MOX-Report No. 56/2017

Infinite dimensional compressed sensing from anisotropic measurements

Alberti, G. S.; Santacesaria, M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Infinite dimensional compressed sensing from anisotropic measurements

Giovanni S. Alberti[*], Matteo Santacesaria[♯]

October 31, 2017

[♯] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
matteo.santacesaria@polimi.it

[*] Dipartimento di Matematica
Università di Genova
Via Dodecaneso 35, 16146 Genova, Italy.
alberti@dima.unige.it

Keywords: Compressed sensing, inverse problems, sparse recovery, coherence, wavelets, nonuniform Fourier sampling, electrical impedance tomography, photoacoustic tomography, thermoacoustic tomography, observability of the wave equation.

AMS Subject Classification: 94A20, 94A08, 42C40, 35R30.

Abstract

In this paper, we consider a compressed sensing problem in which both the measurement and the sparsifying systems are assumed to be frames (not necessarily tight) of the underlying Hilbert space of signals, which may be finite or infinite dimensional. The main result gives explicit bounds on the number of measurements in order to achieve stable recovery, which depends on the mutual coherence of the two systems. As a simple corollary, we prove the efficiency of non-uniform sampling strategies in cases when the two systems are not incoherent, but only asymptotically incoherent, as with the recovery of wavelet coefficients from Fourier samples. This general framework finds applications to several inverse problems in partial differential equations, in which the standard assumptions of compressed sensing are not satisfied: several examples are discussed.

1 Introduction

The problem of the recovery of a sparse signal from a small number of samples is the fundamental question of compressed sensing (CS). A signal is said to be sparse if it can be expressed as a linear combination of a small number of known vectors. The seminal papers [23, 28] have triggered an impressive amount of research in the last decade, from real world applications (MRI, X-ray tomography, etc.) to theoretical generalizations in broader mathematical frameworks.

In the finite dimensional case, the general CS problem can be stated as follows. Given an unknown sparse vector $x \in \mathbb{C}^N$ and a measurement operator represented by a matrix U, we want to reconstruct x from samples of the form $(Ux)_l$, for $l \in \Omega \subset \{1, \dots, N\}$. This is done by solving the following convex optimization problem:

$$\min_{\tilde{x} \in \mathbb{C}^N} \|\tilde{x}\|_{\ell^1} \quad \text{subject to } P_\Omega U \tilde{x} = P_\Omega U x, \tag{1}$$

where P_Ω is the projection matrix on the entries indexed by Ω. It is natural to ask under what conditions the solution of the minimization problem (1) coincides with x. These can be formulated as a lower bound on the number of measurements $m = |\Omega|$, which depends on the sparsity of the signal $s = |\mathrm{supp}(x)|$, the dimension N of the ambient space, and the matrix U. An interesting feature is that the lower bound on m does not guarantee exact recovery for all set of indices $\Omega \subset \{1, \dots, N\}$ with $|\Omega| = m$, but only for *most* of them.

One of the first contributions [23], considered the case where U is the discrete Fourier transform: exact recovery is guaranteed with high probability provided that $\Omega \subseteq \{1, \dots, N\}$ is selected uniformly at random with $m \gtrsim s \log N$. If U is a general orthonormal transformation, the problem has been addressed for the first time in [19], introducing the *coherence* $\mu = \max_{i,j} |U_{ij}|$. In this case, the bound becomes $m \gtrsim s\mu^2 N \log N$.

Similar results have been recently obtained in the infinite dimensional setting, where one considers signals belonging to a separable Hilbert space $\mathcal{H}$ and the measurement operator is represented as a bounded linear map $U \colon \mathcal{H} \to \ell^2(\mathbb{N})$. (Note that U may be expressed by scalar products with a family of vectors $\{\psi_l\}_l \subseteq \mathcal{H}$, namely $(Uf)_l = \langle f, \psi_l \rangle_{\mathcal{H}}$.) The sparsity of a signal $f \in \mathcal{H}$ is characterized by the sparsity of Df, where $D \colon \mathcal{H} \to \ell^2(\mathbb{N})$, $f \mapsto (\langle f, \varphi_j \rangle_{\mathcal{H}})_j$, is the analysis operator associated to a family of vectors $\{\varphi_j\}_j \subseteq \mathcal{H}$. The first results in this framework were presented in [4], in the case where both U and D are unitary operators, i.e. correspond to orthonormal bases; this is the standard assumption taken in virtually all works on CS with deterministic measurements. These results were further extended in [5], introducing the more advanced concepts of asymptotic incoherence, local coherence, and local sparsity. An additional improvement was given in [60], which deals with the case where $\{\varphi_j\}_j$ is a Parseval frame (see also [55, 44, 35]).

In a large number of inverse problems, where one does not have complete freedom in the measurement process, the assumption on U being unitary is not verified, thereby preventing the application of CS to many domains. As a result, two large research areas as inverse problems in partial differential equations and sparse signal recovery have so far been almost completely separated. The purpose of this paper is to provide a solid foundation that is expected to allow a fruitful interaction between these two domains.

In order to do so, in this work we present a very general CS result that deals with any bounded and injective linear operators U and D, defined on any separable complex Hilbert space (finite or infinite dimensional). Equivalently, the families $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are simply required to be frames of $\mathcal{H}$ (not necessarily tight). Since we do not need the measurement operator U to be unitary, our results cover the case of *anisotropic* measurements. These have been already studied in the finite dimensional case using random and not deterministic measurements in [47]. As far as we know, our result is new also in the finite dimensional case.

Another generalization is related to the sampling strategy. Recently, it has been observed in several works [44, 5] that, when precise estimates for the mutual coherence are available, uniform sampling strategies do not give sharp estimates for the minimum number of measurements. Our techniques are also able to cover this case, also known as *structured* sampling, just as a simple corollary of the main result for the uniform sampling. To our knowledge, this is the first sharp result under asymptotic incoherence assumptions, where there is no need to use multi-level sampling strategies and local coherence. This represents only a first step, and we believe that many other interesting estimates may be derived as corollaries of our main general result.

As mentioned above, our main motivation in dealing with the infinite dimensional anisotropic framework comes from inverse problems arising from partial differential equations. These inverse problems are intrinsically infinite dimensional, and often the measurement operator cannot be chosen as a unitary transformation. Moreover, in order to obtain a solution to these problems, an infinite number of measurements is often needed, even when the vector to be recovered belongs to a known finite dimensional subspace. CS can thus provide a rigorous, explicit and numerically viable way to find solutions to such problems when only a finite number of measurements is available. In Section 4 we explore applications of our main result to the problems of (linearized) electrical impedance tomography, nonuniform Fourier sampling and photoacoustic imaging. Many other inverse problems can be tackled with a similar approach and will be the subject of future work.

The plan of the paper is the following. In Section 2 we introduce the mathematical framework of infinite dimensional CS using the language of frames. We define the mutual coherence for general frames as well as the balancing property in this case. The main result is presented in Section 3, which contains also the main corollary about structured sampling and asymptotic incoherence. Section 4 is devoted to the applications of the main results to three inverse problems, while Section 5 contains the main technical propositions needed for the proof of the main result, also included in the same section.

2 Main assumptions

Let $\mathbb{N}$ denote the set of all positive natural numbers. Let $\mathcal{H}$ be a separable complex Hilbert space, representing our signal space, which may be either finite or infinite dimensional. The problem we study in this paper is the recovery of an unknown signal $g_0 \in \mathcal{H}$ from partial measurements of the form $(\langle g_0, \psi_l \rangle_{\mathcal{H}})_l$, under a sparsity assumption on g_0 with respect to a suitable family of vectors $\{\varphi_j\}_j$. The main assumption of this paper is the following: these families of vectors are required to be frames of $\mathcal{H}$ [24, 26, 27].

Hypothesis 1. Let L and J be two index sets[1]. Let $\{\psi_l\}_{l \in L}$ and $\{\varphi_j\}_{j \in J}$ be two frames of $\mathcal{H}$ with frame constants $A_U, B_U > 0$ and $A_D, B_D > 0$, respectively, namely

$$A_U \|g\|_{\mathcal{H}}^2 \le \sum_{l \in L} |\langle g, \psi_l \rangle_{\mathcal{H}}|^2 \le B_U \|g\|_{\mathcal{H}}^2, \qquad A_D \|g\|_{\mathcal{H}}^2 \le \sum_{j \in J} |\langle g, \varphi_j \rangle_{\mathcal{H}}|^2 \le B_D \|g\|_{\mathcal{H}}^2,$$

for every $g \in \mathcal{H}$.

The measurements and the sparsity condition are expressed by the analysis operators $U \colon \mathcal{H} \to \ell^2(L)$ and $D \colon \mathcal{H} \to \ell^2(J)$, defined by

$$(Ug)_l = \langle g, \psi_l \rangle_{\mathcal{H}}, \qquad (Dg)_j = \langle g, \varphi_j \rangle_{\mathcal{H}}.$$

By construction, the dual operators are given by $U^* e_l = \psi_l$ and $D^* e_j = \varphi_j$, where $\{e_i\}_{i \in I}$ is the canonical basis of $\ell^2(I)$. By Hypothesis 1, since $\sum_l |\langle g, \psi_l \rangle_{\mathcal{H}}|^2 = \|Ug\|_2^2$ and $\sum_j |\langle g, \varphi_j \rangle_{\mathcal{H}}|^2 = \|Dg\|_2^2$, we have that U and D are bounded and the operator norms satisfy

$$\|U\| = \|U^*\| \le \sqrt{B_U}, \qquad \|D\| = \|D^*\| \le \sqrt{B_D}. \tag{2}$$

The recovery problem can then be stated as follows: given noisy partial measurements of Ug_0, namely $\zeta = P_\Omega U g_0 + \eta$ for some (finite) set $\Omega \subseteq L$, recover the signal $g_0 \in \mathcal{H}$, under the assumption that Dg_0 is sparse. Here we have used the notation P_Ω for the orthogonal projection onto $\overline{\operatorname{span}\{e_j : j \in \Omega\}}$ (if $\Omega = \{1, \dots, M\}$ we simply write P_M). The classical way to solve this problem is via ℓ^1 minimization, namely

$$\inf_{\substack{g \in \mathcal{H} \\ Dg \in \ell^1(J)}} \|Dg\|_1 \quad \text{subject to } \|P_\Omega U g - \zeta\|_2 \le \varepsilon, \tag{3}$$

where $\varepsilon = \|\eta\|_2$ is the noise level.

[1] We say that $I \subseteq \mathbb{N}$ is an index set if $I = \mathbb{N}$ or $I = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$.

Remark. Equivalently, one may adapt a more abstract point of view, starting from a bounded operator $U\colon \mathcal{H} \to \ell^2(L)$ with bounded inverse. It is immediate to verify that $\psi_l = U^* e_l$ gives rise to a frame, as in Hypothesis 1. The formulation with U allows to consider any linear inverse problem of the form

$$U\colon \mathcal{H} \to \ell^2(L), \qquad Ug = \zeta.$$

The only requirement is that, with full data, the inverse problem should be uniquely and stably solvable. In particular, any invertible operator U may be considered, and not necessarily isometries as in the standard compressed sensing setting. Many linear infinite dimensional inverse problems may be written in this form; see §4.2 for an application to a linearized version of electrical impedance tomography and §4.3 for an application to an inverse source problem for the wave equation.

Remark. The formulation given in (3) of the ℓ^1 optimization problem is the *analysis* approach, because of the minimization of $\|Dg\|_{\ell^1}$, where D is the analysis operator. This is in contrast with the much more popular *synthesis* formulation

$$\inf_{x\in\ell^1(J)} \|x\|_{\ell^1} \quad \text{subject to } \|P_\Omega U D^* x - \zeta\| \le \varepsilon. \tag{4}$$

In general, the two approaches are not equivalent [29]. We have decided to work with the analysis approach since there may be multiple minimizers of (4) if D gives a redundant representation, which complicates the derivation of the estimates.

Remark. When J is infinite, the above minimization problem cannot be implemented numerically. When D and U are unitary operators, it was shown in [4] that this issue may be solved by looking at a corresponding finite-dimensional optimization problem. We expect that the same is true also in our context, and leave this investigation to future work.

Given the generality of our setting, we need to consider the dual frames of $\{\psi_l\}_l$ and $\{\varphi_j\}_j$. By classical frame theory (see [27, Lemma 5.1.5]), the *frame operators* U^*U and D^*D are invertible, and we can consider the *dual frames*

$$\tilde{\psi}_l = (U^*U)^{-1}\psi_l \qquad \text{and} \qquad \tilde{\varphi}_j = (D^*D)^{-1}\varphi_j,$$

which have frame constants B_U^{-1}, A_U^{-1} and B_D^{-1}, A_D^{-1}, respectively. Equivalently, we may write $\tilde{\psi}_l = U^{-1}e_l$ and $\tilde{\varphi}_j = D^{-1}e_j$, where U^{-1} and D^{-1} are the Moore–Penrose pseudoinverses of U and D, respectively, defined as follows:

$$U^{-1} := (U^*U)^{-1}U^* \quad \text{and} \quad D^{-1} := (D^*D)^{-1}D^*.$$

Note that they are left inverses of U and D, respectively. Therefore, $(U^{-1})^*$ and $(D^{-1})^*$ are the analysis operators of the dual frames, and so arguing as in (2) we obtain

$$\|U^{-1}\| = \|U^{-*}\| \le A_U^{-1/2}, \qquad \|D^{-1}\| = \|D^{-*}\| \le A_D^{-1/2}. \tag{5}$$

With an abuse of notation, we have denoted $(U^{-1})^*$ and $(D^{-1})^*$ by U^{-*} and D^{-*}, respectively. It can be immediately checked that they are right inverses of U^* and D^*, i.e. $(U^*)^{-1} = U^{-*}$ and $(D^*)^{-1} = D^{-*}$. For later use, set $\kappa_1 := \max(B_U, A_U^{-1})$ and $\kappa_2 := \max(A_D^{-1}, 1)$, so that by (2) and (5) we obtain

$$\|U\| = \|U^*\| \leq \sqrt{\kappa_1}, \quad \|U^{-1}\| = \|U^{-*}\| \leq \sqrt{\kappa_1}, \quad \|D^{-1}\| = \|D^{-*}\| \leq \sqrt{\kappa_2}. \tag{6}$$

Remark. The frames $\{\tilde{\psi}_l\}_l$ and $\{\tilde{\varphi}_j\}_j$ are the *canonical* dual frames, but in general many other choices are possible. These are in correspondence with all possible bounded left inverses of U and D, and it is possible to give a characterization of all dual frames. The reader is referred to [27, Section 6.3] for the details.

We need to measure the incoherence between the sensing system $\{\psi_l\}_l$ and the representation system $\{\varphi_j\}_j$ or, equivalently, between the measurement operator U and the representation operator D.

Definition 1. The *mutual coherence* of U and D is given by

$$\begin{aligned}\mu &= \sup_{j\in J, l\in L} \max\{|\langle \varphi_j, \psi_l\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \psi_l\rangle_{\mathcal{H}}|, |\langle \varphi_j, \tilde{\psi}_l\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \tilde{\psi}_l\rangle_{\mathcal{H}}|\} \\ &= \sup_{j\in J, l\in L} \max\{|\langle D^* e_j, U^* e_l\rangle_{\mathcal{H}}|, |\langle D^{-1} e_j, U^* e_l\rangle_{\mathcal{H}}|, |\langle D^* e_j, U^{-1} e_l\rangle_{\mathcal{H}}|, |\langle D^{-1} e_j, U^{-1} e_l\rangle_{\mathcal{H}}|\}.\end{aligned}$$

Let us now discuss a particular case.

Example 1. The above construction simplifies considerably if $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are *Parseval* frames, namely if $A_U = B_U = A_D = B_D = 1$, as studied in [60]. In this case the associated analysis operators U and D are isometries, their left inverses simplify to $U^{-1} = U^*$, $D^{-1} = D^*$ and all the operator norms in (6) are simply bounded by 1. The dual frames and the corresponding frames coincide, and the coherence reduces to

$$\mu = \sup_{j\in J, l\in L} |\langle \varphi_j, \psi_l\rangle_{\mathcal{H}}| = \sup_{j\in J, l\in L} |\langle D^* e_j, U^* e_l\rangle_{\mathcal{H}}|, \tag{7}$$

which simply involves scalar products between the elements of the two bases.

As an even more particular case, one may consider orthonormal bases $\{\psi_l\}_l$ and $\{\varphi_j\}_j$, which represents the usual assumption in the classical compressed sensing framework, and in its extension to infinite dimension [4].

A relevant application of this general setting is with nonuniform discrete Fourier sampling (see §4.1), which gives rise to a frame $\{\psi_l\}_l$ in the space of square-integrable compactly supported functions; in this case, the operator U is injective but may not be onto. On the other hand, allowing the system $\{\varphi_j\}_j$ to be a frame, i.e. D is not necessarily invertible, is useful whenever we wish to use a redundant representation to sparsify the signals in $\mathcal{H}$ (e.g. redundant wavelets [32], curvelets [21], ridgelets [20] and shearlets [50, 49, 43]).

The partial measurements $(Ug)_l = \langle g, \psi_l \rangle_{\mathcal{H}}$ are indexed by $l \in \Omega$, where Ω is chosen uniformly at random in $\{1, \ldots, N\}$. Given the infinite dimensionality of the problem, the upper bound N has to be chosen big enough, depending on the sparsity assumptions. This is quantified by the *balancing property*, introduced in [4] and generalized here to the non isometric setting.

Definition 2. (Balancing property) Let $s, M \in \mathbb{N}$ be such that $s \leq M$. We say that $N \in L$ satisfies the balancing property with respect to M and s if for all $\Delta \subseteq \{1, \ldots, M\}$ with $|\Delta| = s$ we have

$$\|P_{\mathcal{W}} U^* P_N^{\perp} (U^*)^{-1} P_{\mathcal{W}}\|_{\mathcal{H} \to \mathcal{H}} \leq \frac{1}{4\sqrt{2\log(s\kappa_2)}}, \tag{8}$$

$$\|P_{\mathcal{W}} U^{-1} P_N^{\perp} U P_{\mathcal{W}}\|_{\mathcal{H} \to \mathcal{H}} \leq \frac{1}{4}, \tag{9}$$

and

$$\|P_{\Delta}^{\perp} D P_{\mathcal{W}}^{\perp} U^* P_N (U^*)^{-1} P_{\mathcal{W}}\|_{\mathcal{H} \to \ell^\infty(J)} \leq \frac{1}{16\sqrt{s\kappa_2}}, \tag{10}$$

$$\|P_{\mathcal{W}} U^{-1} P_N U P_{\mathcal{W}}^{\perp} D^{-1} P_{\Delta}^{\perp}\|_{\ell^2(J) \to \mathcal{H}} \leq 1/2, \tag{11}$$

where $\mathcal{W} := R(D^* P_\Delta) + R(D^{-1} P_\Delta) = \{D^* P_\Delta x + D^{-1} P_\Delta y : x, y \in \ell^2(J)\}$.

Remark 1. If $L = \{1, 2, \ldots, |L|\}$ is finite, it is enough to choose $N = |L|$, since all the norms on the left hand side vanish. If $L = \mathbb{N}$, the existence of a suitable N satisfying the above conditions simply follows by the fact that $P_N \to I$ and $P_N^{\perp} \to 0$ strongly (see [4, Proposition 5.2] for the details of the argument).

For $s, M \in J$, $s \leq M$, we use the notation

$$\sigma_{s,M}(x_0) = \inf\{\|x - x_0\|_{\ell^1(J)} : \operatorname{supp}(x) \subseteq \{1, \ldots, M\},\ |\operatorname{supp}(x)| \leq s\},$$

which measures the compressibility of the signal $x_0 \in \ell^1(J)$ by means of s-sparse signals x. Following [60], for $\Delta \subseteq \{1, \ldots, M\}$ we denote

$$\tilde{B}_\Delta = \max\left\{\|D P_{\mathcal{W}}^{\perp} D^{-1}\|_{\ell^\infty \to \ell^\infty}, \|D^{-*} P_{\mathcal{W}}^{\perp} D^*\|_{\ell^\infty \to \ell^\infty}, 1\right\},$$

$$B(s, M) = \max\{\tilde{B}_\Delta : \Delta \subseteq \{1, \ldots, M\},\ 3 \leq |\Delta| \leq s\},$$

where we have used the notation

$$\|T\|_{\ell^\infty \to \ell^\infty} := \sup_{x \in \ell^2(\mathbb{N}) \setminus \{0\}} \frac{\|Tx\|_{\ell^\infty(J)}}{\|x\|_{\ell^\infty(J)}},$$

for an operator $T \colon \ell^2(J) \to \ell^2(J)$.

Remark 2. It is worth observing that when $\{\varphi_j\}_j$ is an orthonormal basis or, equivalently, when D is a unitary operator, we simply have

$$B(s, M) = 1.$$

Indeed, in view of the identity

$$DP_{\mathcal{W}}^{\perp}D^*x = DP_{\mathcal{W}}^{\perp}D^*(P_\Delta x + P_\Delta^{\perp}x) = DP_{\mathcal{W}}^{\perp}D^*P_\Delta^{\perp}x = DD^*P_\Delta^{\perp}x = P_\Delta^{\perp}x,$$

we obtain $\|DP_{\mathcal{W}}^{\perp}D^*x\|_{\ell^\infty} = \left\|P_\Delta^{\perp}x\right\|_{\ell^\infty} \le \|x\|_{\ell^\infty}$ for every $x \in \ell^2(J)$, so that $\tilde{B}_\Delta = 1$ for every Δ.

In the other extreme case, it may happen that $B(s, M) = +\infty$, even in finite dimension with a Parseval frame, as the following example shows.

Example 2. Consider $\mathcal{H} = \mathbb{R}$ with the Parseval frame

$$\varphi_1 = \varphi_2 = \varphi_3 = 0, \qquad \varphi_{j+3} = f_j,\ j \ge 1,$$

where $f\colon \mathbb{N} \to (0, +\infty)$ is a sequence such that $\sum_j f_j^2 = 1$ and $\sum_j f_j = +\infty$ (here $J = \mathbb{N}$). For $\Delta = \{1, 2, 3\}$ we have $\mathcal{W} = \{0\}$, so that $\tilde{B}_\Delta \ge \|DP_{\mathcal{W}}^{\perp}D^*\|_{\ell^\infty\to\ell^\infty} = \|DD^*\|_{\ell^\infty\to\ell^\infty}$. Thus, setting $x_n = e_1 + \cdots + e_{n+3} \in \ell^2(\mathbb{N})$, by $D^*x_n = \sum_{j=1}^n f_j$ we have

$$\tilde{B}_\Delta \ge |(DD^*x_n)_4| = |\langle D^*x_n, \varphi_4\rangle_{\mathbb{R}}| = f_1 \sum_{j=1}^n f_j \underset{n\to\infty}{\longrightarrow} +\infty,$$

whence $B(s, M) \ge \tilde{B}_\Delta \ge +\infty$ for any s and M.

For $\alpha \in (0, 1]$ let $\tilde{M}(\alpha)$ be the smallest integer such that $\tilde{M}(\alpha) \ge M$ and

$$\sqrt{\kappa_1} \max_{\delta=0,1} \left(\|P_N U D_\delta^* e_j\|_{\ell^2} + \sqrt{\kappa_1}\|P_{\widetilde{\mathcal{W}}} D_\delta^* e_j\|_{\mathcal{H}}\right) < \alpha, \qquad j \in J, j > \tilde{M}(\alpha), \quad (12)$$

where $\widetilde{\mathcal{W}} := R(D^*P_M) + R(D^{-1}P_M)$, $D_0 := D$ and $D_1 := D^{-*}$.

Remark 3. If $J = \{1, 2, \ldots, M\}$ is finite, we simply have $\tilde{M}(\alpha) = M$ for every α. If $J = \mathbb{N}$, $\tilde{M}(\alpha)$ always exists since $D_\delta^* e_j$ tends to zero weakly and P_N and $P_{\widetilde{\mathcal{W}}}$ are compact operators.

Remark 4. In the case when D is associated to an orthonormal basis, the condition $\tilde{M}(\alpha) \ge M$ is implicit, since $\kappa_1\|P_{\widetilde{\mathcal{W}}} D_0^* e_j\|_{\ell^2} = \kappa_1\|D_0^* e_j\|_{\ell^2} = \kappa_1 \ge 1 \ge \alpha$ for $j = 1, \ldots, M$ by definition of $\widetilde{\mathcal{W}}$. Furthermore, condition (12) reduces to

$$\sqrt{\kappa_1}\|P_N U D^* e_j\|_{\ell^2} < \alpha, \qquad j > \tilde{M}(\alpha).$$

As a consequence, note that if $\sup_{l\in L} |\langle \psi_l, \varphi_j\rangle_{\mathcal{H}}| \le C/\sqrt{j}$ for every $j \in J, j > M$ (which is the case in several concrete applications, see §3.3) one has

$$\sqrt{\kappa_1}\|P_N U D^* e_j\|_{\ell^2} \le \sqrt{N\kappa_1}\|P_N U D^* e_j\|_{\ell^\infty} \le \sqrt{N\kappa_1} \sup_{l\in L} |\langle \psi_l, \varphi_j\rangle_{\mathcal{H}}| \le C\sqrt{N\kappa_1/j},$$

and so

$$\tilde{M}(\alpha) \le C^2 \frac{\kappa_1 N}{\alpha^2}.$$

In the case where U is the Fourier transform and D a Wavelet transform in dimension one, a more precise estimate has been derived in [5], namely $\tilde{M}(\alpha) = O(M/\alpha)$.

3 Main results

3.1 Finite and infinite dimensional recovery

We now state the main result of this work. Recall that $\mathcal{H}$ is any separable Hilbert space: we deal with the finite and infinite dimensional case simultaneously.

Theorem 1. *Assume that Hypothesis 1 holds true, and let U and D denote the corresponding analysis operators, satisfying the bounds given in* (6)*. Let $M, s \in J$ and $\omega \geq 1$ be such that $M \geq 5$ and $3 \leq s \leq M$. Let $N \in L$ satisfy the balancing property with respect to M and s, and let $\Omega \subseteq \{1, \dots, N\}$ be chosen uniformly at random with $|\Omega| = m$. Assume that $B(s, M) \leq 2\kappa_1\kappa_2\sqrt{s}$ and that*

$$m \geq C\mu^2 s\kappa_1\kappa_2\omega^2 B(s,M)^2 N \log\left(s\kappa_2 \tilde{M}\big(\tfrac{C'm}{N\sqrt{s\kappa_2}}\big)\right),$$

where $C, C' > 0$ are universal constants.

Let $g_0 \in \mathcal{H}$ and $\eta \in \ell^2(L)$ be such that $\|\eta\|_2 \leq \varepsilon$ for some $\varepsilon \geq 0$. Let $\zeta = P_\Omega U g_0 + \eta$ be the known noisy measurement. Let $g \in \mathcal{H}$ be a minimizer of the minimization problem (3)*. Then, with probability exceeding $1 - e^{-\omega}$, we have*

$$\|g - g_0\|_{\mathcal{H}} \leq 4(2 + \sqrt{\kappa_2})\sigma_{s,M}(Dg_0) + C''\varepsilon\frac{N\sqrt{s}}{m}\sqrt{\kappa_1}\kappa_2\omega$$

where C'' is a universal constant.

Remark. The generality of our construction allows to treat the finite dimensional and the infinite dimensional cases simultaneously. However, in finite dimension the above estimate for m has a simpler form, which is worth to point out. Suppose that $L = \{1, \dots, N\}$ and $J = \{1, \dots, M\}$. By Remarks 1 and 3, we have that N satisfies the balancing property with respect to M and s and that $\tilde{M}\big(\frac{C'm}{N\sqrt{s\kappa_2}}\big) = M$. Thus, the lower bound for the number of measurements m becomes

$$m \geq Cs\kappa_1\kappa_2\omega^2 B(s,M)^2\mu^2 N \log(s\kappa_2 M),$$

which, when $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are both orthonormal bases of $\mathcal{H} = \mathbb{C}^N$, by (6), Example 1 and Remark 2 simply reduces to

$$m \geq Cs\omega^2\mu^2 N \log(sN).$$

Theorem 1 directly generalizes Theorems 6.1, 6.3 and 6.4 of [4] to the case of anisotropic measurements. It also extends the results of [5, 60] to the case of general frames D and U. Although we did not use the concepts of local sparsity and local coherence we believe that our techniques can be extended also in that setting, see §3.2. Our result can also be seen as an infinite dimensional generalization of the finite dimentional result in [47] for anisotropic random measurements. In fact, as far as we know, also the finite dimensional version of Theorem 1 is new.

3.2 Asymptotic incoherence and artificial frames

The above result shows that with random sampling one needs a number of measurements proportional to the sparsity of the signal (up to logarithmic factors and the quantity $B(s, M)$), provided that the coherence is small enough, namely, $\mu = O(1/\sqrt{N})$. While this happens in finite dimension ($\mathcal{H} = \mathbb{C}^N$) in some particular situations, e.g. with signals that are sparse with respect to the Dirac basis $\varphi_j = e_j$ and with Fourier measurements, in many cases of practical interest the above result becomes almost meaningless since the coherence μ is of order one. For instance, this happens when U is the discrete Fourier transform and D the discrete wavelet transform. As it was shown in [5], this is always the case in infinite dimension.

Since the early stages of compressed sensing, it was realized that this issue may be solved by using variable density random sampling [65, 61, 45, 44, 14, 5, 60]. For instance, in the Fourier-Wavelet case, one needs to sample lower frequencies with higher probability than the higher frequencies. We now give a result that deals with this situation; in particular, it takes into account a priori estimates on the coherence and non-uniform sampling. As it is clear from the proof, it follows as a simple corollary of Theorem 1, thanks to the flexibility of its assumptions. More precisely, Theorem 1 is applied to an *artificial frame* $\{\hat{\psi}_l\}_l$ obtained from $\{\psi_l\}_l$ by artificially repeating its elements. More complicated transformations, also involving $\{\varphi_j\}_j$, may be considered (taking into account, for instance, *asymptotic sparsity* [5]): we leave these investigations to future work, and we limit ourselves to an example to show the potential of this framework.

Corollary 1. *Assume that Hypothesis 1 holds true, and let U and D denote the corresponding analysis operators, satisfying the bounds given in* (6). *Assume that there exist $C_1 > 0$ and $w \in \mathbb{R}_+^N$ such that*

$$\sup_{j\in J} \max\{|\langle\varphi_j,\psi_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j,\psi_l\rangle_{\mathcal{H}}|, |\langle\varphi_j,\tilde{\psi}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j,\tilde{\psi}_l\rangle_{\mathcal{H}}|\} \leq C_1 w_l \quad \forall l \leq N, \tag{13}$$

$$\sup_{L\ni l>N,\, j\in J} \max\{|\langle\varphi_j,\psi_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j,\psi_l\rangle_{\mathcal{H}}|, |\langle\varphi_j,\tilde{\psi}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j,\tilde{\psi}_l\rangle_{\mathcal{H}}|\} \leq \frac{C_1}{\sqrt{N}}, \tag{14}$$

where $\{\tilde{\psi}_l\}_l$ and $\{\tilde{\varphi}_j\}_j$ are the dual frames of $\{\psi_l\}_l$ and $\{\varphi_j\}_j$, respectively. Let $M, s \in J$ and $\omega \geq 1$ be such that $M \geq 5$ and $3 \leq s \leq M$. Let $N \in L$ satisfy the balancing property with respect to M and s. Assume that $B(s, M) \leq 2\kappa_1\kappa_2\sqrt{s}$ and that

$$m \geq CC_1^2 s\kappa_1\kappa_2\omega^2 B(s,M)^2(\|w\|^2_{\mathbb{C}^N} + 1)\log\left(s\kappa_2\tilde{M}\left(\tfrac{C'm}{N(\|w\|^2_{\mathbb{C}^N}+1)\sqrt{s\kappa_2}}\right)\right), \tag{15}$$

where $C, C' > 0$ are universal constants. Sample m indices $l_1, \ldots, l_m$ indipendently from $\{1, \ldots, N\}$ according to the probability distribution

$$\nu_l = \tilde{C}_N \lceil N w_l^2 \rceil, \qquad l = 1, \ldots, N,$$

where $\tilde{C}_N = \left(\sum_{l=1}^N \lceil N w_l^2 \rceil\right)^{-1}$*, and set* $\Omega = (l_1, \ldots, l_m)$ *(with possible repetitions to be kept).*

Take $g_0 \in \mathcal{H}$ *and* $\eta \in \mathbb{C}^m$ *such that* $\|\eta\|_w \leq \varepsilon$*, where* $\|\eta\|_w^2 := \sum_{i=1}^m \frac{|\eta_i|^2}{\lceil N w_{l_i}^2 \rceil}$. *Set* $\zeta = P_\Omega U g_0 + \eta$*, i.e.* $\zeta_i = (U g_0)_{l_i} + \eta_i$*. Let* $g \in \mathcal{H}$ *be a minimizer of*

$$\inf_{\substack{g \in \mathcal{H} \\ Dg \in \ell^1(J)}} \|Dg\|_1 \quad \text{subject to } \|P_\Omega U g - \zeta\|_w \leq \varepsilon. \tag{16}$$

Then, with probability exceeding $1 - e^{-\omega}$*, we have*

$$\|g - g_0\|_{\mathcal{H}} \leq 4(2 + \sqrt{\kappa_2})\sigma_{s,M}(Dg_0) + C'' \varepsilon \frac{N(\|w\|_{\mathbb{C}^N}^2 + 1)\sqrt{s}}{m} \sqrt{\kappa_1} \kappa_2 \omega.$$

Remark. This result can be seen as a generalization of [44, Corollary 2.9] to infinite dimension and to the frame case, since $\nu_l = \lceil N w_l^2 \rceil / \sum_{l=1}^N \lceil N w_l^2 \rceil \approx \frac{w_l^2}{\|w\|_{\mathbb{C}^N}^2}$ and $(\|w\|_{\mathbb{C}^N}^2 + 1) \approx \|w\|_{\mathbb{C}^N}^2$.

Remark. Bounds (13) and (14) may be completed by the corresponding decays with respect to the frame $\{\varphi_j\}_j$, namely in the variable j. In this case, $\{\psi_l\}_{l \in L}$ and $\{\varphi_j\}_{j \in J}$ are *asymptotically incoherent* [5]. Under this more restrictive assumption, when D is an orthonormal basis an explicit bound on the factor $\tilde{M}$ may be derived using Remark 4 (see §3.3 for an example).

Proof. Let $r_l = \lceil N w_l^2 \rceil$ for $l \leq N$. We want to apply Theorem 1 to $\{\hat{\psi}_l\}_l$ and $\{\varphi_j\}_j$, where $\{\hat{\psi}_l\}$, with associated operator $\hat{U}$, is given as follows. For $l \leq N$ normalize ψ_l by $\sqrt{r_l}$ and repeat it r_l times, and for $l > N$ leave ψ_l unchanged, namely

$$\left\{\hat{\psi}_l\right\}_l = \Bigg\{\underbrace{\frac{\psi_1}{\sqrt{r_1}}, \ldots, \frac{\psi_1}{\sqrt{r_1}}}_{r_1 \text{ times}}, \ldots, \underbrace{\frac{\psi_N}{\sqrt{r_N}}, \ldots, \frac{\psi_N}{\sqrt{r_N}}}_{r_N \text{ times}}, \psi_{N+1}, \psi_{N+2}, \ldots\Bigg\}. \tag{17}$$

Note that $\{\hat{\psi}_l\}_l$ has the same frame bounds of $\{\psi_l\}_l$, i.e. $\hat{\kappa}_1 = \kappa_1$, by construction, since

$$\sum_{i=1}^{r_l} \left|\left\langle f, \frac{\psi_l}{\sqrt{r_l}} \right\rangle\right|^2 = \sum_{i=1}^{r_l} \frac{1}{r_l} |\langle f, \psi_l \rangle|^2 = |\langle f, \psi_l \rangle|^2. \tag{18}$$

Then we want to prove that $\hat{N} = \sum_{l=1}^N r_l$ satisfies the balancing property with respect to $\hat{U}, D, M$ and s. We first notice that $\hat{U}^* \hat{U} = U^* U$, since

$$\hat{U}^* \hat{U} f = \hat{U}^* \left(\sum_{l=1}^\infty \langle \hat{U} f, e_l \rangle e_l\right) = \sum_{l=1}^\infty \langle \hat{U} f, e_l \rangle \hat{\psi}_l = \sum_{l=1}^\infty \langle f, \hat{\psi}_l \rangle \hat{\psi}_l,$$

so that

$$\hat{U}^*\hat{U}f = \sum_{l=1}^{\infty}\sum_{i=1}^{r_l}\left\langle f, \frac{\psi_l}{\sqrt{r_l}}\right\rangle \frac{\psi_l}{\sqrt{r_l}} = \sum_{l=1}^{\infty}\langle f, \psi_l\rangle\psi_l = U^*Uf.$$

In passing, we remark that this identity tells us that the dual frame $\{\tilde{\hat{\psi}}_l\}_l$ of the artificial frame $\{\hat{\psi}_l\}_l$ coincides with the artificial frame of the dual frame $\{\hat{\tilde{\psi}}_l\}_l$, which is constructed as in (17). Arguing in the same way, and terminating the above sums to $\tilde{N}$ and N, respectively, we readily derive

$$\hat{U}^*P_{\hat{N}}\hat{U}f = \sum_{l=1}^{\hat{N}}\langle f, \hat{\psi}_l\rangle\hat{\psi}_l = \sum_{l=1}^{N}\sum_{i=1}^{r_l}\left\langle f, \frac{\psi_l}{\sqrt{r_l}}\right\rangle \frac{\psi_l}{\sqrt{r_l}} = \sum_{l=1}^{N}\langle f, \psi_l\rangle\psi_l = U^*P_NUf.$$

This immediately yields properties (10) and (11), since $(\hat{U}^*)^{-1} = \hat{U}(\hat{U}^*\hat{U})^{-1}$ and $\hat{U}^{-1} = (\hat{U}^*\hat{U})^{-1}\hat{U}^*$. Similarly, (8) and (9) follow from the identities

$$\hat{U}^*P_{\hat{N}}^{\perp}\hat{U} = \hat{U}^*(I - P_{\hat{N}})\hat{U} = \hat{U}^*\hat{U} - \hat{U}^*P_{\hat{N}}\hat{U} = U^*U - U^*P_NU = U^*P_N^{\perp}U.$$

We have the following straightforward upper bound for $\hat{N}$:

$$\hat{N} = \sum_{l=1}^{N}\lceil Nw_l^2\rceil \le \sum_{l=1}^{N}(Nw_l^2 + 1) = N\sum_{l=1}^{N}(w_l^2 + 1) = N(\|w\|_{\mathbb{C}^N}^2 + 1).$$

The factor $\tilde{M}$ associated to $\hat{U}$ and D, which we denote by $\hat{\tilde{M}}$, verifies $\hat{\tilde{M}}(\alpha) = \tilde{M}(\alpha)$. Indeed, from the definition of $\tilde{M}$, we only need to check that $\|P_{\hat{N}}\hat{U}f\|_2^2 = \|P_NUf\|_2^2$, which follows by (18):

$$\|P_{\hat{N}}\hat{U}f\|_2^2 = \sum_{l=1}^{\hat{N}}|\langle\hat{U}f, e_l\rangle|^2 = \sum_{l=1}^{\hat{N}}|\langle f, \hat{\psi}_l\rangle|^2 = \sum_{l=1}^{N}|\langle f, \psi_l\rangle|^2 = \|P_NUf\|_2^2.$$

The factor $B(s, m)$ does not change since it does not depend on U but only on D, which is left unchanged.

Let us calculate the new coherence

$$\hat{\mu} = \sup_{l,j}\max\{|\langle\varphi_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle\varphi_j, \tilde{\hat{\psi}}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \tilde{\hat{\psi}}_l\rangle_{\mathcal{H}}|\}.$$

For $l > \hat{N}$ there exists $l' > N$ such that $\hat{\psi}_l = \psi_{l'}$. This implies $\tilde{\hat{\psi}}_l = \tilde{\psi}_{l'}$, since $\hat{U}^*\hat{U} = U^*U$, so that

$$\begin{aligned}&\max\{|\langle\varphi_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle\varphi_j, \tilde{\hat{\psi}}_l\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \tilde{\hat{\psi}}_l\rangle_{\mathcal{H}}|\}\\ &= \max\{|\langle\varphi_j, \psi_{l'}\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \psi_{l'}\rangle_{\mathcal{H}}|, |\langle\varphi_j, \tilde{\psi}_{l'}\rangle_{\mathcal{H}}|, |\langle\tilde{\varphi}_j, \tilde{\psi}_{l'}\rangle_{\mathcal{H}}|\} \le \frac{C_1}{\sqrt{N}},\end{aligned}$$

by (13). For $l \leq \hat{N}$ there exists $l' \leq N$ with $\hat{\psi}_l = \psi_{l'}/\sqrt{r_{l'}}$ and $\hat{\tilde{\psi}}_l = \tilde{\psi}_{l'}/\sqrt{r_{l'}}$, so that by (14) we obtain

$$\begin{aligned}
&\max\{|\langle \varphi_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \hat{\psi}_l\rangle_{\mathcal{H}}|, |\langle \varphi_j, \hat{\tilde{\psi}}_l\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \hat{\tilde{\psi}}_l\rangle_{\mathcal{H}}|\} \\
&= \max\{|\langle \varphi_j, \frac{\psi_{l'}}{\sqrt{r_{l'}}}\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \frac{\psi_{l'}}{\sqrt{r_{l'}}}\rangle_{\mathcal{H}}|, |\langle \varphi_j, \frac{\tilde{\psi}_{l'}}{\sqrt{r_{l'}}}\rangle_{\mathcal{H}}|, |\langle \tilde{\varphi}_j, \frac{\tilde{\psi}_{l'}}{\sqrt{r_{l'}}}\rangle_{\mathcal{H}}|\} \leq \frac{C_1 w_{l'}}{\sqrt{r_{l'}}} \leq \frac{C_1}{\sqrt{N}},
\end{aligned}$$

since $r_l = \lceil N w_l^2 \rceil$. Therefore $\hat{\mu} \leq \frac{C_1}{\sqrt{N}}$.

Now, the factor $\hat{\mu}^2 \hat{N}$ in the estimate for m given in Theorem 1 applied to $\hat{U}$ and D becomes:

$$\hat{\mu}^2 \hat{N} \leq \frac{C_1^2}{N} N(\|w\|_{\mathbb{C}^N}^2 + 1) = C_1^2(\|w\|_{\mathbb{C}^N}^2 + 1),$$

so that the estimate on m in Theorem 1 transforms into (15). Finally, note that selecting m elements $\hat{\Omega} = \{\hat{l}_1, \ldots, \hat{l}_m\}$ uniformly at random from $\{\hat{\psi}_{\hat{l}}\}_{\hat{l}=1}^{\hat{N}}$ corresponds to the variable density sampling scheme of the statement (with $\hat{\psi}_{\hat{l}_i} = \psi_{l_i}/\sqrt{r_{l_i}}$). Further, setting $\hat{\zeta} = P_{\hat{\Omega}} \hat{U} g_0 + \left(\eta_i/\sqrt{r_{l_i}}\right)_i$, we have

$$\left\| P_{\hat{\Omega}} \hat{U} g - \hat{\zeta} \right\|_2 = \left\| \left(\left((g, \psi_{l_i}) - \zeta_i \right) / \sqrt{r_{l_i}} \right)_i \right\|_2 = \| P_\Omega U g - \zeta \|_w.$$

This concludes the proof. □

3.3 Recovery of wavelet coefficients from Fourier samples

For $d \in \mathbb{N}$, let $\mathcal{H} = L^2([0, 2\pi]^d)$ be the signal space. Let $\{e_k\}_{k \in \mathbb{Z}^d}$ be the Fourier basis of $\mathcal{H}$, namely

$$e_k(x) = (2\pi)^{-\frac{d}{2}} e^{ik \cdot x}, \qquad x \in [0, 2\pi]^d.$$

Consider a nondecreasing ordering of $\mathbb{Z}^d$, namely a bijective map $\rho \colon \mathbb{N} \to \mathbb{Z}^d$, $l \mapsto k_l$, such that

$$l_1, l_2 \in \mathbb{N},\ l_1 \leq l_2 \quad \Longrightarrow \quad \|\rho(l_1)\| \leq \|\rho(l_2)\|,$$

where $\| \ \|$ is any norm of $\mathbb{R}^d$. Set $\psi_l = e_{k_l}$ for $l \in \mathbb{N}$. Finally, let $\{\varphi_j\}_{j \in \mathbb{N}}$ be a separable wavelet basis of $\mathcal{H}$ (ordered according to the wavelet scales). Note that both systems $\{\psi_l\}_{l \in \mathbb{N}}$ and $\{\varphi_j\}_{j \in \mathbb{N}}$ are orthonormal bases, so that $\tilde{\psi}_l = \psi_l$ and $\tilde{\varphi}_j = \varphi_j$. Under certain decay conditions on the scaling function (which may be relaxed to a condition satisfied by all Daubechies wavelets if one considers a different ordering of the frequencies $k \in \mathbb{Z}^d$), it was shown in [41] that

$$\sup_{l \geq l_0,\, j \in \mathbb{N}} |\langle \psi_l, \varphi_j \rangle_{\mathcal{H}}| \leq \frac{C_1}{\sqrt{l_0}}, \qquad \sup_{l \in \mathbb{N},\, j \geq j_0} |\langle \psi_l, \varphi_j \rangle_{\mathcal{H}}| \leq \frac{C_1}{\sqrt{j_0}}, \qquad j_0, l_0 \in \mathbb{N}$$

for some $C_1 > 0$. In other words, the wavelet basis and the Fourier basis are *asymptotically incoherent.* Thanks to the first of these inequalities, we have that assumptions (13) and (14) of the corollary are satisfied with $w_l = \frac{1}{\sqrt{l}}$, so that $\|w\|^2_{\mathbb{C}^N} + 1 \leq \log N + 2$. Further, by the second of these inequalities and Remark 4 we have $\tilde{M}(\alpha) \leq C_1^2 \frac{N}{\alpha^2}$. As a consequence, estimate (15) of Corollary 1 for the number of measurements m becomes

$$m \geq Cs\,\omega^2 \log^2(Ns) \tag{19}$$

for some constant $C > 0$ depending only on C_1. Up to log factors, the number of measurements required for the success of the recovery using ℓ^1 minimization is directly proportional to the sparsity of the signals, provided that the measurements are chosen at random from $\{1, \ldots, N\}$ with probabilities $\nu_l \propto \lceil N/l \rceil$ for $l = 1, \ldots, N$.

4 Applications

4.1 Nonuniform Fourier sampling

The most classical compressed sensing problem formulated in the continuous setting is the recovery of a function $g \in L^2([0,1]^d)$ from Fourier samples

$$(Ug)(k) := \hat{g}(k) = \int_{[0,1]^d} g(x) e^{-2\pi i k \cdot x}\, dx = \langle g, e^{2\pi i k \cdot} \rangle_{L^2([0,1]^d)}, \qquad k \in \Omega,$$

where $\Omega \subseteq \mathbb{Z}^d$ is a finite set of frequencies where the measurements are taken. Here U is the discrete Fourier transform given by scalar products with the sinusoids $x \mapsto e^{2\pi i k \cdot x}$, which form an orthonormal basis of $L^2([0,1]^d)$. If g is sparse with respect to a suitable orthonormal basis with analysis operator D, this reconstruction problem fits in the framework discussed in the previous section, and g may be recovered by ℓ^1 minimization, provided that enough random measurements Ω are taken. The standard theory of compressed sensing may be applied in this case, since both U and D are unitary operators (see Example 1).

In several applications (such as Magnetic Resonance Imaging, Computed Tomography, geophysical imaging, seismology and electron microscopy), nonuniform Fourier sampling arises naturally, i.e. the frequencies are not taken uniformly in $\mathbb{Z}^d$. In this case, the operator U fails to be an isometry, since the corresponding family of sinusoids may be only a frame, and not an orthonormal basis. The results discussed in the previous section may be directly applied to this case too.

Let us now give a quick overview of nonuniform Fourier frames; we follow [3]. For additional details, the reader is referred to [26, 2] for the one-dimensional case, and to [13, 11, 59, 3] for the multi-dimensional case.

Let $\mathcal{H}$ be the space of square-integrable functions with support contained in a compact, convex and symmetric set $E \subseteq \mathbb{R}^d$, i.e. $\mathcal{H} = L^2(E)$. For $g \in \mathcal{H}$, we consider measurements of the form

$$\hat{g}(k) = \int_E g(x) e^{-2\pi i k \cdot x}\, dx = \langle g, e_k \rangle_{\mathcal{H}}, \qquad k \in Z \subseteq \hat{\mathbb{R}}^d,$$

namely scalar products with the sinusoids

$$e_k(x) = e^{2\pi i k \cdot x}, \qquad x \in E.$$

Instead of considering the case when Z is a cartesian grid of $\hat{\mathbb{R}}^d$, which gives rise to uniform Fourier sampling, we wish to give more general conditions on the set Z so that $\{e_k\}_{k \in Z}$ is a frame of $\mathcal{H}$.

The first of these conditions requires that the samples are fine enough to capture all the frequency information in a given direction.

Definition 3 ([13])**.** We say that the sampling scheme $Z \subseteq \hat{\mathbb{R}}^d$ is *δ-dense* if

$$\delta = \sup_{\hat{y} \in \hat{\mathbb{R}}^d} \inf_{k \in Z} |k - \hat{y}|_{E^\circ},$$

where the norm $|\ |_{E^\circ}$ is given by

$$|\hat{y}|_{E^\circ} = \inf\{a > 0 : x \cdot \hat{y} \leq a \text{ for every } x \in E\}.$$

The second condition limits the concentration of samples, in order to avoid large energies in small frequency regions.

Definition 4. We say that the sampling scheme $Z \subseteq \hat{\mathbb{R}}^d$ is *separated* if there exists a constant $\eta > 0$ such that

$$\inf_{k_1, k_2 \in Z, k_1 \neq k_2} |k_1 - k_2| \geq \eta > 0.$$

We say that Z is *relatively separated* if it is a finite union of separated sets.

Under these conditions, the family of sinusoids e_k with frequencies k in Z forms a frame for $L^2(E)$.

Proposition 2 ([13, 11, 59])**.** *Let $E \subseteq \mathbb{R}^d$ be a compact, convex and symmetric set and take $\delta \in (0, 1/4)$. If $Z \subseteq \hat{\mathbb{R}}^d$ is relatively separated and δ-dense, then $\{e_k\}_{k \in Z}$ is a Fourier frame for $L^2(E)$, namely*

$$A\|g\|^2_{L^2(E)} \leq \sum_{k \in Z} |\langle g, e_k \rangle_{L^2(E)}|^2 \leq B\|g\|^2_{L^2(E)}, \qquad g \in L^2(E)$$

for some $A, B > 0$.

Now, assuming that $\{\varphi_j\}_{j \in \mathbb{N}}$ is a frame for $L^2(E)$, we can apply Theorem 1 and Corollary 1 to this setting. This would provide, to our knowledge, the first result about recovery of a sparse signal from nonuniform Fourier measurements via ℓ^1 minimization. Even if the measurement frame is generally not tight, we can provide explicit bounds for the recovery of the wavelet coefficients from nonuniform Fourier samples.

Some numerical simulations related to this framework are presented in [34, Example 5].

4.2 Electrical impedance tomography

Electrical impedance tomography (EIT) is an imaging technique in which one wants to determine the electrical conductivity $\sigma(x)$ inside a body $\mathcal{O}$ from boundary voltage and current measurements. It is a non-linear inverse boundary value problem whose mathematical formulation was presented for the first time by Calderón [18].

Let $\mathcal{O} \subset \mathbb{R}^d$, $d \geq 2$, be an open bounded domain with Lipschitz boundary and $\sigma \in L^\infty(\mathcal{O})$, $\sigma(x) \geq \sigma_0 > 0$ for almost every $x \in \mathcal{O}$, be the electrical conductivity. Given a voltage $f \in H^{1/2}(\partial\mathcal{O})$ on the boundary of the domain, the associated potential u is the unique $H^1(\mathcal{O})$ solution of the following Dirichlet problem for the conductivity equation:

$$\operatorname{div}(\sigma\nabla u) = 0 \quad \text{in } \mathcal{O}, \qquad u = f \quad \text{on } \partial\mathcal{O}, \tag{20}$$

where H^s, $s > 0$, are the classical Sobolev spaces. The boundary current associated to the voltage f is represented by the trace of the normal derivative of the potential u on $\partial\mathcal{O}$. More precisely, we define the Dirichlet-to-Neumann map $\Lambda_\sigma : H^{1/2}(\partial\mathcal{O}) \to H^{-1/2}(\partial\mathcal{O})$ as

$$\Lambda_\sigma(f) = \sigma \frac{\partial u}{\partial \nu}\bigg|_{\partial\mathcal{O}}, \tag{21}$$

where ν is the outer normal to $\partial\mathcal{O}$ and u is the unique solution of the Dirichlet problem (20).

Calderón's inverse conductivity problem asks if it is possible to determine a conductivity σ from the knowledge of its associated Dirichlet-to-Neumann map Λ_σ. Positive answers to this question have been given since 1987 [63, 58, 57].

If σ is sufficiently smooth, the problem can be reduced to the so-called Gel'fand-Calderón problem for the Schrödinger equation,

$$(-\Delta + q)\tilde{u} = 0, \qquad q = \frac{\Delta\sqrt{\sigma}}{\sqrt{\sigma}}, \tag{22}$$

via the change of variables $u = \tilde{u}/\sqrt{\sigma}$ in the conductivity equation (20). This inverse problem consists in the reconstruction of the potential q from the knowledge of the Dirichlet-to-Neumann map

$$\Lambda_q : \tilde{u}|_{\partial\mathcal{O}} \mapsto \frac{\partial \tilde{u}}{\partial \nu}\bigg|_{\partial\mathcal{O}}. \tag{23}$$

One of the biggest open questions concerning inverse boundary value problems such as Calderón's or Gel'fand-Calderón's is the determination of a conductivity/potential from a finite number of boundary measurements. A priori assumptions on the unknown are needed in this case, and to the best of our knowledge the only result concerns piecewise constant coefficients with discontinuities on a single convex polygon [33]. Several works have studied the general piecewise constant case with infinitely many measurements [8, 12]. In what follows, we will consider finitely many measurements, and present a first result in this direction for the linearized Gel'fand-Calderón problem, using the theory developed in this paper. The main feature of the result is the fairly general assumptions on the unknown potential to be recovered.

In order to linearize the problem, we assume that $q = q_0 + \delta q$ where q_0 is known and δq is small. Given two boundary voltages $f, g \in H^{1/2}(\partial\mathcal{O})$ we have Alessandrini's identity [7]:

$$\langle g, (\Lambda_q - \Lambda_{q_0}) f \rangle_{H^{\frac{1}{2}}(\partial\mathcal{O}) \times H^{-\frac{1}{2}}(\partial\mathcal{O})} = \int_{\mathcal{O}} \delta q \, u_g u_f^0 \, dx,$$

where u_g (resp. u_f^0) solves the Schrödinger equation (22) with potential q (resp. q_0) and Dirichlet data g (resp. f). The quantity on the left of this identity is known since q_0 is known and $\Lambda_q f$ is the boundary measurement corresponding to the chosen potential f (g should be seen as a test function). Since for $\delta q \approx 0$ we have $u_g \approx u_g^0$, the linearization consists in assuming that we can measure the quantity $\int_{\mathcal{O}} \delta q \, u_g^0 u_f^0 \, dx$ for given f, g. Focusing on the solutions themselves instead of on their boundary values, this inverse problem may be rephrased as follows.

Problem (Linearized Gel'fand-Calderón problem)**.** Given a finite number of scalar products of the form $\int_{\mathcal{O}} \delta q \, u_1 u_2 \, dx$, where u_1 and u_2 are solutions of

$$(-\Delta + q_0) u_i = 0 \tag{24}$$

in $\mathcal{O}$, find $\delta q \in L^2(\mathcal{O})$.

Without loss of generality, we can assume that $\mathcal{O} \subseteq \mathbb{T}^d$, where $\mathbb{T} = [0, 2\pi]$. Extend δq by zero to $\mathcal{H} := L^2(\mathbb{T}^d)$ and assume that $q_0 \in H^s(\mathbb{T}^d)$ with $s > d/2$. In the rest of this subsection, several positive constants depending only on d, s and $\|q_0\|_{H^s(\mathbb{T}^d)}$ will be denoted by the letters $c_1, c_2, \dots$. In order to choose the solutions u_i, we make the additional assumption $d \geq 3$, since we will make use of a classical uniqueness result for the Calderón problem for this case. From [63] we have that for every $k \in \mathbb{Z}^d$ and $t \geq c_1$ we can construct solutions $u_i^{k,t}$ of (24) in $\mathbb{T}^d$ of the form

$$u_i^{k,t}(x) = e^{\zeta_i^{k,t} \cdot x}(1 + r(x, \zeta_i^{k,t})), \qquad x \in \mathbb{T}^d$$

where $\zeta_i \in \mathbb{C}^d$ are such that $\zeta_1^{k,t} + \zeta_2^{k,t} = -ik$ and

$$\|r(x, \zeta_i^{k,t})\|_{H^s(\mathbb{T}^d)} \leq \frac{c_2}{t}, \quad i = 1, 2. \tag{25}$$

These solutions $u_i^{k,t}$ are known as exponentially growing solutions, Faddeev-type solutions or complex geometrical optics (CGO) solutions.

We need to consider an ordering of $\mathbb{Z}^d$, namely a bijective map $\rho\colon \mathbb{N} \to \mathbb{Z}^d$, $l \mapsto k_l$. For each $k \in \mathbb{Z}^d$ fix $t_k \geq c_1$ and define the measurement operator $U_{GC}\colon \mathcal{H} \to \ell^2(\mathbb{N})$ by

$$U_{GC}(\delta q) = (\langle \delta q, \psi_l \rangle)_l, \qquad \overline{\psi_l} = u_1^{k_l,t_{k_l}} u_2^{k_l,t_{k_l}}. \tag{26}$$

We now show that U_{GC} is an invertible operator with bounded inverse, provided that the t_k's are chosen big enough.

Lemma 3. *There exists $c_3 > 0$ such that if $t_k \geq \max(1, c_1, c_3|k|^s)$ for every $k \in \mathbb{Z}^d$ then the operator U_{GC} is bounded and invertible and*

$$\|U_{GC}\|_{\mathcal{H}\to\ell^2(\mathbb{N})} \leq \frac{3}{2}, \qquad \|U_{GC}^{-1}\|_{\ell^2(\mathbb{N})\to\mathcal{H}} \leq 2.$$

Proof. Let $F\colon \mathcal{H} \to \ell^2(\mathbb{N})$ denote the discrete Fourier transform defined by $F(\delta q) = (\langle \delta q, e_{k_l}(x) \rangle)_l$, where $e_k(x) = (2\pi)^{-\frac{d}{2}} e^{ik\cdot x}$. Since $\overline{\psi_l} = e^{-ik_l\cdot x}(1 + r(x,\zeta_1^{k_l,t_{k_l}}))(1 + r(x,\zeta_2^{k_l,t_{k_l}}))$, setting $r_i^k = r(\,\cdot\,,\zeta_i^{k,t_k})$ we have

$$\|e_{k_l} - \psi_l\|_{L^2(\mathbb{T}^d)} \leq \|r_1^{k_l}\|_{L^2(\mathbb{T}^d)} + \|r_2^{k_l}\|_{L^2(\mathbb{T}^d)} + \|r_1^{k_l}\|_{L^\infty(\mathbb{T}^d)}\|r_2^{k_l}\|_{L^2(\mathbb{T}^d)} \leq \frac{c_4}{|t_{k_l}|} \leq \frac{c_4}{c_3|k_l|^s}$$

where we used estimate (25) and the Sobolev embedding $H^s(\mathbb{T}^d) \to L^\infty(\mathbb{T}^d)$ for $s > d/2$. This implies $|((U_{GC} - F)\delta q)_l| \leq \|\delta q\|_{L^2(\mathbb{T}^d)} \frac{c_4}{|t_{k_l}|}$, so that

$$\|(U_{GC} - F)q\|_{\ell^2(\mathbb{N})}^2 \leq c_4^2 \|\delta q\|_{L^2(\mathbb{T}^d)}^2 \sum_{k\in\mathbb{Z}^d} \frac{1}{|t_k|^2} \leq \frac{c_4^2}{c_3^2}\|\delta q\|_{L^2(\mathbb{T}^d)}^2 \left(1 + \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \frac{1}{|k|^{2s}}\right)$$

which is finite since $2s > d$. Moreover

$$\|U_{GC} - F\|_{\mathcal{H}\to\ell^2(\mathbb{N})} \leq \frac{c_4}{c_3}\left(1 + \sum_{k\in\mathbb{Z}^d\setminus\{0\}} \frac{1}{|k|^{2s}}\right)^{\frac{1}{2}} \leq \frac{1}{2},$$

provided that c_3 is chosen big enough. From this estimate we immediately obtain $\|U_{GC}\| \leq \|U_{GC} - F\| + \|F\| \leq \frac{3}{2}$, since F is an isometry. Moreover, we have the Neumann series expansion

$$U_{GC}^{-1} = F^{-1}\sum_{k=0}^{+\infty}(-1)^k \left((U_{GC} - F)F^{-1}\right)^k,$$

and so $\left\|U_{GC}^{-1}\right\| \leq 2$, as desired. □

Using Lemma 3, we can apply Theorem 1 and Corollary 1 to obtain a general recipe to recover or approximate a sparse or compressible conductivity from a small number of linearized EIT measurements. The main constraint here is that the sparsifying frame $\{\varphi_j\}_{j\in\mathbb{N}}$ and the measurement frame $\{\psi_l\}_{l\in\mathbb{N}}$ must be incoherent or asymptotically incoherent in order to have a sharp bound on the number of measurements. The study of the incoherence between the frame $\{\psi_l\}_{l\in\mathbb{N}}$ and a wavelet frame $\{\varphi_j\}_{j\in\mathbb{N}}$ will be subject of future work. If that holds, one could argue like in §3.3 to obtain the same explicit bounds on the minimum number of measurements.

We have chosen the functions $u^{k,t}$ from [63] for the sake of simplicity. Other families of functions with similar decay properties might be used as well, leading to similar results as Lemma 3, with lower regularity assumptions on the coefficients to be recovered.

In two dimensions it is unclear if results such as Lemma 3 could hold: for the linearized Calderón problem we cannot use complex geometrical optics solutions to approximate the Fourier transform as in higher dimensions. For the linearized Gel'fand-Calderón problem one could use the Bukhgeim approach [16] to recover pointwise values of a potential via stationary phase type techniques.

More generally, the results of this subsection may be applied to a large class of linearized inverse boundary value problems for which we have families of complex geometrical optics solutions with good decay properties: inverse problems for the Helmoltz equation, the elasticity system and Maxwell's equations, for instance.

4.3 An inverse problem for the wave equation

Our main result can also be applied to another linear infinite dimensional inverse problem, the observability problem for the wave equation [39, 54, 51, 30, 6]. This is a classical inverse problem, and consists in the reconstruction of the initial source of the wave equation from boundary measurements of the solution. In addition to the direct link with control theory, this inverse problem appears in the formulation of thermoacoustic and photoacoustic tomography in a bounded domain [9, 48, 1, 40, 25] (for the free-space formulation, see [46]).

Let $d \geq 2$ and $\mathcal{O} \subseteq \mathbb{R}^d$ be a bounded smooth domain. We consider the following initial value problem for the wave equation[2]

$$\begin{cases} \partial_{tt}p - \Delta p = 0 & \text{in } (0,T)\times\mathcal{O}, \\ p(0,\cdot) = f & \text{in } \mathcal{O}, \\ \partial_t p(0,\cdot) = 0 & \text{in } \mathcal{O}, \\ p = 0 & \text{in } (0,T)\times\partial\mathcal{O}, \end{cases} \tag{27}$$

where $T > 0$ and $f \in H_0^1(\mathcal{O}) := \{u \in H^1(\mathcal{O}) : u = 0 \text{ on } \partial\mathcal{O}\}$ is the unknown initial condition. The above problem admits a unique weak solution $p \in C([0,T]; H_0^1(\mathcal{O}))$ (see [31, Section 7.2] and [15, Theorem 10.14]). The inverse problem of interest may be formulated as follows.

[2]For simplicity, we consider the case of constant sound speed (normalized to 1), but this analysis may be generalized to the case of a spatially varying sound speed c. Similarly, considering a non-homogeneous initial condition for $\partial_t p$ would not add any substantial complications.

Problem (Observability of the wave equation)**.** Supposing that the trace of the normal derivative $\partial_\nu p$ is measured on an open subset Γ of $\partial\mathcal{O}$ for all $t \in (0,T)$, where ν is the exterior unit normal to $\partial\Omega$, find the initial condition f in $\mathcal{O}$.

Observe that the forward problem is always well-posed: by an inequality of Rellich's, the measurement operator

$$V \colon H_0^1(\mathcal{O}) \to L^2((0,T)\times\Gamma), \qquad f \mapsto \partial_\nu p,$$

where p is the solution of (27), is well-defined and bounded [54, (1.20)].

In order to apply our techniques to the inverse problem we need more than continuity, namely injectivity and bounded invertibility of the map V. In this case, f is uniquely and stably determined by the boundary data $Vf = \partial_\nu p$ on $(0,T)\times\Gamma$. This solves the above-mentioned inverse problem when we can perfectly measure $\partial_\nu p$ on the whole $(0,T)\times\Gamma$.

There is a wide literature concerning assumptions on Γ and T that guarantee the invertibility of V (see [10] and references therein). Here we only mention a sufficient condition by Ho [39] and J. L. Lions [54] (see also [30, §5.3.4] and [6, Theorem 2.8]): if $\{x \in \partial\mathcal{O} : (x - x_0)\cdot\nu > 0\} \subseteq \Gamma$ for some $x_0 \in \mathbb{R}^d$ and $T > 2\sup_{x\in\mathcal{O}} |x - x_0|$, then V is invertible with bounded inverse. In the following, we shall assume that V is invertible with bounded inverse.

In order to let compressed sensing come into play, we will make use of the following identity, which follows by a simple integration by parts. For every $v \in L^2((0,T)\times\Gamma)$, we have

$$(Vf, \bar{v})_{L^2((0,T)\times\Gamma)} = \int_{(0,T)\times\Gamma} \partial_\nu p\, v\, dt d\sigma = \langle \partial_t U_v(0,\cdot), f\rangle_{H^{-1}(\mathcal{O}),H_0^1(\mathcal{O})}, \tag{28}$$

where $U_v \in C\left([0,T]; L^2(\mathcal{O})\right) \cap C^1\left([0,T]; H^{-1}(\mathcal{O})\right)$ is the solution of

$$\begin{cases} \partial_{tt}U_v - \Delta U_v = 0 & \text{in } (0,T)\times\mathcal{O}, \\ U_v(T,\cdot) = 0 & \text{in } \mathcal{O}, \\ \partial_t U_v(T,\cdot) = 0 & \text{in } \mathcal{O}, \\ U_v = \chi_\Gamma v & \text{in } (0,T)\times\partial\mathcal{O}, \end{cases} \tag{29}$$

which is defined in the sense of transposition [53, 54], where χ_Γ is the characteristic function of Γ and $H^{-1}(\mathcal{O})$ is the dual of $H_0^1(\mathcal{O})$. Identity (28) shows that we can use the dual solution U_v to *probe* the unknown f: we measure different moments of f by varying v.

Since observability is equivalent to exact controllability [53, 54], we have that for every $h \in H^{-1}(\mathcal{O})$ there exists $v_h \in L^2\left((0,T)\times\Gamma\right)$ such that $\partial_t U_{v_h}(0,\cdot) = h$. The control v_h can be explicitly constructed via an optimization problem. By (28) we obtain:

$$(Vf, \overline{v_h})_{L^2((0,T)\times\Gamma)} = \langle h, f\rangle_{H^{-1}(\mathcal{O}),H_0^1(\mathcal{O})}.$$

The Riesz representation theorem gives the anti-isomorphism $R\colon H^1_0(\mathcal{O}) \to H^{-1}(\mathcal{O})$, defined by $\langle R\psi, g\rangle_{H^{-1}(\mathcal{O}),H^1_0(\mathcal{O})} = (g,\psi)_{H^1_0(\mathcal{O})}$. Inserting this expression into the above identity yields

$$(Vf, \overline{v_{R\psi}})_{L^2((0,T)\times\Gamma)} = (f,\psi)_{H^1_0(\mathcal{O})}, \qquad \psi \in H^1_0(\mathcal{O}). \tag{30}$$

Let $\{\psi_l\}_l$ be a frame of $H^1_0(\mathcal{O})$ and $\{v_l\}_l$ be a family of $L^2((0,T)\times\Gamma)$ such that

$$v_l = \overline{v_{R\psi_l}} \iff \psi_l = R^{-1}\partial_t U_{\overline{v_l}}(0,\cdot). \tag{31}$$

These relations show that one may first choose the frame $\{\psi_l\}_l$ and then construct the related family $\{v_l\}_l$, or viceversa. Define the measurement operator

$$U_{obs} : H^1_0(\mathcal{O}) \to \ell^2(\mathbb{N}), \qquad f \mapsto \big((f,\psi_l)_{H^1_0(\mathcal{O})}\big)_l,$$

which can be measured, thanks to (30). Then, representing f in another frame $\{\varphi_j\}_j$ of $H^1_0(\mathcal{O})$ we can apply Theorem 1 (or Corollary 1) to this setting, provided that $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are incoherent (or asymptotically incoherent). Therefore, via ℓ^1 minimization we can reconstruct f from the partial measurements $\{(f,\psi_l)_{H^1_0(\mathcal{O})}\}_{l\in\Omega}$, for some subsampling subset $\Omega \subseteq \mathbb{N}$, provided that f is sparse with respect to $\{\varphi_j\}_j$.

Note that, in order to measure $(f,\psi_l)_{H^1_0(\mathcal{O})} = (Vf, v_l)_{L^2((0,T)\times\Gamma)}$, in principle we might need to know Vf on the whole $(0,T)\times\Gamma$. The subsampling procedure would then become useless. In order to overcome this issue, one has to choose the functions v_l in such a way that the computation of each $(Vf, v_l)_{L^2((0,T)\times\Gamma)}$ only requires a partial knowledge of Vf. For instance, one could choose compactly supported functions v_l's in order to sample subsets of $(0,T)\times\Gamma$: this would correspond to having sensors only on particular locations of the boundary at specific times. Similarly, scalar products with slowly varying v_l's would correspond to local averages of Vf, which may be obtained with integrating area and line detectors [17, 62, 38].

In summary, the challenge is to construct families $\{\varphi_j\}_j, \{\psi_l\}_l \subseteq H^1_0(\mathcal{O})$ and $\{v_l\}_l \subseteq L^2((0,T)\times\Gamma)$ such that:

- $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are frames of $H^1_0(\mathcal{O})$;
- $\{\psi_l\}_l$ and $\{v_l\}_l$ are related via (31), which involves the solution of the PDE (29);
- $\{\psi_l\}_l$ and $\{\varphi_j\}_j$ are incoherent (or asymptotically incoherent);
- and each scalar product $(Vf, v_l)_{L^2((0,T)\times\Gamma)}$ may be computed with partial measurements of Vf.

A detailed analysis of these issues goes beyond the scope of this paper, and is a very interesting direction for future work, at the interface of applied harmonic analysis and PDE theory.

5 Proof of Theorem 1

The aim of this section is to prove Theorem 1.

5.1 Concentration inequalities

Certain large deviation bounds for sums of vector and matrix valued random variables are required to prove some of the key results. Inspired by the paper of Kueng and Gross [47] we use Bernstein inequalities instead of applying Talagrand as done by Adcock and Hansen [4]. We give a particular vector-valued inequality not depending on the dimension taken from [47] which originally appears in [52] (Chapter 6.3, Eq. (6.12)) with a direct proof in [37].

Theorem 4 (Vector Bernstein inequality)**.** *Let $\{X_k\} \in \mathbb{C}^d$ be a finite sequence of independent random vectors. Suppose that $\mathbb{E}[X_k] = 0$, $\|X_k\| \le B$ a.s. and put $\sigma^2 \ge \sum_k \mathbb{E}\left[\|X_k\|^2\right]$. Then for all $0 \le t \le \frac{\sigma^2}{B}$*

$$\mathbb{P}\left(\left\|\sum_k X_k\right\| > t\right) \le \exp\left(\frac{-t^2}{8\sigma^2} + \frac{1}{4}\right). \tag{32}$$

The matrix-valued deviation estimate that we use is due to Tropp [64] (Theorem 1.6).

Theorem 5 (Matrix Bernstein inequality)**.** *Consider a finite sequence $\{X_k\} \in \mathbb{C}^{d\times d}$ of independent random matrices. Assume that each random matrix satisfies $\mathbb{E}[X_k] = 0$ and $\|X_k\| \le B$ a.s.. Define*

$$\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\left(X_k X_k^*\right)\right\|, \left\|\sum_k \mathbb{E}\left(X_k^* X_k\right)\right\|\right\}.$$

Then for all $t \ge 0$

$$\mathbb{P}\left(\left\|\sum_k X_k\right\| \ge t\right) \le 2d \exp\left(\frac{-t^2/2}{\sigma^2 + Bt/3}\right). \tag{33}$$

5.2 Four useful estimates

Our proofs rely on several estimates. We provide them below, following mostly [4, 47, 60], and using a structure similar to [22]. In order to avoid repetitions and enhance clarity, we summarise here the assumptions we make throughout this subsection:

- Assume that Hypothesis 1 holds true, and let U and D denote the corresponding analysis operators, satisfying the bounds given in (6);
- Let $M \in J$ and $\Delta \subseteq \{1, \dots, M\}$, and set $\mathcal{W} = R(D^* P_\Delta) + R(D^{-1} P_\Delta)$;

- Let $N \in L$ satisfy the balancing property with respect to M and $|\Delta|$;
- For a fixed $\theta \in (0,1)$, let $\{N,\dots,1\} \supseteq \Omega \sim Ber(\theta)$;
- Set $E_\Omega = U^* P_\Omega (U^*)^{-1}$;

The first estimate reads as follows.

Proposition 6 (Off-support incoherence)**.** *For $g \in \mathcal{H}$ and $t \geq \frac{1}{8\sqrt{|\Delta|\kappa_2}}$ we have*

$$\mathbb{P}\left(\|\theta^{-1} P_\Delta^\perp D P_\mathcal{W}^\perp E_\Omega P_\mathcal{W} g\|_{\ell^\infty(J)} > t\|g\|_\mathcal{H}\right) \leq 2\tilde{M}(\tfrac{t\theta}{2}) \exp\left(\frac{-t^2\theta\mu^{-2}}{8\kappa_1 \tilde{B}_\Delta\big(\tilde{B}_\Delta + \frac{\sqrt{2|\Delta|}t}{6}\big)}\right).$$

Proof. Without loss of generality we may assume that $\|g\|_\mathcal{H} = 1$. Let $\{\delta_j\}_{j=1}^N$ be Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \theta$ and such that $\Omega = \{j = 1,\dots,N : \delta_j = 1\}$. We shall need the following inequalities:

$$|\langle U P_\mathcal{W}^\perp D^* e_i, e_k\rangle| \leq |\langle e_i, D P_\mathcal{W}^\perp D^{-1} D U^* e_k\rangle| \leq \tilde{B}_\Delta \|D U^* e_k\|_{\ell^\infty} \leq \tilde{B}_\Delta \mu. \tag{34}$$

$$|\langle U P_\mathcal{W}^\perp D^{-1} e_i, e_k\rangle| \leq |\langle e_i, D^{-*} P_\mathcal{W}^\perp D^* D^{-*} U^* e_k\rangle| \leq \tilde{B}_\Delta \mu. \tag{35}$$

Since $\sum_{k=1}^N e_k e_k^* = P_N$ and $\sum_{k=1}^N \delta_k e_k e_k^* = P_\Omega$ we have

$$\theta^{-1} P_\Delta^\perp D P_\mathcal{W}^\perp E_\Omega P_\mathcal{W} g = \sum_{k=1}^N Y_k + P_\Delta^\perp D P_\mathcal{W}^\perp U^* P_N (U^*)^{-1} P_\mathcal{W} g, \tag{36}$$

where $Y_k = \theta^{-1} P_\Delta^\perp D P_\mathcal{W}^\perp U^* (\delta_k - \theta) e_k e_k^* (U^*)^{-1} P_\mathcal{W} g$. For $i \in \Delta^c$ we define the random variable $X_k^i = \langle Y_k, e_i\rangle$. By the balancing property (10) we need only to bound

$$\mathbb{P}\left(\left\|\sum_{k=1}^N Y_k\right\|_{\ell^\infty} > \frac{t}{2}\right).$$

Let us estimate this quantity by studying the random variables X_k^i via the Bernstein inequality (33) for $d = 1$. In order to do that, first observe that since $\mathbb{E}(\delta_k) = \theta$, then $\mathbb{E}(X_k^i) = 0$. We next study the upper bounds on $\mathbb{E}\left(|X_k^i|^2\right)$ and $|X_k^i|$ for $k = 1,\dots,N$.

On the one hand, by (34) we have

$$\begin{aligned} |\langle D P_\mathcal{W}^\perp U^* e_k e_k^* (U^*)^{-1} P_\mathcal{W} g, e_i\rangle| &= |\langle (U^*)^{-1} P_\mathcal{W} g, e_k\rangle||\langle U P_\mathcal{W}^\perp D^* e_i, e_k\rangle| \\ &\leq \mu \tilde{B}_\Delta |\langle (U^*)^{-1} P_\mathcal{W} g, e_k\rangle|, \end{aligned} \tag{37}$$

so that $\mathbb{E}\left[(\delta_k - \theta)^2\right] = \theta(1-\theta)$ implies for $i \in \Delta^c$

$$\begin{aligned} \mathbb{E}\left(|X_k^i|^2\right) &= \theta^{-2}\mathbb{E}\left((\delta_k - \theta)^2 \left|\langle P_\Delta^\perp D P_\mathcal{W}^\perp U^* e_k e_k^* (U^*)^{-1} P_\mathcal{W} g, e_i\rangle\right|^2\right) \\ &\leq \theta^{-1}(1-\theta)\mu^2 \tilde{B}_\Delta^2 \left|\langle (U^*)^{-1} P_\mathcal{W} g, e_k\rangle\right|^2. \end{aligned}$$

Therefore, since $\| (U^*)^{-1} \| \leq \sqrt{\kappa_1}$, we deduce that

$$\sum_{k=1}^{N} \mathbb{E}\left(|X_k^i|^2\right) \leq \theta^{-1}(1-\theta)\tilde{B}_\Delta^2 \mu^2 \| (U^*)^{-1} P_{\mathcal{W}} g\|^2$$

$$\leq \theta^{-1}\tilde{B}_\Delta^2 \mu^2 \kappa_1 =: \sigma^2. \tag{38}$$

On the other hand, since $\|g\| = 1$,

$$|\langle DP_{\mathcal{W}}^{\perp} U^* e_k e_k^* (U^*)^{-1} P_{\mathcal{W}} g, e_i\rangle| = |\langle g, P_{\mathcal{W}} U^{-1} e_k\rangle||\langle UP_{\mathcal{W}}^{\perp} D^* e_i, e_k\rangle|$$

$$\leq \mu \tilde{B}_\Delta \|P_{\mathcal{W}} U^{-1} e_k\|.$$

We estimate this last term using the identities for $\varepsilon \in \{0,1\}$

$$P_{\mathcal{W}_\varepsilon} = P_{\mathcal{W}_\varepsilon} D_\varepsilon^* P_\Delta D_\varepsilon^{-*}, \quad P_\Delta D_\varepsilon P_{\mathcal{W}_\varepsilon} = P_\Delta D_\varepsilon, \tag{39}$$

where we have set $\mathcal{W}_\varepsilon := R(D_\varepsilon^* P_\Delta)$. Hence, we deduce

$$\|P_{\mathcal{W}_\varepsilon} U^{-1} e_k\|^2 = \langle P_{\mathcal{W}_\varepsilon} D_\varepsilon^* P_\Delta D_\varepsilon^{-*} U^{-1} e_k, P_{\mathcal{W}_\varepsilon} U^{-1} e_k\rangle$$

$$\leq \|P_\Delta D_\varepsilon^{-*} U^{-1} e_k\| \|P_\Delta D_\varepsilon P_{\mathcal{W}_\varepsilon} U^{-1} e_k\|$$

$$= \|P_\Delta D_{1-\varepsilon} U^{-1} e_k\| \|P_\Delta D_\varepsilon U^{-1} e_k\|$$

since $D_\varepsilon^{-*} = D_{1-\varepsilon}$, and we bound each term as follows

$$\|P_\Delta D_\varepsilon U^{-1} e_k\|^2 = \sum_{i\in\Delta} |\langle e_i, D_\varepsilon U^{-1} e_k\rangle|^2 \leq |\Delta|\mu^2.$$

As a consequence, since $\mathcal{W} = \mathcal{W}_0 + \mathcal{W}_1$ we have that

$$\|P_{\mathcal{W}} U^{-1} e_k\|^2 \leq \|P_{\mathcal{W}_0} U^{-1} e_k\|^2 + \|P_{\mathcal{W}_1} U^{-1} e_k\|^2 \leq 2|\Delta|\mu^2. \tag{40}$$

We have obtained that for $i \in \Delta^c$ and $k = 1, \ldots, N$

$$\left|X_k^i\right| \leq \max\{\theta^{-1}(1-\theta), 1\}\tilde{B}_\Delta \mu \|P_{\mathcal{W}} U^{-1} e_k\|$$

$$\leq \theta^{-1}\mu^2 \tilde{B}_\Delta \sqrt{2|\Delta|}\kappa_1 =: B, \tag{41}$$

where in the last inequality we used the fact that $1 \leq \kappa_1$.

Now let $\Gamma \subseteq J$ be a set such that

$$\mathbb{P}\left(\sup_{i\in\Gamma}\left|\sum_{k=1}^{N} X_k^i\right| \geq \frac{t}{2}\right) = 0. \tag{42}$$

Assuming $|\Gamma^c| \leq \tilde{M}(\frac{t\theta}{2})$, by the Bernstein inequality (33) with $d = 1$ we have

$$\mathbb{P}\left(\sup_{i\in\Delta^c}\left|\sum_{k=1}^{N} X_k^i\right| \geq \frac{t}{2}\right) \leq \mathbb{P}\left(\sup_{i\in\Gamma^c}\left|\sum_{k=1}^{N} X_k^i\right| \geq \frac{t}{2}\right)$$

$$\leq 2\tilde{M}(\tfrac{t\theta}{2}) \exp\left(-\frac{t^2\theta}{8\kappa_1\mu^2\tilde{B}_\Delta(\tilde{B}_\Delta + \sqrt{2|\Delta|}t/6)}\right),$$

which is the final estimate.

To finish the proof, we need to show that such Γ^c exists. Note that, because of assumptions (6) and $\|g\| = 1$ we have

$$\begin{aligned}\left|\sum_{k=1}^N X_k^i\right| &= \left|\sum_{k=1}^N \langle \theta^{-1} P_\Delta^\perp D P_\mathcal{W}^\perp U^* (\delta_k - \theta) e_k e_k^* \left(U^*\right)^{-1} P_\mathcal{W} g, e_i\rangle\right| \\ &= \theta^{-1} \left|\langle g, P_\mathcal{W} U^{-1} \left(\textstyle\sum_k (\delta_k - \theta) e_k e_k^*\right) U P_\mathcal{W}^\perp D^* P_\Delta^\perp e_i\rangle\right| \\ &\le \theta^{-1} \|P_\mathcal{W} U^{-1} \left(\textstyle\sum_k (\delta_k - \theta) e_k e_k^*\right) U P_\mathcal{W}^\perp D^* e_i\|_\mathcal{H} \\ &\le \theta^{-1}\sqrt{\kappa_1} \| \left(\textstyle\sum_k (\delta_k - \theta) e_k e_k^*\right) U P_\mathcal{W}^\perp D^* e_i\|_{\ell^2} \\ &\le \theta^{-1}\sqrt{\kappa_1} \|P_N U P_\mathcal{W}^\perp D^* e_i\|_{\ell^2} \\ &\le \theta^{-1}\sqrt{\kappa_1} \left(\|P_N U D^* e_i\|_{\ell^2} + \sqrt{\kappa_1}\|P_{\widetilde{\mathcal{W}}} D^* e_i\|_\mathcal{H}\right).\end{aligned}$$

We then define

$$\Gamma^c = \left\{ i \in J : \theta^{-1}\sqrt{\kappa_1} \left(\|P_N U D^* e_i\|_{\ell^2} + \sqrt{\kappa_1}\|P_{\widetilde{\mathcal{W}}} D^* e_i\|_\mathcal{H}\right) \ge \frac{t}{2} \right\}, \tag{43}$$

which is a finite set and satisfies $|\Gamma^c| \le \tilde{M}(t\theta/2)$ by (12). The proof follows. □

Remark. Observe that in the above proof we have used the full generality of Definition 1: all four terms appear in the derivation.

Proposition 7. *For $g \in \mathcal{W}$ and $(2\sqrt{2\log(|\Delta|k_2)})^{-1} \le t \le 2\kappa_1$ we have*

$$\mathbb{P}\left(\left\|\left(\theta^{-1} P_\mathcal{W} E_\Omega P_\mathcal{W} - P_\mathcal{W}\right) g\right\|_\mathcal{H} > t\|g\|_\mathcal{H}\right) \le \exp\left(\frac{-t^2\theta}{64|\Delta|\mu^2\kappa_1} + \frac{1}{4}\right).$$

Proof. Without loss of generality we assume that $\|g\| = 1$. Let $\{\delta_j\}_{j=1}^N$ be random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \theta$ and such that $\Omega = \{j = 1, \dots, N : \delta_j = 1\}$. For $k \in L$, let

$$\xi_k = \left(U P_\mathcal{W}\right)^* e_k, \qquad \alpha_k = \left(\left(U^*\right)^{-1} P_\mathcal{W}\right)^* e_k.$$

We first make the following observations which will be useful along the proof.

$$\|\xi_k\|^2 = \|P_\mathcal{W} U^* e_k\|^2 \le 2\mu^2|\Delta|, \tag{44}$$

$$\|\alpha_k\|^2 = \|P_\mathcal{W} U^{-1} e_k\|^2 \le 2\mu^2|\Delta|, \tag{45}$$

$$\sum_{k=1}^N |\langle \alpha_k, g\rangle|^2 = \sum_{k=1}^N \left|\langle e_k, U^{-*} P_\mathcal{W} g\rangle\right|^2 \le \|U^{-*} P_\mathcal{W} g\|^2 \le \|U^{-*}\|^2 \le \kappa_1. \tag{46}$$

Note that (45) was already proved in (40) and the derivation of (44) is analogous.

For $u, v \in \mathcal{W}$, let $u \otimes \overline{v}$ denote the continuous operator $\mathcal{W} \to \mathcal{W}$ defined by $(u \otimes \overline{v})(w) = \langle w, v \rangle u$ for $w \in \mathcal{W}$ (note that $u \otimes \overline{v}$ is linear in u and antilinear in v). We have that

$$\sum_{k=1}^{N} \xi_k \otimes \overline{\alpha}_k = P_{\mathcal{W}} U^* P_N U^{-*} P_{\mathcal{W}}, \qquad \sum_{k \in L \setminus \{1,\dots,N\}} \xi_k \otimes \overline{\alpha}_k = P_{\mathcal{W}} U^* P_N^{\perp} U^{-*} P_{\mathcal{W}},$$

$$\theta^{-1} \sum_{k=1}^{N} \delta_k (\xi_k \otimes \overline{\alpha}_k) = \theta^{-1} P_{\mathcal{W}} U^* P_{\Omega} U^{-*} P_{\mathcal{W}}, \qquad P_{\mathcal{W}} = \sum_{k \in L} \xi_k \otimes \overline{\alpha}_k.$$

Hence, we have

$$\begin{aligned}
&\left\| \left(\theta^{-1} P_{\mathcal{W}} U^* P_{\Omega} \left(U^*\right)^{-1} P_{\mathcal{W}} - P_{\mathcal{W}} \right) g \right\| \\
&= \left\| \left(\sum_{k=1}^{N} (\theta^{-1} \delta_k - 1)(\xi_k \otimes \overline{\alpha}_k) \right) g - \sum_{k \in L \setminus \{1,\dots,N\}} (\xi_k \otimes \overline{\alpha}_k) g \right\| \\
&\leq \left\| \left(\sum_{k=1}^{N} (\theta^{-1} \delta_k - 1)(\xi_k \otimes \overline{\alpha}_k) \right) g \right\| + \left\| \left(P_{\mathcal{W}} U^* P_N^{\perp} \left(U^*\right)^{-1} P_{\mathcal{W}} \right) g \right\|.
\end{aligned}$$

Therefore, by the balancing property (8) it follows that

$$\begin{aligned}
&\mathbb{P} \left(\left\| \left(\theta^{-1} P_{\mathcal{W}} U^* P_{\Omega} \left(U^*\right)^{-1} P_{\mathcal{W}} - P_{\mathcal{W}} \right) g \right\| > t \right) \\
&\leq \mathbb{P} \left(\left\| \left(\theta^{-1} P_{\mathcal{W}} U^* P_{\Omega} \left(U^*\right)^{-1} P_{\mathcal{W}} - P_{\mathcal{W}} \right) g \right\| > \frac{t}{2} + \left\| P_{\mathcal{W}} U^* P_N^{\perp} \left(U^*\right)^{-1} P_{\mathcal{W}} \right\| \right) \\
&\leq \mathbb{P} \left(\left\| \sum_{k=1}^{N} (\theta^{-1} \delta_k - 1)(\xi_k \otimes \overline{\alpha}_k) g \right\| > \frac{t}{2} \right)
\end{aligned}$$

for $t \geq (2\sqrt{2 \log(|\Delta| k_2)})^{-1}$.

Let us estimate the above probability by using the vector Bernstein inequality (32). For that, we define

$$X_k = (\theta^{-1} \delta_k - 1)(\xi_k \otimes \overline{\alpha}_k) g \in \mathcal{W} \cong \mathbb{C}^d$$

with $d = \dim \mathcal{W}$. First note that $\mathbb{E}(X_k) = 0$. Next, observe that

$$\begin{aligned}
\|X_k\|^2 &= (\theta^{-1} \delta_k - 1)^2 |\langle g, \alpha_k \rangle|^2 \|\xi_k\|^2 \\
&\leq (\theta^{-1} \delta_k - 1)^2 \|\alpha_k\|^2 \|\xi_k\|^2.
\end{aligned}$$

Thus, by (44) and (45) it follows that

$$\begin{aligned}
\|X_k\| &\leq \max\{\theta^{-1} - 1, 1\} \|\xi_k\| \|\alpha_k\| \\
&\leq 2\theta^{-1} |\Delta| \mu^2 =: B.
\end{aligned}$$

In addition, since $\mathbb{E}(\theta^{-1}\delta_k - 1)^2 = \theta^{-1} - 1$, by (44) and (46) we obtain

$$\begin{aligned}\sum_{k=1}^{N} \mathbb{E}\left(\|X_k\|^2\right) &\leq 2(\theta^{-1}-1)|\Delta|\mu^2 \sum_{k=1}^{N} |\langle \alpha_k, g\rangle|^2 \\ &\leq 2\theta^{-1}|\Delta|\mu^2\kappa_1 =: \sigma^2.\end{aligned}$$

Therefore, applying the Vector Bernstein inequality (32) for $(2\sqrt{2\log(|\Delta|k_2)})^{-1} \leq t \leq 2\kappa_1$ we get the desired estimate. □

The next result will play an important role in order to guarantee property (i) of Proposition 10. Thus, instead of working with the operator $U^* P_\Omega (U^*)^{-1}$ as in the previous two results, now we will deal with a matrix operator containing $U^{-1}P_\Omega U$.

Proposition 8 (Local isometry)**.** *We have*

$$\mathbb{P}\left(\left\|(\theta^{-1}P_{\mathcal{W}}U^{-1}P_\Omega U P_{\mathcal{W}} - P_{\mathcal{W}})\right\|_{\mathcal{H}\to\mathcal{H}} > \frac{1}{2}\right) \leq 4|\Delta|\exp\left(\frac{-3\theta}{208|\Delta|\mu^2\kappa_1}\right).$$

Proof. Let $\{\delta_j\}_{j=1}^N$ be the random Bernoulli variables with $\mathbb{P}(\delta_j = 1) = \theta$ and such that $\Omega = \{j = 1, \ldots, N : \delta_j = 1\}$. We consider $\xi_k = (UP_{\mathcal{W}})^* e_k$, $\alpha_k = \left((U^*)^{-1} P_{\mathcal{W}}\right)^* e_k$, and arguing as in Proposition 7, we arrive to

$$\begin{aligned}&\mathbb{P}\left(\left\|(\theta^{-1}P_{\mathcal{W}}U^{-1}P_\Omega U P_{\mathcal{W}} - P_{\mathcal{W}})\right\| > \frac{1}{2}\right) \\ &\leq \mathbb{P}\left(\left\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_\Omega U P_{\mathcal{W}} - P_{\mathcal{W}}\right\| > \frac{1}{4} + \left\|P_{\mathcal{W}}U^{-1}P_N^{\perp}UP_{\mathcal{W}}\right\|\right) \\ &\leq \mathbb{P}\left(\left\|\sum_{k=1}^{N}(\theta^{-1}\delta_k - 1)(\alpha_k \otimes \overline{\xi}_k)\right\| > \frac{1}{4}\right),\end{aligned}$$

where the last probability of the above inequality will be estimated by using the matrix Bernstein inequality (33).

Let us define now

$$X_k = (\theta^{-1}\delta_k - 1)(\alpha_k \otimes \overline{\xi}_k)\colon \mathcal{W} \to \mathcal{W}.$$

Since $\mathcal{W}$ is finite dimensional, X_k may be identified with an element in $\mathbb{C}^{d\times d}$, where $d = \dim \mathcal{W} \leq 2|\Delta|$. We have $\mathbb{E}(X_k) = 0$. Further, since $1 \leq \kappa_1$ and by (44), (45), it follows that

$$\begin{aligned}\|X_k\| &\leq \max\{\theta^{-1}-1, 1\}\|\alpha_k\|\|\xi_k\| \\ &\leq 2\theta^{-1}\mu^2|\Delta|\kappa_1 =: B.\end{aligned}$$

We next study $\mathbb{E}(X_k^* X_k)$ and $\mathbb{E}(X_k X_k^*)$. Since $X_k^* = (\theta^{-1}\delta_k - 1)\xi_k \otimes \overline{\alpha}_k$, we have

$$\begin{aligned}X_k^* X_k &= (\theta^{-1}\delta_k - 1)^2(\xi_k \otimes \overline{\alpha}_k)(\alpha_k \otimes \overline{\xi}_k) = (\theta^{-1}\delta_k - 1)^2\|\alpha_k\|^2\xi_k \otimes \overline{\xi}_k \\ X_k X_k^* &= (\theta^{-1}\delta_k - 1)^2(\alpha_k \otimes \overline{\xi}_k)(\xi_k \otimes \overline{\alpha}_k) = (\theta^{-1}\delta_k - 1)^2\|\xi_k\|^2\alpha_k \otimes \overline{\alpha}_k\end{aligned}$$

and since

$$\sum_{k=1}^{N} \xi_k \otimes \overline{\xi}_k = P_{\mathcal{W}} U^* P_N U P_{\mathcal{W}}, \qquad \sum_{k=1}^{N} \alpha_k \otimes \overline{\alpha}_k = P_{\mathcal{W}} U^{-1} P_N \left(U^*\right)^{-1} P_{\mathcal{W}},$$

using that the norms of U, U^*, U^{-1} and $(U^*)^{-1}$ are upper bounded by $\sqrt{\kappa_1}$, we obtain

$$\left\| \sum_{k=1}^{N} \mathbb{E}(X_k^* X_k) \right\| \leq 2(\theta^{-1} - 1)\mu^2 |\Delta| \| P_{\mathcal{W}_\delta} U^* P_N U P_{\mathcal{W}_\delta} \| \leq 2(\theta^{-1} - 1)\mu^2 |\Delta| \kappa_1,$$

$$\left\| \sum_{k=1}^{N} \mathbb{E}(X_k X_k^*) \right\| \leq 2(\theta^{-1} - 1)\mu^2 |\Delta| \| P_{\mathcal{W}_\delta} U^{-1} P_N \left(U^*\right)^{-1} P_{\mathcal{W}_\delta} \| \leq 2(\theta^{-1} - 1)\mu^2 |\Delta| \kappa_1.$$

Hence we can choose

$$\sigma^2 := 2\theta^{-1}\mu^2 |\Delta| \kappa_1$$

and applying the Bernstein inequality (33) for $t = \frac{1}{4}$, as well as the balancing property, we deduce the wanted estimate. The proof follows. □

We conclude this subsection with the following estimate.

Proposition 9 (Uniform off-support incoherence)**.** *If* $\tilde{B}_\Delta \leq 2\sqrt{2|\Delta|}\kappa_1\kappa_2$ *then*

$$\mathbb{P}\left(\max_{j \in \Delta^c} \| \theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j \| > 1 \right) \leq \tilde{M}(\theta) \exp\left(\frac{-\theta}{64\mu^2 |\Delta| \kappa_1 \kappa_2} + \frac{1}{4} \right).$$

Proof. Fix $j \in \Delta^c$. We have

$$\theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j = \sum_{k=1}^{N} Y_k + P_{\mathcal{W}} U^{-1} P_N U P_{\mathcal{W}}^\perp D^{-1} e_j,$$

where $Y_k = \theta^{-1} P_{\mathcal{W}} U^{-1} (\delta_k - \theta) e_k e_k^* U P_{\mathcal{W}}^\perp D^{-1} e_j$. Note that $\mathbb{E}(Y_k) = 0$. Since

$$\begin{aligned} \| \theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j \| &\leq \left\| \sum_{k=1}^{N} Y_k \right\| + \| P_{\mathcal{W}} U^{-1} P_N U P_{\mathcal{W}}^\perp D^{-1} e_j \| \\ &\leq \left\| \sum_{k=1}^{N} Y_k \right\| + \| P_{\mathcal{W}} U^{-1} P_N U P_{\mathcal{W}}^\perp D^{-1} P_\Delta^\perp \|_{\ell^2(J) \to \mathcal{H}}, \end{aligned}$$

by the balancing property (11) we obtain

$$\begin{aligned} &\mathbb{P}\left(\| \theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j \| > 1 \right) \\ &\leq \mathbb{P}\left(\| \theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j \| > \frac{1}{2} + \frac{1}{2} \right) \\ &\leq \mathbb{P}\left(\| \theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j \| > \frac{1}{2} + \| P_{\mathcal{W}} U^{-1} P_N U P_{\mathcal{W}}^\perp D^{-1} P_\Delta^\perp \|_{\ell^2 \to \mathcal{H}} \right) \\ &\leq \mathbb{P}\left(\left\| \sum_{k=1}^{N} Y_k \right\| > \frac{1}{2} \right). \end{aligned}$$

Next, by (35) and (40) we have

$$\begin{aligned}\|Y_k\| &= \theta^{-1}|\langle e_k, UP_{\mathcal{W}}^{\perp}D^{-1}e_j\rangle|\|P_{\mathcal{W}}U^{-1}(\delta_k-\theta)e_k\|\\ &\leq \max\{1,(1-\theta)/\theta)\}\mu^2\tilde{B}_{\Delta}\sqrt{2|\Delta|}\\ &\leq \theta^{-1}\mu^2\tilde{B}_{\Delta}\sqrt{2|\Delta|} =: B.\end{aligned}$$

In addition, using again (40) yields

$$\begin{aligned}\sum_{k=1}^{N}\mathbb{E}(\|Y_k\|^2) &\leq 2(\theta^{-1}-1)\mu^2|\Delta|\sum_{k=1}^{N}|\langle e_k, UP_{\mathcal{W}}^{\perp}D^{-1}e_j\rangle|^2\\ &\leq 2(\theta^{-1}-1)\mu^2|\Delta|\kappa_1\kappa_2\\ &\leq 2\theta^{-1}\mu^2|\Delta|\kappa_1\kappa_2 =: \sigma^2.\end{aligned}$$

Now assume that there exists a non-empty set $\Gamma\subseteq J$ such that

$$\mathbb{P}\left(\sup_{j\in\Gamma}\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}e_j\|>1\right)=0. \tag{47}$$

Assuming $|\Gamma^c|\leq\tilde{M}(\theta)$, using the Vector Bernstein inequality (32) for $t=\frac{1}{2}$ and the union bound, we obtain

$$\begin{aligned}&\mathbb{P}\left(\sup_{j\in\Delta^c}\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}e_j\|>1\right)\\ &\quad=\mathbb{P}\left(\sup_{j\in\Delta^c\cap\Gamma^c}\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}e_j\|>1\right)\\ &\quad\leq\tilde{M}(\theta)\exp\left(\frac{-\theta}{64\mu^2|\Delta|\kappa_1\kappa_2}+\frac{1}{4}\right),\end{aligned}$$

which is our final estimate.

We only have to show the existence of Γ and provide a bound on $|\Gamma^c|$. Note that

$$\begin{aligned}&\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}e_j\|\leq\theta^{-1}\sqrt{\kappa_1}\|P_NUP_{\mathcal{W}}^{\perp}D^{-1}e_j\|_{\ell^2}\\ &\leq\theta^{-1}\sqrt{\kappa_1}\left(\|P_NUD^{-1}e_j\|_{\ell^2}+\sqrt{\kappa_1}\|P_{\widetilde{\mathcal{W}}}D^{-1}e_j\|_{\mathcal{H}}\right).\end{aligned}$$

We then define as in Proposition 6

$$\Gamma=\left\{j\in J:\theta^{-1}\sqrt{\kappa_1}\left(\|P_NUD^{-1}e_j\|_{\ell^2}+\sqrt{\kappa_1}\|P_{\widetilde{\mathcal{W}}}D^{-1}e_j\|_{\mathcal{H}}\right)\leq 1\right\}, \tag{48}$$

which satisfies $|\Gamma^c|\leq\tilde{M}(\theta)$ by (12) (since $D^{-1}=D_1^*$). The proof follows. □

5.3 The dual certificate

We now show how the existence of a *dual certificate* ρ satisfying certain properties guarantees exact recovery up to measurement noise. We follow closely [60, Proposition 6.1].

Proposition 10. *Assume that Hypothesis 1 holds true, and let U and D denote the corresponding analysis operators, satisfying the bounds given in* (6)*. Let $\Delta \subseteq J$ and Ω be a finite subset of L. Let $g_0 \in \mathcal{H}$ and $\eta \in \ell^2(L)$ be such that $\|\eta\|_2 \leq \varepsilon$ for some $\varepsilon \geq 0$. Let $\zeta = P_\Omega U g_0 + \eta$ be the known noisy measurement. Assume that there exist $\rho = U^* P_\Omega \rho'$ for some $\rho' \in \ell^2(L)$, $Q > 0$ and $0 < \theta \leq 1$ with the following properties:*

(i) $\| \left(\theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}\right)^{-1} \|_{\mathcal{W}\to\mathcal{W}} \leq 2$,

(ii) $\|P_{\mathcal{W}}\rho - D^* \mathrm{sgn}(P_\Delta D g_0)\|_{\mathcal{H}} \leq \frac{1}{8}$,

(iii) $\|P_\Delta^\perp D^{-*} P_{\mathcal{W}}^\perp \rho\|_{l^\infty(J)} \leq \frac{1}{4}$,

(iv) $\max_{j\in\Delta^c} \|\theta^{-1} P_{\mathcal{W}} U^{-1} P_\Omega U P_{\mathcal{W}}^\perp D^{-1} e_j\|_{\mathcal{H}} \leq 1$,

(v) $\|\rho'\|_{\ell^2(L)} \leq Q\sqrt{\kappa_1\kappa_2|\Delta|}$.

Let $g \in \mathcal{H}$ be a minimizer of the problem

$$\inf_{\substack{\tilde g \in \mathcal{H}\\ D\tilde g \in \ell^1(J)}} \|D\tilde g\|_{\ell^1} \quad \text{subject to } \|P_\Omega U \tilde g - \zeta\|_2 \leq \varepsilon.$$

Then

$$\|g-g_0\|_{\mathcal{H}} \leq 4(2+\sqrt{\kappa_2})\|P_\Delta^\perp D g_0\|_{\ell^1} + \varepsilon\sqrt{\kappa_1}\left(4\theta^{-1} + (2+\sqrt{\kappa_2})(\theta^{-1} + 4L\sqrt{\kappa_2|\Delta|})\right).$$

Proof. We start from the following identity, for any $\tilde g \in \mathcal{H}$:

$$P_{\mathcal{W}}^\perp \tilde g = P_{\mathcal{W}}^\perp D^{-1} D\tilde g = P_{\mathcal{W}}^\perp D^{-1} P_\Delta^\perp D\tilde g. \tag{49}$$

This follows from the fact that $D^{-1}D$ is the identity and that $P_{\mathcal{W}}$ is the orthogonal projection on $R(D^{-1}P_\Delta) + R(D^* P_\Delta)$. From the estimates (6) we obtain, for any $\tilde g \in \mathcal{H}$,

$$\|\tilde g\|_{\mathcal{H}} \leq \|P_{\mathcal{W}}\tilde g\|_{\mathcal{H}} + \|P_{\mathcal{W}}^\perp \tilde g\|_{\mathcal{H}} \leq \|P_{\mathcal{W}}\tilde g\|_{\mathcal{H}} + \sqrt{\kappa_2}\|P_\Delta^\perp D\tilde g\|_{\ell^1}. \tag{50}$$

From the last inequality applied to $h = g - g_0$, we see that it is enough to bound $\|P_{\mathcal{W}} h\|_{\mathcal{H}}$ and $\|P_\Delta^\perp D h\|_{\ell^1}$ in order to finish the proof. Let us start from $\|P_{\mathcal{W}} h\|_{\mathcal{H}}$.

First note that since g is a minimizer we have

$$\|P_\Omega U h\|_2 \leq 2\varepsilon. \tag{51}$$

From this and (i) we find

$$\begin{aligned}\|P_{\mathcal{W}}h\| &= \|(P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}})^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}h\| \\ &\leq 2\theta^{-1}\|P_{\mathcal{W}}U^{-1}P_{\Omega}U(h-P_{\mathcal{W}}^{\perp}h)\| \\ &\leq 2(2\varepsilon\sqrt{\kappa_1}\theta^{-1}+\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}h\|).\end{aligned}$$

We bound the last term as follows:

$$\begin{aligned}\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}h\| &= \|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}P_{\Delta}^{\perp}Dh\| \\ &\leq \max_{j\in\Delta^c}\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_{\Omega}UP_{\mathcal{W}}^{\perp}D^{-1}e_j\|\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \\ &\leq \|P_{\Delta}^{\perp}Dh\|_{\ell^1},\end{aligned}$$

where we used identity (49) and (iv). Thus we have found

$$\|P_{\mathcal{W}}h\| \leq 2(2\varepsilon\sqrt{\kappa_1}\theta^{-1}+\|P_{\Delta}^{\perp}Dh\|_{\ell^1}). \tag{52}$$

We now pass to the estimate of $\|P_{\Delta}^{\perp}Dh\|_{\ell^1}$. Note that

$$\begin{aligned}\|Dg\|_{\ell^1} &= \|P_{\Delta}^{\perp}D(g_0+h)\|_{\ell^1}+\|P_{\Delta}D(g_0+h)\|_{\ell^1} \\ &\geq \|P_{\Delta}^{\perp}Dh\|_{\ell^1}-\|P_{\Delta}^{\perp}Dg_0\|_{\ell^1}+\|P_{\Delta}Dg_0\|_{\ell^1} \\ &\quad+\operatorname{Re}\langle P_{\Delta}Dh,\operatorname{sgn}(P_{\Delta}Dg_0)\rangle \\ &= \|P_{\Delta}^{\perp}Dh\|_{\ell^1}-2\|P_{\Delta}^{\perp}Dg_0\|_{\ell^1}+\|Dg_0\|_{\ell^1}-|\langle P_{\Delta}Dh,\operatorname{sgn}(P_{\Delta}Dg_0)\rangle|.\end{aligned}$$

Singe g is a minimizer, we find that

$$\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \leq 2\|P_{\Delta}^{\perp}Dg_0\|_{\ell^1}+|\langle P_{\Delta}Dh,\operatorname{sgn}(P_{\Delta}Dg_0)\rangle|. \tag{53}$$

Now, since $\rho = U^*P_{\Omega}\rho'$ and identity (49) we have

$$\begin{aligned}|\langle P_{\Delta}Dh,\operatorname{sgn}(P_{\Delta}Dg_0)\rangle| &= |\langle h, D^*\operatorname{sgn}(P_{\Delta}Dg_0)\rangle| \\ &\leq |\langle h, D^*\operatorname{sgn}(P_{\Delta}Dg_0)-P_{\mathcal{W}}\rho\rangle|+|\langle h,\rho\rangle|+|\langle h,P_{\mathcal{W}}^{\perp}\rho\rangle| \\ &\leq \|P_{\mathcal{W}}h\|\|D^*\operatorname{sgn}(P_{\Delta}Dg_0)-P_{\mathcal{W}}\rho\|+\|P_{\Omega}Uh\|\|\rho'\|+|\langle D^{-1}P_{\Delta}^{\perp}Dh,P_{\mathcal{W}}^{\perp}\rho\rangle| \\ &\leq \frac{1}{8}\|P_{\mathcal{W}}h\|+2\varepsilon Q\sqrt{\kappa_1\kappa_2|\Delta|}+\frac{1}{4}\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \\ &\leq \varepsilon\sqrt{\kappa_1}\left(\frac{\theta^{-1}}{2}+2Q\sqrt{\kappa_2|\Delta|}\right)+\frac{1}{2}\|P_{\Delta}^{\perp}Dh\|_{\ell^1},\end{aligned}$$

where we have used also (52). Thus we have obtained

$$\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \leq 4\|P_{\Delta}^{\perp}Dg_0\|+\varepsilon\sqrt{\kappa_1}\left(\theta^{-1}+4Q\sqrt{\kappa_2|\Delta|}\right), \tag{54}$$

which yields the final estimate

$$\begin{aligned}\|h\|_{\mathcal{H}} &\leq \|P_{\mathcal{W}}h\|+\sqrt{\kappa_2}\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \leq 4\varepsilon\sqrt{\kappa_1}\theta^{-1}+(2+\sqrt{\kappa_2})\|P_{\Delta}^{\perp}Dh\|_{\ell^1} \\ &\leq 4(2+\sqrt{\kappa_2})\|P_{\Delta}^{\perp}Dg_0\|_{\ell^1}+\varepsilon\sqrt{\kappa_1}\left(4\theta^{-1}+(2+\sqrt{\kappa_2})(\theta^{-1}+4Q\sqrt{\kappa_2|\Delta|})\right).\square\end{aligned}$$

By using the results of §5.2, we now show that the dual certificate ρ can be constructed. The proof is based on a *golfing scheme* [37, 36].

Proposition 11. *Assume that Hypothesis 1 holds true, and let U and D denote the corresponding analysis operators, satisfying the bounds given in* (6)*. Let $M \in J$ and $\omega \geq 1$ be such that $M \geq 5$. Let $\Delta \subseteq \{1, \ldots, M\}$ be such that $|\Delta| \geq 3$ and $\tilde{B}_\Delta \leq 2\sqrt{2|\Delta|}\kappa_1\kappa_2$. Let N satisfy the balancing property with respect to M and $|\Delta|$. Let $\Omega \subseteq \mathbb{N}$ be chosen uniformly at random with $|\Omega| = m$. Take $g_0 \in \mathcal{H}$. If*

$$m \geq C\mu^2|\Delta|\kappa_1\kappa_2\omega^2\tilde{B}_\Delta^2 N \log\left(|\Delta|\kappa_2\tilde{M}\left(\frac{C'm}{N\sqrt{|\Delta|\kappa_2}}\right)\right),$$

then, with probability exceeding $1 - e^{-\omega}$, there exist $\rho = U^*P_\Omega\rho'$ for some $\rho' \in \ell^2(L)$ and $Q \leq C''\omega\frac{N}{m}$ satisfying the properties (i)-(v) of Proposition 10, with $\theta = m/N$, where $C, C', C'' > 0$ are universal constants.

Proof. The proof is based on a recursive procedure to construct a sequence of vectors $\{Y_i\}$ converging to the dual certificate ρ with high probability.

The set $\Omega \subseteq \{1, \ldots, N\}$ is chosen uniformly at random with $|\Omega| = m$. It is well known that we may, without loss of generality, replace this way of choosing Ω with the model that $\{1, \ldots, N\} \supset \Omega \sim Ber(\theta)$ for $\theta = m/N$ (θ will have this value throughout the proof). This is equivalent to choosing Ω as

$$\Omega = \Omega_1 \cup \ldots \cup \Omega_{l'}$$

with $\Omega_{l'}$ following a Bernoulli distribution as explained below. The main difference with the golfing scheme in [4] is that the number l' of sampled sets is greater than l, the number of iterations in our recursive scheme (both to be determined later). In fact, given q_i for $i = 1, \ldots, l$, we will sample $l' \geq l$ sets distributed as $Ber(q_i)$, for some $i = 1, \ldots, l$ and will keep only l of them for the construction of the certificate.

To initialize the iterations, set

$$Y_0 = 0,$$

and define

$$Z_i = D^*\mathrm{sgn}(P_\Delta Dg_0) - P_\mathcal{W}Y_i, \qquad 0 \leq i \leq l. \tag{55}$$

Let us define the sequence $\{Y_i\}_{i=1}^l$ iteratively as follows. Given q_i, for $j = 1, 2, \ldots$ we choose $\Omega_i^j \subseteq \{1, \ldots, N\}$ at random such that $\Omega_i^j \sim Ber(q_i)$. Let $E_{\Omega_i^j} = U^*P_{\Omega_i^j}(U^*)^{-1}$. We repeat the choice for $j = 1, 2, \ldots$ until the conditions

$$\left\|\left(P_\mathcal{W} - q_i^{-1}P_\mathcal{W}E_{\Omega_i^j}P_\mathcal{W}\right)Z_{i-1}\right\| \leq \alpha_i\|Z_{i-1}\|, \tag{56}$$

$$\left\|q_i^{-1}P_\Delta^\perp D^{-*}P_\mathcal{W}^\perp E_{\Omega_i^j}Z_{i-1}\right\|_{l^\infty} \leq \beta_i\|Z_{i-1}\|, \tag{57}$$

hold for some parameters $\alpha_i, \beta_i \in \mathbb{R}$ that will be chosen later. We set

$$r_i = \min\{j = 1, 2, \cdots : (56) \text{ and } (57) \text{ are satisfied}\},$$

namely r_i denotes the number of repetitions of the i-th step. We also set

$$\Omega = \bigcup_{i=1}^{l} \bigcup_{j=1}^{r_i} \Omega_i^j, \qquad \Omega_i = \Omega_i^{r_i}, \qquad Y_i = \sum_{k=1}^{i} q_k^{-1} E_{\Omega_k} Z_{k-1},$$

and

$$\rho := Y_l, \qquad \rho' = \sum_{i=1}^{l} q_i^{-1} P_{\Omega_i} (U^*)^{-1} Z_{i-1}, \tag{58}$$

so that $\rho = U^* P_\Omega \rho'$.

The identities (55) and $Y_i = Y_{i-1} + q_i^{-1} E_{\Omega_i} Z_{i-1}$ yield

$$Z_i = Z_{i-1} - q_i^{-1} P_{\mathcal{W}} E_{\Omega_i} P_{\mathcal{W}} Z_{i-1} = (P_{\mathcal{W}} - q_i^{-1} P_{\mathcal{W}} E_{\Omega_i} P_{\mathcal{W}}) Z_{i-1}.$$

Thus by (56) it follows that

$$\|Z_i\| \le \alpha_i \|Z_{i-1}\| \le \prod_{j=1}^{i} \alpha_j \|Z_0\| \le \sqrt{\kappa_2 |\Delta|} \prod_{j=1}^{i} \alpha_j,$$

which together with $Z_l = D^* \mathrm{sgn}(P_\Delta D g_0) - P_{\mathcal{W}} \rho$ gives

$$\|D^* \mathrm{sgn}(P_\Delta D g_0) - P_{\mathcal{W}} \rho\| \le \sqrt{\kappa_2 |\Delta|} \prod_{i=1}^{l} \alpha_i. \tag{59}$$

Moreover by (57) we have

$$\|P_\Delta^\perp D^{-*} P_{\mathcal{W}}^\perp \rho\|_{l^\infty} \le \sum_{i=1}^{l} \left\| q_i^{-1} P_\Delta^\perp D^{-*} P_{\mathcal{W}}^\perp E_{\Omega_i} Z_{i-1} \right\|_{l^\infty} \le \sqrt{|\Delta| \kappa_2} \sum_{i=1}^{l} \beta_i \prod_{j=1}^{i-1} \alpha_j$$

and

$$\|\rho'\| \le \sum_{i=1}^{l} q_i^{-1} \left\| U^{-*} Z_{i-1} \right\| \le \sqrt{\kappa_1} \sum_{i=1}^{l} q_i^{-1} \|Z_{i-1}\| \le \sqrt{|\Delta| \kappa_1 \kappa_2} \sum_{i=1}^{l} q_i^{-1} \prod_{j=1}^{i-1} \alpha_j.$$

(For $i = 1$ we take $\Pi_{j=1}^{i-1} \alpha_j = 1$.) We next choose the parameters l, α_i and β_i in a suitable way to show that (ii), (iii) and (v) in Proposition 10 are satisfied. Letting

$$l = \left\lceil \log_2(\sqrt{|\Delta| \kappa_2}) + 2 \right\rceil, \qquad \alpha_1 = \alpha_2 = \frac{1}{\sqrt{8 \log |\Delta| \kappa_2}}, \qquad \beta_1 = \beta_2 = \frac{1}{7\sqrt{|\Delta| \kappa_2}}$$

and for $i \geq 3$

$$\alpha_i = \frac{1}{2}, \qquad \beta_i = \frac{\log(|\Delta|\kappa_2)}{7\sqrt{|\Delta|\kappa_2}},$$

from the above estimates we readily derive

$$\|D^* \mathrm{sgn}(P_\Delta D g_0) - P_{\mathcal{W}}\rho\| \leq \frac{1}{8}, \quad \|P_\Delta^\perp D^{-*} P_{\mathcal{W}}^\perp \rho\|_{l^\infty} \leq \frac{1}{4}, \quad \|\rho'\| \leq \sqrt{|\Delta|\kappa_1\kappa_2} Q,$$

where $Q = \sum_{i=1}^{l} q_i^{-1} \prod_{j=1}^{i-1} \alpha_j$ will be estimated at the end of the proof.

Next, we need to establish that the total number of sampled Ω_i^j remains small with high probability. More precisely, we will bound the probability

$$p_3 = \mathbb{P}\left((r_1 > 1) \quad \text{or} \quad (r_2 > 1) \quad \text{or} \quad \sum_{i=1}^{l} r_i > l' \right)$$

for some l' to be chosen later. To that end, denote $p_1(i)$ the probability that (56) fails in the i-th step and $p_2(i)$ the probability of failure for (57). We want to use Propositions 7 and 6 to bound these probabilities. Proposition 7 for $t = \alpha_i$ gives the estimate

$$p_1(i) \leq \exp\left(\frac{-\alpha_i^2 q_i}{64|\Delta|\mu^2\kappa_1} + \frac{1}{4} \right).$$

Thus if

$$q_i \geq \frac{64\mu^2|\Delta|\kappa_1}{\alpha_i^2}(\omega + \log(\gamma) + \frac{1}{4}),$$

then $p_1(i) \leq \frac{1}{\gamma}e^{-\omega}$. Similarly, Proposition 6 for $t = \beta_i$ yields

$$p_2(i) \leq 2\tilde{M}(q_i\beta_i/2) \exp\left(\frac{-\beta_i^2 q_i}{8\kappa_1\mu^2\tilde{B}_\Delta(\tilde{B}_\Delta + \sqrt{2|\Delta|}\beta_i/6)} \right).$$

Thus if

$$q_i \geq \frac{8\mu^2\kappa_1\tilde{B}_\Delta(\tilde{B}_\Delta + \sqrt{2|\Delta|}\beta_i/6)}{\beta_i^2}(\omega + \log(2\tilde{M}(q_i\beta_i/2)\gamma)),$$

then $p_2(i) \leq \frac{1}{\gamma}e^{-\omega}$.

Assume q_i are chosen as follows:

$$q_1 = q_2 \geq 512\mu^2\kappa_1\kappa_2|\Delta|\tilde{B}_\Delta^2(\omega + 2.05)\log\left(|\Delta|\kappa_2\tilde{M}(q_1\beta_1/2) \right), \tag{60}$$

$$q_i \geq 406\mu^2\kappa_1\kappa_2|\Delta|\tilde{B}_\Delta^2(\omega + 2.17)\frac{\log \tilde{M}(q_i\beta_i/2)}{\log(|\Delta|\kappa_2)}, \quad i \geq 3. \tag{61}$$

Since $\tilde{B}_\Delta, \kappa_2 \geq 1$, $M \geq 5$ and $|\Delta|\kappa_2 \geq 3$, with this choice we have $p_1(i), p_2(i) \leq \frac{1}{6}e^{-\omega} \leq \frac{1}{16}$ for $i \geq 1$, so that

$$\mathbb{P}((56) \text{ and } (57) \text{ are satisfied}) \geq \quad 7/8, \qquad i \geq 1.$$

As a consequence, since $\sum_{i=1}^{l} r_i > l'$ if and only if fewer than l of the first l' samplings satisfied both (56) and (57), we have

$$\mathbb{P}\left(\sum_{i=1}^{l} r_i > l'\right) \leq \mathbb{P}(X < l) = \mathbb{P}(X \leq l-1), \qquad X \sim Bin\left(l', \frac{7}{8}\right),$$

(see equation (45) in [36]). Thus we need to bound the probability of obtaining less than l outcomes in a binomial process with l' repetitions and individual success probability $\frac{7}{8}$. Following [37] and [47] we bound this quantity using a standard concentration bound from [56]

$$\mathbb{P}\left(Bin(n,p) - np \leq -\tau\right) \leq e^{-2\tau^2/n},$$

which implies

$$\mathbb{P}\left(\sum_{i=1}^{l} r_i > l'\right) \leq \exp\left(\frac{-2\left(\frac{7}{8}l' - l + 1\right)^2}{l'}\right).$$

Therefore, choosing $l' = \frac{16}{7}(l-1) + \frac{32}{49}(\omega + \log 6)$, we get

$$\mathbb{P}\left(\sum_{i=1}^{l} r_i > l'\right) \leq \frac{1}{6}e^{-\omega},$$

and, as a consequence, we obtain

$$p_3 \leq p_1(1) + p_2(1) + p_1(2) + p_2(2) + \frac{1}{6}e^{-\omega} \leq \frac{5}{6}e^{-\omega}.$$

Let us now consider property (i) of Proposition 10. Our aim is to show that

$$p_4 = \mathbb{P}\left(\left\|\theta^{-1}P_{\mathcal{W}}U^{-1}P_\Omega U P_{\mathcal{W}} - P_{\mathcal{W}}\right\| > \frac{1}{2}\right) \leq \frac{1}{12}e^{-\omega}.$$

From Proposition 8 we immediately obtain that if θ satisfies

$$\theta \geq 70|\Delta|\mu^2\kappa_1(\omega + \log|\Delta| + \log 48), \tag{62}$$

then $p_4 \leq \frac{1}{12}e^{-\omega}$.

Now, let p_5 be the probability that property (iv) of Proposition 10 fails. We want to show that

$$p_5 = \mathbb{P}\left(\max_{j \in \Delta^c} \|\theta^{-1}P_{\mathcal{W}}U^{-1}P_\Omega U P_{\mathcal{W}}^{\perp}D^{-1}e_j\| > 1\right) \leq \frac{1}{12}e^{-\omega}.$$

By Proposition 9, if θ satisfies

$$\theta \geq 64|\Delta|\mu^2\kappa_1\kappa_2(\omega + \log \tilde{M}(\theta) + 1/4 + \log 12), \tag{63}$$

we have $p_5 \leq \frac{1}{12}e^{-\omega}$.

In order to finish the proof we need to give a bound on m (or, equivalently θ) and construct q_i such that conditions (60), (61), (62) and (63) are satisfied.

Let θ satisfy

$$\theta \geq 10902\mu^2|\Delta|\kappa_1\kappa_2\omega^2\tilde{B}_\Delta^2 \log\left(|\Delta|\kappa_2\tilde{M}\left(\frac{643}{10902\omega}\beta_1\theta\right)\right). \tag{64}$$

(We have made no attempt to optimize the constant.) Then conditions (62) and (63) are clearly satisfied (using $\tilde{M}(\frac{643}{10902\omega}\beta_1\theta) \geq \tilde{M}(\theta)$). Now recall that at each iteration i we sampled r_i sets $\Omega_i^j \sim Ber(q_i)$ and we stopped after $\sum_{i=1}^{l} r_i \leq l'$ sampling. Since

$$\Omega = \bigcup_{i=1}^{l}\bigcup_{j=1}^{r_i}\Omega_i^j, \qquad \Omega \sim Ber(\theta), \qquad \Omega_i^j = Ber(q_i),$$

we have the identity $\Pi_{i=1}^{l}(1-q_i)^{r_i} = 1-\theta$, which yields the constraint

$$\sum_{i=1}^{l} r_i q_i \geq \theta. \tag{65}$$

Define

$$q_1 = q_2 = \frac{3.05 \cdot 512}{10902}\theta, \qquad q = q_i = 1 - \left(\frac{1-\theta}{(1-q_1)(1-q_2)}\right)^{\frac{1}{r_3+\cdots+r_l}}, \; i \geq 3.$$

By (64), condition (60) is satisfied (using also $\tilde{M}(\frac{643}{10902\omega}\beta_1\theta) = \tilde{M}(\frac{643}{3.05\cdot 256\omega}\beta_1 q_1/2) \geq \tilde{M}(\beta_1 q_1/2)$). By (65), since $r_1 = r_2 = 1$ and assuming $l' \geq \sum_i r_i$, we have

$$(l'-2)q \geq \sum_{i=3}^{l} r_i q_i \geq \theta(1-2q_1) = \theta\left(1 - \frac{6.1\cdot 512}{10902}\right).$$

As a consequence, since

$$\begin{aligned} l' - 2 &= \frac{16}{7}\lceil\log_2\sqrt{|\Delta|\kappa_2} + 1\rceil + \frac{32}{49}(\omega + \log 6) - 2 \\ &\leq \frac{16}{7}\log_2\sqrt{|\Delta|\kappa_2} + \frac{32}{7} + \frac{32}{49}(\omega + \log 6) - 2 \\ &= \frac{8}{7}\log_2 e \log|\Delta|\kappa_2 + \frac{18}{7} + \frac{32}{49}(\omega + \log 6), \end{aligned}$$

by using (64) it is straightforward to check that condition (61) is satisfied as well. Here we have also used the fact that $\beta_i q_i/2 \geq \frac{643}{10902\omega}\beta_1\theta$ for $i \geq 3$.

We can now estimate the constant $Q = \sum_{i=1}^{l} q_i^{-1} \prod_{j=1}^{i-1} \alpha_j$. We have:

$$
\begin{aligned}
Q &= q_1^{-1} + q_2^{-1}\alpha_1 + \sum_{i=3}^{l} q^{-1}\alpha_1\alpha_2 \prod_{j=3}^{i-1} \alpha_j \\
&= q_1^{-1}\left(1 + \frac{1}{\sqrt{8\log(|\Delta|\kappa_2)}}\right) + \frac{q^{-1}}{8\log(|\Delta|\kappa_2)} \sum_{i=3}^{l} \frac{1}{2^{i-3}} \\
&\leq q_1^{-1}\left(1 + \frac{1}{\sqrt{8\log(|\Delta|\kappa_2)}}\right) + \frac{q^{-1}}{4\log(|\Delta|\kappa_2)} \\
&\leq q_1^{-1}\left(1 + \frac{1}{2\sqrt{2}}\right) + \frac{q^{-1}}{8\log(|\Delta|\kappa_2)} \\
&\leq C_1\theta^{-1} + C_2\theta^{-1}\frac{(l'-2)}{\log(|\Delta|\kappa_2)} \\
&\leq C''\theta^{-1}\omega,
\end{aligned}
$$

where we have used the fact that $|\Delta| \geq 3$, the definition of q_1, q_2 and the inequalities above involving $q, \theta, l' - 2$ (here C_1, C_2 and C'' are universal constants).

Finally, the union bound gives $p_3 + p_4 + p_5 \leq e^{-\omega}$, which finishes the proof of the proposition. □

5.4 Proof of Theorem 1

The proof is now immediate. By Proposition 11, under our assumptions with high probability there exists a dual certificate. Thus, by Proposition 10 we have

$$
\|g - g_0\| \leq 4(2 + \sqrt{\kappa_2})\|P_\Delta^\perp D g_0\|_{\ell^1} + \varepsilon\sqrt{\kappa_1}\frac{N}{m}\left(4 + (2 + \sqrt{\kappa_2})(1 + C''\omega\sqrt{\kappa_2 s})\right)
$$

for every $\Delta \subseteq \{1, \ldots, M\}$ such that $|\Delta| = s$. Observing that

$$
\begin{aligned}
\sigma_{s,M}(Dg) &= \inf\{\|x - Dg\|_{\ell^1} : \operatorname{supp}(x) \subseteq \{1, \ldots, M\}, |\operatorname{supp}(x)| \leq s\} \\
&= \inf\{\|x - Dg\|_{\ell^1} : \operatorname{supp}(x) \subseteq \Delta \subseteq \{1, \ldots, M\}, |\Delta| = s\} \\
&= \inf\{\|x - P_\Delta Dg\|_{\ell^1} + \left\|P_\Delta^\perp Dg\right\|_{\ell^1} : \operatorname{supp}(x) \subseteq \Delta \subseteq \{1, \ldots, M\}, |\Delta| = s\} \\
&= \inf\{\left\|P_\Delta^\perp Dg\right\|_{\ell^1} : \Delta \subseteq \{1, \ldots, M\}, |\Delta| = s\},
\end{aligned}
$$

and that

$$
4 + (2 + \sqrt{\kappa_2})(1 + C''\omega\sqrt{\kappa_2 s}) \leq C'''\sqrt{s}\kappa_2\omega
$$

for some absolute constant $C''' > 0$, gives the desired estimate.

Acknowledgments

The authors express their sincere thanks to Miren Zubeldia for having largely contributed to this project with many helpful comments and useful insights.

They also thank Maximilian März for suggesting an important simplification in the minimization problem. M.S. also thanks Eric Bonnetier for stimulating discussions about electrical impedance tomography and compressed sensing that have influenced the present work. G.S.A. acknowledges support from the ETH Zürich Postdoctoral Fellowship Program as well as from the Marie Curie Actions for People COFUND Program. The first conversations about this project took place during the trimester *Program on Inverse Problems*, held at the Institut Henri Poincaré of Paris in 2015: the authors would like to thank the institute and the organizers of the program for the wonderful opportunity offered.

References

[1] Sebastián Acosta and Carlos Montalto. Multiwave imaging in an enclosure with variable wave speed. *Inverse Problems*, 31(6):065009, 12, 2015.

[2] Ben Adcock, Milana Gataric, and Anders Hansen. On stable reconstructions from nonuniform Fourier measurements. *SIAM J. Imaging Sci.*, 7(3):1690–1723, 2014.

[3] Ben Adcock, Milana Gataric, and Anders C. Hansen. Weighted frames of exponentials and stable recovery of multidimensional functions from nonuniform Fourier samples. *Appl. Comput. Harmon. Anal.*, 42(3):508–535, 2017.

[4] Ben Adcock and Anders C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, pages 1–61, 2015.

[5] Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: a new theory for compressed sensing. *Forum Math. Sigma*, 5:e4, 84, 2017.

[6] G. S. Alberti and Y. Capdeboscq. *Lectures on elliptic methods for hybrid inverse problems.* Cours Spécialisés [Specialized Courses]. Société Mathématique de France, Paris, to appear.

[7] Giovanni Alessandrini. Stable determination of conductivity by boundary measurements. *Applicable Analysis*, 27(1-3):153–172, 1988.

[8] Giovanni Alessandrini and Sergio Vessella. Lipschitz stability for the inverse conductivity problem. *Adv. in Appl. Math.*, 35(2):207–241, 2005.

[9] Habib Ammari, Emmanuel Bossy, Vincent Jugnon, and Hyeonbae Kang. Mathematical modeling in photoacoustic imaging of small absorbers. *SIAM Rev.*, 52(4):677–695, 2010.

[10] Claude Bardos, Gilles Lebeau, and Jeffrey Rauch. Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary. *SIAM J. Control Optim.*, 30(5):1024–1065, 1992.

[11] John J. Benedetto and Hui C. Wu. Nonuniform sampling and spiral mri reconstruction, 2000.

[12] Elena Beretta and Elisa Francini. Lipschitz stability for the electrical impedance tomography problem: the complex case. *Comm. Partial Differential Equations*, 36(10):1723–1749, 2011.

[13] Arne Beurling. Local harmonic analysis with some applications to differential operators. pages 109–125, 1966.

[14] Jérémie Bigot, Claire Boyer, and Pierre Weiss. An analysis of block sampling strategies in compressed sensing. *IEEE Trans. Inform. Theory*, 62(4):2125–2139, 2016.

[15] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[16] Alexander L Bukhgeim. Recovering a potential from cauchy data in the two-dimensional case. *Journal of Inverse and Ill-posed Problems jiip*, 16(1):19–33, 2008.

[17] P. Burgholzer, C. Hofer, G. Paltauf, M. Haltmeier, and O. Scherzer. Thermoacoustic tomography with integrating area and line detectors. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 52(9):1577–1583, Sept 2005.

[18] A.-P. Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980)*, pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

[19] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969, 2007.

[20] Emmanuel J. Candès and David L. Donoho. Ridgelets: a key to higher-dimensional intermittency? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 357(1760):2495–2509, 1999.

[21] Emmanuel J. Candès and David L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[22] Emmanuel J. Candès and Yaniv Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 57(11):7235–7254, 2011.

[23] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[24] Peter G. Casazza. The art of frame theory. *Taiwanese J. Math.*, 4(2):129–201, 2000.

[25] Olga Chervova and Lauri Oksanen. Time reversal method with stabilizing boundary conditions for photoacoustic tomography. *Inverse Problems*, 32(12):125004, 16, 2016.

[26] Ole Christensen. Frames, Riesz bases, and discrete Gabor/wavelet expansions. *Bull. Amer. Math. Soc. (N.S.)*, 38(3):273–291 (electronic), 2001.

[27] Ole Christensen. *An introduction to frames and Riesz bases*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], second edition, 2016.

[28] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[29] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007.

[30] Sylvain Ervedoza and Enrique Zuazua. The wave equation: control and numerics. In *Control of partial differential equations*, volume 2048 of *Lecture Notes in Math.*, pages 245–339. Springer, Heidelberg, 2012.

[31] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[32] J. E. Fowler. The redundant discrete wavelet transform and additive noise. *IEEE Signal Processing Letters*, 12(9):629–632, Sept 2005.

[33] Avner Friedman and Victor Isakov. On the uniqueness in the inverse conductivity problem with one measurement. *Indiana University Mathematics Journal*, 38(3):563–579, 1989.

[34] Milana Gataric and Clarice Poon. A practical guide to the recovery of wavelet coefficients from Fourier measurements. *SIAM J. Sci. Comput.* ., 2016.

[35] Raja Giryes. Sampling in the analysis transform domain. *Applied and Computational Harmonic Analysis*, 40(1):172–185, 2016.

[36] D. Gross, F. Krahmer, and R. Kueng. A partial derandomization of PhaseLift using spherical designs. *J. Fourier Anal. Appl.*, 21(2):229–266, 2015.

[37] David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, 57(3):1548–1566, 2011.

[38] Markus Haltmeier, Thomas Berer, Sunghwan Moon, and Peter Burgholzer. Compressed sensing and sparsity in photoacoustic tomography. *Journal of Optics*, 18(11):114004, 2016.

[39] Lop Fat Ho. Observabilité frontière de l'équation des ondes. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(12):443–446, 1986.

[40] B. Holman and L. Kunyansky. Gradual time reversal in thermo- and photo-acoustic tomography within a resonant cavity. *Inverse Problems*, 31(3):035008, 25, 2015.

[41] A. Jones, B. Adcock, and A. Hansen. Analyzing the structure of multidimensional compressed sensing problems through coherence. *ArXiv e-prints*, October 2016.

[42] Alexander Daniel Jones, Ben Adcock, and Anders C Hansen. On asymptotic incoherence and its implications for compressed sensing of inverse problems. *IEEE Transactions on Information Theory*, 62(2):1020–1037, 2016.

[43] Pisamai Kittipoom, Gitta Kutyniok, and Wang-Q Lim. Construction of compactly supported shearlet frames. *Constr. Approx.*, 35(1):21–72, 2012.

[44] Felix Krahmer, Deanna Needell, and Rachel Ward. Compressive sensing with redundant dictionaries and structured measurements. *SIAM Journal on Mathematical Analysis*, 47(6):4606–4629, 2015.

[45] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Process.*, 23(2):612–622, 2014.

[46] Peter Kuchment and Leonid Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 1117–1167. Springer, New York, 2015.

[47] R. Kueng and D. Gross. RIPless compressed sensing from anisotropic measurements. *Linear Algebra Appl.*, 441:110–123, 2014.

[48] L. Kunyansky, B. Holman, and B. T. Cox. Photoacoustic tomography in a rectangular reflecting cavity. *Inverse Problems*, 29(12):125010, 20, 2013.

[49] Gitta Kutyniok and Wang-Q Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

[50] Demetrio Labate, Wang-Q Lim, Gitta Kutyniok, and Guido Weiss. Sparse multidimensional representation using shearlets. *Proc. SPIE*, 5914:59140U–59140U–9, 2005.

[51] Irena Lasiecka and Roberto Triggiani. *Control theory for partial differential equations: continuous and approximation theories. II*, volume 75 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2000. Abstract hyperbolic-like systems over a finite time horizon.

[52] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[53] J.-L. Lions. *Contrôlabilité exacte, perturbations et stabilisation de systèmes distribués. Tome 1*, volume 8 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1988. Contrôlabilité exacte. [Exact controllability], With appendices by E. Zuazua, C. Bardos, G. Lebeau and J. Rauch.

[54] J.-L. Lions. Exact controllability, stabilization and perturbations for distributed systems. *SIAM Rev.*, 30(1):1–68, 1988.

[55] Yulong Liu, Tiebin Mi, and Shidong Li. Compressed sensing with general frames via optimal-dual-based ℓ^1-analysis. *IEEE Transactions on information theory*, 58(7):4201–4214, 2012.

[56] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[57] A. I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Annals of Mathematics*, 143:71–96, 1996.

[58] R.G. Novikov. A multidimensional inverse spectral problem for the equation $-\Delta\psi + (v(x) - Eu(x))\psi = 0$. *Functional Analysis and Its Applications*, 22(4):263–272, 1988.

[59] Alexander Olevskii and Alexander Ulanovskii. On multi-dimensional sampling and interpolation. *Anal. Math. Phys.*, 2(2):149–170, 2012.

[60] Clarice Poon. Structure dependent sampling in compressed sensing: Theoretical guarantees for tight frames. *Appl. Comput. Harmon. Anal.*, 42(3):402–451, 2017.

[61] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Signal Processing Letters*, 18(10):595–598, Oct 2011.

[62] M. Sandbichler, F. Krahmer, T. Berer, P. Burgholzer, and M. Haltmeier. A novel compressed sensing scheme for photoacoustic tomography. *SIAM J. Appl. Math.*, 75(6):2475–2494, 2015.

[63] J. Sylvester and G. Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Annals of Mathematics*, 125:153–169, 1987.

[64] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

[65] Yaakov Tsaig and David L. Donoho. Extensions of compressed sensing. *Signal Processing*, 86(3):549 – 571, 2006. Sparse Approximations in Signal and Image Processing.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

54/2017 Dede', L; Quarteroni, A.
Isogeometric Analysis of a Phase Field Model for Darcy Flows with Discontinuous Data

53/2017 Bertagna, L.; Deparis, S.;Formaggia, L.; Forti, D.; Veneziani A.
The LifeV library: engineering mathematics beyond the proof of concept

55/2017 Agosti, A.; Cattaneo, C.; Giverso, C.; Ambrosi, D.; Ciarletta, P.
A computational platform for the personalized clinical treatment of glioblastoma multiforme

50/2017 Formaggia, F.; Vergara, C.
Defective boundary conditions for PDEs with applications in haemodynamics

51/2017 Gerbi, A.; Dede', L.; Quarteroni, A.
A monolithic algorithm for the simulation of cardiac electromechanics in the human left ventricle

52/2017 Beretta, E.; Ratti, L.; Verani, M.
A phase-field approach for the interface reconstruction in a nonlinear elliptic problem arising from cardiac electrophysiology

49/2017 Antonietti, P. F.,; Pennesi, G.
V-cycle multigrid algorithms for discontinuous Galerkin methods on non-nested polytopic meshes

48/2017 Regazzoni, F.; Dedè, L.; Quarteroni, A.
Active contraction of cardiac cells: a model for sarcomere dynamics with cooperative interactions

47/2017 Menghini, F.; Dede, L.; Forti, D.; Quarteroni, A.
Hemodynamics in a left atrium based on a Variational Multiscale-LES numerical model

46/2017 Agosti, A.; Gower, A.L.; Ciarletta, P.
The constitutive relations of initially stressed incompressible Mooney-Rivlin materials