



# Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all

Yann Traonmilin, Rémi Gribonval

## ► To cite this version:

Yann Traonmilin, Rémi Gribonval. Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all. *Applied and Computational Harmonic Analysis*, 2018, 45 (1), pp.170–205. 10.1016/j.acha.2016.08.004 . hal-01207987v5

**HAL Id: hal-01207987**

**<https://inria.hal.science/hal-01207987v5>**

Submitted on 5 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all

Yann Traonmilin<sup>a,\*</sup>, Rémi Gribonval<sup>a</sup>

<sup>a</sup>*INRIA Rennes - Bretagne Atlantique,  
Campus de Beaulieu  
35042 Rennes Cedex, France*

---

## Abstract

Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low-dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the 1-norm or low rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). In this paper, we establish generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. These guarantees are illustrated on selected examples. For block structured sparsity in the infinite-dimensional setting, we use the guarantees for a family of regularizers which efficiency in terms of RIP constant can be controlled, leading to stronger and sharper guarantees than the state of the art.

*Keywords:* sparse recovery; low dimensional models; restricted isometry; structured sparsity; regularization

---

This work was partly funded by the European Research Council, PLEASE project (ERC-StG-2011-277906).

## 1. Introduction

Linear inverse problems have become ubiquitous in signal processing and machine learning, in particular in the most extreme form of underdetermined linear inverse problems. In such settings, where some unknown

---

\*Corresponding author  
Email address: [yann.traonmilin@gmail.com](mailto:yann.traonmilin@gmail.com) (Yann Traonmilin)

data is to be estimated from underdetermined observations, a low-dimensional model is needed to complete these observations with some prior knowledge on the “regularity” of the unknown. Sparse and low-rank models have become particularly popular and provide a powerful theoretical and algorithmic framework to perform this estimation when the unknown is well approximated by a sparse vector or a low-rank matrix.

*Low-dimensional model sets.* A series of recent advances in low-dimensional regularization of inverse problems have established that stable reconstruction guarantees are also possible with more general low complexity models (beyond sparsity and low-rank) which provide the flexibility to model many types of data. Consider underdetermined observations  $y = Mx$  of a signal  $x \in \mathcal{H}$  ( $\mathcal{H}$  is a Hilbert space, finite or infinite-dimensional;  $x$  may be a vector, a matrix, a tensor, a function ...) by a linear operator  $M$  taking values in another Hilbert space  $\mathcal{F}$ . The “regularity” or “low-complexity” of  $x$  can be defined by the fact that  $x \in \Sigma$ , where  $\Sigma \subset \mathcal{H}$  is a *model set*. The two most traditional model sets encoding the low complexity of  $x$  are the set of sparse vectors with at most  $K$  non-zero coefficients, and the set of low-rank matrices with at most  $r$  non-zero singular values. Many more model sets incorporating various other structures have been considered, most of them being unions of low-dimensional subspaces [13, 27].

This paper focuses on general model sets with a special emphasis on models that are either *homogeneous* ( $t \cdot z \in \Sigma$  for any  $t \in \mathbb{R}$  and  $z \in \Sigma$ ) or *positively homogeneous* ( $t \cdot z \in \Sigma$  for any  $t \geq 0$  and  $z \in \Sigma$ ). Homogeneous model sets are in fact unions of (possibly infinitely many) subspaces, and will be referred to as *UoS*. Positively homogeneous models sets are called *cones* and cover, e.g., the set of low-rank positive semi-definite matrices, the positive orthant, etc..

*Uniform recovery guarantees and the RIP.* Given a model set  $\Sigma$ , a widely studied fundamental question is the construction of decoders (i.e., functions taking the observations  $y$  as input and yielding an estimation of  $x$  as output) that recovers any  $x \in \Sigma$  from observations  $y = Mx$  in the noiseless case, with stability guarantees in the presence of noise, and robustness when  $x$  is only approximately in  $\Sigma$ . Such decoders are called *instance-optimal*, see Section 3.2 for details. Many efficient instance-optimal algorithms have been exhibited for classical sparse/low-rank model sets under appropriate restricted isometry properties (RIP). One can extend the notion of RIP of a linear operator  $M$  to a general model set  $\Sigma$  as follows (the set  $\Sigma - \Sigma := \{x - x' : x, x' \in \Sigma\}$  is called the *secant set* of  $\Sigma$ ):

**Definition 1.1** (RIP).  *$M : \mathcal{H} \rightarrow \mathcal{F}$  has the RIP on  $\Sigma - \Sigma \subset \mathcal{H}$  with constant  $\delta$  if for all  $x \in \Sigma - \Sigma$ :*

$$(1 - \delta)\|x\|_{\mathcal{H}}^2 \leq \|Mx\|_{\mathcal{F}}^2 \leq (1 + \delta)\|x\|_{\mathcal{H}}^2 \quad (1)$$

where  $\|\cdot\|_{\mathcal{H}}$  and  $\|\cdot\|_{\mathcal{F}}$  are the Euclidean norms of  $\mathcal{H}$  and  $\mathcal{F}$ .

It is now well understood [15] that a lower RIP (the left hand side inequality in (1)) is necessary for the existence of instance-optimal decoders with respect to a given arbitrary model set  $\Sigma$  in an arbitrary Banach space, and that a RIP is also sufficient for the existence of a decoder. However, such a decoder may not be easily computable. Further, the existence of linear operators  $M : \mathcal{H} \rightarrow \mathbb{R}^m$  satisfying the RIP on  $\Sigma - \Sigma$  with an arbitrary model set  $\Sigma \subset \mathcal{H}$  has been established with general probabilistic arguments:  $m$  need only be of the order of the *covering dimension of the normalized secant set* of  $\Sigma$ , which may be finite even when  $\mathcal{H}$  is infinite-dimensional [25, 23, 33].

*Decoding by constrained minimization.* For a selection of model sets  $\Sigma$  and (often convex) regularizers  $f$ , a huge body of work [24, 20, 34, 19, 21, 29] has established that a RIP with small enough  $\delta$  implies exact recovery guarantees for the particular decoder corresponding to the (convex) optimization program

$$\operatorname{argmin}_{x \in \mathcal{H}} f(x) \text{ s.t. } Mx = Mx_0. \quad (2)$$

In the presence of bounded noise  $e$  ( $\|e\|_{\mathcal{F}} \leq \eta$ ), decoders of the following form were shown to be instance-optimal:

$$\operatorname{argmin}_{x \in \mathcal{H}} f(x) \text{ s.t. } \|Mx - (Mx_0 + e)\|_{\mathcal{F}} \leq \epsilon. \quad (3)$$

A celebrated example [17] in  $\mathcal{H} = \mathbb{R}^n$  is the combination of the  $\ell^1$ -norm decoder ( $f(x) := \|x\|_1$ ) with Gaussian random projections ( $M$  is  $m \times n$  with i.i.d. Gaussian entries): when  $m \gtrsim K \log(n/K)$ , the matrix  $M$  satisfies a RIP on the set  $\Sigma - \Sigma$  of  $2K$ -sparse vectors that is sufficient to guarantee that  $\ell^1$  minimization is instance-optimal on  $K$ -sparse vectors. Most RIP results aim at either generalizing the area of application [12, 18] or the sharpness of hypotheses. For example the classical sufficient RIP condition  $\delta < \sqrt{2} - 1 \approx 0.414$  [17] for  $\ell^1$  minimization was improved by several authors (see e.g. historical notes in [29, Chapter 6]). Cai and Zhang [16] closed this game for sparse vectors and low rank matrices by establishing recovery guarantees when  $\delta < 1/\sqrt{2} \approx 0.707$ , a result which is sharp in light of the work of Davies and Gribonval [22].

*Convex regularization.* Convex regularizers are of practical interest. The decoder (3) is indeed convex when  $f$  is convex and most particularly when  $f$  is a norm. Chandrasekaran et al. [21] gave general results for the recovery of sparse signals under random Gaussian observations (compressed sensing, random  $M$ ) with atomic norms. These results give a unified understanding of a wide variety of norms used for low

complexity recovery :  $\ell^1$ -norm, nuclear norm ... They give a way to calculate *non-uniform* recovery results by studying the relation between the geometry of the atomic norm and random observations. These results were extended to understand the phenomenon of phase transition [7] in compressed sensing. However such generic results are not available for *uniform* recovery.

*General objective.* These observations lead us to the general objective of this paper. Our objective is to address the following question: *given a general model set  $\Sigma$  and a regularizer  $f$ , what RIP condition is sufficient to ensure that solving the optimization problem (3) yields an instance-optimal decoder?*

We summarize how we answer this question in this article in the following Section 1.1.

### 1.1. Contributions and outline

*Admissible RIP constants for stable recovery with an arbitrary regularizer  $f$ .* In this article, we give an admissible RIP constant  $\delta_\Sigma(f)$  on  $\Sigma - \Sigma$  that guarantees stable recovery of elements of  $\Sigma$  with a regularizer  $f$ . As the notation suggests, the constant  $\delta_\Sigma(f)$  only depends on the model  $\Sigma$  and the regularizer  $f$ . It is related to the geometry of the model set  $\Sigma$  through the *atomic norm*  $\|\cdot\|_\Sigma$  (a function which construction depends only on the model  $\Sigma$ , see Section 2.2) and to the geometry of the descent vectors of  $f$  at points of  $\Sigma$  (which union is denoted  $\mathcal{T}_f(\Sigma)$ , its precise definition is given in Section 2). This constant is expressed as

$$\delta_\Sigma(f) := \inf_{z \in \mathcal{T}_f(\Sigma) \setminus \{0\}} \delta_\Sigma(z) \quad (4)$$

where  $\delta_\Sigma(z) := \sup_{x \in \Sigma} \delta_\Sigma(x, z)$ . Depending whether  $\Sigma$  is a UoS or a cone:

$$\delta_\Sigma(x, z) = \delta_\Sigma^{UoS}(x, z) \quad := \quad \frac{-\mathcal{R}e\langle x, z \rangle}{\|x\|_{\mathcal{H}} \sqrt{\|x + z\|_\Sigma^2 - \|x\|_{\mathcal{H}}^2} - 2\mathcal{R}e\langle x, z \rangle} \quad (5)$$

$$\delta_\Sigma(x, z) = \delta_\Sigma^{cone}(x, z) \quad := \quad \frac{-2\mathcal{R}e\langle x, z \rangle}{\|x + z\|_\Sigma^2 - 2\mathcal{R}e\langle x, z \rangle}. \quad (6)$$

We will show that in most conventional examples these constants can be bounded by estimates that are sharp with respect to a collection of models (e.g.  $\delta_\Sigma(f) \geq 1/\sqrt{2}$  for sparse vectors and  $\ell^1$ -norm). Moreover, we can derive better than state-of-the art constants in other cases (block structured sparsity). For the cone case, we actually characterize a sharper constant with a more complicated expression, which will be exhibited precisely in Theorem 3.1.

Section 3 is dedicated to the proof of our main result. This result builds on the sharp RIP condition of Cai and Zhang for stable recovery of sparse vectors (respectively low rank matrices) by  $\ell^1$  norm (respectively nuclear norm) minimization, establishing that it can be extended to this general setting.

**Theorem 1.1** (RIP condition for exact recovery). *Assume that  $\Sigma$  is either a UoS or a cone. Then, for any  $M$  continuous on  $\mathcal{H}$  that satisfies the RIP on  $\Sigma - \Sigma$  with constant  $\delta < \delta_\Sigma(f)$  we have: for all  $x_0 \in \Sigma$ ,  $x_0$  is the unique minimizer of (2).*

When  $\Sigma$  is not a cone (consider, e.g., a point cloud, see Section 5), we can consider the cone generated by  $\Sigma$ ,  $\Sigma' := \mathbb{R}_+\Sigma$  instead. Since  $\Sigma' \supset \Sigma$  is a cone, a RIP hypothesis on  $\Sigma' - \Sigma'$  allows to apply the above Theorem and to provide recovery guarantees on  $\Sigma$ .

A stable recovery formulation for minimization (3) is also established.

**Theorem 1.2** (RIP condition for stable recovery). *Assume that  $\Sigma$  is either a UoS or a cone. Then, for any  $M$  continuous on  $\mathcal{H}$  that satisfies the RIP on  $\Sigma - \Sigma$  with constant  $\delta < \delta_\Sigma(f)$  we have: for all  $x_0 \in \Sigma$ ,  $\|e\|_{\mathcal{F}} \leq \eta \leq \epsilon$ , with  $x^*$  the result of minimization (3),*

$$\|x^* - x_0\|_{\mathcal{H}} \leq C_\Sigma(f, \delta) \cdot (\eta + \epsilon) \quad (7)$$

where  $C_\Sigma(f, \delta) < +\infty$ .

The constant  $C_\Sigma(f, \delta)$  is explicit in most classical examples (see also Section 4 for an apparently new explicit value for block structured sparsity). These two theorems are direct consequences of the more technical Theorem 3.1 proven in Section 3. As described in Section 3.2, the hypotheses of Theorem 1.2 also imply robustness to modeling error (stable and robust instance optimality) for positively homogeneous, non-negative, convex functions  $f$ . In these cases, the level of modeling error is not needed to set the parameter  $\epsilon$ . Such knowledge would be needed if the modeling error was considered as an observation noise. We will show on various classical examples how the additional hypotheses on  $f$  are met.

*Low complexity compressed sensing framework.* These results give a general framework to check if a given regularizer  $f$  is a good candidate to perform low complexity recovery and get a compressed sensing result when combining with the work of Dirksen et al. [23] and Puy et al. [33]. Given a UoS (or a cone)  $\Sigma$  and a function  $f$ , we can get a uniform recovery result for random observations of arbitrary signals in  $\Sigma$  by establishing that (with high probability) a random linear measurement operator satisfies a RIP on the secant set  $\Sigma - \Sigma$ .

Let  $\Sigma$  be a model set and  $f$  a candidate as a regularizer. We can follow these steps to get a potential low complexity compressed sensing result :

1. Calculate (or bound)  $\delta_\Sigma(f)$  and  $C_\Sigma(f, \delta)$ .
2. Construct a dimension-reducing matrix  $M$  satisfying the desired RIP on  $\Sigma - \Sigma$ , e.g., by applying techniques from [23] and [33] involving random subgaussian matrices.
3. Apply Theorem 1.2 to conclude that regularization with  $f$  leads to stable recovery of vectors from  $\Sigma$  from their noisy low-dimensional linear measurements with  $M$ .

This framework is illustrated on a few examples.

- **Block Structured Sparsity in Infinite dimension.** We use our theorems in Section 4 to give uniform recovery results under RIP conditions for the set  $\Sigma$  of block group sparse vectors, using regularization with structured sparsity inducing norms (group norms, without overlap between groups).

In a finite-dimensional setting, Ayaz et al. [9] established uniform recovery guarantees for compressed sensing of such structured sparse vectors under a RIP hypothesis. In this case, the regularizer is a mixed  $\ell^1 - \ell^2$  norm (also called group norm, it can be interpreted as an atomic norm, see Section 2.2.1). Given any matrix  $M$  with a RIP constant  $\delta < \sqrt{2} - 1$  for vectors in  $\Sigma - \Sigma$ , this regularizer recovers vectors from  $\Sigma$ .

Adcock and Hansen [2] propose a generalized sampling strategy to recover (structured) sparse signals in an infinite-dimensional setting. This setting is important to model the acquisition of analog signals, e.g. in magnetic resonance imaging (MRI). Bastounis et al. [11] showed that the RIP constant must necessarily depend on the ratio of sparsity in each block and the number of blocks for the  $\ell^1$  norm and block sparsity.

With our new general framework:

- we improve the RIP constant on group sparse vectors of Ayaz et al. [9] to the sharp  $\frac{1}{\sqrt{2}}$ ;
- we generalize the results of [9] to (non-overlapping) group norms and block sparsity in an *infinite-dimensional context* which encompasses the generalized sampling of Adcock and Hansen;
- we exhibit a particular weighting of the group norm that removes the dependency of admissible RIP constant on the ratio of sparsity between blocks.

- **Other models and regularizers** In Section 5, we show how some other classical examples can be handled by this framework (including *low rank matrix recovery*, where our results match the sharp RIP results of Cai and Zhang [16] on which they are built).

The theoretical derivations in this article highlight the role of a particular *atomic norm*  $\|\cdot\|_\Sigma$  constructed directly from the model  $\Sigma$  (the definition and properties of this norm can be found in Section 2). This norm can be found in the literature [8, 35] for some specific models but cannot be directly used as a regularizer with uniform recovery guarantees on  $\Sigma$  in general. In Section 6, we study  $f = \|\cdot\|_\Sigma$  as a candidate for regularization:

- we first show that this norm enables stable recovery under a RIP assumption with constant  $\delta < \frac{2(1-\mu(\Sigma))}{3+2\mu(\Sigma)}$  when the cone model  $\Sigma$  is an *arbitrary finite union of one-dimensional half-spaces* with coherence parameter  $\mu(\Sigma)$  (see Section 6.1).
- we further show that the set  $\Sigma$  of (non-negative multiples of) permutation matrices can be stably recovered with  $O(n \log(n))$  subgaussian linear measurements by using the regularizer  $\|\cdot\|_\Sigma$ , the norm induced by the Birkhoff polytope of bi-stochastic matrices. This is the consequence of a sufficient RIP constant  $\delta < \frac{2}{3}$  for recovery. Similar results were established regarding *non-uniform*<sup>1</sup> recovery of permutation matrices [21] with the Gaussian ensemble, with applications to the recovery of rankings from low-dimensional projections. Here we obtain *uniform*<sup>2</sup> recovery guarantees with qualitatively as many linear measurements.
- however, this norm cannot perform stable recovery in the general case. We show that the set  $\Sigma$  of  $K$ -sparse vectors *cannot be uniformly recovered* by regularization with  $\|\cdot\|_\Sigma$  (called the  $K$ -support norm in [8]).

Finally, we discuss these results and conclude in Section 7, especially, future directions for the design of regularizers.

## 2. Preliminaries and definitions

We consider a real or complex Hilbert space  $\mathcal{H}$  equipped with a Hermitian inner product  $\langle \cdot, \cdot \rangle$  and a Euclidean norm  $\|x\|_{\mathcal{H}}^2 := \langle x, x \rangle$ . The Pythagorean property  $\|x + x'\|_{\mathcal{H}}^2 = \|x\|_{\mathcal{H}}^2 + 2\operatorname{Re}\langle x, x' \rangle + \|x'\|_{\mathcal{H}}^2$  will play

<sup>1</sup>For every given matrix  $x \in \Sigma$ , the probability of drawing a (Gaussian) matrix  $M$  yielding exact recovery of this particular  $x$  is high.

<sup>2</sup>The probability of drawing a (subgaussian) matrix  $M$  yielding exact recovery of *every*  $x \in \Sigma$  is high.



a major role in the proofs. Convergence of sequences, of sums, closure of sets and continuity of functions are all defined with respect to  $\|\cdot\|_{\mathcal{H}}$ . For  $\gamma \in \mathbb{R}_+ = \{a \in \mathbb{R} : a \geq 0\}$ , we define the sphere and the closed ball of radius  $\gamma$ ,

$$\begin{aligned}\mathcal{S}(\gamma) &= \{x \in \mathcal{H} : \|x\|_{\mathcal{H}} = \gamma\}, \\ \mathcal{B}(\gamma) &= \{x \in \mathcal{H} : \|x\|_{\mathcal{H}} \leq \gamma\}.\end{aligned}\tag{8}$$

The model set is some subset  $\Sigma \subset \mathcal{H}$ . The operator  $M$  is a linear map from  $\mathcal{H}$  to a Hilbert space  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ , thus  $\|\cdot\|_{\mathcal{F}}$  is as well Euclidean. Its domain  $\mathcal{H}_M := \text{dom}(M) \subset \mathcal{H}$  may be strictly smaller than  $\mathcal{H}$ .

The regularizer  $f$  is a generalized function from  $\mathcal{H}$  to  $\mathbb{R} \cup \{+\infty\}$  which is finite on the set  $\mathcal{H}_f := \text{dom}(f) \subset \mathcal{H}$ , and we assume that  $\Sigma \subset \mathcal{H}_f \cap \mathcal{H}_M$  so that for  $x \in \Sigma$ ,  $f(x) < \infty$  and  $Mx$  is well-defined.

### 2.1. Descent sets

The notion of descent vector plays an important role in the analysis conducted in this paper.

**Definition 2.1** (Descent vectors). *For any  $x \in \mathcal{H}$ , the collection of descent vectors of  $f$  at  $x$  is*

$$\mathcal{T}_f(x) := \{z \in \mathcal{H} : f(x+z) \leq f(x)\}\tag{9}$$

In the noiseless case, the main question investigated in this paper is whether for all  $x \in \Sigma$ ,

$$\{x\} = \underset{\tilde{x} \in \mathcal{H}_M}{\text{argmin}} f(\tilde{x}) \text{ s.t. } M\tilde{x} = Mx,\tag{10}$$

i.e., the minimizer is unique and matches  $x$ .

This is easily seen to be equivalent to  $\ker M \cap \mathcal{T}_f(x) = \{0\}$ , for all  $x \in \Sigma$ , where  $\ker M := \{z \in \mathcal{H}_M, Mz = 0\}$  is the null space of  $M$ . In other words, the property we wish to establish under a RIP assumption is

$$\ker M \cap \mathcal{T}_f(\Sigma) = \{0\}.\tag{11}$$

where as a shorthand the union of all descent vectors on points of the model  $\Sigma$ , later also called descent set for brevity, is denoted

$$\mathcal{T}_f(\Sigma) := \bigcup_{x \in \Sigma} \mathcal{T}_f(x).\tag{12}$$

## 2.2. Atomic norms

A particular family of *convex* regularizers  $f$  is defined using the notion of *atomic “norm”* [21] (which in fact is not always a norm). Considering a set  $\mathcal{A} \subset \mathcal{H}$ , commonly called the set of *atoms*, the corresponding *atomic “norm”* is built using the convex hull of  $\mathcal{A}$ .

**Definition 2.2** (Convex hull). *The convex hull of a set  $\mathcal{A}$  is :*

$$\text{conv}(\mathcal{A}) := \left\{ x = \sum c_i a_i : a_i \in \mathcal{A}, c_i \in \mathbb{R}_+, \sum c_i = 1 \right\} \quad (13)$$

**Definition 2.3** (Atomic norm). *The atomic “norm” induced by the set  $\mathcal{A}$  is defined as:*

$$\|x\|_{\mathcal{A}} := \inf \{ t \in \mathbb{R}_+ : x \in t \cdot \overline{\text{conv}}(\mathcal{A}) \} \quad (14)$$

where  $\overline{\text{conv}}(\mathcal{A})$  is the closure of  $\text{conv}(\mathcal{A})$  in  $\mathcal{H}$ . This “norm” is finite only on

$$\mathcal{E}(\mathcal{A}) := \mathbb{R}_+ \cdot \overline{\text{conv}}(\mathcal{A}) = \{ x = t \cdot y, t \in \mathbb{R}_+, y \in \overline{\text{conv}}(\mathcal{A}) \} \subset \mathcal{H}. \quad (15)$$

It is extended to  $\mathcal{H}$  by setting  $\|x\|_{\mathcal{A}} = +\infty$  if  $x \notin \mathcal{E}(\mathcal{A})$ .

The main properties of  $\|\cdot\|_{\mathcal{A}}$  depend on the considered set  $\mathcal{A}$  through those of the convex set  $C := \overline{\text{conv}}(\mathcal{A})$ . In fact, the atomic “norm”  $\|\cdot\|_{\mathcal{A}}$  is the gauge function associated with the convex set  $C$  [38]. Table 1 summarizes the properties of such gauges. The proof of these results in infinite dimension can be found in [6], the finite dimension is covered by [37].

Moreover, from [14, 28],  $\mathcal{E}(\mathcal{A})$  is continuously embedded in  $\mathcal{H}$  if  $\mathcal{A}$  is *bounded* (because  $\mathcal{E}(\mathcal{A} \cup -\mathcal{A})$  is continuously embedded in  $\mathcal{H}$ ). Thus we will suppose  $\mathcal{A}$  bounded from now on.

By abuse of notation, we define  $\mathcal{T}_{\mathcal{A}}(x) := \mathcal{T}_{\|\cdot\|_{\mathcal{A}}}(x)$  and  $\mathcal{T}_{\mathcal{A}}(\Sigma) = \bigcup_{x \in \Sigma} \mathcal{T}_{\mathcal{A}}(x)$ .

### 2.2.1. Some well known atomic norms

As pointed out in [21], many well know norms used for low complexity recovery are atomic norms with atoms of interest. Atoms are often *normalized*. We say that a vector  $u$  is *normalized* if  $\|u\|_{\mathcal{H}} = 1$  :

- $\ell^1$  norm :  $\mathcal{A}$  is the set of canonical orthonormal basis vectors multiplied by a complex scalar with modulus 1, i.e. the normalized 1-sparse vectors.
- Nuclear norm :  $\mathcal{A}$  is the set of normalized rank one matrices (Section 5).

Hypotheses on $C := \overline{\text{conv}}(\mathcal{A})$	Properties of $\ \cdot\ _{\mathcal{A}}$ in $\mathcal{E}(\mathcal{A}) = \mathbb{R}_+ \cdot C$
No hypotheses	Non-negative, lower semi-continuous and sublinear
$0 \in C$	Positively homogeneous
$0 \in \text{int}C$	Continuous
$C$ is bounded (and closed by definition)	Coercive ( $\lim_{\ x\ _{\mathcal{H}} \rightarrow +\infty} \ x\ _{\mathcal{A}} = +\infty$ )
$-C = C$	Is a norm

Table 1: Link between  $\mathcal{A}$  and  $\|\cdot\|_{\mathcal{A}}$  using the properties of gauges.

- Mixed  $\ell^1 - \ell^2$  norms (covered extensively in Section 4) :  $\mathcal{A}$  is the set of normalized 1-group-sparse vectors.
- Gauge generated by a finite polytope :  $\mathcal{A}$  is composed of the vertices of a polytope (Section 6).
- Spectral norm :  $\mathcal{A}$  is the set of normalized orthogonal matrices.

### 2.2.2. Duality

The dual of an atomic norm, defined as

$$\|x\|_{\mathcal{A}}^* := \sup_{\|u\|_{\mathcal{A}}=1} |\langle x, u \rangle| \quad (16)$$

is

$$\|x\|_{\mathcal{A}}^* = \sup_{a \in \mathcal{A}} |\langle x, a \rangle|. \quad (17)$$

and satisfies the usual property  $\|x\|_{\mathcal{H}}^2 \leq \|x\|_{\mathcal{A}} \cdot \|x\|_{\mathcal{A}}^*$  for all  $x \in \mathcal{H}$ .

### 2.2.3. Alternate characterization of atomic norms with normalized atoms

The following bound, which is valid in particular when  $\mathcal{A}$  is made of *normalized atoms*, will be useful in the proof of our main results.

**Fact 2.1.** *Suppose  $\mathcal{A} \subset \mathcal{B}(1)$ . We define*

$$\|x\|'_{\mathcal{A}} := \inf \left\{ \sqrt{\sum \lambda_i \|u_i\|_{\mathcal{H}}^2} : \lambda_i \in \mathbb{R}_+, \sum \lambda_i = 1, u_i \in \mathbb{R}_+ \mathcal{A}, x = \sum \lambda_i u_i \right\}. \quad (18)$$

Then for any  $x \in \mathcal{E}(\mathcal{A})$  we have

$$\|x\|_{\mathcal{A}} \geq \lim_{\gamma \rightarrow 0_+} \inf\{\|\tilde{x}\|'_{\mathcal{A}} : \tilde{x} \in t \cdot \text{conv}(\mathcal{A}), \|\tilde{x} - x\|_{\mathcal{H}} \leq \gamma\}. \quad (19)$$

When  $\text{conv}(\mathcal{A})$  is closed we have in fact  $\|x\|_{\mathcal{A}} = \|x\|'_{\mathcal{A}}$  for any  $x$ .

*Proof.* By definition of  $\text{conv}(\mathcal{A})$ , if  $x \in t \cdot \text{conv}(\mathcal{A})$  there are vectors  $a_i \in \mathcal{A}$  and scalars  $\lambda_i \in \mathbb{R}_+$  such that  $x = t \sum \lambda_i a_i$  and  $\sum_i \lambda_i = 1$ . Denoting  $u_i := ta_i$ , we have  $x = \sum_i \lambda_i u_i$  and  $u_i \in \mathbb{R}_+ \mathcal{A}$  and we can indeed define  $\|x\|'_{\mathcal{A}}$ . Since  $\mathcal{A} \subset \mathcal{B}(1)$  we have  $\|u_i\|_{\mathcal{H}} = \|ta_i\|_{\mathcal{H}} \leq t$ , hence  $\sum \lambda_i \|u_i\|_{\mathcal{H}}^2 \leq t^2$ , yielding  $\|x\|'_{\mathcal{A}} \leq t$ . Finally, observe that  $u_i \in \mathbb{R}_+ \mathcal{A}$  implies  $\|u_i\|_{\mathcal{A}} \leq \|u_i\|_{\mathcal{H}}$ . Hence, by convexity of  $\|\cdot\|_{\mathcal{A}}^2$ , for any  $\lambda_i \in \mathbb{R}_+$  and  $u_i \in \mathbb{R}_+ \mathcal{A}$  such that  $\sum \lambda_i = 1$  and  $\sum \lambda_i u_i = x$  we thus have

$$\|x\|_{\mathcal{A}}^2 = \left\| \sum_i \lambda_i u_i \right\|_{\mathcal{A}}^2 \leq \sum_i \lambda_i \|u_i\|_{\mathcal{A}}^2 \leq \sum_i \lambda_i \|u_i\|_{\mathcal{H}}^2 \quad (20)$$

The infimum of the right side establishes that  $\|x\|_{\mathcal{A}} \leq \|x\|'_{\mathcal{A}}$ .

Consider now an arbitrary  $x \in \mathcal{H}$ , and  $t > 0$  such that  $x \in t \cdot \overline{\text{conv}}(\mathcal{A})$ . By definition of  $\overline{\text{conv}}(\mathcal{A})$ , for any  $\gamma > 0$  there is  $\tilde{x} \in t \cdot \text{conv}(\mathcal{A})$  such that  $\|x - \tilde{x}\|_{\mathcal{H}} \leq \gamma$ . As shown above, for any such  $\gamma$  and  $\tilde{x}$ , since  $\tilde{x} \in t \cdot \text{conv}(\mathcal{A})$  we have  $t \geq \|\tilde{x}\|'_{\mathcal{A}}$ , hence

$$t \geq \sup_{\gamma > 0} \inf\{\|\tilde{x}\|'_{\mathcal{A}} : \tilde{x} \in t \cdot \text{conv}(\mathcal{A}), \|\tilde{x} - x\|_{\mathcal{H}} \leq \gamma\} = \lim_{\gamma \rightarrow 0_+} \inf\{\|\tilde{x}\|'_{\mathcal{A}} : \tilde{x} \in t \cdot \text{conv}(\mathcal{A}), \|\tilde{x} - x\|_{\mathcal{H}} \leq \gamma\}. \quad (21)$$

This holds for any  $t > 0$  such that  $x \in t \cdot \overline{\text{conv}}(\mathcal{A})$ . We conclude by observing that  $\|x\|_{\mathcal{A}}$  is precisely defined as the infimum of such  $t$ .

When  $\text{conv}(\mathcal{A})$  is closed, the above argument simplifies since  $x \in t \cdot \overline{\text{conv}}(\mathcal{A})$  is equivalent to  $x \in t \cdot \text{conv}(\mathcal{A})$ , implying that  $\|x\|_{\mathcal{A}} \leq \|x\|'_{\mathcal{A}} \leq t$ . Taking the infimum over such  $t$  yields  $\|x\|_{\mathcal{A}} = \|x\|'_{\mathcal{A}}$ .  $\square$

#### 2.2.4. Atomic norms associated to cones

Given a cone  $\Sigma \subset \mathcal{H}$ , the norm associated to *normalized atoms*

$$\mathcal{A}(\Sigma) := \Sigma \cap \mathcal{S}(1) \quad (22)$$

will be of particular interest for the RIP analysis conducted in this paper. As a shorthand we define:

$$\|\cdot\|_{\Sigma} := \|\cdot\|_{\Sigma \cap \mathcal{S}(1)}. \quad (23)$$

From the general properties of atomic norms (Table 1) we have:

**Fact 2.2** (The “norm”  $\|\cdot\|_{\Sigma}$ ). *Given any cone  $\Sigma$  we have:*

1. *For all  $z \in \mathcal{H}$ ,  $\|z\|_{\Sigma} \geq \|z\|_{\mathcal{H}}$ .*
2. *For all  $x \in \Sigma$ ,  $\|x\|_{\Sigma} = \|x\|_{\mathcal{H}}$ .*
3.  *$\|\cdot\|_{\Sigma}$  is non-negative, lower semi-continuous, sublinear, coercive and positively homogeneous.*
4. *When  $\Sigma$  is homogeneous (i.e., a UoS),  $\|\cdot\|_{\Sigma}$  is indeed a norm (since  $\Sigma = -\Sigma$ ).*

### 2.3. Atomic norms are often sufficient

Ultimately, our aim is to find a convex regularizer that enables the uniform recovery of  $\Sigma$  for some linear observation operator  $M$ . We show below that if such a convex regularizer exists, there exists an atomic norm with atoms  $\mathcal{A} \subset \Sigma$  (independent of  $M$ ) that also ensures uniform recovery of  $\Sigma$ . This means that the search for a convex regularizer can be limited to the subset of such atomic norms. In fact, Lemma 2.1 below implies that we can restrict our study to such atomic norms, as soon as there exists some “well-behaved” regularizer  $f$  with uniform recovery guarantees.

**Lemma 2.1.** *Let  $\Sigma$  be a cone and  $f$  be a proper coercive continuous regularizer. Assume that for some  $t > f(0)$ , the level set  $\mathcal{L}(f, t) = \{y \in \mathcal{H} : f(y) \leq t\}$  is a convex set. Then there exists a family of atoms  $\mathcal{A} \subset \Sigma$  such that :*

$$\mathcal{T}_{\mathcal{A}}(\Sigma) \subset \mathbb{R}_+^* \cdot \mathcal{T}_f(\Sigma) \quad (24)$$

where  $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$ .

*Proof.* We define  $\mathcal{A} := \mathcal{L}(f, t) \cap \Sigma = \{x \in \Sigma : f(x) \leq t\}$ . Let  $z \in \mathcal{T}_{\mathcal{A}}(\Sigma)$ . By definition, there exists  $x \in \Sigma$  such that  $\|x + z\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}}$ . We have  $f(0 \cdot x) = f(0) < t$ , and since  $f$  is coercive  $f(\lambda x) \xrightarrow{\lambda \rightarrow +\infty} +\infty$ . Thus, by the continuity of  $f$  there is  $\lambda_0 > 0$  such that  $f(\lambda_0 x) = t$ . Since  $\Sigma$  is a cone, the vector  $x' = \lambda_0 x$  belongs to  $\Sigma$  and, since  $f(x') = t$ , by definition of  $\mathcal{A}$  we have indeed  $x' \in \mathcal{A}$  and  $\|x'\|_{\mathcal{A}} \leq 1$ . Furthermore, we have  $\|x' + \lambda_0 z\|_{\mathcal{A}} = \lambda_0 \|x + z\|_{\mathcal{A}} \leq \lambda_0 \|x\|_{\mathcal{A}} = \|x'\|_{\mathcal{A}}$ .

We now observe that, on the one hand, the level set  $\mathcal{L}(\|\cdot\|_{\mathcal{A}}, 1) = \overline{\text{conv}}(\mathcal{A})$  is the smallest closed convex set containing  $\mathcal{A}$ ; on the other hand  $\mathcal{A} \subset \mathcal{L}(f, t)$  and  $\mathcal{L}(f, t)$  is convex and closed (by the continuity of  $f$ ).

Thus  $\mathcal{L}(\|\cdot\|_{\mathcal{A}}, 1) \subset \mathcal{L}(f, t)$  and the fact that  $\|x' + \lambda_0 z\|_{\mathcal{A}} \leq \|x'\|_{\mathcal{A}} \leq 1$  implies

$$f(x' + \lambda_0 z) \leq t = f(x'). \quad (25)$$

This shows that  $\lambda_0 z \in \mathcal{T}_f(\Sigma)$  and finally that  $z \in \mathbb{R}_+^* \cdot \mathcal{T}_f(\Sigma)$ . □

**Corollary 2.1.** *With such a set of atoms, we have  $\delta_{\Sigma}(\|\cdot\|_{\mathcal{A}}) \geq \delta_{\Sigma}(f)$ .*

*Proof.* First remark that for all  $\lambda > 0$ ,  $\delta_{\Sigma}(\lambda z) = \delta_{\Sigma}(z)$  both in the UoS and cone cases. Thus, because  $\mathcal{T}_{\mathcal{A}}(\Sigma) \subset \mathbb{R}_+^* \cdot \mathcal{T}_f(\Sigma)$ , we have

$$\delta_{\Sigma}(\|\cdot\|_{\mathcal{A}}) = \inf_{z \in \mathcal{T}_{\mathcal{A}}(\Sigma) \setminus \{0\}} \delta_{\Sigma}(z) \geq \inf_{z \in \mathbb{R}_+^* \cdot \mathcal{T}_f(\Sigma)} \delta_{\Sigma}(z) = \inf_{z \in \mathcal{T}_f(\Sigma) \setminus \{0\}} \delta_{\Sigma}(z) = \delta_{\Sigma}(f) \quad (26)$$

□

In other words this corollary shows that given a sufficiently well behaved regularizer  $f$  (in terms of recovery guarantees Theorem 1.2 under a RIP assumption on  $\Sigma - \Sigma$ ), we can find an atomic norm that behaves at least as well, possibly with an even less restrictive RIP condition.

#### 2.4. Tools to calculate admissible RIP constants

In all examples, explicit bounds on  $\delta_{\Sigma}(f)$  are achieved by decomposing descent vectors  $z \in \mathcal{T}_f(\Sigma)$  as  $z = (x + z) - x$ , with  $x \in \Sigma$ . The “goodness” of such a decomposition is a trade-off between:

- the incoherence between the two components, measured by the *quasi-orthogonality constant*

$$\rho(x, z) := \frac{\operatorname{Re}\langle x, x + z \rangle}{\|x\|_{\mathcal{H}}^2}; \quad (27)$$

- the fact that  $z$  is (quasi) a descent vector for the  $\|\cdot\|_{\Sigma}$  norm at  $x \in \Sigma$ , as measured by the *quasi-descent constant*

$$\alpha_{\Sigma}(x, z) := \frac{\|x + z\|_{\Sigma}^2}{\|x\|_{\Sigma}^2}. \quad (28)$$

For most examples in Section 4 and Section 5, we bound  $\delta_{\Sigma}(f)$  by exhibiting for each descent vector  $z \in \mathcal{T}_f(\Sigma) \setminus \{0\}$  an element of the model,  $x \in \Sigma$ , such that  $\rho(x, z) = 0$  and  $\alpha_{\Sigma}(x, z) \leq \alpha_{\Sigma}(f)$ , where  $\alpha_{\Sigma}(f)$  is a constant that depends only on  $f$  and  $\Sigma$ . When this is possible we conclude that  $\delta_{\Sigma}(f) \geq 1/\sqrt{1 + \alpha_{\Sigma}(f)}$ .

### 3. Stable and robust recovery of cones under a RIP condition

This section is dedicated to the precise statement and proof of our main result. Applications of these results will be discussed in later sections.

#### 3.1. Exact recovery and stability to noise

The main part of the argument is a generalization of the sharp RIP result by Cai and Zhang [16] whose RIP theorem states that  $\delta < \sqrt{1/2}$  on  $\Sigma - \Sigma$  implies stable recovery of sparse vectors (when  $\Sigma$  is the set of  $K$ -sparse vectors) with the  $\ell^1$ -norm, and of low rank matrices (when  $\Sigma$  is the set of rank- $r$  matrices) with the nuclear norm. This RIP theorem is sharp because for any  $\epsilon > 0$  there exists a dimension  $N$  of  $\mathcal{H} = \mathbb{R}^N$ , a sparsity model  $\Sigma_K$  and matrix  $M$  satisfying the RIP with constant  $\delta \geq \sqrt{1/2} + \epsilon$  for which uniform recovery of by  $\ell^1$  minimization is impossible [22].

Theorem 1.1 and 1.2 are immediate corollaries of Theorem 3.1 below. The proof of Theorem 3.1 relies on the following identity, where we recall that  $\|\cdot\|_{\mathcal{F}}$  is the Euclidean norm.

**Fact 3.1.** *Let  $\lambda_i \in \mathbb{R}_+$  such that  $\sum \lambda_i = 1$ , and  $h_i \in \mathcal{H}$  such that the  $\|Mh_i\|_{\mathcal{F}}$  are uniformly bounded.*

*Then*

$$\sum_i \lambda_i \|Mh_i\|_{\mathcal{F}}^2 = 4 \sum_i \lambda_i \left( \sum_j \lambda_j Mh_j \right) - \frac{1}{2} \|Mh_i\|_{\mathcal{F}}^2. \quad (29)$$

*Proof.* The boundedness of the  $\|Mh_i\|_{\mathcal{F}}$  imply that  $\sum_j \lambda_j Mh_j$  and  $\sum_i \lambda_i \|Mh_i\|_{\mathcal{F}}^2$  converge. Moreover, by convexity  $\|\sum_j \lambda_j Mh_j\|_{\mathcal{F}}^2 \leq \sum_j \lambda_j \|Mh_j\|_{\mathcal{F}}^2 \leq \sup_j \|Mh_j\|_{\mathcal{F}}^2$ . Hence  $\sum_i \lambda_i \|\sum_j \lambda_j Mh_j\|_{\mathcal{F}}^2$  converges. Consequently, all sums in the following converge. We simply develop

$$\begin{aligned} \sum_i \lambda_i \left( \sum_j \lambda_j Mh_j \right) - \frac{1}{2} \|Mh_i\|_{\mathcal{F}}^2 &= \sum_i \lambda_i \left( \sum_j \lambda_j Mh_j \right) - 2\text{Re} \langle \sum_j \lambda_j Mh_j, \sum_i \lambda_i \frac{1}{2} Mh_i \rangle_{\mathcal{F}} \\ &\quad + \frac{1}{4} \sum_i \lambda_i \|Mh_i\|_{\mathcal{F}}^2 \\ &= \frac{1}{4} \sum_i \lambda_i \|Mh_i\|_{\mathcal{F}}^2. \end{aligned}$$

□

**Theorem 3.1** (RIP condition for stable recovery). *Assume that  $\Sigma$  is a UoS or a cone and define  $\delta_{\Sigma}(f) :=$*

$$\inf_{z \in \mathcal{T}_{\Sigma}(\Sigma) \setminus \{0\}} \delta_{\Sigma}(z), \text{ with } \delta_{\Sigma}(z) := \sup_{x \in \Sigma} \delta_{\Sigma}(x, z), \text{ where}$$

- **UoS setting**

$$\delta_{\Sigma}(x, z) = \delta_{\Sigma}^{UoS}(x, z) := \frac{-\mathcal{R}e\langle x, z \rangle}{\|x\|_{\mathcal{H}} \sqrt{\|x + z\|_{\Sigma}^2 - \|x\|_{\mathcal{H}}^2 - 2\mathcal{R}e\langle x, z \rangle}} \quad (30)$$

- **Cone setting**

$$\delta_{\Sigma}(x, z) = \delta_{\Sigma}^{cone, sharp}(x, z) := \begin{cases} \delta_{\Sigma}^{UoS}(x, z) & \text{if } \mathcal{R}e\langle x, z \rangle \geq \|x + z\|_{\Sigma}^2/2; \\ \frac{-2\mathcal{R}e\langle x, z \rangle}{\|x + z\|_{\Sigma}^2 - 2\mathcal{R}e\langle x, z \rangle}, & \text{if } \mathcal{R}e\langle x, z \rangle < \|x + z\|_{\Sigma}^2/2 \end{cases} \quad (31)$$

If  $M$  is continuous on  $\mathcal{H}$  and satisfies the RIP on  $\Sigma - \Sigma$  with constant  $\delta < \delta_{\Sigma}(f)$ , then for all  $x_0 \in \Sigma$ ,  $\|e\|_{\mathcal{F}} \leq \eta \leq \epsilon$  we have, with  $x^*$  the result of minimization (3),

$$\|x^* - x_0\|_{\mathcal{H}} \leq \|x^* - x_0\|_{\Sigma} \leq \frac{2\sqrt{1+\delta}}{D_{\Sigma}(f, \delta)} \cdot (\eta + \epsilon) \quad (32)$$

where

$$D_{\Sigma}(f, \delta) := \inf_{z \in \mathcal{T}_f(\Sigma) \setminus \{0\}} \sup_{x \in \Sigma} D(x, z, \delta) > 0 \quad (33)$$

with

- **UoS setting:**

$$D_{\Sigma}(x, z, \delta) = D_{\Sigma}^{UoS}(x, z, \delta) := \frac{-2\mathcal{R}e\langle x, z \rangle - \delta\|x\|_{\mathcal{H}} \sqrt{\|x + z\|_{\Sigma}^2 - \|x\|_{\mathcal{H}}^2 - 2\mathcal{R}e\langle x, z \rangle}}{\|x\|_{\mathcal{H}} + \|x + z\|_{\Sigma}}. \quad (34)$$

- **Cone setting:** We have  $D_{\Sigma}(x, z, \delta) = D_{\Sigma}^{cone, sharp}(x, z, \delta)$  with

$$D_{\Sigma}^{cone, sharp}(x, z, \delta) := \begin{cases} D_{\Sigma}^{UoS}(x, z, \delta) & \text{if } \mathcal{R}e\langle x, z \rangle \geq \|x + z\|_{\Sigma}^2/2; \\ \frac{-2\mathcal{R}e\langle x, z \rangle - \delta(\|x + z\|_{\Sigma}^2 - 2\mathcal{R}e\langle x, z \rangle)}{\|x\|_{\mathcal{H}} + \|x + z\|_{\Sigma}}, & \text{if } \mathcal{R}e\langle x, z \rangle < \|x + z\|_{\Sigma}^2/2 \end{cases} \quad (35)$$

**Remark 3.1.** This theorem directly implies Theorem 1.1 and Theorem 1.2 because  $\delta^{cone} \leq \delta^{cone, sharp}$ : when they are different, we have  $\delta^{cone, sharp}(x, z)/\delta^{cone}(x, z) = \frac{\sqrt{\|x + z\|_{\Sigma}^2/\|x\|_{\mathcal{H}}^2 - 2\mathcal{R}e\langle x, z \rangle/\|x\|_{\mathcal{H}}^2} - 1}{\|x + z\|_{\Sigma}^2/\|x\|_{\mathcal{H}}^2 - 2\mathcal{R}e\langle x, z \rangle/\|x\|_{\mathcal{H}}^2} \leq 1$ . Similarly we have  $D_{\Sigma}^{UoS}(x, z, \delta) \geq D_{\Sigma}^{cone, sharp}(x, z, \delta)$ .



The proof of Theorem 3.1 mainly relies on Lemma 3.1 which will soon be stated. This Lemma gives a bound of the norm of elements of  $\mathcal{T}_f(\Sigma)$ . We start by the proof of the theorem then state and prove Lemma 3.1.

*Proof of Theorem 3.1.* Let  $x_0 \in \Sigma$  and  $e \in \mathcal{F}$  with  $\|e\|_{\mathcal{F}} \leq \eta \leq \epsilon$ . Choose any

$$x^* \in \operatorname{argmin}_{x \in \mathcal{H}} f(x) \text{ s.t. } \|Mx - (Mx_0 + e)\|_{\mathcal{F}} \leq \epsilon \quad (36)$$

Let  $z = x^* - x_0$ . By definition, we have  $f(x^*) \leq f(x_0)$  i.e.  $f(x_0 + z) \leq f(x_0)$ , hence  $z \in \mathcal{T}_f(\Sigma)$ . We use Lemma 3.1:

$$\|x^* - x_0\|_{\Sigma} = \|z\|_{\Sigma} \leq 2 \frac{\sqrt{1 + \delta}}{D_{\Sigma}(f, \delta)} \|Mz\|_{\mathcal{F}}. \quad (37)$$

Following the classical proof of stable  $\ell^1$  recovery [17], observe that (with the triangle inequality):

$$\|Mz\|_{\mathcal{F}} \leq \eta + \epsilon. \quad (38)$$

We conclude

$$\|x^* - x_0\|_{\mathcal{H}} \leq \|x^* - x_0\|_{\Sigma} \leq 2 \frac{\sqrt{1 + \delta}}{D_{\Sigma}(f, \delta)} (\eta + \epsilon). \quad (39)$$

□

**Lemma 3.1.** *Under the hypotheses of Theorem 3.1, for any  $z \in \mathcal{T}_f(\Sigma)$ , we have*

$$\|z\|_{\Sigma} \leq 2 \frac{\sqrt{1 + \delta}}{D_{\Sigma}(f, \delta)} \|Mz\|_{\mathcal{F}}. \quad (40)$$

*Proof.* Let  $z \in \mathcal{T}_f(\Sigma)$ . If  $z = 0$  the result is trivial. From now on we assume  $z \neq 0$ . Consider an arbitrary  $x \in \Sigma$  and  $\rho := \operatorname{Re}\langle x, x + z \rangle / \|x\|_{\mathcal{H}}^2$ ,  $\alpha := \|x + z\|_{\Sigma}^2 / \|x\|_{\mathcal{H}}^2$ . To exploit Fact 2.1, we consider  $\gamma' > 0$  and  $z'$  such that  $(x + z') \in \mathbb{R}_+ \cdot \operatorname{conv}(\Sigma \cap S(1))$  and  $\|x + z - (x + z')\|_{\mathcal{H}} = \|z - z'\|_{\mathcal{H}} \leq \gamma'$ , and  $\lambda_i \in \mathbb{R}_+$ ,  $u_i \in \Sigma$ , such that  $\sum \lambda_i = 1$  and  $x + z' = \sum \lambda_i u_i$ .

Consider an arbitrary  $\beta \in \mathbb{R}$  and  $h_i := u_i - (1 + \beta)x$ . Remark that  $\sum_j \lambda_j h_j - \frac{1}{2} h_i = \frac{1}{2} ((1 - \beta)x - u_i) + z'$ .

We apply Fact 3.1 to obtain the identity

$$\begin{aligned}
\sum \lambda_i \|M(u_i - (1 + \beta)x)\|_{\mathcal{F}}^2 &= \sum \lambda_i \|M((1 - \beta)x - u_i + 2z')\|_{\mathcal{F}}^2 \\
&= \sum \lambda_i \left\{ \|M((1 - \beta)x - u_i)\|^2 + 4\operatorname{Re}\langle M((1 - \beta)x - u_i), Mz' \rangle + 4\|Mz'\|_{\mathcal{F}}^2 \right\} \\
&= \sum \lambda_i \|M((1 - \beta)x - u_i)\|^2 - 4\operatorname{Re}\langle M(\beta x + z'), Mz' \rangle + 4\|Mz'\|_{\mathcal{F}}^2 \\
&= \sum \lambda_i \|M((1 - \beta)x - u_i)\|^2 - 4\beta \operatorname{Re}\langle Mx, Mz' \rangle.
\end{aligned} \tag{41}$$

Consider the vectors  $h_i = u_i - (1 + \beta)x$  and  $(1 - \beta)x - u_i$ . When  $\Sigma$  is a UoS, these vectors are in  $\Sigma - \Sigma$  for any  $\beta \in \mathbb{R}$ . This is still the case when  $\Sigma$  is a cone, provided that  $\beta \in (-1, 1)$ . In both cases, they are in  $\Sigma - \Sigma$  provided that  $\beta \in I$  where  $I := \mathbb{R}$  when  $\Sigma$  is a UoS, and  $I := (-1, 1) \subset \mathbb{R}$  when  $\Sigma$  is a cone.

Thus, when  $\beta \in I$ , we apply the RIP hypothesis on both vectors (lower RIP on the left side, upper RIP on the right side) and (41) yields

$$\begin{aligned}
(1 - \delta) \sum \lambda_i \|u_i - (1 + \beta)x\|_{\mathcal{H}}^2 &\leq \sum \lambda_i \|M((1 - \beta)x - u_i)\|_{\mathcal{F}}^2 - 4\beta \operatorname{Re}\langle Mx, Mz' \rangle \\
&\leq (1 + \delta) \sum \lambda_i \|(1 - \beta)x - u_i\|_{\mathcal{H}}^2 - 4\beta \operatorname{Re}\langle Mx, Mz' \rangle.
\end{aligned} \tag{42}$$

We calculate:

$$\begin{aligned}
&(1 + \delta) \sum \lambda_i \|(1 - \beta)x - u_i\|_{\mathcal{H}}^2 - (1 - \delta) \sum \lambda_i \|u_i - (1 + \beta)x\|_{\mathcal{H}}^2 \\
&= (1 + \delta) \sum \lambda_i \left( (1 - \beta)^2 \|x\|_{\mathcal{H}}^2 + \|u_i\|_{\mathcal{H}}^2 \right) - (1 - \delta) \sum \lambda_i \left( \|u_i\|_{\mathcal{H}}^2 + (1 + \beta)^2 \|x\|_{\mathcal{H}}^2 \right) \\
&\quad + 2(-1 + \delta)(1 - \beta) + (1 - \delta)(1 + \beta)) \cdot \operatorname{Re}\langle x, \sum \lambda_i u_i \rangle \\
&= \sum \lambda_i \left( (1 + \delta)(1 - \beta)^2 - (1 - \delta)(1 + \beta)^2 \right) \|x\|_{\mathcal{H}}^2 + 2\delta \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 + 4(\beta - \delta) \operatorname{Re}\langle x, x + z' \rangle \\
&= (2\delta\beta^2 - 4\beta + 2\delta) \cdot \|x\|_{\mathcal{H}}^2 + 2\delta \cdot \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 + 4(\beta - \delta) \operatorname{Re}\langle x, x + z' \rangle.
\end{aligned} \tag{43}$$

Combining with (42), since  $\operatorname{Re}\langle x, x + z' \rangle = \operatorname{Re}\langle x, x + z \rangle + O(\gamma')$  and (by continuity of  $M$ )  $\operatorname{Re}\langle Mx, Mz' \rangle = \operatorname{Re}\langle Mx, Mz \rangle + O(\gamma')$ , we obtain using Fact 2.1 (taking the limit when  $\gamma' \rightarrow 0_+$  of the infimum over  $z'$  and  $\lambda_i, u_i$  of the right hand side in (43)) that

$$(2\delta\beta^2 - 4\beta + 2\delta) \cdot \|x\|_{\mathcal{H}}^2 + 2\delta \cdot \|x + z\|_{\Sigma}^2 + 4(\beta - \delta) \operatorname{Re}\langle x, x + z \rangle - 4\beta \operatorname{Re}\langle Mx, Mz \rangle \geq 0.$$

Since  $x \in \Sigma$ , we have  $\|x\|_{\mathcal{H}}^2 = \|x\|_{\Sigma}^2$ . Dividing the above inequality by  $2\|x\|_{\mathcal{H}}^2$  we get:

$$\delta\beta^2 - 2\beta(1 - \rho) + \delta(1 + \alpha - 2\rho) - 2\beta \frac{\operatorname{Re}\langle Mx, Mz \rangle}{\|x\|_{\mathcal{H}}^2} \geq 0.$$

Denoting  $g_1(\beta|\rho, \alpha, \delta) := \delta\beta^2 - 2\beta(1 - \rho) + \delta(1 + \alpha - 2\rho)$ , the Cauchy-Schwartz inequality yields

$$-g_1(\beta|\rho, \alpha, \delta) \leq \frac{2|\beta|\|Mx\|_{\mathcal{F}}\|Mz\|_{\mathcal{F}}}{\|x\|_{\mathcal{H}}^2}. \quad (44)$$

Using the upper RIP we have  $\|Mx\|_{\mathcal{F}} \leq \sqrt{1 + \delta}\|x\|_{\mathcal{H}}$  hence using (38) we obtain for  $\beta \neq 0$

$$\frac{1}{|\beta|} \cdot [-g_1(\beta|\rho, \alpha, \delta)] \leq 2\sqrt{1 + \delta} \frac{\|Mz\|_{\mathcal{H}}}{\|x\|_{\mathcal{H}}} \quad (45)$$

The rest of the proof consists first in optimizing the choice of  $\beta \in I$ , then that of  $x \in \Sigma$  to best exploit the above inequality.

**Choice of  $\beta$ .** We now compute the supremum of the left hand side,  $g(\beta) := \frac{1}{|\beta|} [-g_1(\beta|\rho, \alpha, \delta)] = -\delta|\beta| + 2\frac{\beta}{|\beta|}(1 - \rho) - \delta(1 + \alpha - 2\rho)\frac{1}{|\beta|}$ , over admissible  $\beta \in I \setminus \{0\}$ . Since  $g(-|\beta|) \geq g(|\beta|)$ , and since the set  $I$  is symmetric around zero, we need only consider the maximization of the function  $g$  over positive  $\beta$ .

Since the function  $g$  is bounded from above, we have  $1 + \alpha - 2\rho \geq 0$ . Straightforward calculus shows that  $g$  is increasing for  $0 < \beta < \beta_{\text{opt}}$  and decreasing for  $\beta > \beta_{\text{opt}}$  with  $\beta_{\text{opt}} := \sqrt{1 + \alpha - 2\rho}$ .

- **UoS setting.**  $I = \mathbb{R}$ . The supremum is reached at  $\beta_{\text{opt}}$ . Thus

$$\begin{aligned} G_{\text{UoS}}(\rho, \alpha, \delta) &:= \sup_{\beta \in I} g(\beta) = g(\beta_{\text{opt}}) = -\delta\sqrt{1 + \alpha - 2\rho} + 2(1 - \rho) - \delta\sqrt{1 + \alpha - 2\rho} \\ &= 2(1 - \rho) - 2\delta\sqrt{1 + \alpha - 2\rho}. \end{aligned}$$

Moreover,  $G_{\text{UoS}}(\rho, \alpha, \delta) > 0$  if and only if  $\delta < \delta_{\text{UoS}}(\rho, \alpha) := \frac{1 - \rho}{\sqrt{1 + \alpha - 2\rho}}$ .

- **Cone setting.**  $I = (-1, 1)$ . We compute  $G_{\text{cone}}(\rho, \alpha, \delta) := \sup_{\beta \in I} g(\beta)$  by considering two cases:

- if  $\rho \geq \alpha/2$  then  $\beta_{\text{opt}} \leq 1$  and  $g(\beta)$  is again maximized at  $\beta = \beta_{\text{opt}}$  yielding  $G_{\text{cone}}(\rho, \alpha, \delta) = G_{\text{UoS}}(\rho, \alpha, \delta)$ , which is positive when  $\delta < \delta_{\text{UoS}}(\rho, \alpha)$ ;
- if  $\rho < \frac{\alpha}{2}$ , then  $g(\beta)$  is maximized at  $\beta = 1$ , yielding  $G_{\text{cone}}(\rho, \alpha, \delta) = g(1) = 2(1 - \rho) - \delta(2 + \alpha - 2\rho)$ , which is positive if and only if  $\delta < \frac{2(1 - \rho)}{2 + \alpha - 2\rho}$ .

Overall, we obtain

$$G_{\text{cone}}(\rho, \alpha, \delta) := \sup_{\beta \in I} g(\beta) = \begin{cases} 2(1 - \rho) - 2\delta \sqrt{1 + \alpha - 2\rho}, & \text{when } \rho \geq \alpha/2; \\ 2(1 - \rho) - \delta(2 + \alpha - 2\rho), & \text{when } \rho < \alpha/2. \end{cases} \quad (46)$$

This is positive if and only if  $\delta < \delta_{\text{cone}}(\rho, \alpha)$ , with

$$\delta_{\text{cone}}(\rho, \alpha) := \begin{cases} \frac{(1-\rho)}{\sqrt{1+\alpha-2\rho}}, & \text{when } \rho \geq \alpha/2; \\ \frac{2(1-\rho)}{2+\alpha-2\rho}, & \text{when } \rho < \alpha/2. \end{cases} \quad (47)$$

From now on when  $\Sigma$  is a UoS (resp. a cone) we use the notation  $G_{\Sigma}(\cdot)$  as a shorthand for  $G_{\text{UoS}}(\cdot)$  (resp.  $G_{\text{cone}}(\cdot)$ ), and a similar convention for  $\delta_{\Sigma}(\cdot)$ .

*Conclusion of the proof.* Under the theorem hypothesis, simple algebraic manipulations of the expressions of  $\delta_{\Sigma}(\cdot)$  and  $G_{\Sigma}(\cdot)$  above show that we have, with the notations of the theorem,

$$\delta < \delta_{\Sigma}(z) = \sup_{x \in \Sigma} \delta_{\Sigma}(\rho(x, z), \alpha_{\Sigma}(x, z)).$$

Hence there is indeed at least one  $x \in \Sigma$  such that  $\delta < \delta_{\Sigma}(\rho, \alpha)$  with  $\rho = \rho(x, z)$ ,  $\alpha = \alpha_{\Sigma}(x, z)$ , or equivalently so that  $G_{\Sigma}(\rho, \alpha, \delta) > 0$ . For any such  $x$ , with the triangle inequality, we obtain

$$\|z\|_{\Sigma} \leq \|x + z\|_{\Sigma} + \|x\|_{\Sigma} \leq (1 + \sqrt{\alpha})\|x\|_{\Sigma} = (1 + \sqrt{\alpha})\|x\|_{\mathcal{H}}. \quad (48)$$

Combined with (45), this yields  $\|z\|_{\Sigma} \leq 2\sqrt{1 + \delta}\|Mz\|_{\mathcal{F}}/D_{\Sigma}(x, z, \delta)$  with

$$D_{\Sigma}(x, z, \delta) := \frac{G_{\Sigma}(\rho, \alpha, \delta)}{1 + \sqrt{\alpha}} \quad (49)$$

Maximizing  $D_{\Sigma}(x, z, \delta)$  over  $x$  yields the best stability constant for the considered descent vector  $z$

$$D_{\Sigma}(z, \delta) := \sup_{x \in \Sigma} \frac{G_{\Sigma}(\rho, \alpha, \delta)}{1 + \sqrt{\alpha}}. \quad (50)$$

This gives

$$\|x^* - x_0\|_{\Sigma} = \|z\|_{\Sigma} \leq 2 \frac{\sqrt{1 + \delta}}{D_{\Sigma}(f, \delta)} \|Mz\|_{\mathcal{F}} \quad (51)$$

with

$$D_{\Sigma}(f, \delta) := \inf_{z \in \mathcal{F}_f(\Sigma) \setminus \{0\}} \sup_{x \in \Sigma} \frac{G_{\Sigma}(\rho, \alpha, \delta)}{1 + \sqrt{\alpha}}. \quad (52)$$

□

### 3.2. Robustness to model error - Instance optimality

An important question is whether the results of the previous section are robust to model error, i.e., when the observed signal  $x_0$  is not *exactly* in  $\Sigma$ . For arbitrary  $f$ , it is not difficult to derive such robust instance optimality results by directly using Theorem 1.2 and considering modeling error as observation noise. However, in this case, we would need to know the level  $d(x_0, \Sigma)$  of modeling error in order to “tune”  $\epsilon$  in the optimization problem (2).

Instead, we establish below robustness results that do not require such prior knowledge at the price of certain assumptions on the regularizer  $f$ . We first give an upper bound on the reconstruction error in Lemma 3.2 then show in Theorem 3.2 that this bound implies instance optimality for  $f$  with certain properties. We will give examples in classical cases after the statement of our result.

Lemma 3.2 uses the notion of  $M$ -norm [15]: given a constant  $C$ , the  $M$ -norm is defined by

$$\|\cdot\|_{M,C} := C \cdot \|M \cdot\|_{\mathcal{F}} + \|\cdot\|_{\mathcal{H}}. \quad (53)$$

**Lemma 3.2.** *Let  $\Sigma$  be a cone or a UoS. Consider a continuous linear operator  $M$  with RIP  $\delta < \delta_{\Sigma}(f)$  on  $\Sigma - \Sigma$ , and a noise level  $\eta \leq \epsilon$ . Let  $C_{\Sigma} = C_{\Sigma}(f, \delta) = \frac{2\sqrt{1+\delta}}{D_{\Sigma}(f, \delta)}$ . Then for all  $x_0 \in \mathcal{H}$ ,  $e \in \mathcal{F}$ , such that  $\|e\|_{\mathcal{H}} \leq \eta \leq \epsilon$ , any minimizer  $x^*$  of (3) satisfies*

$$\|x^* - x_0\|_{\mathcal{H}} \leq C_{\Sigma} \cdot (\eta + \epsilon) + \inf_{z' \in \mathcal{F}_f(\Sigma)} \|x^* - x_0 - z'\|_{M, C_{\Sigma}}. \quad (54)$$

**Remark 3.2.** *Note that contrary to Theorem 3.1, the unknown  $x_0$  is no longer restricted to  $\Sigma$ . The constant  $C_{\Sigma} = C_{\Sigma}(f, \delta)$  has the same definition as in Theorem 3.1.*

*Proof.* Let  $x_0 \in \mathcal{H}$ ,  $e \in \mathcal{F}$ . Let

$$x^* \in \operatorname{argmin}_{x \in \mathcal{H}} f(x) \text{ s.t. } \|Mx - (Mx_0 + e)\|_{\mathcal{F}} \leq \epsilon. \quad (55)$$

Let  $z = x^* - x_0$ . We have  $f(x_0 + z) = f(x^*) \leq f(x_0)$ . Consider any  $z' \in \mathcal{T}_f(\Sigma)$ . With Lemma 3.1,

$$\|z'\|_{\Sigma} \leq C_{\Sigma} \cdot \|Mz'\|_{\mathcal{F}}. \quad (56)$$

Thus

$$\|z\|_{\mathcal{H}} \leq \|z'\|_{\mathcal{H}} + \|z - z'\|_{\mathcal{H}} \leq \|z'\|_{\Sigma} + \|z - z'\|_{\mathcal{H}} \leq C_{\Sigma} \cdot \|Mz'\|_{\mathcal{F}} + \|z - z'\|_{\mathcal{H}} \leq C_{\Sigma} \cdot \|Mz\|_{\mathcal{F}} + \|z - z'\|_{\mathcal{H}} + C_{\Sigma} \|M(z - z')\|_{\mathcal{F}}. \quad (57)$$

As in Equation (38),  $\|Mz\|_{\mathcal{F}} \leq \eta + \epsilon$  and we obtain

$$\|z\|_{\mathcal{H}} \leq C_{\Sigma} \cdot (\eta + \epsilon) + \|z - z'\|_{M, C_{\Sigma}}. \quad (58)$$

Taking the infimum with respect to  $z' \in \mathcal{T}_f(\Sigma)$  yields the result.  $\square$

To go further, we need to replace the bound involving the  $M$ -norm with an estimate directly measuring a modeling error. For this, given a regularizer  $f$  we use the symmetrized “distance” of a vector to the model with respect to  $f$ :

$$d_f(x_0, \Sigma) = \inf_{\tilde{x} \in \Sigma} \frac{f(x_0 - \tilde{x}) + f(\tilde{x} - x_0)}{2} \quad (59)$$

Instance optimality is a direct consequence of Lemma 3.2 using an appropriate *robustness constant*.

**Definition 3.1.** A constant  $D \in \mathbb{R}_+ \cup \{+\infty\}$  is called a *robustness constant of the regularizer  $f$  for the recovery of elements of  $\Sigma$  from their measurement with  $M$*  if, under the assumptions of Lemma 3.2, we have: for all  $x_0 \in \mathcal{H}$ ,

$$\inf_{z' \in \mathcal{T}_f(\Sigma)} \|x^* - x_0 - z'\|_{M, C_{\Sigma}} \leq D \cdot d_f(x_0, \Sigma).$$

The following theorem provides a first general route to finding and exploiting such robustness constants assuming that the (positively homogeneous, non-negative and convex) regularizer  $f$  dominates the  $M$ -norm. We will provide more tailored results for structured sparsity in Section 4.4.2.

**Theorem 3.2.** Let  $\Sigma$  be a cone or a UoS. Let  $f$  be positively homogeneous, non-negative and convex such that  $f(x) < +\infty$  for  $x \in \Sigma$ . Consider a continuous linear operator  $M$  with RIP  $\delta < \delta_{\Sigma}(f)$  on  $\Sigma - \Sigma$  and a noise level  $\eta \leq \epsilon$ . Let  $C_{f, M, \Sigma} < \infty$  such that for all  $u \in \mathcal{H}$ ,  $\|u\|_{M, C_{\Sigma}} \leq C_{f, M, \Sigma} \cdot f(u)$ . Then

- $D := 2C_{f, M, \Sigma}$  is a robustness constant as defined above.

- As a consequence, for all  $x_0 \in \mathcal{H}$ ,  $e \in \mathcal{F}$ , such that  $\|e\|_{\mathcal{H}} \leq \eta \leq \epsilon$ , any minimizer  $x^*$  of (3) satisfies

$$\|x^* - x_0\|_{\mathcal{H}} \leq C_{\Sigma} \cdot (\eta + \epsilon) + 2C_{f,M,\Sigma} \cdot d_f(x_0, \Sigma). \quad (60)$$

*Proof.* If  $d_f(x_0, \Sigma) = \infty$  the result is trivial. We now assume that  $d_f(x_0, \Sigma) < \infty$ . Let  $x_0 \in \mathcal{H}$ ,  $e \in \mathcal{F}$  and  $x^*$  the result of minimization (3). Let  $z = x^* - x_0$ . We have  $f(x_0 + z) \leq f(x_0)$ .

Consider  $x_1 \in \Sigma$ : with the sub-additivity of  $f$  (which follows from convexity),

$$f(x_1 + z) \leq f(x_0 + z) + f(x_1 - x_0) \leq f(x_0) + f(x_1 - x_0) \leq f(x_1) + f(x_0 - x_1) + f(x_1 - x_0). \quad (61)$$

Consider now  $x_2 \in \Sigma$  such that

$$f(x_2 + z) \leq f(x_2) + f(x_0 - x_1) + f(x_1 - x_0). \quad (62)$$

For example<sup>3</sup> choose  $x_2 = x_1$ . Since  $d_f(x_0, \Sigma) < \infty$ , both  $f(x_0 - x_1)$  and  $f(x_1 - x_0)$  are finite. Let

$$y := \frac{f(x_2)}{f(x_2) + f(x_0 - x_1) + f(x_1 - x_0)} \cdot (x_2 + z); \quad \text{and} \quad z' := y - x_2. \quad (63)$$

With the positive homogeneity of  $f$ , we have  $f(x_2 + z') = f(y) = \frac{f(x_2)}{f(x_2) + f(x_0 - x_1) + f(x_1 - x_0)} \cdot f(x_2 + z) \leq f(x_2)$ .

This means that

$$z' \in \mathcal{T}_f(x_2) \subset \mathcal{T}_f(\Sigma). \quad (64)$$

Moreover remark that

$$\begin{aligned} z - z' &= x_2 + z - y = \left(1 - \frac{f(x_2)}{f(x_2) + f(x_0 - x_1) + f(x_1 - x_0)}\right) \cdot (x_2 + z) \\ &= \frac{f(x_0 - x_1) + f(x_1 - x_0)}{f(x_2) + f(x_0 - x_1) + f(x_1 - x_0)} \cdot (x_2 + z). \end{aligned} \quad (65)$$

Thus, with homogeneity

$$f(z - z') = \frac{f(x_0 - x_1) + f(x_1 - x_0)}{f(x_2) + f(x_0 - x_1) + f(x_1 - x_0)} \cdot f(x_2 + z) \leq f(x_0 - x_1) + f(x_1 - x_0). \quad (66)$$

---

<sup>3</sup>The use of other choices for this intermediate  $x_2$  will be useful for the derivation of better robustness constants in particular cases in Section 4.4.2.

Using the property of  $C_{f,M,\Sigma}$  it follows that

$$\inf_{\tilde{z} \in \mathcal{T}_f(\Sigma)} \|x^* - x_0 - \tilde{z}\|_{M,C_\Sigma} \leq \|z - z'\|_{M,C_\Sigma} \leq C_{f,M,\Sigma} \cdot f(z - z') \leq 2C_{f,M,\Sigma} \cdot \frac{f(x_0 - x_1) + f(x_1 - x_0)}{2}. \quad (67)$$

Minimizing over  $x_1 \in \Sigma$  yields the first claim. Using Lemma 3.2 yields the second one.  $\square$

We can derive robustness constants for the following cases:

- Consider  $f$  a convex gauge induced by any bounded closed convex set containing 0 in a finite-dimensional vector space  $\mathcal{H}$ .

From Table 1,  $f$  is positively homogeneous, non-negative, lower semi-continuous, convex and coercive. With the positive homogeneity of  $f$ , we have  $\inf_{u \in \mathcal{H} \setminus \{0\}} \frac{f(u)}{\|u\|_{M,C_\Sigma}} = \inf_{v \in S_{\|\cdot\|_{M,C_\Sigma}}} f(v)$  where  $S_{\|\cdot\|_{M,C_\Sigma}}$  is the unit sphere with respect to  $\|\cdot\|_{M,C_\Sigma}$ . The sphere  $S_{\|\cdot\|_{M,C_\Sigma}}$  is compact because  $\mathcal{H}$  is finite-dimensional and  $\|\cdot\|_{M,C_\Sigma}$  is lower semi-continuous. Thus  $f$  admits a minimum value  $\beta \geq 0$  on  $S_{\|\cdot\|_{M,C_\Sigma}}$  because is bounded and closed (because  $\|\cdot\|_{M,C_\Sigma}$  is lower semi-continuous). This minimum value is strictly positive because  $f$  cannot take the value 0 outside of the origin (otherwise  $f$  being positively homogeneous would not be coercive), and we conclude that  $C_{f,M,\Sigma} = 1/\beta < \infty$  hence  $D = 2C_{f,M,\Sigma}$  is a robustness constant and we can apply Theorem 3.2.

- Consider  $f = \|\cdot\|_1$  and  $\Sigma = \Sigma_K$  the set of  $K$ -sparse vectors in a finite-dimensional space  $\mathcal{H}$ . Here,  $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_2$  the usual  $\ell^2$ -norm. A proof technique for the general case of block structured sparsity, developed in Theorem 4.6 from the next section, establishes that  $D := 2 \frac{1 + \sqrt{1 + \delta \cdot C_\Sigma}}{\sqrt{K}}$  is a robustness constant for the considered problem. Lemma 3.2 thus implies

$$\|x^* - x_0\|_2 \leq C_\Sigma \cdot (\epsilon + \eta) + D \cdot d_{\|\cdot\|_1}(x_0, \Sigma). \quad (68)$$

In the literature, the distance  $d_{\|\cdot\|_1}(x_0, \Sigma)$  is often written as  $\sigma_K(x_0)_1 = \|x_{0,T^c}\|_1$ , i.e. the best  $K$ -term approximation for the  $\ell^1$ -norm. The constant  $D$  is of the same order  $O(1/\sqrt{K})$  found in classical results [29].

- Consider  $f = \|\cdot\|_*$  the nuclear norm and  $\Sigma = \Sigma_r$  the set of matrices of rank at most  $r$ . The norm  $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_F$  is the Frobenius norm. As explained in Remark 4.1 in the next section,  $D = 2 \frac{1 + \sqrt{1 + \delta \cdot C_\Sigma}}{\sqrt{r}}$



is a robustness constant for the considered problem. Lemma 3.2 thus implies

$$\|x^* - x_0\|_F \leq C_\Sigma \cdot (\epsilon + \eta) + D \cdot d_{\|\cdot\|_*}(x_0, \Sigma_r). \quad (69)$$

In this case,  $d_{\|\cdot\|_*}(x_0, \Sigma_r)$  is reached by zeroing the  $r$  largest singular values of  $x_0$ .

In the previous examples  $\mathcal{H}$  is a finite-dimensional space. Theorem 3.2 can be used to derive results in some infinite-dimensional setting as well. Such result will be shown in the case of block structured sparsity, in the next section.

#### 4. Application to block structured sparsity in infinite dimension

Structured sparse models generalize sparse models by including constraints on the (sparse) support of the signal [30, 10, 26]. Adding a notion of block sparsity has applications for simultaneous noise and signal sparse modeling and in imaging [3, 39, 40]. We begin by applying our general results to these models, then we discuss how the resulting framework improves and extends previous known results.

##### 4.1. The finite group-sparse model in infinite dimension

Here, we suppose that  $\mathcal{H}$  is separable. Thus there exists an orthonormal Hilbert basis  $(e_i)_{i \in \mathbb{N}}$ . Let  $G$  be a finite collection of  $|G| < +\infty$  non overlapping finite *groups*, i.e. supports subsets  $g \subset \mathbb{N}$  with  $|g| < \infty$  and  $g \cap g' = \emptyset$ ,  $g \neq g'$ . The restriction of the vector  $x \in \mathcal{H}$  to the group  $g$  is  $x_g := \sum_{i \in g} \langle x, e_i \rangle e_i$ . A group support  $H$  is a subset of  $G$  and the restriction of  $x$  to  $H$  is  $x_H := \sum_{g \in H} x_g$ . The group support of  $x \in \mathcal{H}$ , denoted  $\text{supp}(x)$ , is the smallest  $H \subset G$  such that  $x_H = x$ .

Given an integer  $K$ , the  $K$ -group-sparse model is defined as

$$\Sigma_K := \{x \in \mathcal{H}, |\text{supp}(x)| \leq K\}. \quad (70)$$

Considering the atoms

$$\mathcal{A} := \Sigma_1 \cap \mathcal{S}(1) \quad (71)$$

the corresponding atomic norm is associated to the finite-dimensional space

$$\mathcal{E}(\mathcal{A}) = \text{span}(\{e_i\}_{i \in \bigcup_{g \in G} g})$$

and simply given by

$$\|x\|_{\mathcal{A}} = \begin{cases} \sum_{g \in G} \|x_g\|_{\mathcal{H}}, & x \in \mathcal{E}(\mathcal{A}); \\ +\infty, & x \notin \mathcal{E}(\mathcal{A}) \end{cases} \quad (72)$$

and its dual norm is

$$\|x\|_{\mathcal{A}}^* = \max_{g \in G} \|x_g\|_{\mathcal{H}}. \quad (73)$$

Theorem 1.2 can be leveraged to show that  $\|\cdot\|_{\mathcal{A}}$  yields stable recovery under a RIP hypothesis. To establish this, we show in the following section that for any  $z \in \mathcal{T}_{\mathcal{A}}(\Sigma) \setminus \{0\}$  we can find decompositions  $z = x + z - x$  such that : a)  $\rho(x, z) = 0$  and b)  $\alpha_{\Sigma}(x, z) = 1$ .

Because  $\|x\|_{\mathcal{A}} = +\infty$  if  $x \notin \mathcal{E}(\mathcal{A})$ , we can do the following proofs exclusively in the finite-dimensional space  $\mathcal{E}(\mathcal{A})$  where  $\|\cdot\|_{\mathcal{A}}$  matches the classical  $\ell^1 - \ell^2$  norm [41] : then  $\|x\|_{\mathcal{A}} = \sum_{g \in G} \|x_g\|_2$ . Yet, we can use Theorem 3.2 to obtain robustness results in the whole space  $\mathcal{H}$ , by appropriately defining the behaviour of the regularizer on elements of  $\mathcal{H} \setminus \mathcal{E}(\mathcal{A})$  (Section 4.4.2).

#### 4.2. Decompositions for a “sharp” admissible RIP constant

We will show that the decompositions defined by the following sets  $\mathcal{Y}_{\Sigma}(z, \mathcal{A})$  lead to a bound on  $\delta_{\Sigma}(f)$  that matches the sharp bound of Cai and Zhang and extends it to the considered general setting of structured sparsity. Let

$$\mathcal{Y}_{\Sigma}(z, \mathcal{A}) := \operatorname{argmin}_{x \in \Sigma} \{\|x + z\|_{\mathcal{A}} - \|x\|_{\mathcal{A}}\} \subset \Sigma. \quad (74)$$

We first characterize  $\mathcal{Y}_{\Sigma}(z, \mathcal{A})$  and show that  $x \in \mathcal{Y}_{\Sigma}(z, \mathcal{A})$  implies  $\rho(x, z) = 0$ .

**Proposition 4.1.** *Consider  $z \in \mathcal{E}(\mathcal{A})$ . The set  $\mathcal{Y}_{\Sigma}(z, \mathcal{A})$  is exactly the collection of all vectors  $x = -z_H$  where  $H$  is some group support made of  $K$ -groups with largest individual Euclidean norms :  $|H| = K$  and*

$$\|z_g\|_{\mathcal{H}} \geq \|z_{g'}\|_{\mathcal{H}}, \quad \forall g \in H, \quad \forall g' \notin H.$$

As a consequence,  $\mathcal{Y}_{\Sigma}(z, \mathcal{A}) \neq \emptyset$  and any  $x \in \mathcal{Y}_{\Sigma}(z, \mathcal{A})$  satisfies

$$\min_{g \in H} \|x_g\|_{\mathcal{H}} \geq \max_{g' \notin H} \|z_{g'}\|_{\mathcal{H}} = \|x + z\|_{\mathcal{A}}^*.$$

and  $\rho(x, z) = 0$ .

*Proof.* Let  $z \in \mathcal{E}(\mathcal{A})$ . We minimize the following expression with respect to  $x \in \Sigma_K$  :

$$\|x + z\|_{\mathcal{A}} - \|x\|_{\mathcal{A}} = \sum_{g \in G} (\|x_g + z_g\|_{\mathcal{H}} - \|x_g\|_{\mathcal{H}}) \quad (75)$$

For a given group support  $H$ , we first minimize this expression over the value of  $x$  under the constraint  $\text{supp}(x) = H$ . Because there is no overlap, we can minimize this sum over each group separately. Denoting  $\tilde{x}_H$  a minimizer, each summand is lower bounded (by a triangle inequality) by  $-\|z_g\|_{\mathcal{H}}$ , a value that is reached by setting  $\tilde{x}_g := -z_g$ . Thus

$$\|\tilde{x}_H + z\|_{\mathcal{A}} - \|\tilde{x}_H\|_{\mathcal{A}} = \sum_{g \in H^c} \|z_g\|_{\mathcal{H}} - \sum_{g \in H} \|z_g\|_{\mathcal{H}}. \quad (76)$$

We minimize over  $H$  under the constraint  $|H| \leq K$  and obtain the result.  $\square$

We now give a bound on  $\alpha_{\Sigma}(x, z)$  for such decompositions. We need the following lemma which extends the sparse decomposition of polytopes from [16] to the general case where  $\Sigma_K$  is made of combinations of  $K$  pairwise orthogonal elements of  $\mathcal{A}$ .

**Lemma 4.1.** *Let  $\mathcal{A}$  be a set of normalized atoms and  $u = \sum_{i=1,L} c_i a_i$  with  $c_i \in \mathbb{R}, a_i \in \mathcal{A}$ , such that the  $a_i$  are pairwise orthogonal. Let  $\Sigma$  be any model set containing all such combinations for  $L = K$  (e.g., consider  $\Sigma = \Sigma_K(\mathcal{A})$ ). Then*

$$\|u\|_{\Sigma} \leq \max\left(\sum_{i=1,L} |c_i| / \sqrt{K}, \max_{i=1,L} |c_i| \sqrt{K}\right).$$

*Proof.* This proof is a direct extension of the proof of Cai [16]. We proceed by induction on  $L$ .

For  $L \leq K$ , we have  $u \in \Sigma$  thus

$$\|u\|_{\Sigma}^2 = \|u\|_{\mathcal{H}}^2 = \sum |c_i|^2 \leq \left(\sum_{i=1,L} |c_i|\right) \left(\max_{i=1,L} |c_i|\right) \leq \max\left(\sum_{i=1,L} |c_i| / \sqrt{K}, \max_{i=1,L} |c_i| \sqrt{K}\right)^2. \quad (77)$$

Suppose now the statement is true for a given  $L \geq K$ . Let  $u = \sum_{i=1}^{L+1} c_i a_i$  with all  $c_i \neq 0$  and the  $a_i$  pairwise orthogonal. Note  $u_i = c_i a_i$ . Denoting  $\alpha = \max(\sum_{i=1}^{L+1} |c_i| / \sqrt{K}, \max_{i=1,L+1} |c_i| \sqrt{K})$ , we have  $\sum_{i=1}^{L+1} |c_i| \leq \alpha \sqrt{K}$  and  $\max_{i=1,L+1} |c_i| \sqrt{K} \leq \alpha / \sqrt{K}$ . Without loss of generality we (re)order the decomposition such that  $|c_1| \geq |c_2| \geq \dots \geq |c_{L+1}| > 0$ .

For  $1 \leq j \leq L+1$  define  $\sigma_j := \sum_{k=j}^{L+1} |c_k|$ . By definition, for  $j = 1$  we have  $\sigma_j = \sigma_1 = \sum_{i=1}^{L+1} |c_i| \leq \alpha \sqrt{K} = \frac{K}{\sqrt{K}} \alpha$ . Since  $L \geq K$ , we obtain  $\sigma_1 \leq (L+1-j)\alpha / \sqrt{K}$ . For  $j = L+1$  we have  $\sigma_j = |c_{L+1}| > 0 = (L+1-j)\alpha / \sqrt{K}$ .

Considering the largest  $1 \leq j^* \leq L$  such that  $\sigma_j \leq (L+1-j)\frac{\alpha}{\sqrt{K}}$  for  $1 \leq j \leq j^*$ : we have

$$\begin{aligned} \sum_{k=j^*}^{L+1} |c_k| &\leq (L+1-j^*) \frac{\alpha}{\sqrt{K}} \\ \sum_{k=j^*+1}^{L+1} |c_k| &> (L-j^*) \frac{\alpha}{\sqrt{K}}. \end{aligned} \tag{78}$$

Define  $\beta := \frac{1}{L+1-j^*} \sum_{k=j^*}^{L+1} |c_k|$ . With inequations (78), we have  $\beta \leq \alpha / \sqrt{K}$  and

$$\begin{aligned} \max_{j^* \leq i \leq L+1} |c_i| &= |c_{j^*}| = \sum_{k=j^*}^{L+1} |c_k| - \sum_{k=j^*+1}^{L+1} |c_k| \\ &< (L+1-j^*)\beta - (L-j^*) \frac{\alpha}{\sqrt{K}} = \beta + (L-j^*)(\beta - \alpha / \sqrt{K}) \\ &\leq \beta \leq \frac{\alpha}{\sqrt{K}} \end{aligned} \tag{79}$$

For any index  $j^* \leq i \leq L+1$ , defining  $\lambda_i := \frac{\beta - |c_i|}{\beta}$  we thus have  $\lambda_i > 0$ , and

$$\sum_{i=j^*}^{L+1} \lambda_i = L+2-j^* - \frac{\sum_{i=j^*}^{L+1} |c_i|}{\beta} = L+2-j^* - \frac{(L+1-j^*)\beta}{\beta} = 1. \tag{80}$$

Defining further

$$v_i := \sum_{k=1}^{j^*-1} u_k + \beta \sum_{k=j^*; k \neq i}^{L+1} \frac{u_k}{|c_k|}, \tag{81}$$

we have

$$\begin{aligned} \sum_{i=j^*}^{L+1} \lambda_i v_i &= \sum_{k=1}^{j^*-1} u_k + \sum_{i=j^*}^{L+1} (\beta - |c_i|) \sum_{k=j^*; k \neq i}^{L+1} \frac{u_k}{|c_k|} \\ &= \sum_{k=1}^{j^*-1} u_k + \sum_{i=j^*}^{L+1} (\beta - |c_i|) \left( -\frac{u_i}{|c_i|} + \sum_{k=j^*}^{L+1} \frac{u_k}{|c_k|} \right) \\ &= \sum_{k=1}^{j^*-1} u_k + \sum_{i=j^*}^{L+1} u_i + \sum_{i=j^*}^{L+1} \left( -\beta \frac{u_i}{|c_i|} + (\beta - |c_i|) \sum_{k=j^*}^{L+1} \frac{u_k}{|c_k|} \right). \end{aligned} \tag{82}$$

We calculate the last term. Denoting  $w = \sum_{i=j^*}^{L+1} \frac{u_i}{|c_i|}$ , we have

$$\sum_{i=j^*}^{L+1} \left( -\beta \frac{u_i}{|c_i|} + (\beta - |c_i|) \sum_{k=j^*}^{L+1} \frac{u_k}{|c_k|} \right) = -\beta w + (L+2-j^*)\beta w - (L+1-j^*)\beta w = 0. \quad (83)$$

Along with equation (82), this yields

$$\sum_{i=j^*}^{L+1} \lambda_i v_i = u. \quad (84)$$

Moreover,

$$\sum_{k=1}^{j^*-1} |c_k| + \beta \sum_{k=j^*; k \neq i}^{L+1} \frac{|c_k|}{|c_k|} = \sum_{k=1}^{j^*-1} |c_k| + \sum_{k=j^*}^{L+1} |c_k| \leq \alpha \sqrt{K} \quad (85)$$

and

$$\max(\max_{1 \leq k \leq j^*-1} |c_k|, \beta) \leq \max(\alpha / \sqrt{K}, \beta) \stackrel{(79)}{\leq} \alpha / \sqrt{K} \quad (86)$$

Finally, we have shown that  $v_i = \sum_{k=1}^L c'_k a'_k$ , with pairwise orthogonal  $a'_k$  and  $\sum_{k=1}^L |c'_k| \leq \alpha \sqrt{K}$  and  $\max_{k=1,L} |c'_k| \leq \alpha / \sqrt{K}$ . We can use the induction hypothesis to get  $\|v_i\|_{\Sigma} \leq \alpha$  and use convexity to obtain from (84) :

$$\|u\|_{\Sigma} \leq \sum_{i=j^*}^{L+1} \lambda_i \|v_i\|_{\Sigma} \leq \sum_{i=j^*}^{L+1} \lambda_i \alpha \leq \alpha. \quad (87)$$

□

The above lemma has a simple and practical consequence for group-sparsity.

**Corollary 4.1.** *Let  $\mathcal{A}$  be the set of normalized 1-group sparse vectors. Then for all  $u \in \mathcal{E}(\mathcal{A})$  we have*

$$\|u\|_{\Sigma_K} \leq \max \left( \frac{\|u\|_{\mathcal{A}}}{\sqrt{K}}, \sqrt{K} \|u\|_{\mathcal{A}}^* \right). \quad (88)$$

*Proof.* Let  $u = \sum_i c_i a_i$  such that  $\sum |c_i| = \|u\|_{\mathcal{A}}$  and  $\langle a_i, a_j \rangle = 0$ . Apply Lemma 4.1 □

With these results we are now equipped to bound  $\alpha(x, z)$  for  $x \in \mathcal{Y}_{\Sigma}(z, \mathcal{A})$  when  $z$  is a descent vector for the group-norm  $\|\cdot\|_{\mathcal{A}}$  with respect to  $\Sigma_K$ .

**Proposition 4.2.** *For any  $z \in \mathcal{T}_{\mathcal{A}}(\Sigma_K) \setminus \{0\}$ ,  $x \in \mathcal{Y}_{\Sigma}(z, \mathcal{A})$ ,*

$$\alpha(x, z) \leq 1 \quad (89)$$

*Proof.* Let  $z \in \mathcal{T}_{\mathcal{A}}(\Sigma_K) \setminus \{0\}$  and  $x \in \mathcal{Y}_{\Sigma}(z, \mathcal{A})$ . By definition of  $\mathcal{T}_{\mathcal{A}}(\Sigma_K)$  there is  $\tilde{x} \in \Sigma_K$  such that  $\|\tilde{x} + z\|_{\mathcal{A}} \leq \|\tilde{x}\|_{\mathcal{A}}$ , hence by definition of  $\mathcal{Y}_{\Sigma}(z, \mathcal{A})$  we have  $\|x + z\|_{\mathcal{A}} - \|x\|_{\mathcal{A}} = \min_{\tilde{x} \in \Sigma_K} (\|\tilde{x} + z\|_{\mathcal{A}} - \|\tilde{x}\|_{\mathcal{A}}) \leq 0$ . Let  $H$  be the group support of  $x$ . By Proposition 4.1,

$$K\|x + z\|_{\mathcal{A}}^* \leq K \min_{g \in H} \|x_g\|_{\mathcal{H}} \leq \sum_{g \in H} \|x_g\|_{\mathcal{H}} = \|x\|_{\mathcal{A}}.$$

Combining the above estimate with  $\|x + z\|_{\mathcal{A}} \leq \|x\|_{\mathcal{A}}$  and Lemma 4.1 yields

$$\|x + z\|_{\Sigma}^2 \leq \max(\|x + z\|_{\mathcal{A}} / \sqrt{K}, \|x + z\|_{\mathcal{A}}^* \sqrt{K})^2 \leq \|x\|_{\mathcal{A}}^2 / K.$$

Since  $x$  is  $K$ -group-sparse, we have

$$\|x\|_{\mathcal{A}} = \sum_{i=1}^K \|x_{g_i}\|_{\mathcal{H}} \leq \sqrt{K} \sqrt{\sum_{i=1}^K \|x_{g_i}\|_{\mathcal{H}}^2} = \sqrt{K} \|x\|_{\mathcal{H}}$$

hence  $\|x + z\|_{\Sigma}^2 \leq \|x\|_{\mathcal{H}}^2$ . □

To summarize, when  $z \in \mathcal{T}_{\mathcal{A}}(\Sigma_K) \setminus \{0\}$  we can exhibit  $x \in \Sigma_K$  such that  $\rho(x, z) = 0$  and  $\alpha_{\Sigma_K}(x, z) \leq 1$ . It follows that

$$\alpha_{\Sigma_K}(f) := \sup_{z \in \mathcal{T}_{\mathcal{A}}(\Sigma_K) \setminus \{0\}} \inf_{x \in \Sigma_K : \rho(x, z) = 0} \alpha_{\Sigma_K}(x, z) \leq 1. \quad (90)$$

and we obtain our main theorem for the  $K$ -group-sparse model.

**Theorem 4.1.** *For the  $K$ -group-sparse model  $\Sigma_K \subset \mathcal{H}$  defined in (70), and the atomic norm  $f(\cdot) = \|\cdot\|_{\mathcal{A}}$  defined in (72), we have*

$$\delta_{\Sigma_K}(\|\cdot\|_{\mathcal{A}}) \geq \frac{1}{\sqrt{2}}.$$

#### 4.3. Extension to block structured sparsity

We now show that we can extend the above results to *block structured sparsity*. Consider a collection of  $J$  finite-dimensional orthogonal spaces  $E_j \subset \mathcal{H}$  each equipped with a  $K_j$ -group-sparse model  $\Sigma_j$  as defined in (70) (each with its set of groups  $G_j$ ). For each  $j$ , we have the associated structured norm  $\|\cdot\|_{\mathcal{A}_j}$  as defined in (72). Since the subspaces are orthogonal, there is a natural isomorphism between their direct sum and

their Cartesian product. It is simpler to work with the latter, and the model

$$\Sigma := \left\{ x \in \mathcal{H}, x = \sum_{j=1}^J x_j, x_j \in \Sigma_j \right\} \quad (91)$$

is identified to the Cartesian product of the models  $\Sigma_1 \times \Sigma_2 \times \dots \times \Sigma_J$ . A natural regularizer for this *block structured sparsity* model is defined as follows in  $E_1 \times \dots \times E_J$ :

$$f_w : (x_1, \dots, x_J) \mapsto w_1 \|x_1\|_{\mathcal{A}_1} + \dots + w_J \|x_J\|_{\mathcal{A}_J} \quad (92)$$

with weights  $w_j > 0$ . We show that for  $w_{opt,i} = 1/\sqrt{K_i}$ , the admissible RIP constant  $\delta_\Sigma(f_{w_{opt}})$  does not depend on the ratios  $K_i/K_j$  contrary to the result for  $w_i = 1$  and simple sparsity from [11].

First we need to characterize the atomic norm  $\|\cdot\|_\Sigma$  for this model.

**Lemma 4.2** (The atomic norm  $\|\cdot\|_\Sigma$  for Cartesian products of models). *Consider  $J$  spaces  $\mathcal{H}_j$  and cone models  $\Sigma_j \subset \mathcal{H}_j$ , and their Cartesian product  $\Sigma = \Sigma_1 \times \Sigma_2 \times \dots \times \Sigma_J \subset \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_J = \mathcal{H}$ . The space  $\mathcal{H}$  is naturally equipped with the inner product:*

$$\langle (x_1, \dots, x_J), (y_1, \dots, y_J) \rangle = \sum_j \langle x_j, y_j \rangle_{\mathcal{H}_j}. \quad (93)$$

We have the characterization

$$\|(x_1, \dots, x_J)\|_\Sigma^2 = \sum_j \|x_j\|_{\Sigma_j}^2. \quad (94)$$

*Proof.* Proving the result for  $J = 2$  is enough to obtain it by induction for any  $J > 2$ . We use Fact 2.1, assuming that all convex hulls are closed to keep the proof simple. First we get a lower bound

$$\begin{aligned} \|(x_1, x_2)\|_\Sigma^2 &= \inf \left\{ \sum \lambda_i \|(u_i, v_i)\|_{\mathcal{H}}^2 : \lambda_i \in \mathbb{R}_+, \sum \lambda_i = 1, (u_i, v_i) \in \Sigma, (x_1, x_2) = \sum \lambda_i (u_i, v_i) \right\} \\ &= \inf \left\{ \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 + \sum \lambda_i \|v_i\|_{\mathcal{H}}^2 : \lambda_i \in \mathbb{R}_+, \sum \lambda_i = 1, (u_i, v_i) \in \Sigma, (x_1, x_2) = \sum \lambda_i (u_i, v_i) \right\} \\ &\geq \inf \left\{ \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 : \lambda_i \in \mathbb{R}_+, \sum \lambda_i = 1, u_i \in \Sigma_1, x_1 = \sum \lambda_i u_i \right\} \\ &\quad + \inf \left\{ \sum \lambda_i \|v_i\|_{\mathcal{H}}^2 : \lambda_i \in \mathbb{R}_+, \sum \lambda_i = 1, v_i \in \Sigma_2, x_2 = \sum \lambda_i v_i \right\} \\ &= \|x_1\|_{\Sigma_1}^2 + \|x_2\|_{\Sigma_2}^2. \end{aligned}$$

To prove the converse bound, consider  $(u_i)_{i \in I} \in \Sigma_1, (v_j)_{j \in J} \in \Sigma_2, \lambda_i > 0, \mu_j > 0$  such that  $\sum \lambda_i = 1, \sum \mu_j = 1$ .

We show that there exist  $\gamma_k, u'_k \in \Sigma_1, v'_k \in \Sigma_2$  such that  $\sum_k \gamma_k = 1$  and

$$\sum \gamma_k (\|u'_k\|_{\mathcal{H}}^2 + \|v'_k\|_{\mathcal{H}}^2) = \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 + \sum \mu_j \|v_j\|_{\mathcal{H}}^2.$$

Without loss of generality, up to renumbering, we can assume that  $J \subset I \subset \mathbb{N}$  (the case  $I \subset J$  uses the same proof). Then, with  $\gamma_k = \lambda_k, u'_k = u_k$  and  $v'_k = \sqrt{\frac{\mu_k}{\lambda_k}} v_k$  if  $k \in J$ , 0 otherwise. We have

$$\sum \gamma_k (\|u'_k\|_{\mathcal{H}}^2 + \|v'_k\|_{\mathcal{H}}^2) = \sum \lambda_k \|u_k\|_{\mathcal{H}}^2 + \sum \lambda_k \frac{\mu_k}{\lambda_k} \|v_k\|_{\mathcal{H}}^2 = \sum \lambda_i \|u_i\|_{\mathcal{H}}^2 + \sum \mu_j \|v_j\|_{\mathcal{H}}^2.$$

Thus  $\|(x_1, x_2)\|_{\Sigma}^2 \leq \|x_1\|_{\Sigma_1}^2 + \|x_2\|_{\Sigma_2}^2$ . □

For block structured sparsity, we are now equipped to bound  $\delta_{\Sigma}(f_w)$ .

**Theorem 4.2.** *Consider the block-group-sparse model  $\Sigma = \Sigma_1 \times \dots \times \Sigma_J$ ,  $J \geq 2$ , and the regularizer*

$$f_w(x_1, \dots, x_J) = \sum_{j=1}^J w_j \|x_j\|_{\mathcal{A}_j}.$$

Denoting  $\kappa_w = \left( \frac{\max(w_j \sqrt{K_j})}{\min(w_j \sqrt{K_j})} \right)$ , we have

$$\delta_{\Sigma}(f_w) \geq \frac{1}{\sqrt{2 + J\kappa_w^2}}.$$

In particular when  $w_j = \frac{1}{\sqrt{K_j}}$ , we have  $\delta_{\Sigma}(f_w) \geq \frac{1}{\sqrt{2+J}}$ .

*Proof.* Let  $z = (z_j)_{j=1,J} \in \mathcal{T}_{f_w}(\Sigma) \setminus \{0\}$ . For all  $j$ , Let  $x_j \in \mathcal{Y}_{\Sigma_j}(z_j, \mathcal{A}_j)$ . Considering  $x = (x_j)_j$ , we have  $z \in \mathcal{T}_{f_w}(x)$  and

$$\sum_{j=1,J} w_j \|x_j + z_j\|_{\mathcal{A}_j} \leq \sum_{j=1,J} w_j \|x_j\|_{\mathcal{A}_j} \leq \sum_{j=1,J} w_j \sqrt{K_j} \|x_j\|_{\mathcal{H}} \leq \max(w_j \sqrt{K_j}) \sqrt{J} \|x\|_{\mathcal{H}}. \quad (95)$$



Moreover, by Proposition 4.1,  $K_j \|x_j + z_j\|_{\mathcal{A}_j}^* \leq \|x_j\|_{\mathcal{A}_j} \leq \sqrt{K_j} \|x_j\|_{\mathcal{H}_j}$ , and by Corollary 4.1 we have :

$$\begin{aligned}
\|x + z\|_{\Sigma}^2 &= \sum_{j=1, J} \|x_j + z_j\|_{\Sigma_j}^2 \leq \sum_{j=1, J} \max \left( \|x_j + z_j\|_{\mathcal{A}_j}^2 / K_j, K_j (\|x_j + z_j\|_{\mathcal{A}_j}^*)^2 \right) \\
&\leq \sum_{j=1, J} \left( w_j^2 \|x_j + z_j\|_{\mathcal{A}_j}^2 / (w_j^2 K_j) + K_j (\|x_j + z_j\|_{\mathcal{A}_j}^*)^2 \right) \\
&\leq \left( \sum_{j=1, J} w_j \|x_j + z_j\|_{\mathcal{A}_j} / (w_j \sqrt{K_j}) \right)^2 + \sum_{j=1, J} \|x_j\|_{\mathcal{H}_j}^2 \\
&\leq \left( J \left( \frac{\max(w_j \sqrt{K_j})}{\min(w_j \sqrt{K_j})} \right)^2 + 1 \right) \|x\|_{\mathcal{H}}^2.
\end{aligned} \tag{96}$$

Thus  $\rho(x, z) = \langle x + z, x \rangle = 0$ ,  $\alpha_{\Sigma}(f_w) \leq \alpha_{\Sigma}(x, z) \leq 1 + J\kappa_w^2$ , and  $\delta_{\Sigma}(f_w) \geq 1 / \sqrt{1 + \alpha_{\Sigma}(f_w)} = 1 / \sqrt{2 + J\kappa_w^2}$ .  $\square$

Even with adjusted weights, our lower bound  $\delta_{\Sigma}(f_{w_{opt}})$  depends on  $J$ . In light of Bastounis et al. [11], this dependency could be necessary even for  $w = w_{opt}$ .

#### 4.4. Consequences and discussion

We just calculated an admissible RIP constant for the block structured sparse model (91) and regularizer  $f_w$  (92). We now specify recovery theorems and compare them to previous state of the art results.

##### 4.4.1. Dimension reduction and the RIP

Ayaz et al. [9] gave a uniform recovery result with the mixed  $\ell^1 - \ell^2$ -norm for structured compressed sensing under a RIP hypothesis. They show that a RIP constant  $\delta < \sqrt{2} - 1$  for vectors in  $\Sigma_{2K} = \Sigma_K - \Sigma_K$  guarantees the recovery of vectors from  $\Sigma_K$ . We just showed that the RIP constant of Ayaz et al. can be improved to the sharp  $\frac{1}{\sqrt{2}}$ . In [1], a model of sparsity in levels is proposed. This is in fact a block sparsity model in  $\mathcal{E}(\mathcal{A})$  with classical sparsity in each block, which is covered by the model of Section 4.3. In [11], Bastounis et al. show in the case of block sparsity that  $f(\cdot) = \sum \|\cdot\|_{\mathcal{A}_j} = \|\cdot\|_1$  (i.e., with weights  $w_j = 1$ , in this case we write  $\kappa_w = \kappa_1$  and  $\kappa$  represent the ratio of sparsity between blocks) and RIP (called there RIP in levels)  $\delta = 1 / \sqrt{J(\kappa_1 + 0.25)^2 + 1}$  on  $\Sigma - \Sigma$  guarantees recovery. This constant is improved by our constant  $\delta_{\Sigma}(f_w) \geq 1 / \sqrt{2 + J}$  when appropriately weighting the norm of each block. Our result further extends the work of Bastounis et al. to general structured sparsity. The following theorem summarizes our result:

**Theorem 4.3.** Let  $f(x_1, \dots, x_J) = \sum_{j=1}^J \|x_j\|_{\mathcal{A}_j} / \sqrt{K_j}$ . For any  $M$  that satisfies the RIP on  $\Sigma - \Sigma = \Sigma_{2K_1} \times \dots \times \Sigma_{2K_J}$  with constant  $\delta < \delta_0$  we have: for all  $x_0 \in \Sigma$ ,  $\|e\|_{\mathcal{F}} \leq \eta \leq \epsilon$ , with  $x^*$  the result of minimization (3),

$$\|x^* - x_0\|_{\mathcal{H}} \leq \|x^* - x_0\|_{\Sigma} \leq C_{\Sigma}(f, \delta)(\eta + \epsilon) \quad (97)$$

where :

- For  $J = 1$ ,  $\delta_0 = \frac{1}{\sqrt{2}}$  and  $C_{\Sigma}(f, \delta) \leq \frac{2\sqrt{1+\delta}}{1-\delta\sqrt{2}}$ .
- For  $J \geq 2$ ,  $\delta_0 = \sqrt{\frac{1}{2+J}}$  and  $C_{\Sigma}(f, \delta) \leq \frac{(1+\sqrt{1+J})\sqrt{1+\delta}}{1-\delta\sqrt{2+J}}$ .

The comparison of RIP constants is summarized in Figure 1.

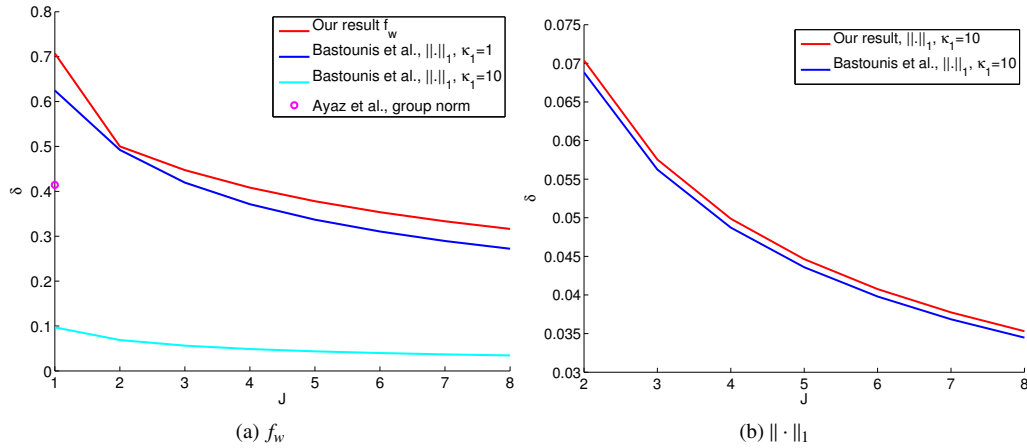


Figure 1: Comparison of our admissible RIP constant and the state of the art (Higher is better). (a) Our result is for weighted group norm with our proposed weights and block group sparsity. Bastounis et al. is for  $\ell^1$  norm and simple sparsity. Ayaz et al. is for the group norm. (b)  $\kappa_1 = 10$ . Our result is for the  $\ell^1$  norm and block sparsity, Bastounis et al. is for the  $\ell^1$  norm and block sparsity.

Ayaz et al. [9] then proceed to show that subgaussian matrices of appropriate dimensions satisfy the RIP on vectors from  $\Sigma_K - \Sigma_K$  with high probability, thus providing a sufficient number of observations for guaranteed uniform recovery of structured sparse signals.

Even in an infinite-dimensional Hilbert space, it has recently been established [33] that one can construct random linear measurement operators  $M : \mathcal{H} \rightarrow \mathbb{R}^m$  satisfying the RIP on the secant set  $\Sigma - \Sigma$  with high probability, when  $m$  is large enough compared to the covering dimension of the normalized secant set  $(\Sigma - \Sigma) \cap S(1)$ . The covering number  $N(\epsilon)$  of a set is the minimum number of ball of radius  $\epsilon$  sufficient to cover the set. If  $N(\epsilon) \leq \epsilon^{-s}$  for all  $0 < \epsilon \leq \epsilon_0$ , where  $\epsilon_0 \leq 1/2$  and  $s > 0$ , then subgaussian matrices of size  $m \times n$

satisfy the RIP with constant  $\delta$  on  $\Sigma - \Sigma$  with high probability, provided that  $m \geq \delta^{-2} O(s \log(1/\epsilon_0))$ .

For the set  $\Sigma_K$  of  $K$ -group sparse vectors, the covering number of  $(\Sigma_K - \Sigma_K) \cap S(1) = \Sigma_{2K} \cap S(1)$  is calculated in [9] when the groups have fixed size  $|g| = r$ , for all  $g \in G$ . In the Hilbert space setting considered here, for groups of possibly different sizes, the covering number of  $(\Sigma_K - \Sigma_K) \cap S(1)$  can be bounded using the maximum group size  $r := \max\{|g| : g \in G\}$ . For  $\epsilon \leq 1/2$ , we obtain  $N(\epsilon) \leq \left(\frac{3e^{r|G|}}{K\epsilon}\right)^K$ .

Observing that  $N(\epsilon) \leq (C/\epsilon)^{s_0} = \epsilon^{-2s_0} (C\epsilon)^{s_0}$  for  $\epsilon \leq 1/2$  implies  $N(\epsilon) \leq \epsilon^{-2s_0}$  for  $\epsilon \leq \epsilon_0 = \min(1/2, 1/C)$  gives us immediately the following theorem for structured sparsity with only one block ( $J = 1$ ).

**Theorem 4.4.** *For  $J = 1$ , consider  $\Sigma = \Sigma_K$  the set of  $K$ -group sparse vectors with groups  $G$ . Denote  $r = \max\{|g| : g \in G\}$ . One can construct a random (subgaussian) linear measurement operator  $M : \mathcal{H} \rightarrow \mathbb{R}^m$  that satisfies the RIP with constant  $\delta$  on vectors from  $\Sigma_{2K}$  with high probability if :*

$$m \geq \delta^{-2} O\left(Kr + K \log\left(\frac{3e|G|}{K}\right)\right). \quad (98)$$

In the case of block structured sparsity, we extend this to any number of blocks  $J$  to determine a sufficient number of (subgaussian) measurements to ensure the RIP on  $\Sigma - \Sigma$  holds with high probability.

**Theorem 4.5.** *For  $J \geq 1$ , consider  $\Sigma = \Sigma_1 \times \dots \times \Sigma_J$  with  $\Sigma_j = \Sigma_{K_j}$  the set of  $K_j$  group-sparse vectors with group  $G_j$ . Denote  $r_j = \max\{|g| : g \in G_j\}$ . One can construct a random (subgaussian) linear measurement operator  $M : \mathcal{H} \rightarrow \mathbb{R}^m$  that satisfies the RIP with constant  $\delta$  on vectors from  $\Sigma - \Sigma$  with high probability if :*

$$m \geq \delta^{-2} O\left(\sum_{j=1}^J \left(K_j r_j + K_j \log\left(\frac{3e|G_j|}{K_j}\right)\right)\right). \quad (99)$$

*Proof.* We can bound the covering number of the set  $(\Sigma - \Sigma) \cap S(1)$  by that of  $((\Sigma_1 - \Sigma_1) \cap B(1)) \times \dots \times ((\Sigma_J - \Sigma_J) \cap B(1))$ : take  $(x_1, \dots, x_J) \in (\Sigma - \Sigma) \cap S(1)$ . Then for all  $j$ ,  $\|x_j\|_{\mathcal{H}} \leq 1$ . Thus  $x_j \in (\Sigma_j - \Sigma_j) \cap B(1)$ . Note that we use the covering number of  $(\Sigma_j - \Sigma_j) \cap B(1)$  to bound the covering number of  $(\Sigma_j - \Sigma_j) \cap S(1)$ . As the covering number of a Cartesian product of sets is the product of their covering numbers [36], we have when  $\epsilon \leq 1/2$ :

$$\begin{aligned} N(\epsilon) &\leq \left(\frac{3e^{r_1|G_1|}}{K_1\epsilon}\right)^{K_1} \times \dots \times \left(\frac{3e^{r_J|G_J|}}{K_J\epsilon}\right)^{K_J} \\ &= \left(\frac{1}{\epsilon}\right)^{\sum K_j} \left(\left(\frac{3e^{r_1|G_1|}}{K_1}\right)^{K_1} \times \dots \times \left(\frac{3e^{r_J|G_J|}}{K_J}\right)^{K_J}\right)^{1/(\sum K_j) \sum K_j}. \end{aligned} \quad (100)$$

Hence  $N(\epsilon) \leq \epsilon^{-s}$  for  $0 < \epsilon \leq \epsilon_0$  with  $s = 2 \sum K_j$  and

$$\log 1/\epsilon_0 = O \left( \frac{\sum_j \left( K_j r_j + \log \left( \frac{3|G_j|}{K_j} \right) \right)}{\sum K_j} \right).$$

yields the desired result.  $\square$

Exploiting Theorem 4.5 in the context of Theorem 4.3 considering  $\delta < 1/\sqrt{2+J}$  yields a sufficient number of (subgaussian) measurements to ensure the appropriate RIP for recovery with  $f_{w_{opt}}$  and the following number of measurements

$$m \geq O \left( J \sum_{j=1}^J \left( K_j r_j + K_j \log \left( \frac{3e|G_j|}{K_j} \right) \right) \right).$$

The factor  $J$  might seem pessimistic, and we attribute its presence to the generality of the result. Should the structure of the observation matrix  $M$  be taken into account, better results shall be achieved. In fact, if  $M$  is a block diagonal matrix where each block  $M_j$  has size  $m_j \times n_j$ , uniform recovery guarantees with the  $\ell^1$ -norm hold if and only if uniform recovery holds on each block: this is possible as soon as each block  $M_j$  of  $M$  satisfies the RIP with some constant  $\delta_j < \frac{1}{\sqrt{2}}$  on  $\Sigma_j - \Sigma_j$ , which is in turn exactly equivalent to the RIP with constant  $\delta < \frac{1}{\sqrt{2}}$  on  $\Sigma - \Sigma$ . The above observation suggests that sharper results for block structured sparsity could be achievable under some structure assumption on  $M$ . A possible structure assumption could correspond to assuming the existence of a constant  $c$  such that for all  $z = (z_1, \dots, z_J) \in \Sigma_1 \times \dots \times \Sigma_J$ :

$$\frac{\|Mz\|_{\mathcal{H}}^2}{\sum \|Mz_j\|_{\mathcal{H}}^2} \leq c \quad (101)$$

With such a constraint, the dependence of  $\delta$  on the number of blocks  $J$  may be substantially decreased or even disappear in Theorem 4.3.

#### 4.4.2. Robustness to model error and infinite-dimensional context

In a series of papers [2, 4, 1, 5, 3], Adcock and Hansen propose a sampling strategy (called generalized sampling) to perform compressed sensing in an infinite-dimensional setting. With our new general framework, we naturally extend the notion of structured sparsity to this infinite-dimensional setting. In fact, we observe that, whereas the ambient space has infinite dimension, the low complexity recovery happens in the finite-dimensional space  $\mathcal{E}(\mathcal{A})$ . This space  $\mathcal{E}(\mathcal{A})$  characterizes the arbitrarily high resolution at which the problem is considered.

The main advantage of this infinite-dimensional setting is the instance optimality result where modeling error is measured in the infinite-dimensional space.

Note that generally, the  $M$ -norm needs to be bounded by  $f$  to obtain (robust) instance optimality with Lemma 3.2 from Section 3.2.

**Theorem 4.6** (Instance optimality for block structured sparsity). *Let us consider the regularizer  $f = f_w$  with the adapted weights  $w_j = 1/\sqrt{K_j}$ . Suppose  $M$  has the RIP with constant  $\delta < \delta_\Sigma(f)$  on  $\Sigma$ . Then for all  $x_0 \in \mathcal{H}$ ,  $\|e\|_{\mathcal{F}} \leq \eta \leq \epsilon$  and  $x^*$  the result of minimization (3), we have*

$$\|x^* - x_0\|_{\mathcal{H}} \leq C_\Sigma(f, \delta)(\epsilon + \eta) + D \cdot d_f(x_0, \Sigma). \quad (102)$$

We have

- for  $J = 1$

$$D \cdot d_f(x_0, \Sigma_K) := 2(1 + \sqrt{1 + \delta C_\Sigma}) \cdot \frac{d_{\|\cdot\|_{\mathcal{A}}}(x_0, \Sigma_K)}{\sqrt{K}} \quad (103)$$

- for  $J \geq 2$

$$D := 2\sqrt{2}(1 + \sqrt{1 + \delta C_\Sigma}). \quad (104)$$

*Proof.* We use a similar argument as in Theorem 3.2. Let  $z = x^* - x_0$ . Take  $x_1 = x_{0,T}$  where  $T$  is the support of the  $K$  greatest groups of  $x_0$  in each block, and let  $x_2 = -z_{T'}$  where  $T'$  is the support of the  $K$  greatest groups of  $z$  in each block. The vector  $-x_1$  is made of the best  $K_j$ -group approximation to  $x_0$  in each block, while the vector  $-x_2$  is similarly the best approximation to  $z$ . As a result  $f(x_2 + z) - f(x_2) \leq f(x_1 + z) - f(x_1)$ . As in the proof of Theorem 3.2 it follows that

$$f(x_2 + z) \leq f(x_2) + f(x_0 - x_1) + f(x_1 - x_0) = f(x_2) + 2f(x_0 - x_1). \quad (105)$$

This is classically known as the cone constraint in the case of classical sparsity, [17].

Given the above properties, we use the definition (63) of  $z'$  and obtain as in the proof of Theorem 3.2 that  $z' \in \mathcal{T}_f(\Sigma)$  and  $z - z' = \frac{2f(x_0 - x_1)}{f(x_2) + 2f(x_0 - x_1)} \cdot (x_2 + z)$ .

The rest of the proof will now deviate from that of Theorem 3.2. With the fact that for any  $u \in \mathcal{H}$ ,  $\|Mu\|_2 \leq \sqrt{1 + \delta}\|u\|_\Sigma$  (using the RIP of  $M$ , see e.g. [15, Section IV, eq. (50)]), we have

$$\|z - z'\|_{M, C_\Sigma} = \|z - z'\|_{\mathcal{H}} + C_\Sigma \|M(z - z')\|_{\mathcal{F}} \leq (1 + \sqrt{1 + \delta C_\Sigma})\|z - z'\|_\Sigma. \quad (106)$$

We will conclude by establishing the bound  $\|z - z'\|_\Sigma \leq 2\sqrt{2}f(x_0 - x_1)$ .

With the definition of  $x_2$ , by Proposition 4.1 we have  $K_j\|(x_2 + z)_j\|_{\mathcal{A}_j}^* \leq \|(x_2)_j\|_{\mathcal{A}_j}$ . Using the result from Section 4 that  $\|\cdot\|_\Sigma^2 \leq \sum_{j=1,J} \max(\|\cdot\|_{\mathcal{A}_j}/\sqrt{K_j}, \sqrt{K_j}\|\cdot\|_{\mathcal{A}_j}^*)^2$  (Corollary 4.1) and the expression of  $z - z'$ :

$$\begin{aligned}
\|z - z'\|_\Sigma^2 &\leq \frac{4f(x_0 - x_1)^2}{(f(x_2) + 2f(x_0 - x_1))^2} \sum_{j=1,J} \max(\|(x_2 + z)_j\|_{\mathcal{A}_j}/\sqrt{K_j}, \sqrt{K_j}\|(x_2 + z)_j\|_{\mathcal{A}_j}^*)^2 \\
&\leq \frac{4f(x_0 - x_1)^2}{(f(x_2) + 2f(x_0 - x_1))^2} \left( \sum_{j=1,J} \|(x_2 + z)_j\|_{\mathcal{A}_j}^2/K_j + \|(x_2)_j\|_{\mathcal{A}_j}^2/K_j \right) \\
&\leq \frac{4f(x_0 - x_1)^2}{(f(x_2) + 2f(x_0 - x_1))^2} \left( \left( \sum_{j=1,J} \|x_2 + z\|_{\mathcal{A}_j}/\sqrt{K_j} \right)^2 + \left( \sum_{j=1,J} \|x_2\|_{\mathcal{A}_j}/\sqrt{K_j} \right)^2 \right) \quad (107) \\
&= \frac{4f(x_0 - x_1)^2}{(f(x_2) + 2f(x_0 - x_1))^2} \cdot (f(x_2 + z)^2 + f(x_2)^2) \\
&\stackrel{(105)}{\leq} 8f(x_0 - x_1)^2
\end{aligned}$$

When  $J = 1$  we let the reader check that the constant 8 above can be replaced with 4.  $\square$

**Remark 4.1.** *The above proof extends to rank- $r$  matrices by considering proper adaptations of the choice of  $x_1$  and  $x_2$  and Proposition 5.1 in the next section. With  $f$  the nuclear norm, this yields the robustness constant*

$$D := 2(1 + \sqrt{1 + \delta C_\Sigma})/\sqrt{r}. \quad (108)$$

Beyond these examples, to make sense of Theorem 3.2 in the infinite-dimensional setting, the domain where  $f$  is finite must be extended outside of  $\mathcal{E}(\Sigma)$  while keeping a finite robustness constant  $D$ . This can be done on a case-by-case basis when properties of  $M$  and  $f$  allow to conclude.

For example, as Adcock and Hansen in [5], consider the following setting:  $\mathcal{H} = \ell^2(\mathbb{N})$  with Hilbert basis  $(e_i)_{i=1,+\infty}$ . Consider  $\Sigma$  a block sparsity model in  $(e_1, \dots, e_N)$ . Let  $f = \|\cdot\|_1$ . Then  $f$  is an extension of the definition of  $f_w$  in  $\mathcal{E}(\Sigma)$  to the whole space  $\mathcal{H}$  (with  $w_j = 1$  for all  $j$ ). In [5],  $M$  is a collection of (Fourier) measurements that have a *strong balancing property*. The important fact here is that this property requires  $\|M^H M\|_\infty \leq C'$  where  $\|\cdot\|_\infty$  is the maximum of the  $\ell^\infty$ -norms of the coefficients of  $M^H M$  (where  $M^H$  is the Hermitian conjugate of  $M$ ). With such an hypothesis, for any  $u \in \mathcal{H}$ , we have:  $\|Mu\|_2^2 = |\langle u, M^H Mu \rangle| \leq \|M^H Mu\|_\infty \|u\|_1 \leq \|M^H M\|_\infty \|u\|_1 \|u\|_1 \leq C' \|u\|_1^2$ . Thus in this case the  $M$ -norm is bounded by the  $\ell^1$ -norm :  $\|\cdot\|_{M,C} \leq (1 + C\sqrt{C'})\|u\|_1$ .

## 5. Worked examples

As pointed out in Section 4, the sharp RIP conditions  $\delta \leq 1/\sqrt{2}$  for classical sparse recovery with the  $\ell^1$  norm are covered by Theorem 1.2. We now demonstrate the flexibility of the general framework of this paper by applying it to other examples.

### 5.1. Stable recovery from linear subspaces

Let  $\mathcal{H}$  be a Hilbert space and  $\Sigma \subset \mathcal{H}$  be a linear subspace of  $\mathcal{H}$ . With such a model, it is easy to characterize the atomic norm  $\|\cdot\|_\Sigma$ : we have  $\|u\|_\Sigma = \|u\|_{\mathcal{H}}$  if  $u \in \mathcal{E}(\Sigma) = \Sigma$ ,  $\|u\|_\Sigma = +\infty$  otherwise. As a result we can characterize  $\delta_\Sigma(f)$  for two natural regularizers.

*Regularizing with the indicator function.* Let  $f$  be the indicator function of  $\Sigma$ :  $f(u) = 0$  if  $u \in \Sigma$ ,  $f(u) = +\infty$  otherwise. One can easily check that the set of descent vectors is  $\mathcal{T}_f(\Sigma) = \Sigma - \Sigma = \Sigma + \Sigma = \Sigma$ , hence for any  $z \in \mathcal{T}_f(\Sigma) \setminus \{0\}$ , we can take  $x = -z \in \Sigma$  which yields  $\alpha(x, z) = 0$  and  $\rho(x, z) = 0$ . We conclude that  $\delta_\Sigma(f) = 1$ .

*Regularizing with the atomic norm  $\|\cdot\|_\Sigma$ .* Consider  $z \in \mathcal{T}_\Sigma(\Sigma)$ : by definition, there is  $x \in \Sigma$  such that  $\|x + z\|_\Sigma \leq \|x\|_\Sigma = \|x\|_{\mathcal{H}}$ . Since  $\|u\|_\Sigma = +\infty$  when  $u \notin \Sigma$ , this implies  $x + z \in \Sigma$  and  $z \in \Sigma - \Sigma$ . Thus,  $\mathcal{T}_\Sigma(\Sigma) \subset \Sigma - \Sigma = \Sigma$  and, as above, we conclude that  $\delta_\Sigma(\|\cdot\|_\Sigma) = 1$ .

### 5.2. Low rank matrix recovery

Let  $\mathcal{H}$  be the set of  $m \times n$  (real or complex) matrices equipped with the Frobenius norm and the associated inner product  $\langle x, y \rangle = \text{trace}(xy^T)$  (respectively  $\text{Re}(\text{trace}(xy^H))$  for the complex case),  $\Sigma = \Sigma_r$  be the model set of  $m \times n$  matrices of rank at most  $r$ , and  $\mathcal{A} = \Sigma_1 \cap S(1)$  the set of normalized rank one-matrices. We consider the nuclear norm  $f(\cdot) = \|\cdot\|_{\mathcal{A}} = \|\cdot\|_*$  as the regularizer and verify that the low-rank matrix recovery result of [16] is covered by the general framework of this paper.

We decompose any matrix  $z$  as  $z = (z - z_r) + z_r$  where  $z_r$  is obtained by truncating the singular value decomposition (SVD) of  $z$  to its  $r$  largest singular values ( $z_r$  is not unique if some singular values are equal). We first show that we have  $\|z - z_r\|_* - \|z_r\|_* \leq 0$  as soon as  $z$  is a descent matrix,  $z \in \mathcal{T}_f(\Sigma)$ . We use this property to prove that  $\delta_\Sigma(f) \geq 1/\sqrt{2}$ .

**Proposition 5.1.** *Consider a matrix  $z \in \mathcal{T}_f(\Sigma)$  and its SVD  $z = USV$ . Let  $z_r = US_rV$  where  $S_r$  is a restriction of the matrix  $S$  to its  $r$  largest diagonal entries (usually the  $r$  first if there is no tie). Then  $\|z - z_r\|_* - \|z_r\|_* \leq 0$ .*

*Proof.* By definition, since  $z \in \mathcal{T}_f(\Sigma) \setminus \{0\}$  there exists  $x \in \Sigma$  such that  $\|x+z\|_* - \|x\|_* \leq 0$ . For any  $m \times n$  matrix  $A$ , denote  $\sigma_i(A)$  the  $i$ -th singular value ( $\sigma_1(A) \geq \dots \geq \sigma_n(A) \geq 0$ ). From [31, Chapter 7], for any matrices  $A, B$ , the singular values have the following property:  $\|A - B\|_* = \sum_{i=1}^n \sigma_i(A - B) \geq \sum_{i=1}^n |\sigma_i(A) - \sigma_i(B)|$ . Taking  $A = z$  and  $B = -x$  and observing that  $\sigma_i(x) = \sigma_i(-x) = 0$  for  $i > r$  yields

$$\begin{aligned} \|x+z\|_* - \|x\|_* &= \sum_{i=1,n} \sigma_i(x+z) - \sum_{i=1,n} \sigma_i(x) \\ &\geq \sum_{i=1,n} |\sigma_i(z) - \sigma_i(-x)| - \sum_{i=1,n} \sigma_i(x) \\ &= \sum_{i=1,r} |\sigma_i(z) - \sigma_i(-x)| - \sum_{i=1,r} \sigma_i(x) + \sum_{i=r+1,n} \sigma_i(z) \\ &\geq - \sum_{i=1,r} \sigma_i(z) + \sum_{i=r+1,n} \sigma_i(z) = \|z - z_r\|_* - \|z_r\|_*. \end{aligned}$$

□

**Proposition 5.2.** *With  $\Sigma = \Sigma_r$  the set of rank  $r$  matrices and  $\|\cdot\|_*$  the nuclear norm, we have  $\delta_\Sigma(\|\cdot\|_*) \geq 1/\sqrt{2}$ .*

*Proof.* Let  $z \in \mathcal{T}_{\|\cdot\|_*}(\Sigma_r)$  a descent matrix for the nuclear norm, and  $x = -z_r$ . With Proposition 5.1, we have  $\|z - z_r\|_* - \|z_r\|_* \leq 0$ . Since  $\Sigma$  is a UoS we use (30) to compute

$$\delta_\Sigma(x, z) = \frac{\|x\|_2^2}{\|x\|_2 \sqrt{\|x+z\|_\Sigma^2 - \|x\|_2^2 + 2\|x\|_2^2}} = \frac{1}{\sqrt{\|z - z_r\|_\Sigma^2 / \|z_r\|_2^2 + 1}} \quad (109)$$

Recalling that the operator norm of a matrix  $A$  is  $\|A\|_{op} = \max_i \sigma_i(A)$  and that the SVD gives an optimal atomic decomposition for the nuclear norm with pairwise orthogonal atoms, Lemma 4.1 yields

$$\|z - z_r\|_\Sigma \leq \max(\|z - z_r\|_* / \sqrt{r}, \|z - z_r\|_{op} \sqrt{r}). \quad (110)$$

We have  $r\|z - z_r\|_{op} \leq \|z_r\|_*$  (similarly to Proposition 4.1) and  $\|z - z_r\|_* \leq \|z_r\|_*$ , hence  $\|z - z_r\|_\Sigma \leq \|z_r\|_* / \sqrt{r} \leq \|z_r\|_2$  and we conclude that  $\delta_\Sigma(x, z) \geq 1/\sqrt{2}$ . Consequently,  $\delta_\Sigma(\|\cdot\|_*) \geq 1/\sqrt{2}$  □

With Proposition 5.2, we have verified: when  $M$  satisfies the RIP with constant  $\delta < \frac{1}{\sqrt{2}}$  on  $\Sigma_r - \Sigma_r = \Sigma_{2r}$  (the set of matrices with rank at most  $2r$ ), nuclear norm regularization is guaranteed to yield stable uniform recovery of matrices in  $\Sigma_r$ .



## 6. Regularizing with the model's atomic norm $\|\cdot\|_\Sigma$ ?

The atomic norm  $\|\cdot\|_\Sigma$  intrinsically associated to the considered model set  $\Sigma$  appears in the definition of the RIP constant  $\delta_\Sigma(f)$ . Can this norm be directly used as a regularizer  $f$ , with recovery guarantees ?

In simple cases such as  $k$ -sparse vector recovery with  $k = 1$ , we have seen that regularizing with  $f(\cdot) = \|\cdot\|_\Sigma$  indeed yields uniform recovery. We show below that this “natural” regularizer can still be used for arbitrary cone models somehow close to the set of one-sparse vectors: unions of one-dimensional half-lines. We then show examples where it is impossible to achieve uniform recovery with  $\|\cdot\|_\Sigma$ .

### 6.1. Recovery of a finite union of one-dimensional half-spaces

Consider  $\Sigma$  a cone such that  $\Sigma \cap S(1)$  finite:  $\Sigma$  is a finite union of one-dimensional half-spaces (half-lines). A particular example is the set of  $k$ -sparse vectors with  $k = 1$ . In such a simple setting, for any  $z \in \mathcal{E}(\Sigma)$  there is an optimum atomic decomposition  $c_i \in \mathbb{R}_+, a_i \in \Sigma \cap S(1)$  such that  $\sum c_i a_i = z$  and  $\|z\|_\Sigma = \sum c_i$ . Because  $\Sigma \cap S(1)$  is finite, the coherence

$$\mu(\Sigma) := \max_{x,y \in \Sigma \cap S(1), x \neq \pm y} |\langle x, y \rangle| \quad (111)$$

satisfies  $\mu(\Sigma) < 1$ . We use it to show that uniform recovery is possible by regularization with  $\|\cdot\|_\Sigma$ .

**Proposition 6.1.** *If  $\Sigma$  is a cone such that  $\Sigma \cap S(1)$  is a finite subset of  $S(1)$ , then*

$$\delta_\Sigma(\|\cdot\|_\Sigma) \geq \frac{2(1 - \mu(\Sigma))}{3 + 2\mu(\Sigma)} > 0. \quad (112)$$

*If additionally  $\Sigma$  is a UoS we have*

$$\delta_\Sigma(\|\cdot\|_\Sigma) \geq \frac{1 - \mu(\Sigma)}{\sqrt{2(1 + \mu(\Sigma))}} > 0. \quad (113)$$

*Proof.* Consider a descent vector  $z \in \mathcal{T}_\Sigma(\Sigma) \setminus \{0\}$ . By definition there is  $x \in \Sigma$  such that  $\|x+z\|_\Sigma \leq \|x\|_\Sigma = \|x\|_{\mathcal{H}}$ , and we can decompose  $x+z$  as a positive linear combination of  $a_i$  (otherwise we would have  $\|x+z\|_\Sigma = +\infty$ ). Let  $x+z = \sum c_i a_i$  be an optimal atomic decomposition of  $x+z$  such that  $\|x+z\|_\Sigma = \sum c_i$ ,  $c_i \geq 0$  and  $a_{i_1} \neq \pm a_{i_2}$  for any  $i_1, i_2$ . Let  $i_0$  such that  $x = \|x\|_{\mathcal{H}} a_{i_0}$  and  $x' := x - c_{i_0} a_{i_0} = (\|x\|_{\mathcal{H}} - c_{i_0}) a_{i_0}$ . Since  $x' + z = \sum_{i \neq i_0} c_i a_i$  we have

$$\|x' + z\|_\Sigma \leq \sum_{i \neq i_0} c_i \leq \sum_i c_i - c_{i_0} = \|x+z\|_\Sigma - c_{i_0} \leq \|x\|_{\mathcal{H}} - c_{i_0}. \quad (114)$$

Thus  $\|x\|_{\mathcal{H}} \geq c_{i_0}$  and  $x' \in \Sigma$  with  $\|x\|_{\mathcal{H}} - c_{i_0} = \|x'\|_{\mathcal{H}} \geq \|x' + z\|_{\Sigma}$ . Moreover

$$|\langle x', x' + z \rangle| = |\langle x', \sum_{i \neq i_0} c_i a_i \rangle| \leq \sum_{i \neq i_0} c_i |\langle x', a_i \rangle| \leq (\sum_{i \neq i_0} c_i) \mu \|x'\|_{\mathcal{H}} \leq \mu \|x'\|_{\mathcal{H}}^2. \quad (115)$$

We have thus found  $x' \in \Sigma$  such that  $(1 + \mu)\|x'\|_{\mathcal{H}}^2 \geq -\langle x', z \rangle \geq (1 - \mu)\|x'\|_{\mathcal{H}}^2$  and, with the expression of  $\delta_{\Sigma}(\|\cdot\|_{\Sigma})$  for cones (Equation (6)):

$$\delta_{\Sigma}(\|\cdot\|_{\Sigma}) \geq \frac{2(1 - \mu)\|x'\|_{\mathcal{H}}^2}{\|x' + z\|_{\Sigma}^2 + 2(1 + \mu)\|x'\|_{\mathcal{H}}^2} \geq \frac{2(1 - \mu)}{3 + 2\mu}. \quad (116)$$

If additionally,  $\Sigma$  is a UoS, we have (Equation (5))

$$\delta_{\Sigma}(\|\cdot\|_{\Sigma}) \geq \frac{(1 - \mu)\|x'\|_{\mathcal{H}}^2}{\|x'\|_{\mathcal{H}} \sqrt{\|x' + z\|_{\Sigma}^2 - \|x'\|_{\mathcal{H}}^2 + 2(1 + \mu)\|x'\|_{\mathcal{H}}^2}} \geq \frac{1 - \mu}{\sqrt{2(1 + \mu)}}. \quad (117)$$

□

When  $\mu = 0$ ,  $\delta_{\Sigma}(\|\cdot\|_{\Sigma}) \geq 2/3$  for cones and  $\delta_{\Sigma}(\|\cdot\|_{\Sigma}) \geq 1/\sqrt{2}$  for UoS. This last case is exactly the result for 1-sparse vectors recovery and the  $\ell^1$ -norm ( $\|\cdot\|_{\Sigma}$  is the  $\ell^1$ -norm).

## 6.2. Recovery of a finite point cloud

Let  $\Sigma$  be a finite point cloud containing  $r$  points in a Hilbert space  $\mathcal{H}$ . Consider  $\mathcal{A} = (\mathbb{R}_+ \Sigma) \cap S(1)$ , the set of normalized elements of  $\Sigma$ . Then  $\|\cdot\|_{\mathcal{A}} = \|\cdot\|_{\Sigma'}$  where  $\Sigma' = \mathbb{R}_+ \Sigma$ . We just saw that elements of  $\Sigma' \supset \Sigma$  can be stably recovered by regularization with  $\|\cdot\|_{\Sigma'}$  provided that the measurement operator  $M$  satisfies the RIP on the secant set  $\Sigma' - \Sigma' = \mathbb{R}_+ \Sigma - \mathbb{R}_+ \Sigma$  with an appropriate constant  $\delta$ . From [33], we just need to calculate the covering number of the normalized secant set  $(\mathbb{R}_+ \Sigma - \mathbb{R}_+ \Sigma) \cap S(1)$  to construct random subgaussian measurements satisfying the desired RIP on this set with high probability.

This normalized secant set is the intersection of the unit sphere with a finite union of  $\binom{r}{2}$  two-dimensional subspaces. Thus for  $0 < \epsilon \leq 1/2$ ,  $(\mathbb{R}_+ \Sigma - \mathbb{R}_+ \Sigma) \cap S(1)$  is covered by  $N(\epsilon)$  balls with :

$$N(\epsilon) \leq \binom{r}{2} \left(\frac{C}{\epsilon}\right)^2 \leq \left(\frac{re}{2}\right)^2 \left(\frac{C}{\epsilon}\right)^2.$$

Consequently, from [33], it is possible to construct a random linear operator  $M : \mathcal{H} \rightarrow \mathbb{R}^m$  with  $m \geq O(\log(r)/\delta^2)$  satisfying with high probability the RIP with constant  $\delta$  on the secant set. This number of

measurements is of the same order as in the non-uniform guarantees from [21] provided that the required RIP constant  $\delta$  does not depend on the dimensions of the problem (see next section for an example), which is not obvious given the possible dependency on  $\mu(\Sigma')$ . In the Johnson-Lindestrauss Lemma, the existence of measurement operators with RIP on  $\Sigma - \Sigma$  (as opposed to  $\Sigma' - \Sigma'$ ) is guaranteed provided  $m \geq O(\log(r)/\delta^2)$  [32]. In our case, the same dimension relations are required. We however give a *convex* decoder  $\|\cdot\|_{\Sigma'}$  that can recover elements of the point cloud. This decoder may not be easily computable if no other hypothesis is available on the model set.

### 6.3. Recovery of permutation matrices

Based on the above generic analysis for finite point clouds, we can deduce a number of measurements sufficient for uniform recovery of permutation matrices in  $\mathbb{R}^{n \times n}$  (thus  $\mathcal{H}$  is a real Hilbert space in this case). In this setting,  $\Sigma'_n = \mathbb{R}_+ \cdot \Sigma_n$  where the set  $\Sigma_n$  is made of  $r = n!$  points, and we deduce that  $m \geq O(n \log(n)/\delta^2)$  measurements are sufficient for uniform recovery. It leaves us however to determine  $\delta_{\Sigma'_n}(\|\cdot\|_{\Sigma'_n})$ .

If we use the generic correlation bound of  $\delta_{\Sigma'_n}(\|\cdot\|_{\Sigma'_n})$  from Proposition 6.1, the coherence depends on the dimension of the problem: we let the reader check that  $\mu(\Sigma'_n) = 1 - 2/n$  and that  $\delta_{\Sigma'_n}(\|\cdot\|_{\Sigma'_n}) = O(1/n)$ , suggesting  $m \geq O(n^3 \log(n))$  which is even larger than the ambient dimension  $n^2$ .

In fact, we can bound  $\delta_n = \delta_{\Sigma'_n}(\|\cdot\|_{\Sigma'_n})$  more precisely by a constant independent of the ambient dimension, showing that stable uniform recovery of permutation matrices with the regularizer  $\|\cdot\|_{\Sigma'_n}$  is possible with  $O(n \log(n))$  measurements. This qualitatively extends to *uniform stable recovery* the non-uniform recovery results of [21], and also shows that the dimension argument holds beyond Gaussian measurements.

**Proposition 6.2.** *Let  $\Sigma$  be the set of  $n \times n$  permutation matrices and  $\Sigma' = \mathbb{R}^+ \Sigma$ . We have*

$$\delta_{\Sigma'}(\|\cdot\|_{\Sigma'}) \geq \frac{2}{3}. \quad (118)$$

*Proof.* Here  $\Sigma'$  is a cone, and  $\|\cdot\|_{\Sigma'}$  is the gauge generated by the Birkhoff polytope of bi-stochastic matrices (non-negative matrices which sums of values over each row and column equal 1). The norm  $\|x\|_{\Sigma'}$  is thus infinite unless  $x = \alpha u$  with  $\alpha \geq 0$  and  $u$  is bi-stochastic, in which case  $\|x\|_{\Sigma'} = \alpha$  is the sum of the entries of  $x$  over the first row (or any other row or column since all such sums are equal).

Let  $z \in \mathcal{T}_{\Sigma'}(\Sigma') \setminus \{0\}$ . There is  $x \in \Sigma'$  such that  $\|x + z\|_{\Sigma'} \leq \|x\|_{\Sigma'} = \|x\|_{\mathcal{H}} < \infty$  hence  $x + z = \alpha u$  for  $0 \leq \alpha \leq \|x\|_{\mathcal{H}}$  and some bi-stochastic matrix  $u$ . In particular we must have  $z(i, j) \geq -x(i, j)$  for all row and column indices  $i, j$ , as well as  $\sum_j (x(i, j) + z(i, j)) = \|x + z\|_{\Sigma'} \leq \|x\|_{\Sigma'} = \sum_j x(i, j)$  for all  $i$ . This implies

$\sum_j z(i, j) \leq 0$  for all  $i$ . Thus each line (and similarly each column) of  $z$  has exactly one negative coordinate and  $z = \alpha u - \beta v$  where  $u$  is bi-stochastic,  $v$  is a permutation matrix and  $\alpha \leq \beta = \|x\|_{\Sigma'}$ . Reciprocally, all such  $z$  can be checked to be elements of  $\mathcal{T}_{\Sigma'}(\Sigma')$ . Without loss of generality we can consider the case  $\beta = 1$ ,  $0 < \alpha \leq 1$  and (up to permuting the lines and columns of all considered matrices) we can assume that  $v$  is the identity matrix.

We are now going to choose  $x_0 = \beta_0 v$  to maximize  $\delta_{\Sigma'}(x_0, z)$ . First, we wish to ensure that  $\|x_0 + z\|_{\Sigma'} < +\infty$ . This is equivalent to  $\beta_0 v + \alpha u - v \geq 0$  entry-wise. The only constraint is on the diagonal and reads  $\beta_0 - 1 + \alpha u(i, i) \geq 0$  for all  $i$ , which we rewrite  $\beta_0 \geq 1 - \alpha u_{\min}$  with  $u_{\min} := \min_i u(i, i)$ . We then compute using (6)

$$\begin{aligned} \delta_{\Sigma'}(x_0, z) &= \frac{-2\operatorname{Re}\langle x_0, z \rangle}{\|x_0 + z\|_{\Sigma'}^2 - 2\operatorname{Re}\langle x_0, z \rangle} = \frac{2(\beta_0\|v\|_2^2 - \beta_0\alpha\langle v, u \rangle)}{(\beta_0 - 1 + \alpha)^2 + 2(\beta_0\|v\|_2^2 - \beta_0\alpha\langle v, u \rangle)} \\ &= \frac{2\beta_0(n - \alpha\operatorname{tr}(u))}{(\beta_0 - 1 + \alpha)^2 + 2\beta_0(n - \alpha\operatorname{tr}(u))} = \frac{n - \alpha\operatorname{tr}(u)}{\frac{(\beta_0 - 1 + \alpha)^2}{2\beta_0} + (n - \alpha\operatorname{tr}(u))}. \end{aligned}$$

Since  $\beta_0 \mapsto (\beta_0 - 1 + \alpha)^2/\beta_0$  is minimized for  $\beta_0 = 1 - \alpha$  and increasing for  $\beta_0 > 1 - \alpha$ , and since  $1 - \alpha u_{\min} \geq 1 - \alpha$ , the above expression is minimized at  $\beta_0 = 1 - \alpha u_{\min}$  yielding

$$\delta_{\Sigma}(z) \geq \frac{n - \alpha\operatorname{tr}(u)}{\frac{(\alpha(1 - u_{\min}))^2}{2(1 - \alpha u_{\min})} + (n - \alpha\operatorname{tr}(u))} = \frac{1}{\frac{1}{2} \cdot \frac{(\alpha(1 - u_{\min}))^2}{(1 - \alpha u_{\min})(n - \alpha\operatorname{tr}(u))} + 1}$$

We look for the supremum of  $\frac{(\alpha(1 - u_{\min}))^2}{(1 - \alpha u_{\min})(n - \alpha\operatorname{tr}(u))}$  over bi-stochastic  $u$  and  $0 \leq \alpha \leq 1$ , which is obviously the supremum for  $0 < \alpha \leq 1$ . Considering  $\gamma = 1/\alpha \in [1, +\infty)$  we observe that

$$\left( \frac{(\alpha(1 - u_{\min}))^2}{(1 - \alpha u_{\min})(n - \alpha\operatorname{tr}(u))} \right)^{-1} = \frac{n(\gamma - u_{\min})(\gamma - \operatorname{tr}(u)/n)}{(1 - u_{\min})^2}$$

is a second degree polynomial with roots  $0 \leq u_{\min} \leq \operatorname{tr}(u)/n \leq 1$ . Thus this expression reaches its minimum at  $\gamma = 1$ , where its value is

$$\frac{n - \operatorname{tr}(u)}{1 - u_{\min}} = \frac{1 - u_{\min} + n - 1 - (\operatorname{tr}(u) - u_{\min})}{1 - u_{\min}} = 1 + \frac{n - 1 - (\operatorname{tr}(u) - u_{\min})}{1 - u_{\min}} \geq 1 \quad (119)$$

where we used the fact that for all  $i$ ,  $u(i, i) \leq 1$  which implies  $\text{tr}(u) - u_{\min} \leq n - 1$ . Finally,

$$\delta_{\Sigma'}(z) \geq \frac{1}{\frac{1}{2} \cdot \frac{1}{\inf_{\gamma \geq 1} \frac{n(\gamma - u_{\min})(\gamma - \text{tr}(u)/n)}{(1 - u_{\min})^2}} + 1} \geq \frac{1}{\frac{1}{2} + 1} = \frac{2}{3}. \quad (120)$$

Since this holds for any  $z \in \mathcal{T}_{\Sigma'}(\Sigma')$  we conclude that  $\delta_{\Sigma'}(\|\cdot\|_{\Sigma'}) \geq 2/3$ .  $\square$

#### 6.4. Limits of $\|\cdot\|_{\Sigma}$ : structured sparsity and low rank matrix recovery

In two examples –structured sparsity and low rank matrix recovery– we now show that we cannot use  $\|\cdot\|_{\Sigma}$  as a regularizer to perform uniform recovery. With the set  $\mathcal{T}_{\Sigma}(\Sigma)$ , we can characterize when recovery with  $f(\cdot) = \|\cdot\|_{\Sigma}$  in minimization (2) and at least one dimension-reducing linear operator  $M$  is *impossible*: if  $\mathcal{T}_{\Sigma}(\Sigma) = \mathcal{E}(\Sigma)$  then for any  $M$  with a non-trivial null space we have  $(\ker M \setminus 0) \cap \mathcal{T}_f(\Sigma) \neq \emptyset$  and recovery with  $M$  is not possible.

As we will see, this is the case for  $k$ -sparse vectors with  $k \geq 2$ , establishing that the norm  $\|\cdot\|_{\Sigma}$  considered in [8] (named  $k$ -support norm there) does *not* permit uniform recovery of all such sparse vectors. We show the unit ball of this norm in Figure 2. We observe that the span of the set of descent vectors at 1-sparse vectors are half-spaces, preventing any hope of recovery of these vectors. The norm  $\|\cdot\|_{\Sigma}$  has also been considered in [35] for simultaneously sparse and low rank matrix recovery. The limit case of the low rank matrices shows that uniform recovery with  $\|\cdot\|_{\Sigma}$  is not generally possible in this case either.

**Proposition 6.3** (Structured sparsity and the norm  $\|\cdot\|_{\Sigma}$ ). *For  $K \geq 2$ , uniform recovery of  $K$ -group sparse vectors with the norm  $\|\cdot\|_{\Sigma}$  as a regularizer is impossible.*

*Proof.* If  $K \geq 2$ , we prove that  $\mathcal{T}_{\Sigma}(\Sigma) = \mathcal{E}(\Sigma)$ . By contradiction, suppose there is  $z \in \mathcal{T}_{\Sigma}(\Sigma)^c$ . Consider a group  $H = \{h\}$  and  $x_h$  a normalized 1-group sparse vector. We have  $\|x_h\|_{\Sigma} = \|x_h\|_{\mathcal{H}} = 1$ , and  $x_h \in \Sigma_1 \subset \Sigma_K \subset \Sigma$ . Since  $z \in \mathcal{T}_{\Sigma}(\Sigma)^c$  we have for all  $\lambda \in \mathbb{R} \setminus \{0\}$ :

$$1 = \|x_h\|_{\Sigma} < \|x_h + \lambda z\|_{\Sigma}.$$

We now upper bound  $\|x_h + \lambda z\|_{\Sigma}$  by exhibiting an atomic decomposition of  $x_h + \lambda z$ . Let  $x_h + \lambda z = v_h + \sum_{g \in [H^c]} v_g$  be the canonical decomposition of  $x_h + \lambda z$  into 1-group sparse vectors :  $v_h, v_g \in \Sigma_1$ . Remark that  $v_h = x_h + \lambda z_h$  and  $v_g = z_g$ . For each  $g_i \in H^c$  (with  $1 \leq i \leq |H^c|$ ), let  $u_i \in \Sigma_2 \subset \Sigma_K = \Sigma$  such that  $u_i = \frac{1}{|H^c|} v_h + v_{g_i}$ . Then

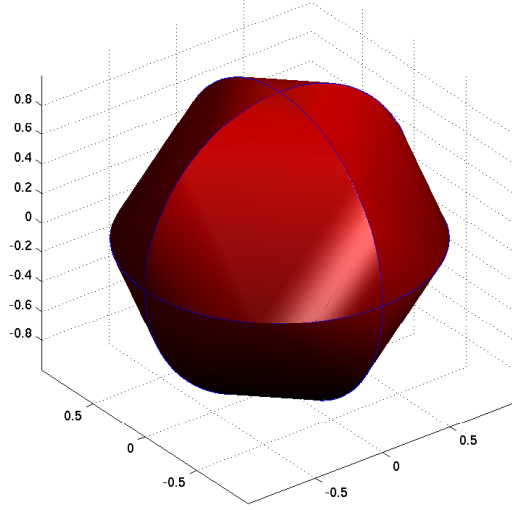


Figure 2: The unit ball of the norm  $\|\cdot\|_\Sigma$  for 2-sparse vectors in 3D (K-support norm).

$x_h + \lambda z = \sum_{i=1}^{|H^c|} u_i$  hence  $\|x_h + \lambda z\|_\Sigma \leq \sum_{i=1}^{|H^c|} \|u_i\|_{\mathcal{H}}$  and

$$\begin{aligned}
 1 &< \sum_{i=1}^{|H^c|} \|u_i\|_{\mathcal{H}} = \frac{1}{|H^c|} \sum_{i=1}^{|H^c|} \|x_h + \lambda z_h + \lambda |H^c| z_{g_i}\|_{\mathcal{H}} \\
 &= \frac{1}{|H^c|} \sum_{i=1}^{|H^c|} \sqrt{\|x_h + \lambda z_h\|_{\mathcal{H}}^2 + |\lambda|^2 |H^c|^2 \|z_{g_i}\|_{\mathcal{H}}^2} \\
 &= \frac{1}{|H^c|} \sum_{i=1}^{|H^c|} \sqrt{\|x_h\|_{\mathcal{H}}^2 + 2\lambda \langle x_h, z_h \rangle + \lambda^2 (\|z_h\|_{\mathcal{H}}^2 + |H^c|^2 \|z_{g_i}\|_{\mathcal{H}}^2)} := F(\lambda)
 \end{aligned} \tag{121}$$

We calculate the derivative  $F'(0)$  :

$$\begin{aligned}
 F'(\lambda) &= -\frac{1}{|H^c|} \sum_{i=1}^{|H^c|} \frac{2\langle x_h, z_h \rangle + 2\lambda (\|z_h\|_{\mathcal{H}}^2 + |H^c|^2 \|z_{g_i}\|_{\mathcal{H}}^2)}{2\sqrt{\|x_h\|_{\mathcal{H}}^2 + 2\lambda \langle x_h, z_h \rangle + \lambda^2 (\|z_h\|_{\mathcal{H}}^2 + |H^c|^2 \|z_{g_i}\|_{\mathcal{H}}^2)}} \\
 F'(0) &= -\frac{\langle x_h, z_h \rangle}{\|x_h\|_{\mathcal{H}}}
 \end{aligned} \tag{122}$$

Besides, since  $F(\lambda) > 1 = F(0)$  for all  $\lambda \neq 0$ , we must have  $F'(0) = 0$ , i.e.  $\langle x_h, z_h \rangle = 0$ .

This being true for all normalized 1-group sparse vectors with support  $H = \{h\}$ ,  $x_h$ , we must have  $z = z_{H^c}$ . This being true for all 1-group sparse supports  $H$ , implies  $z = 0$ . However  $z = 0$  is *not* in  $\mathcal{T}_\Sigma(\Sigma)^c$  therefore

contradicting our assumption. Consequently,  $\mathcal{T}_\Sigma(\Sigma)^c = \emptyset$  and  $\mathcal{T}_\Sigma(\Sigma) = \mathcal{E}(\Sigma)$ .  $\square$

The same fact holds with low rank matrices.

**Proposition 6.4** (Low rank matrix recovery and the norm  $\|\cdot\|_\Sigma$ ). *For  $r \geq 2$ , uniform recovery of  $\Sigma_r$  (the set of matrices of rank at most  $r$ ) with  $\|\cdot\|_\Sigma$  is impossible.*

*Proof.* The proof is almost identical to that of Proposition 6.3, by using the singular value decomposition of  $z$  and considering  $x + z$  where  $x$  is the rank-one matrix associated with the largest singular value of  $z$ .  $\square$

## 7. Discussion and future work

In this article, we gave a framework to assess the compatibility of a regularizer  $f$  with a given generic model set  $\Sigma$  (a cone). The number of examples that are treated show the flexibility and the effectiveness of this framework. Some additional questions emerge naturally from this work.

*Sharpness of  $\delta_\Sigma(f)$ .* The sharpness of RIP constants can be considered in two ways : a weak sharpness or a strong sharpness.

Given a model  $\Sigma$  and a regularizer  $f$ , we say that the RIP constant  $\delta_\Sigma^{strong}(f)$  has *strong sharpness* if the RIP with  $\delta < \delta_\Sigma^{strong}(f)$  implies uniform recovery from  $\Sigma$  with  $f$ , and there exists linear operators with RIP constant arbitrarily close to  $\delta_\Sigma^{strong}(f)$  for which uniform recovery from  $\Sigma$  with  $f$  is impossible.

In the context of  $K$ -sparse recovery in dimension  $N$ , we consider a family of models  $\Sigma_K^N$  and of regularizers  $f_N(\cdot) = \|\cdot\|_{\ell_1^N}$ . The constant  $\delta^{weak}$  has *weak sharpness* over this family of models and regularizers if the RIP with  $\delta < \delta^{weak}$  implies uniform recovery from  $\Sigma_K^N$  with  $f_N$  for any  $K, N$ , and there exists dimensions  $K, N$  and matrices with RIP constant arbitrarily close to  $\delta^{weak}$  for which uniform recovery from  $\Sigma_K^N$  with  $f_N$  is impossible. The notion of weak sharpness extends to many other families of models and regularizers, e.g. for  $r$ -rank recovery with the nuclear norm, or the recovery of permutation matrices with the gauge generated by the Birkhoff polytope of bi-stochastic matrices.

For classical families of models and regularizers (sparse recovery with the  $\ell^1$  norm and low-rank matrix recovery with the nuclear norm), as well as for structured sparsity and the associated mixed-norm, Theorem 1.1 and Theorem 1.2 are weakly sharp. Indeed, we know that there exist RIP matrices with constant arbitrarily close to  $1/\sqrt{2}$  which do not permit uniform recovery [22], and since  $\delta_\Sigma(f) \geq 1/\sqrt{2}$  one must have  $\delta^{weak} = 1/\sqrt{2}$ . Thus for these families of models, our results are weakly sharp in the sense that

$$\delta_\Sigma(f) \geq \delta^{weak} = 1/\sqrt{2}.$$

Considering block sparsity, further analysis would be needed to check weak sharpness  $\delta_\Sigma(f)$ . However one can already observe that it complies with the necessary dependency on the number of blocks  $J$  [11].

The question of strong sharpness remains, i.e. do we have the pointwise equality (for *each*  $f$  and  $\Sigma$ )

$$\delta_\Sigma(f) = \delta_\Sigma^{strong}(f) ?$$

A related open question for the family of sparse models is to determine whether the pointwise equality  $\delta_{\Sigma_K^N}^{strong}(f) = \delta^{weak}$  holds. Answering this would close the discussion on RIP constants for such models.

*Optimal decoders with respect to a model set  $\Sigma$ .* Some new regularizer design appeared in Section 4. We showed that adequately chosen weights for the block structured norm improve the RIP recovery guarantees for block structured sparse vectors. The admissible RIP constant  $\delta_\Sigma(f)$  introduces a hierarchy in potential decoders that is also consistent with the quality of stability constants. From an algorithm design perspective, and given a constrained function class  $f \in C$ , we can define :

$$\delta_\Sigma(C) := \sup_{f \in C} \delta_\Sigma(f) \tag{123}$$

The design of regularizers under this perspective could be addressed by studying the following problems:

1. Is  $\delta_\Sigma(C) > 0$ ?
2. If yes, is the optimum reached at a function  $f_0 \in C$ ?
3. Can we characterize  $f_0$ ? Is it unique?

*Convex regularizer design.* It is particularly interesting to study the class  $C_{\text{convex}}$  of convex regularizers. We showed in Section 2 that for (almost) any convex regularizer  $f$  such that  $\delta_\Sigma(f) \geq \delta$ , there is an atomic norm  $\|\cdot\|_{\mathcal{A}}$  such that  $\delta_\Sigma(\|\cdot\|_{\mathcal{A}}) \geq \delta_\Sigma(f) \geq \delta$ . Thus we can restrict the study to the class  $C_{\text{atomic}}$  of atomic norm regularizers  $f$ . Considering the function  $z \mapsto \delta_\Sigma(z)$ , we can define the set  $T_\Sigma(\delta) := \{z \in \mathcal{H} : \delta_\Sigma(z) \geq \delta\}$ . Then  $\delta_\Sigma(f) \geq \delta$  is equivalent to  $\mathcal{T}_f \subset T_\Sigma(\delta)$ . This constraint on the descent set characterizes regularizers  $f$  such that recovery with RIP  $\delta$  is possible. In the case of 1-sparse vectors this condition can be characterized exactly and yields atomic norms that are close to the  $\ell^1$  norm. When  $\delta_\Sigma(f) \geq 1/\sqrt{2}$  is required,  $f$  is necessarily a multiple of the  $\ell^1$  norm. This route looks promising for the design of convex regularizers for model sets where such regularizers are unknown or sub-optimal.



## Acknowledgments

The authors would like to thank B. Adcock, M. Davies, N. Keriven, G. Kutyniok, G. Obozinski and the anonymous reviewers for their insightful comments which helped us improve this paper at various stages of its preparation.

## References

- [1] B. Adcock, A. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *arXiv preprint arXiv:1310.1141*, 2013.
- [2] B. Adcock and A. C. Hansen. A generalized sampling theorem for stable reconstructions in arbitrary bases. *Journal of Fourier Analysis and Applications*, 18(4):685–716, 2012.
- [3] B. Adcock and A. C. Hansen. Generalized sampling and infinite dimensional compressed sensing. *Foundations of Computational Mathematics (to appear)*, 2015.
- [4] B. Adcock, A. C. Hansen, and C. Poon. Beyond Consistent Reconstructions: Optimality and Sharp Bounds for Generalized Sampling, and Application to the Uniform Resampling Problem. *SIAM Journal on Mathematical Analysis*, 45(5):3132–3167, 2013.
- [5] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *arXiv preprint arXiv:1302.0561*, 2013.
- [6] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis, A Hitchhiker’s Guide*. Springer, 2006.
- [7] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, page iau005, 2014.
- [8] A. Argyriou, R. Foygel, and N. Srebro. Sparse Prediction with the k-Support Norm. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems* 25, pages 1457–1465. Curran Associates, Inc., 2012.
- [9] U. Ayaz, S. Dirksen, and H. Rauhut. Uniform recovery of fusion frame structured sparse signals. *Applied and Computational Harmonic Analysis*, 2016.
- [10] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *Information Theory, IEEE Transactions on*, 56(4):1982–2001, 2010.

- [11] A. Bastounis and A. C. Hansen. On random and deterministic compressed sensing and the restricted isometry property in levels. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 297–301. IEEE, 2015.
- [12] T. Blumensath. Sampling and reconstructing signals from a union of linear subspaces. *Information Theory, IEEE Transactions on*, 57(7):4660–4671, 2011.
- [13] T. Blumensath and M. E. Davies. Sampling Theorems for Signals From the Union of Finite-Dimensional Linear Subspaces. *Information Theory, IEEE Transactions on*, 55(4):1872–1882, Apr. 2009.
- [14] F. F. Bonsall. A general atomic decomposition theorem and banach’s closed range theorem. *The Quarterly Journal of Mathematics*, 42(1):9–14, 1991.
- [15] A. Bourrier, M. Davies, T. Peleg, P. Perez, and R. Gribonval. Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems. *Information Theory, IEEE Transactions on*, 60(12):7928–7946, Dec 2014.
- [16] T. Cai and A. Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *Information Theory, IEEE Transactions on*, 60(1):122–132, Jan 2014.
- [17] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9-10):589–592, May 2008.
- [18] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.
- [19] E. J. Candes and Y. Plan. Matrix Completion With Noise. *Proceedings of the IEEE*, 98(6):925–936, June 2010.
- [20] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, Feb. 2006.
- [21] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

- [22] M. E. Davies and R. Gribonval. Restricted isometry constants where  $\ell^p$  sparse recovery can fail for  $0 < p \leq 1$ . *Information Theory, IEEE Transactions on*, 55(5):2203–2214, 2009.
- [23] S. Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Foundations of Computational Mathematics*, pages 1–30, 2015.
- [24] D. L. Donoho. For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, June 2006.
- [25] A. Eftekhari and M. B. Wakin. New analysis of manifold embeddings and signal recovery from compressive measurements. *Applied and Computational Harmonic Analysis*, 39(1):67–109, 2015.
- [26] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *Signal Processing, IEEE Transactions on*, 58(6):3042–3054, 2010.
- [27] Y. C. Eldar and M. Mishali. Robust Recovery of Signals From a Structured Union of Subspaces. *Information Theory, IEEE Transactions on*, 55(11):5302–5316, Nov. 2009.
- [28] H. G. Feichtinger and G. Zimmermann. An exotic minimal banach space of functions. *Mathematische Nachrichten*, 239-240(1):42–61, 2002.
- [29] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [30] R. Gribonval and M. Nielsen. Beyond sparsity: Recovering structured representations by  $\ell^1$  minimization and greedy algorithms. *Advances in Computational Mathematics*, 28(1):23–41, 2008.
- [31] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1990.
- [32] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.
- [33] G. Puy, M. E. Davies, and R. Gribonval. Recipes for stable linear embeddings from hilbert spaces to  $\mathbb{R}^m$ . *arXiv preprint arXiv:1509.06947*, 2015.
- [34] B. Recht, M. Fazel, and P. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, 52(3):471–501, 2010.

- [35] E. Richard, G. R. Obozinski, and J.-P. Vert. Tight convex relaxations for sparse matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3284–3292. Curran Associates, Inc., 2014.
- [36] J. C. Robinson. *Dimensions, embeddings, and attractors*, volume 186. Cambridge University Press, 2010.
- [37] R. T. Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970.
- [38] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- [39] C. Studer and R. G. Baraniuk. Stable restoration and separation of approximately sparse signals. *Applied and Computational Harmonic Analysis*, Sept. 2013.
- [40] Y. Traonmilin, S. Ladjal, and A. Almansa. Robust multi-image processing with optimal sparse regularization. *Journal of Mathematical Imaging and Vision*, 51(3):413–429, 2015.
- [41] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.