COMPRESSED SENSING AND BEST APPROXIMATION FROM UNIONS OF SUBSPACES: BEYOND DICTIONARIES

Tomer Peleg[†], *Rémi Gribonval*[‡], *and Mike E. Davies*[∓]

[†] Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[‡] INRIA Rennes – Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France
[∓] School of Engineering and Electronics, The University of Edinburgh, Edinburgh EH9 3JL, UK
tomerfa@tx.technion.ac.il, remi.gribonval@inria.fr, mike.davies@ed.ac.uk

ABSTRACT

We propose a theoretical study of the conditions guaranteeing that a decoder will obtain an *optimal* signal recovery from an underdetermined set of linear measurements. This special type of performance guarantee is termed instance optimality and is typically related with certain properties of the dimensionality-reducing matrix **M**. Our work extends traditional results in sparse recovery, where instance optimality is expressed with respect to the set of sparse vectors, by replacing this set with an arbitrary finite union of subspaces. We show that the suggested instance optimality is equivalent to a generalized null space property of **M** and discuss possible relations with generalized restricted isometry properties.

Index Terms— Instance optimality, null space property, restricted isometry property, union-of-subspaces

1. INTRODUCTION

Traditional results in sparse recovery relate certain properties of a dimensionality-reducing matrix **M**, considered as an *encoder*, to performance guarantees of certain explicit or implicit decoders Δ. A popular family of performance guarantees is coined *instance optimality*: a decoder is *instance optimal* at order k with respect to **M** and two norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ if, for *every* vector **x**

$$\|\Delta(\mathbf{Mx}) - \mathbf{x}\|_X \leq C\sigma_k(\mathbf{x})_Y \quad (1)$$

where $\sigma_k(\mathbf{x})_Y$ measures the *distance* (in the sense of the Y norm) from **x** to the set of k-sparse vectors:

$$\Sigma_k \doteq \{\mathbf{z},\ \|\mathbf{z}\|_0 \leq k\}\,. \quad (2)$$

Cohen *et al* [1] have shown that, for a given matrix **M**, the existence of an instance optimal decoder is equivalent to a null space property (NSP),

$$\|\mathbf{h}\|_X \leq C'\sigma_{2k}(\mathbf{h})_Y\,,\ \forall \mathbf{h} \in \mathrm{Ker}\mathbf{M}, \quad (3)$$

TP is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. RG acknowledges the support of the European Research Council, PLEASE project (ERC-StG-2011-277906).

where $\mathrm{Ker}\mathbf{M} \doteq \{\mathbf{z},\ \mathbf{Mz} = 0\}$.

While this equivalence serves as a fundamental result in sparse recovery, it is often too abstract to handle. Therefore, researchers typically turn to simplified setups. First of all, they focus on specific choices of the norms X and Y. It is well known from [1] that the NSP for $X = Y = \ell_2$ is too restrictive in a dimensionality-reducing context. This means that the corresponding instance optimality (with a constant C independent of the dimension) is only possible if **M** does not significantly decrease dimension. In light of this "negative" result, typical choices are $X = \ell_2$, $Y = \ell_1$, or $X = \ell_p$, $Y = \ell_q$ with $1 \leq q \leq p < 2$.

Furthermore, the decoder implied by the NSP is somewhat inconvenient to handle. To prove the instance optimality of more convenient decoders such as the minimum ℓ_1 norm decoder, one typically exploits restricted isometry properties (RIP) [2,3]. We say that **M** satisfies the RIP of order k if there exists a constant $0 < \delta_k < 1$ such that

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Mx}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2 \quad (4)$$

holds for all $\mathbf{x} \in \Sigma_k$. Typical results in sparse recovery assume that the RIP holds at various orders with small enough constants δ_k. Based on these assumptions, which obviously serve as stronger conditions than the NSP, the instance optimality of explicit and/or computationally efficient decoders is proven.

In recent years, the cosparse analysis model – an alternative and distinct viewpoint to sparse and redundant representations – has drawn considerable attention [4, 5]. This model relies on a matrix $\mathbf{\Omega} \in \mathbf{R}^{p\times d}$, referred to as the analysis operator. The key property of the analysis model is our expectation that the analysis representation vector $\mathbf{\Omega x} \in \mathbf{R}^p$ should be sparse with ℓ zeros. The locations of the zeros in this sparse representation vector carve out the low-dimensional subspace that the signal belongs to, i.e. each analysis subspace is orthogonal to ℓ rows of the analysis operator $\mathbf{\Omega}$. To prove instance optimality of explicit decoders in the analysis setup, several extensions to the classical RIP have been suggested, such as **D**-RIP [6] and $\mathbf{\Omega}$-RIP [7].

In this work our main goal is to replace Σ_k, which is a finite union of k-dimensional linear subspaces, by an *arbitrary finite union of low-dimensional linear subspaces (UoS)* denoted by Σ. One special case of interest covered by our study is that of the cosparse analysis model mentioned above. In this model the set Σ is the union of low-dimensional spaces, each orthogonal to ℓ rows of the analysis operator $\mathbf{\Omega}$. A popular example is the $\mathbf{\Omega}_{DIF}$ operator, associated to finite differences on the edges of a graph, e.g., a 2D regular grid [5]. This operator generates a UoS Σ covering the family of piecewise constant 2D signals. Our study also applies to structured sparsity models, where aside from being sparse in some representation, the signals of interest exhibit an additional structure, such as tree structure [8, 9] and block sparse signal models [9, 10].

An extension of traditional results in sparse recovery to general UoS has already been considered in a recent paper by Blumensath [11], which explores specific types of instance optimality under a bi-Lipschitz condition involving the operator $\mathbf{M}$ and the UoS Σ.

In this work we address the broader class of generalized NSP and prove that it is equivalent to instance optimality in Section 2. Then we show that under seemingly "mild" assumptions on the general UoS Σ, the NSP for $X = Y = \ell_2$ is too restrictive in a dimensionality-reducing context. This is the topic of Section 3. Finally, we discuss the connections with the work of Blumensath [11]. Specifically, we show in Section 4 that the results of Blumensath can be viewed as a generalized RIP that implies a generalized NSP for the norm $X = \ell_2$ and a specific choice of the Y norm.

2. GENERALIZED NULL SPACE PROPERTY AND ITS EQUIVALENT INSTANCE OPTIMALITY

In this section we extend the results of Cohen *et al* [1] who proved equivalence between instance optimality (Eq. (1)) and NSP (Eq. (3)) for the UoS Σ_k consisting of all k-sparse vectors in $\mathbf{R}^d$. We provide two types of extensions to the classical result by Cohen *et al*. The first is replacing the set of k-sparse vectors by an arbitrary UoS Σ residing in $\mathbf{R}^d$. The second extension is inserting a linear operator $\mathbf{A}$ to the left-hand side of the instance optimality and NSP.

This second extension is useful when we are interested in measuring the decoding error (between the signal and the decoded version computed from its linear measurements) not in the domain of $\mathbf{x}$, but in another domain which constitutes a linear mapping of the original domain. For example, consider a setup where we measure a signal $\mathbf{x}$ through $\mathbf{M}$, but we are only interested in estimating its spectrum in a certain frequency band, and $\mathbf{A}$ is the corresponding partial Fourier matrix. Nevertheless, for the rest of the paper we shall assume that $\mathbf{A}$ is the identity matrix.

The equivalence between generalized versions of Eq. (1) and Eq. (3) is stated below,

Theorem 1. *Given matrices $\boldsymbol{A}$, $\boldsymbol{M}$ and norms $\|\cdot\|_X$, $\|\cdot\|_Y$, the following are equivalent:*
I. Instance optimality – there exists some decoder Δ such that

$$\|\Delta(\boldsymbol{Mx}) - \boldsymbol{Ax}\|_X \leq C d\,(\boldsymbol{x}, \Sigma)_Y\,,\ \forall \boldsymbol{x} \tag{5}$$

II. Generalized NSP:

$$\|\boldsymbol{Ah}\|_X \leq C' d\,(\boldsymbol{h}, \Sigma + \Sigma)_Y\,,\ \forall \boldsymbol{h} \in \mathit{Ker}\boldsymbol{M} \tag{6}$$

where the constants C, C' are related by a factor at most 2, $\Sigma + \Sigma = \{\mathbf{z} :\ \mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2,\ \mathbf{z}_1, \mathbf{z}_2 \in \Sigma\}$ and $d(\mathbf{x}, \Sigma)_Y$ is the distance (in the sense of the Y norm) from $\mathbf{x}$ to the set Σ.

At first sight, a more relevant instance optimality is one where the matrix $\mathbf{A}$ appears in both sides of the equation. In fact, this type of instance optimality is already covered by our result, as stated below,

Corollary 1. *Given some norm Y', by choosing $\|\cdot\|_Y = \|\boldsymbol{A}(\cdot)\|_{Y'}$, we obtain from Theorem 1,*

$$\|\Delta(\boldsymbol{Mx}) - \boldsymbol{Ax}\|_X \leq C d\,(\boldsymbol{Ax}, \boldsymbol{A}\Sigma)_{Y'}\,,\ \forall \boldsymbol{x}. \tag{7}$$

Proof. First, we prove that the NSP is a necessary condition with $C' = C$. Let Δ be a decoder such that (5) holds, let $\mathbf{h} \in \mathrm{Ker}\mathbf{M}$ and let us denote by $\mathbf{h}_0$ the best approximation to $\mathbf{h}$ in $\Sigma + \Sigma$. Let us also split $\mathbf{h}_0$ into a sum of two vectors, each in Σ, namely $\mathbf{h}_0 = \mathbf{h}_1 + \mathbf{h}_2,\ \mathbf{h}_1, \mathbf{h}_2 \in \Sigma$, and denote the residual with respect to $\mathbf{h}$ by $\mathbf{h}_3 = \mathbf{h} - \mathbf{h}_0$. Now since $\mathbf{h}_2 \in \Sigma$ and $\mathbf{h} \in \mathrm{Ker}\mathbf{M}$, we have

$$\mathbf{Ah}_2 = \Delta(\mathbf{Mh}_2) = -\Delta(\mathbf{M}(\mathbf{h}_1 + \mathbf{h}_3)) \tag{8}$$

where the first equality holds from Eq. (5) and the second from the decomposition of $\mathbf{h}$ as a sum of $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$.
We now derive an upper bound on $\|\mathbf{Ah}\|_X$:

$$\begin{aligned} \|\mathbf{Ah}\|_X &= \|\mathbf{A}(\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3)\|_X \\ &= \|\Delta(\mathbf{M}(\mathbf{h}_1 + \mathbf{h}_3)) - \mathbf{A}(\mathbf{h}_1 + \mathbf{h}_3)\|_X \\ &\leq C \operatorname*{Inf}_{\mathbf{z} \in \Sigma} \|\mathbf{h}_1 + \mathbf{h}_3 - \mathbf{z}\|_Y \\ &\leq C\|\mathbf{h}_3\|_Y = C\|\mathbf{h} - \mathbf{h}_0\|_Y \\ &= C d\,(\mathbf{h}, \Sigma + \Sigma)_Y \end{aligned} \tag{9}$$

where the second equality holds from Eq. (8), in the first inequality we apply Eq. (5) for the vector $\mathbf{h}_1 + \mathbf{h}_3$, in the second inequality we consider a specific vector in Σ, namely $\mathbf{z} = \mathbf{h}_1$, and the last equality holds from the definition of $\mathbf{h}_0$.

Next, we prove that the NSP is a sufficient condition with $C' = C/2$. Let us define a decoder $\Delta(\mathbf{y}) = \mathbf{A}\Delta_0(\mathbf{y})$, where

$$\Delta_0(\mathbf{y}) = \operatorname*{Argmin}_{\mathbf{z}:\ \mathbf{Mz} = \mathbf{y}} d(\mathbf{z}, \Sigma)_Y. \tag{10}$$

For a vector $\mathbf{x}$ let us denote $\mathbf{h} = \Delta_0(\mathbf{Mx}) - \mathbf{x}$. From the definition of $\Delta_0(\cdot)$ we have $\mathbf{h} \in \mathrm{Ker}\mathbf{M}$. Using the basic properties of the norm $\|\cdot\|_Y$, the symmetry of the UoS Σ (i.e. $\mathbf{z}\in\Sigma$ implies $-\mathbf{z}\in\Sigma$) and the definition of $d(\cdot,\cdot)_Y$ we derive an upper bound on $d(\mathbf{h},\Sigma+\Sigma)_Y$:

$$\begin{aligned} d(\mathbf{h},\Sigma+\Sigma)_Y &= \operatorname*{Inf}_{\mathbf{z}\in\Sigma+\Sigma}\|\Delta_0(\mathbf{Mx})-\mathbf{x}-\mathbf{z}\|_Y \\ &= \operatorname*{Inf}_{\mathbf{z}_1\in\Sigma,\mathbf{z}_2\in\Sigma}\|\Delta_0(\mathbf{Mx})-\mathbf{x}-\mathbf{z}_1-\mathbf{z}_2\|_Y \\ &\le \operatorname*{Inf}_{\mathbf{z}_1\in\Sigma}\|\mathbf{x}-\mathbf{z}_1\|_Y+\operatorname*{Inf}_{\mathbf{z}_2\in\Sigma}\|\Delta_0(\mathbf{Mx})-\mathbf{z}_2\|_Y \\ &= d(\mathbf{x},\Sigma)_Y+d(\Delta_0(\mathbf{Mx}),\Sigma)_Y. \qquad (11)\end{aligned}$$

Finally, we derive an upper bound on $\|\Delta(\mathbf{Mx})-\mathbf{Ax}\|_X$:

$$\begin{aligned}\|\Delta(\mathbf{Mx})-\mathbf{Ax}\|_X &= \|\mathbf{Ah}\|_X \le C' d(\mathbf{h},\Sigma+\Sigma)_Y \\ &\le C'\left(d(\mathbf{x},\Sigma)_Y+d(\Delta_0(\mathbf{Mx}),\Sigma)_Y\right) \\ &\le 2C' d(\mathbf{x},\Sigma)_Y = C d(\mathbf{x},\Sigma)_Y, \qquad (12)\end{aligned}$$

where in the first inequality we apply Eq. (6) for the vector $\mathbf{h}$, the second holds from Eq. (11) and the third from the definition of $\Delta_0(\cdot)$ in Eq. (10). □

3. THE CASE $X=Y=\ell_2$

We now turn to the special case where X,Y are both the ℓ_2 norm. In this case the NSP becomes

$$\forall\mathbf{h}\in\text{Ker}\mathbf{M},\ \|\mathbf{h}\|_2^2\le C^2\operatorname*{Inf}_{\mathbf{z}\in\Sigma+\Sigma}\|\mathbf{h}-\mathbf{z}\|_2^2. \qquad (13)$$

In this section we extend the result of Cohen *et al* [1] on this setup and show that just like the NSP for the set of k-sparse vectors Σ_k, the generalized NSP for arbitrary UoS Σ is too restrictive in a dimensionality-reducing context.

It is useful to exploit the fact that any vector $\mathbf{v}\in\mathbf{R}^d$ can be decomposed as

$$\mathbf{v}=\left(\mathbf{q}^T\mathbf{v}\right)\mathbf{q}+\left(\mathbf{I}-\mathbf{q}\mathbf{q}^T\right)\mathbf{v}, \qquad (14)$$

for any unit-norm vector $\mathbf{q}$ in $\Sigma+\Sigma$. We denote the set consisting of all such vectors $\mathbf{q}$ by Φ. Thus, we can further simplify the right-hand side of Eq. (13):

$$\begin{aligned}\operatorname*{Inf}_{\mathbf{z}\in\Sigma+\Sigma}\|\mathbf{h}-\mathbf{z}\|_2^2 &\le \operatorname*{Inf}_{\mathbf{q}\in\Phi}\left\|\left(\mathbf{I}-\mathbf{q}\mathbf{q}^T\right)\mathbf{h}\right\|_2^2 \\ &= \|\mathbf{h}\|_2^2-\operatorname*{Sup}_{\mathbf{q}\in\Phi}\left(\mathbf{q}^T\mathbf{h}\right)^2, \qquad (15)\end{aligned}$$

where the first inequality holds since $(\mathbf{q}^T\mathbf{h})\mathbf{q}$ is the best approximation for $\mathbf{h}$ in Span $\{\mathbf{q}\}$ and the last equality holds from (14). Combining (13) and (15) we get

$$\forall\mathbf{h}\in\text{Ker}\mathbf{M},\ \forall\mathbf{q}\in\Phi,\ \left(\mathbf{q}^T\mathbf{h}\right)^2\le\left(1-\frac{1}{C^2}\right)\|\mathbf{h}\|_2^2. \qquad (16)$$

Let us denote by $\mathbf{P}_{\mathcal{N}}$ the projection matrix onto the null space of $\mathbf{M}$. Applying (16) for the vector $\mathbf{h}=\mathbf{P}_{\mathcal{N}}\mathbf{q}$, where $\mathbf{q}$ is an arbitrary vector in Φ, we have

$$\forall\mathbf{q}\in\Phi,\ \left((\mathbf{P}_{\mathcal{N}}\mathbf{q})^T\mathbf{q}\right)^2\le\left(1-\frac{1}{C^2}\right)\|\mathbf{P}_{\mathcal{N}}\mathbf{q}\|_2^2, \qquad (17)$$

which leads to

$$\forall\mathbf{q}\in\Phi,\ \|\mathbf{P}_{\mathcal{N}}\mathbf{q}\|_2^2\le 1-\frac{1}{C^2}. \qquad (18)$$

Let us denote the number of measurements (the number of rows in $\mathbf{M}$) by m and let $\mathbf{v}_1,\dots,\mathbf{v}_{d-m}$ be an orthonormal basis for Ker$\mathbf{M}$, then we have

$$\forall\mathbf{q}\in\Phi,\ \sum_{j=1}^{d-m}\left(\mathbf{q}^T\mathbf{v}_j\right)^2\le 1-\frac{1}{C^2}. \qquad (19)$$

This brings us to the main result of this section:

Lemma 1. *Assume there exists a set of d unit-norm vectors $\boldsymbol{q}^{(1)},\dots,\boldsymbol{q}^{(d)}\in\Sigma+\Sigma$ that satisfy*

$$\forall\boldsymbol{z}\in\mathbf{R}^d,\ \sum_{i=1}^{d}\left((\boldsymbol{q}^{(i)})^T\boldsymbol{z}\right)^2\ge\tilde{C}\|\boldsymbol{z}\|_2^2, \qquad (20)$$

for some constant $\tilde{C}>0$, namely the vectors $\left\{\boldsymbol{q}^{(i)}\right\}_{i=1}^d$ form a frame in $\mathbf{R}^d$ with a lower bound $\tilde{C}$. Then the number of measurements m required for satisfying Eq. (13) must be at least $\left(1-\frac{1}{\tilde{C}}\left(1-\frac{1}{C^2}\right)\right)d$.

Proof. Applying Eq. (19) on the vectors $\mathbf{q}^{(1)},\dots,\mathbf{q}^{(d)}\in\Phi$,

$$\sum_{j=1}^{d-m}\sum_{i=1}^{d}\left((\mathbf{q}^{(i)})^T\mathbf{v}_j\right)^2\le d\left(1-\frac{1}{C^2}\right). \qquad (21)$$

Assuming the condition of Eq. (20) holds for this set of vectors with constant $\tilde{C}>0$, we have

$$\begin{aligned} d-m=\sum_{i=1}^{d-m}\|\mathbf{v}_j\|_2^2 &\le \frac{1}{\tilde{C}}\sum_{j=1}^{d-m}\sum_{i=1}^{d}\left((\mathbf{q}^{(i)})^T\mathbf{v}_j\right)^2 \\ &\le \frac{d}{\tilde{C}}\left(1-\frac{1}{C^2}\right). \qquad (22)\end{aligned}$$

Straightforward considerations lead to the lower bound on m. □

An immediate consequence from Lemma 1 is that it is sufficient to find d vectors $\mathbf{q}^{(1)},\dots,\mathbf{q}^{(d)}\in\Sigma$ that are orthogonal, as they then form an orthonormal basis in $\mathbf{R}^d$, and the condition of Eq. (20) is then trivially met with a constant $\tilde{C}=1$. In the synthesis case $\Sigma=\Sigma_k$, the set Φ contains the canonical basis, hence the condition of Eq. (20) is met.

We now turn to discuss the implications of Lemma 1 on the analysis cosparse model. In the analysis case each vector $\mathbf{q}$ in the set Φ defined above is associated with an index subset[1] $\Lambda\subseteq\{1,\dots,p\}$ such that $\mathbf{\Omega}_\Lambda\mathbf{q}=0$, i.e., $\mathbf{q}$ is orthogonal to the rows of $\mathbf{\Omega}_\Lambda$ (see [5] for more details on analysis subspaces). Note that the number of vectors in the set Φ equals

[1] Note that here the subscript Λ selects rows, not columns as usual.

the number of sets Λ consisting of linearly independent rows in $\mathbf{\Omega}$. This measure is in fact directly connected to the signature of the matrix $\mathbf{\Omega}^T$, which is defined in [12] as the ratio of linearly independent sets of k columns out of all possible sets of size k.

From the description above it is obvious that the number of vectors in Φ depends on the analysis operator $\mathbf{\Omega}$. We now consider two special cases. First, let us assume that the rows in $\mathbf{\Omega}$ are in general-position, implying that every subset of d or less rows are necessarily linearly independent. In this case we have for example p choose $d-1$ vectors, each corresponding to a one-dimensional analysis subspace, and when dealing with a redundant analysis operator (with p greater than d), this number is much greater than d. Hence, for any such reasonable operator it should be possible to find d linearly independent vectors in the set Φ and since they form a basis in $\mathbf{R}^d$, the condition of Eq. (20) is met with some constant $\tilde{C}$. Note however that since this basis is not necessarily orthogonal, the constant $\tilde{C}$ is not controlled, i.e. it can be very small.

Second, in the special case of the analysis operator $\mathbf{\Omega}_{DIF}$ mentioned in Section 1 we have the following result:

Corollary 2. *For the analysis operator $\mathbf{\Omega}_{DIF}$, ℓ_2-ℓ_2 instance optimality implies that **M** does not reduce the dimension.*

To see that this is true, take $\left\{\mathbf{q}^{(i)}\right\}_{i=1}^d$ to be the canonical basis. Since $\mathbf{\Omega}_{DIF}\mathbf{q}^{(i)}$ has few nonzero entries, $\mathbf{q}^{(i)}$ belongs to Σ, hence they belong also to $\Sigma+\Sigma$. Thus, the condition of Eq. (20) is trivially met.

4. GENERALIZED RESTRICTED ISOMETRY PROPERTY

In light of the "negative" result derived in the previous section, it seems natural to ask the following question: If one would still like to guarantee instance optimality with respect to $X=\ell_2$, what are the norms $\|\cdot\|_Y$ that can be used in a dimensionality reduction context? Clearly, ℓ_1 or ℓ_p norms used in previous works will not generalize to arbitrary UoS models, since they are intimately aligned to the ambient basis. In this section we shall suggest a different norm that applies for any UoS.

We begin by taking a closer look at a recent work by Blumensath [11] who also explored instance optimality for general UoS. This work makes the following assumption on $\mathbf{M}$ and the UoS Σ – there exist constants $0<\alpha\leq\beta$ such that

$$\alpha\|\mathbf{x}_1+\mathbf{x}_2\|_2^2 \leq \|\mathbf{M}(\mathbf{x}_1+\mathbf{x}_2)\|_2^2 \leq \beta\|\mathbf{x}_1+\mathbf{x}_2\|_2^2 \tag{23}$$

holds for all $\mathbf{x}_1,\mathbf{x}_2\in\Sigma$. This condition for $\mathbf{M}$ is termed in [11] a bi-Lipschitz condition on the set Σ. Note that it can also be viewed as a generalized RIP on the set $\Sigma+\Sigma$:

$$\forall \mathbf{x}\in\Sigma+\Sigma, \alpha\|\mathbf{x}\|_2^2 \leq \|\mathbf{M}\mathbf{x}\|_2^2 \leq \beta\|\mathbf{x}\|_2^2. \tag{24}$$

With $\alpha=1-\delta$ and $\beta=1+\delta$, this corresponds exactly to the $\mathbf{D}$-RIP [6] when $\Sigma=\mathbf{D}\Sigma_k$, and to the $\mathbf{\Omega}$-RIP [7] when Σ is the union of analysis subspaces, each orthogonal to ℓ rows of $\mathbf{\Omega}$. We propose to call it a $(\Sigma+\Sigma)$-RIP.

In his work Blumensath addresses a noisy measurement setup. Restricting his results to a noise-free setup, like the one considered here, we can formulate the following instance optimality, which is derived only from the left-hand side of Eq. (23),

$$\|\Delta_{opt}(\mathbf{M}\mathbf{x})-\mathbf{x}\|_2 \leq \frac{2}{\sqrt{\alpha}}\|\mathbf{M}(\mathbf{x}-\mathbf{x}_\Sigma)\|_2+\|\mathbf{x}-\mathbf{x}_\Sigma\|_2. \tag{25}$$

In this equation $\mathbf{x}_\Sigma$ is the best approximation for $\mathbf{x}$ in Σ,

$$\mathbf{x}_\Sigma = \underset{\mathbf{z}\in\Sigma}{\operatorname{Argmin}}\ \|\mathbf{x}-\mathbf{z}\|_2, \tag{26}$$

namely this is the vector minimizing $d(\mathbf{x},\Sigma)_2$. Similarly, $\Delta_{opt}(\mathbf{M}\mathbf{x})$ is the vector in the set Σ that best explains the measurements $\mathbf{M}\mathbf{x}$,

$$\Delta_{opt}(\mathbf{M}\mathbf{x}) = \underset{\mathbf{z}\in\Sigma}{\operatorname{Argmin}}\ \|\mathbf{M}\mathbf{x}-\mathbf{M}\mathbf{z}\|_2. \tag{27}$$

The right-hand side of Eq. (25) motivates the definition of the following norm, hereafter referred to as the $\ell_{\mathbf{M}}$ norm[2],

$$\|u\|_{\mathbf{M}} \doteq \|u\|_2+\frac{\|\mathbf{M}u\|_2}{\sqrt{\alpha}}. \tag{28}$$

We now focus on the lower $(\Sigma+\Sigma)$-RIP (the left-hand side of Eq. (24)) and show that it implies a generalized NSP with respect to the norms $X=\ell_2$ and $Y=\ell_{\mathbf{M}}$,

Theorem 2. *Assume that **M** satisfies the lower $(\Sigma+\Sigma)$-RIP with constant α. Then this implies the generalized NSP,*

$$\forall \boldsymbol{h}\in \mathit{Ker}\boldsymbol{M}, \|\boldsymbol{h}\|_2 \leq d(\boldsymbol{h},\Sigma+\Sigma)_{\boldsymbol{M}}. \tag{29}$$

Proof. By the lower $(\Sigma+\Sigma)$-RIP we have for $\mathbf{x}\in\Sigma+\Sigma$ and $\mathbf{h}\in\operatorname{Ker}\mathbf{M}$,

$$\begin{aligned}\|\mathbf{h}\|_2 &\leq \|\mathbf{h}-\mathbf{x}\|_2+\|\mathbf{x}\|_2 \leq \|\mathbf{h}-\mathbf{x}\|_2+\frac{\|\mathbf{M}\mathbf{x}\|_2}{\sqrt{\alpha}} \\ &= \|\mathbf{h}-\mathbf{x}\|_2+\frac{\|\mathbf{M}(\mathbf{h}-\mathbf{x})\|_2}{\sqrt{\alpha}}\end{aligned} \tag{30}$$

Taking the infimum over $\mathbf{x}\in\Sigma+\Sigma$ yields the desired generalized NSP. □

From Theorem 1 and Theorem 2 we have that the lower $(\Sigma+\Sigma)$-RIP implies that the following decoder,

$$\Delta_{consist}(\mathbf{y}) = \underset{\mathbf{z}:\ \mathbf{M}\mathbf{z}=\mathbf{y}}{\operatorname{Argmin}}\ d(\mathbf{z},\Sigma)_M, \tag{31}$$

[2]The specific choice of the constants in this norm will become more clear in the sequel, when we shall explore the generalized NSP implied by the lower $(\Sigma+\Sigma)$-RIP.

obtains for every $\mathbf{x} \in \mathbf{R}^d$,

$$\|\Delta_{consist}(\mathbf{Mx}) - \mathbf{x}\|_2 \leq 2d(\mathbf{x}, \Sigma)_{\mathbf{M}}. \tag{32}$$

Despite its apparent similarity with the ℓ_2-ℓ_2 instance optimality, this type of instance optimality can actually hold for highly dimensionality-reducing $\mathbf{M}$, as demonstrated in a previous work [13] which further explored the implications of the bi-Lipschitz condition (or generalized RIP).

Note that in the synthesis setting, where $\Sigma = \Sigma_k$, the NSP with the $\mathbf{M}$-norm can be upper bounded (when the upper $(\Sigma + \Sigma)$-RIP holds) with the ℓ_1 norm, and it actually serves as an intermediate and somehow stronger step to get the well known results of Candès [2, 3].

Comparing the two decoders defined in Eq. (27) and Eq. (31), we notice several significant differences. While the second is consistent with the measurements, namely $\mathbf{M}\Delta(\mathbf{Mx}) = \mathbf{Mx}$, the first is not necessarily so. Moreover, the first is in the set Σ, whereas the second is not restricted to Σ. Finally, in the first decoder optimality is achieved with respect to the ℓ_2 norm, while the second is optimal with respect to the $\ell_{\mathbf{M}}$ norm.

Since $\mathbf{x}_\Sigma$ minimizies $d(\mathbf{x}, \Sigma)_2$ and not $d(\mathbf{x}, \Sigma)_{\mathbf{M}}$ we have

$$\begin{aligned} 2d(\mathbf{x}, \Sigma)_{\mathbf{M}} &\leq \frac{2}{\sqrt{\alpha}}\|\mathbf{M}(\mathbf{x} - \mathbf{x}_\Sigma)\|_2 + 2\|\mathbf{x} - \mathbf{x}_\Sigma\|_2 \qquad (33) \\ &= \left(\frac{2}{\sqrt{\alpha}}\|\mathbf{M}(\mathbf{x} - \mathbf{x}_\Sigma)\|_2 + \|\mathbf{x} - \mathbf{x}_\Sigma\|_2\right) + \|\mathbf{x} - \mathbf{x}_\Sigma\|_2. \end{aligned}$$

It is clear from Eq. (33) that the upper bounds on the decoding error for the two decoders (the right-hand side of Eq. (25) and Eq. (32)) are different from each other. However, it is not conclusive if one instance optimality is more strict.

To conclude this section, our main contribution here is to draw connections between generalized RIPs on UoS (in the same lines of the $\mathbf{D}$-RIP and $\mathbf{\Omega}$-RIP suggested in previous works) and generalized NSPs for the standard choice of $X = \ell_2$ and a feasible (in the context of a dimensionality-reducing matrix $\mathbf{M}$) choice of the Y norm.

5. CONCLUSIONS

In this work we extend classical results in the sparse recovery field, by replacing the union of all k-sparse vectors with an arbitrary finite union of low-dimensional subspaces. We derive a generalized version of the null space property for the measurement matrix $\mathbf{M}$ and prove that it is equivalent to a generalized instance optimality. Then we show that just as for sparse approximation, the observation that ℓ_2-ℓ_2 instance optimality is not feasible if $\mathbf{M}$ decreases dimension remains valid also for an arbitrary UoS. Finally, we draw connections between generalized NSP and generalized RIP for a specific and feasible choice of norms.

6. REFERENCES

[1] A. Cohen, W. Dahmen, and R.A. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.

[2] E.J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] E.J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, , no. 346, pp. 589–592, 2008.

[4] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *IOP Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[5] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "The cosaprse analysis model and algorithms," *Appl. Comp. Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[6] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comp. Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[7] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M.E. Davies, "Greedy-like algorithms for the cosparse analysis model," 2012, available at arXiv:1207.2456.

[8] C. La and M. N. Do, "Tree-based orthogonal matching pursuit algorithm for signal reconstruction," in *Proc. SPIE Conf., Wavelet Applications in Signal and Image Processing XI*, San Diego, California, 2005.

[9] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[10] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[11] T. Blumensath, "Sampling and reconstructing signals from a union of linear subspaces," *IEEE Trans. Info. Theory*, vol. 57, no. 7, pp. 4660–4671, 2011.

[12] M. Elad, "Sparse representations are most likely to be the sparsest possible," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–12, 2006.

[13] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional subspaces," *IEEE Trans. Info. Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.