lionheart, S. McDonald, and P. Withers. SparseBeads data: Benchmarking sparsity-regularized computed tomography. *Meas. Sci. Technol.*, 28, 09 2017.
- [69] J. S. Jørgensen, S. B. Coban, W. R. B. Lionheart, S. A. McDonald, and P. J. Withers. SparseBeads data: benchmarking sparsity-regularized computed tomography. *Meas. Sci. Technol.*, 28(12):124005, nov 2017.
- [70] J. S. Jørgensen and E. Y. Sidky. How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography. *Phil. Trans. R. Soc. A.*, 373(2043):20140387, 2015.
- [71] J. S. Jørgensen, E. Y. Sidky, P. C. Hansen, and X. Pan. Quantitative study of undersampled recoverability for sparse images in computed tomography. 2012.
- [72] J. Kaipio and E. Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.*, 198(2):493–504, 2007.
- [73] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*, volume 120 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2011.
- [74] F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 25–29, 2015.
- [75] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

- [76] P. E. Z. Larson, S. Hu, M. Lustig, A. B. Kerr, S. J. Nelson, J. Kurhanewicz, J. M. Pauly, and V. D. B. Fast dynamic 3D MR spectroscopic imaging with compressed sensing and multiband excitation pulses for hyperpolarized  $^{13}\text{C}$  studies. *Magn. Reson. Med.*, 65(3):610–9, 2011.
- [77] C. Li and B. Adcock. Compressed sensing with local structure: uniform recovery guarantees for the sparsity in levels class. *Appl. Comput. Harmon. Anal.*, 46(3):453–477, 2019.
- [78] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, 2007.
- [79] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.
- [80] S. Mallat. *A Wavelet Tour of Signal Processing. The Sparse Way*. Elsevier/Academic Press, Amsterdam, 3rd edition, 2009.
- [81] Y. Meyer. *Wavelets and Operators*. Cambridge University Press, Cambridge, 1992.
- [82] J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*, volume 10 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012.
- [83] F. Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, 2001.
- [84] F. Natterer. X-ray tomography. In *Inverse problems and imaging*, volume 1943 of *Lecture Notes in Math.*, pages 17–34. Springer, Berlin, 2008.
- [85] H. Ohlsson, A. Yang, R. Dong, and S. Sastry. CPRL – An Extension of Compressive Sensing to the Phase Retrieval Problem. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1367–1375. Curran Associates, Inc., 2012.
- [86] A. C. Polak, M. F. Duarte, and D. L. Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *IEEE Transactions on Signal Processing*, 63(11):2877–2887, 2015.
- [87] C. Poon. Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames. *Appl. Comput. Harmon. Anal.*, 42(3):402–451, 2017.
- [88] Z. Purisha, S. Karhula, J. Ketola, J. Rimpeläinen, M. Nieminen, S. Saarakkala, H. Kröger, and S. Siltanen. An automatic regularization method: An application for 3-D X-ray micro-CT reconstruction using sparse data. *IEEE Trans. Med. Imaging*, 38:1–1, 08 2018.
- [89] A. Rastogi. Tikhonov regularization with oversmoothing penalty for nonlinear statistical inverse problems. *Commun. Pure Appl. Anal.*, 19(8):4111–4126, 2020.
- [90] A. Rastogi, G. Blanchard, and P. Mathé. Convergence analysis of Tikhonov regularization for non-linear statistical inverse problems. *Electron. J. Stat.*, 14(2):2798 – 2841, 2020.
- [91] H. Rauhut and W. Rachel. Interpolation via weighted  $\ell^1$  minimization. *Appl. Comput. Harmon. Anal.*, 40(2):321–351, 2016.
- [92] O. Scherzer, editor. *Handbook of Mathematical Methods in Imaging. Vol. 1, 2, 3*. Springer, New York, second edition, 2015.
- [93] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [94] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical X-ray tomography with few radiographs: I. General theory. *Phys. Med. Biol.*, 48(10):1437, 2003.
- [95] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.
- [96] Y. Tsaig and D. L. Donoho. Extensions of compressed sensing. *Signal Processing*, 86(3):549–571, 2006. Sparse Approximations in Signal and Image Processing.

- [97] G. Uhlmann. Electrical impedance tomography and Calderón's problem. *Inverse Problems*, 25(12):123011, 39, 2009.
- [98] V. A. Zorich. *Mathematical Analysis II*. Universitext, Springer, Berlin, Heidelberg, 2016.

## APPENDIX A. PROOF OF THEOREM 5.7

We first recall a variant of the Bernstein-Talagrand inequality that will be needed later.

**Lemma A.1 (Bernstein inequality [54, Theorem 8.42]).** *Let  $f_x: \mathbb{C}^M \rightarrow \mathbb{R}$ , be a countable family of functions indexed by  $x \in \mathcal{B}_*$ . Let  $Y_1, \dots, Y_m$  be independent random variables such that the following conditions are satisfied for every  $k = 1, \dots, m$  and for  $x \in \mathcal{B}_*$ :*

$$\mathbb{E}f_x(Y_k) = 0; \quad |f_x(Y_k)| \leq K \text{ a.s.}; \quad \mathbb{E}|f_x(Y_k)|^2 \leq \Sigma^2.$$

Consider the random variable

$$Z := \sup_{x \in \mathcal{B}_*} \sum_{k=1}^m f_x(Y_k).$$

Then, for every  $\varepsilon > 0$ ,

$$\mathbb{P}(Z \geq \mathbb{E}Z + \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2(m\Sigma^2 + 2K\mathbb{E}Z) + 2K\varepsilon/3}\right).$$

**A.1. The case  $f_\nu = 1$ .** Let us first consider the case where  $f_\nu = 1$ , so that  $\mu = \nu$  is a probability measure. Assumption 3.1 thus becomes

$$(37) \quad \|F_t \phi_i\|_{\mathcal{H}_2} \leq B\omega_i, \quad t \in \mathcal{D}, \quad i \in \Gamma.$$

Furthermore, we have  $Q = I$ .

Consider the following seminorm on  $\mathbb{C}^{M \times M}$ :

$$\|B\|_{G, \omega, \lambda} := \sup_{x \in \mathcal{B}} |\langle Bx, x \rangle|,$$

where we set

$$\mathcal{B} := \{x \in \mathbb{C}^M : \|Gx\|_2 \leq 1, \|x\|_{0, \omega} \leq \lambda\}.$$

A standard argument allows us to characterize the smallest  $\delta > 0$  for which  $A$  satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$  as

$$(38) \quad \delta_* := \|A^*A - G^*G\|_{G, \omega, \lambda}.$$

Moreover, let

$$X_t := F_t \Phi^* \iota_\Lambda, \quad t \in \mathcal{D}.$$

Elementary computations show that

$$(39) \quad A^*A = \frac{1}{m} \sum_{k=1}^m X_{t_k}^* X_{t_k},$$

$$(40) \quad G^*G = \mathbb{E}X_t^* X_t.$$

The goal is now to provide a bound for  $\mathbb{E}\delta_* = \mathbb{E}\|A^*A - G^*G\|_{G,\omega,\lambda}$ . To this aim, let us introduce a Rademacher vector  $(\varepsilon_1, \dots, \varepsilon_m)$ , independent of the samples  $(t_1, \dots, t_m)$ , and consider the following random variable, for  $u \in \mathcal{B}$ :

$$Z_u := \sum_{k=1}^m \varepsilon_k \|X_{t_k} u\|_{\mathcal{H}_2}^2.$$

It is straightforward to check that  $\{Z_u : u \in \mathcal{B}\}$  is a centered, symmetric, sub-Gaussian stochastic process [54, Section 8.6] with respect to the following pseudo-metric:

$$d(u, v) := \left( \sum_{k=1}^m (\|X_{t_k} u\|_{\mathcal{H}_2}^2 - \|X_{t_k} v\|_{\mathcal{H}_2}^2)^2 \right)^{1/2}.$$

We are thus in the position to use Dudley's inequality.

**Proposition A.2.** *The following estimate holds:*

$$\mathbb{E}\delta_* \leq \frac{16\sqrt{2}}{m} \mathbb{E} \int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d, y)} dy.$$

*Proof.* In light of (38), (39) and (40), we have

$$\mathbb{E}\delta_* = \mathbb{E} \left\| \frac{1}{m} \sum_{k=1}^m (X_{t_k}^* X_{t_k} - \mathbb{E}X_{t_k}^* X_{t_k}) \right\|_{G,\omega,\lambda}.$$

Using a standard symmetrization trick (see, for instance, [54, Lemma 8.4]), we get

$$\mathbb{E}\delta_* \leq \frac{2}{m} \mathbb{E} \left\| \sum_{k=1}^m \varepsilon_k X_{t_k}^* X_{t_k} \right\|_{G,\omega,\lambda} = \frac{2}{m} \mathbb{E} \sup_{u \in \mathcal{B}} \left| \sum_{k=1}^m \varepsilon_k \|X_{t_k} u\|_{\mathcal{H}_2}^2 \right|.$$

The desired conclusion directly follows from Dudley's inequality.  $\square$

In order to get a manageable estimate for the covering numbers, we first slightly modify the underlying metric. Before, we need to prove a technical estimate involving the constant

$$\tau := B^2 \|G^{-1}\|^2 \lambda.$$

**Lemma A.3.** *Assume that (37) holds. Then*

$$\|X_t x\|_{\mathcal{H}_2} \leq \sqrt{\tau}, \quad t \in \mathcal{D}, x \in \mathcal{B}.$$

*Proof.* Let  $t \in \mathcal{D}$ ,  $x \in \mathcal{B}$ . Let  $S := \text{supp}(x)$ . Using (37), we obtain

$$\begin{aligned} \|F_t \Phi^* x\|_{\mathcal{H}_2} &\leq \sum_{i \in S} |x_i| \|F_t \Phi^* e_i\|_{\mathcal{H}_2} \leq \sum_{i \in S} |x_i| B \omega_i \\ &\leq \|x\|_2 B \sqrt{\omega(S)} \leq \|G^{-1}\| \|Gx\|_2 B \sqrt{\lambda} \leq \sqrt{\tau}. \end{aligned} \quad \square$$

For  $1 < p < \infty$ , we define the following seminorm:

$$\|u\|_{X,p} := \left( \sum_{k=1}^m \|X_{t_k} u\|_{\mathcal{H}_2}^{2p} \right)^{1/2p}, \quad u \in \mathbb{C}^M.$$

We denote the corresponding pseudo-distance by  $d_{X,p}$ .

**Lemma A.4.** *Let  $1 < r < \infty$  and  $r'$  such that  $\frac{1}{r} + \frac{1}{r'} = 1$ . Then we have*

$$d(u, v) \leq 2\tau^{(r-1)/2r} m^{1/2r} \|A^* A\|_{G,\omega,\lambda}^{1/2r} d_{X,r'}(u, v), \quad u, v \in \mathcal{B}.$$

*Proof.* By Hölder's inequality, we have that

$$\begin{aligned} d(u, v) &= \left( \sum_{k=1}^m (\|X_{t_k} u\|_{\mathcal{H}_2}^2 - \|X_{t_k} v\|_{\mathcal{H}_2}^2)^2 \right)^{1/2} \\ &= \left( \sum_{k=1}^m (\|X_{t_k} u\|_{\mathcal{H}_2} + \|X_{t_k} v\|_{\mathcal{H}_2})^2 (\|X_{t_k} u\|_{\mathcal{H}_2} - \|X_{t_k} v\|_{\mathcal{H}_2})^2 \right)^{1/2} \\ &\leq \left( \sum_{k=1}^m (\|X_{t_k} u\|_{\mathcal{H}_2} + \|X_{t_k} v\|_{\mathcal{H}_2})^{2r} \right)^{1/2r} \left( \sum_{k=1}^m (\|X_{t_k} u\|_{\mathcal{H}_2} - \|X_{t_k} v\|_{\mathcal{H}_2})^{2r'} \right)^{1/2r'} \\ &\leq 2 \sup_{w \in \mathcal{B}} \left( \sum_{k=1}^m \|X_{t_k} w\|_{\mathcal{H}_2}^{2r} \right)^{1/2r} \left( \sum_{k=1}^m (\|X_{t_k}(u-v)\|_{\mathcal{H}_2})^{2r'} \right)^{1/2r'} \\ &= 2 \sup_{w \in \mathcal{B}} \left( \sum_{k=1}^m \|X_{t_k} w\|_{\mathcal{H}_2}^{2r} \right)^{1/2r} d_{X,r'}(u, v). \end{aligned}$$

By Lemma A.3 and (39), we get

$$\sup_{w \in \mathcal{B}} \left( \sum_{k=1}^m \|X_{t_k} w\|_{\mathcal{H}_2}^{2r} \right)^{1/2r} = \sup_{w \in \mathcal{B}} \left( \sum_{k=1}^m \|X_{t_k} w\|_{\mathcal{H}_2}^2 \|X_{t_k} w\|_{\mathcal{H}_2}^{2(r-1)} \right)^{1/2r}$$

$$\begin{aligned}
&\leq \tau^{(r-1)/2r} \sup_{w \in \mathcal{B}} \left( \sum_{k=1}^m \|X_{t_k} w\|_{\mathcal{H}_2}^2 \right)^{1/2r} \\
&= \tau^{(r-1)/2r} m^{1/2r} \|A^* A\|_{G, \omega, \lambda}^{1/2r}. \quad \square
\end{aligned}$$

The relationship between  $d(u, v)$  and  $d_{X,r}(u, v)$  just proved reflects into an inequality for the corresponding covering numbers.

**Proposition A.5.** *For any  $r > 1$ , we have*

$$\int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d, y)} \, dy \leq C(r) \int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d_{X,r'}, y)} \, dy,$$

where  $C(r) := 2\tau^{(r-1)/2r} m^{1/2r} \|A^* A\|_{G, \omega, \lambda}^{1/2r}$ .

*Proof.* Lemma A.4 implies that

$$d(u, v) \leq C(r) d_{X,r'}(u, v), \quad u, v \in \mathcal{B}.$$

Standard properties of covering numbers (see, for instance, [54]) imply that

$$\mathcal{N}(\mathcal{B}, d, y) \leq \mathcal{N}(\mathcal{B}, C(r) d_{X,r'}, y) = \mathcal{N}(\mathcal{B}, d_{X,r'}, y/C(r)),$$

and therefore

$$\int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d, y)} \, dy \leq \int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d_{X,r'}, y/C(r))} \, dy.$$

The substitution  $y' = y/C(r)$  finally proves the claim.  $\square$

As a consequence of the modification of the metric, we accordingly change the underlying space. Consider the set

$$T_\omega^{\lambda, M} := \{x \in \mathbb{C}^M : \|x\|_2 \leq 1, \|x\|_{0, \omega} \leq \lambda\}.$$

It is easy to prove that

$$\mathcal{B} \subseteq \|G^{-1}\| T_\omega^{\lambda, M}.$$

**Proposition A.6.** *The following estimate holds:*

$$\int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{B}, d_{X,r'}, y)} \, dy \leq \|G^{-1}\| \int_0^\infty \sqrt{\log \mathcal{N}(T_\omega^{\lambda, M}, d_{X,r'}, y)} \, dy.$$

*Proof.* It suffices to apply the identity  $\mathcal{N}(\mathcal{B}, Cd, y) = \mathcal{N}(\mathcal{B}, d, y/C)$  to  $\mathcal{B} \subseteq \|G^{-1}\| T_\omega^{\lambda, M}$  with respect to  $d_{X,r'}$ , then consider the substitution  $y' = y/\|G^{-1}\|$ .  $\square$

The efforts to modify the underlying structure of the covering numbers so far is repaid by the reduction to an easier setting, which in turn leads us to a convenient bound. In fact, the proof of the following result is essentially identical to that given in the proof of [91, Theorem 5.2], up to minimal adjustments, hence is omitted.

**Proposition A.7.** *The following estimate holds:*

$$\int_0^\infty \sqrt{\log \mathcal{N}(T_\omega^{\lambda, M}, d_{X, r'}, y)} dy \leq C_1 \sqrt{r' m^{1/r'} \lambda \log^2 \lambda \log M},$$

where  $C_1 > 0$  is a universal constant.

At this stage, we are ready to prove the desired bound for  $\mathbb{E}\delta_*$ .

**Proposition A.8.** *The following estimate holds:*

$$\mathbb{E}\delta_* \leq C_2 \sqrt{\frac{\tau \log^3 \tau \log M}{m}} \sqrt{\mathbb{E}\delta_* + 1},$$

where  $C_2 > 0$  is a universal constant.

*Proof.* Combining Proposition A.2 with all the estimates proved so far, we ultimately get

$$\mathbb{E}\delta_* \leq C'_2 m^{-1+1/2r+1/2r'} \tau^{(r-1)/2r} \|G^{-1}\| (r' \lambda \log^2 \lambda \log M)^{1/2} \mathbb{E} \|A^* A\|_{G, \omega, \lambda}^{1/2r},$$

where  $C'_2 > 0$  is a universal constant. Note that  $m^{-1+1/2r+1/2r'} = m^{-1/2}$ . Given the arbitrariness in the choice of  $r$ , let us conveniently set

$$r = 1 + (\log \tau)^{-1} \quad \Rightarrow \quad r' = 1 + \log \tau.$$

As a result, we obtain

$$\tau^{(r-1)/2r} = \tau^{1/2r'} = \tau^{1/2(1+\log \tau)} \leq e^{1/2},$$

and, using that  $\log \tau \geq \log 3 \geq 1$ ,

$$\sqrt{r'} = \sqrt{1 + \log \tau} \leq \sqrt{2} \sqrt{\log \tau}.$$

Finally, note that

$$\|A^* A\|_{G, \omega, \lambda} \leq \|A^* A - G^* G\|_{G, \omega, \lambda} + \|G^* G\|_{G, \omega, \lambda}.$$

In particular, for  $x \in \mathcal{B}$  we have

$$|\langle G^* G x, x \rangle| = \|G x\|_2^2 \leq 1,$$

hence  $\|G^* G\|_{G, \omega, \lambda} \leq 1$ . Therefore,

$$\begin{aligned} \|A^* A\|_{G, \omega, \lambda}^{1/2r} &\leq (\|A^* A - G^* G\|_{G, \omega, \lambda} + 1)^{1/2r} \\ &\leq (\|A^* A - G^* G\|_{G, \omega, \lambda} + 1)^{1/2}. \end{aligned}$$

By Jensen's inequality and (38), we conclude that

$$\mathbb{E} \|A^* A\|_{G, \omega, \lambda}^{1/2r} \leq \mathbb{E} (\|A^* A - G^* G\|_{G, \omega, \lambda} + 1)^{1/2} \leq \sqrt{\mathbb{E}\delta_* + 1}. \quad \square$$

Given the bound just proved, it is easy to derive a lower bound for the number of samples  $m$  needed in order to make  $\mathbb{E}\delta_*$  sufficiently small.

**Proposition A.9.** *There exists a universal constant  $C > 0$  such that, for any  $\delta \in (0, 1)$ ,*

$$m \geq C\delta^{-2}\tau \log^3 \tau \log M \quad \Rightarrow \quad \mathbb{E}\delta_* \leq \delta/2.$$

*Proof.* It is clear that for any  $C \geq 16C_2^2$ , the assumptions imply that

$$C_2 \sqrt{\frac{\tau \log^3 \tau \log M}{m}} \leq \frac{\delta}{4}.$$

By Proposition A.8, this implies that

$$(41) \quad \mathbb{E}\delta_* \leq \frac{\delta}{4} \sqrt{\mathbb{E}\delta_* + 1}.$$

A straightforward computation shows that, if  $a \in (0, 1)$  and  $x > 0$ , then

$$x \leq a\sqrt{x+1} \quad \Rightarrow \quad x \leq 2a.$$

Therefore, from (41) we infer

$$\mathbb{E}\delta_* \leq \delta/2,$$

that is the claim. □

The claim of Theorem 5.7 will be finally proved once a bound on the deviation of  $\delta_*$  is established, in the spirit of the concentration inequality reported in Lemma A.1. To this aim, notice first that the map

$$\mathbb{C}^M \ni x \mapsto |\langle (A^*A - G^*G)x, x \rangle| \in \mathbb{R}$$

is continuous. Therefore, using  $\mathcal{B}'$  to denote a countable dense subset of  $\mathcal{B}$ , we get

$$\delta_* = \sup_{x \in \mathcal{B}'} |\langle (A^*A - G^*G)x, x \rangle|.$$

Moreover, setting  $\mathcal{P} := \{\pm 1\}$  allows us to write

$$\delta_* = \sup_{(x,p) \in \mathcal{B}' \times \mathcal{P}} p \langle (A^*A - G^*G)x, x \rangle,$$

and (39) further yields

$$(42) \quad \delta_* = \frac{1}{m} \sup_{(x,p) \in \mathcal{B}' \times \mathcal{P}} \sum_{k=1}^m f_{x,p}(X_{t_k}),$$

where we set  $f_{x,p}(B) := p \langle (B^*B - G^*G)x, x \rangle$ .

We are now in the position to obtain a tail bound for  $\delta_*$ . Indeed, the assumptions of the Bernstein-type inequality in Lemma A.1 are satisfied, as shown below.

**Lemma A.10.** *For every  $k = 1, \dots, m$  and for  $(x, p) \in \mathcal{B}' \times \mathcal{P}$ ,*

$$\mathbb{E}f_{x,p}(X_{t_k}) = 0, \quad |f_{x,p}(X_{t_k})| \leq \tau, \quad \mathbb{E}|f_{x,p}(X_{t_k})|^2 \leq \tau.$$

*Proof.* The zero mean property of  $f_{x,p}(X_{t_k})$  follows from (40).

Concerning the bound for  $|\mathbb{E}f_{x,p}(X_{t_k})|$ , Lemma A.3 implies that

$$\|X_{t_k}x\|_{\mathcal{H}_2}^2 \leq \tau, \quad x \in \mathcal{B}.$$

On the other hand, we have that

$$\|Gx\|_2^2 \leq 1 \leq \tau, \quad x \in \mathcal{B}.$$

Therefore, for  $(x, p) \in \mathcal{B}' \times \mathcal{P}$ , we get that

$$\begin{aligned} |f_{x,p}(X_{t_k})| &\leq \max(\|X_{t_k}x\|_{\mathcal{H}_2}^2 - \|Gx\|_2^2, \|Gx\|_2^2 - \|X_{t_k}x\|_{\mathcal{H}_2}^2) \\ &\leq \max(\|X_{t_k}x\|_{\mathcal{H}_2}^2, \|Gx\|_2^2) \leq \tau. \end{aligned}$$

Finally, we use the inequalities for  $x \in \mathcal{B}$  already proved, namely

$$\|X_{t_k}x\|_{\mathcal{H}_2}^2 \leq \tau; \quad \mathbb{E}\|X_{t_k}x\|_{\mathcal{H}_2}^2 = \|Gx\|_2^2 \leq 1.$$

For  $(x, p) \in \mathcal{B}' \times \mathcal{P}$ , we thus get

$$\begin{aligned} \mathbb{E}|f_{x,p}(X_{t_k})|^2 &= \mathbb{E}\|\|X_{t_k}x\|_{\mathcal{H}_2}^2 - \|Gx\|_2^2\|^2 \\ &= \mathbb{E}\|X_{t_k}x\|_{\mathcal{H}_2}^4 - (\mathbb{E}\|X_{t_k}x\|_{\mathcal{H}_2}^2)^2 \\ &\leq \tau \mathbb{E}\|X_{t_k}x\|_{\mathcal{H}_2}^2 - \|Gx\|_2^4 \leq \tau. \end{aligned} \quad \square$$

Resorting to Lemma A.1 gives the desired concentration bound.

**Proposition A.11.** *There exists a universal constant  $C > 0$  such that, for any  $\delta, \gamma \in (0, 1)$ , if*

$$m \geq C\delta^{-2}\tau \log(1/\gamma) \quad \text{and} \quad \mathbb{E}\delta_* \leq \delta/2,$$

then  $\mathbb{P}(\delta_* \geq \delta) \leq \gamma$ .

*Proof.* Since  $\mathbb{E}\delta_* \leq \delta/2$  by assumption, we have

$$\mathbb{P}(\delta_* \geq \delta) \leq \mathbb{P}(\delta_* \geq \mathbb{E}\delta_* + \delta/2) = \mathbb{P}(m\delta_* \geq \mathbb{E}(m\delta_*) + \varepsilon),$$

where we set  $\varepsilon = m\delta/2$ . Recalling (42), combining Lemma A.1 (for  $m\delta_*$ ) with the estimates from Lemma A.10, we get

$$\begin{aligned} \mathbb{P}(m\delta_* \geq \mathbb{E}(m\delta_*) + \varepsilon) &\leq \exp\left(-\frac{\delta^2 m^2/4}{2(m\tau + m\tau\delta) + m\tau\delta/3}\right) \\ &= \exp\left(-\frac{1}{4} \frac{1}{2(1+\delta) + \delta/3} \frac{\delta^2 m}{\tau}\right) \\ &\leq \exp\left(-\frac{1}{C} \frac{\delta^2 m}{\tau}\right), \end{aligned}$$

where  $C \geq 18$ . The claim thus follows from the assumptions on  $m$ .  $\square$

**A.2. The general case.** We are ready to conclude the proof of Theorem 5.7 in full generality. So far, we have proved that if  $f_\nu \equiv 1$ , namely if  $\mu = \nu$  is a probability measure, then by Proposition A.9 and Proposition A.11 we get  $\delta_* \leq \delta$  with probability exceeding  $1 - \gamma$ . In particular,  $A$  satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$  with overwhelming probability.

Consider now the case where  $f_\nu \not\equiv 1$ . We introduce the normalized measurement operators  $\tilde{F}_t$  defined by

$$\tilde{F}_t := f_\nu(t)^{-1/2} F_t.$$

Assumption 3.1 now reads as follows:

$$\|F_t \phi_i\|_{\mathcal{H}_2} \leq B \sqrt{f_\nu(t)} \omega_i, \quad t \in \mathcal{D}, \quad i \in \Gamma,$$

the latter being equivalent to

$$\|\tilde{F}_t \phi_i\|_{\mathcal{H}_2} \leq B \omega_i, \quad t \in \mathcal{D}, \quad i \in \Gamma.$$

Moreover, note that  $\tilde{F}_t$  satisfies the following identity with respect to  $d\nu = f_\nu d\mu$  for every  $u \in \mathcal{H}_1$ :

$$\int_{\mathcal{D}} \|\tilde{F}_t u\|_{\mathcal{H}_2}^2 d\nu(t) = \int_{\mathcal{D}} f_\nu(t)^{-1} \|F_t u\|_{\mathcal{H}_2}^2 f_\nu(t) d\mu(t) = \|F u\|_{L_\mu^2(\mathcal{D}; \mathcal{H}_2)}^2.$$

The proof thus reduces again to the case where  $f_\nu \equiv 1$  for the measurement operators  $\tilde{F}_t$  and the companion forward map  $F$ . In particular, with probability exceeding  $1 - \gamma$ , we obtain that  $\tilde{A}$  satisfies the g-RIP with respect to  $(G, \omega, \delta, \lambda)$ , where

$$\tilde{A} := \left( \frac{1}{\sqrt{m}} \tilde{F}_{t_k} \Phi^* \iota_\Lambda \right)_{k=1}^m = \left( \frac{1}{\sqrt{m}} f_\nu(t_k)^{-1/2} F_{t_k} \Phi^* \iota_\Lambda \right)_{k=1}^m = QA.$$

MALGA CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GENOA, VIA DODECANESO 35, 16146 GENOVA, ITALY.

*Email address:* giovanni.alberti@unige.it

MALGA CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GENOA, VIA DODECANESO 35, 16146 GENOVA, ITALY.

*Email address:* alessandro.felisi@edu.unige.it

MALGA CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GENOA, VIA DODECANESO 35, 16146 GENOVA, ITALY.

*Email address:* matteo.santacesaria@unige.it

DEPARTMENT OF MATHEMATICAL SCIENCES ‘‘G. L. LAGRANGE’’, POLITECNICO DI TORINO, CORSO DUCA DEGLI ABRUZZI, 24, 10129 TORINO, ITALY

*Email address:* salvatore.trapasso@polito.it