Sampling and reconstructing signals from a union of linear subspaces

Thomas Blumensath

Abstract

In this note we study the problem of sampling and reconstructing signals which are assumed to lie on or close to one of several subspaces of a Hilbert space. Importantly, we here consider a very general setting in which we allow infinitely many subspaces in infinite dimensional Hilbert spaces. This general approach allows us to unify many results derived recently in areas such as compressed sensing, affine rank minimisation and analog compressed sensing.

Our main contribution is to show that a conceptually simple iterative projection algorithms is able to recover signals from a union of subspaces whenever the sampling operator satisfies a bi-Lipschitz embedding condition. Importantly, this result holds for all Hilbert spaces and unions of subspaces, as long as the sampling procedure satisfies the condition for the set of subspaces considered. In addition to recent results for finite unions of finite dimensional subspaces and infinite unions of subspaces in finite dimensional spaces, we also show that this bi-Lipschitz property can hold in an analog compressed sensing setting in which we have an infinite union of infinite dimensional subspaces living in infinite dimensional space.

I. INTRODUCTION

To motivate the general setting of this paper, we start with a review of the compressed sensing signal model in finite dimensions. In compressed sensing, sparse signals are considered. A class of N-dimensional signals f in a Hilbert space is said to be K-sparse, if there is an orthonormal basis $\{\psi_i\}$, such that the N-dimensional vector $\mathbf{x} = [\langle f, \psi_i \rangle]_i$ has at most K non-zero elements. More generally, if $\mathbf{x}_K$ is the best approximation to $\mathbf{x}$ with no more than K non-zero elements, then if $\mathbf{x} - \mathbf{x}_K$ is small, $\mathbf{x}$ is said informally to be approximately K-sparse.

In compressed sensing, a sparse signal is sampled by taking M linear measurements $\tilde{y}_j = \langle f, \phi_j \rangle$. In matrix notation, this can be written as

$$\tilde{\mathbf{y}} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where $\tilde{\mathbf{y}}$ is the vector of measurements $\langle f, \phi_j \rangle$ and where $\mathbf{\Phi}$ is the matrix with entries $[\mathbf{\Phi}]_{j,i} = \langle \psi_i, \phi_j \rangle$. In practice, the measurement process is never perfect and we have to account for measurement noise and inaccuracies. We thus assume that the measurements (or samples) are of the form

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}, \tag{2}$$

where $\mathbf{e}$ is a measurement error.

Traditional sampling theory would predict that we require N samples to be able to reconstruct $\mathbf{x}$ form the measurements. However, if $\mathbf{x}$ is K-sparse or approximately K-sparse, then we can often take less samples and still reconstruct $\mathbf{x}$ with near optimal precision [1] [2]. Importantly, reconstructing $\mathbf{x}$ from $\mathbf{y}$ can often be done using fast polynomial time algorithms. One of the conditions that has been shown to be sufficient for the reconstruction of $\mathbf{x}$ with many different fast algorithms is that the measurement process satisfies what is known as the Restricted Isometry Condition of a given order, where the order of the condition is related to the sparsity K.

The Restricted Isometry Constant of order K is generally defined as the smallest quantity δ_K that satisfies the condition

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2, \tag{3}$$

for *all* K sparse vectors $\mathbf{x}$.

The sparse compressed sensing model defines a set of subspaces associated with the set of K-sparse vectors. Fixing the location of the K non-zero elements in a vector $\mathbf{x}$ defines a K-dimensional subspace of $\mathbb{R}^N$. There are $\binom{N}{K}$ such K dimensional subspaces, each for a different sparsity pattern. All K-sparse vectors, that is, all vectors with no more than K non-zero elements, thus lie in the union of these $\binom{N}{K}$ subspaces. This interpretation of the sparse model led to the consideration of more general union of subspaces (UoS) as in [3], [4] and [5]. Such a generalization offers many advantages. For example, many types of data are known to be sparse in some representation, but also exhibit additional structure. These are so called structured sparse signals, an example of which are images, which are not only approximately sparse in the wavelet domain but also have wavelet coefficients that exhibit tree structures [6], [7]. Apart from tree structured sparse models, structured sparse models include block sparse signal models [8], [9], [10] and the simultaneous sparse approximation problem [11], [12], [13], [14], [15]. All of these models can be readily seen as UoS models.

However, the idea of UoS is applicable beyond constrained sparse models. For example, signals sparse in an over-complete dictionary [16], [10], the union of statistically independent subspaces as considered by Fletcher et. [17] or signals sparse in an analysis frame [18] can all be understood from this general viewpoint. All of these examples were of finite unions of subspaces in finite dimensional space. But there is nothing that stops us from considering infinite dimensional spaces and infinite unions. In this case, the UoS model also incorporates signal models such as the finite rate of innovation model [19], the low rank matrix approximation model [20] and the analog compressed sensing model [21].

We here consider this general setting where we allow infinite unions. In this setting, we derive a conceptually simple and efficient computational strategy to solve linear inverse problems. To achieve this, we build on previous work of [3] and [4], where theoretical properties of UoS models were studied. Of importance are also the computational strategies previously suggested in [5] (where the authors studied block-sparse models) and in [8] (where structured sparse signals were considered).

We here make the following contribution. We show that, if the sampling strategy satisfies a certain bi-Lipschitz embedding property (closely related to the Restricted Isometry Property known in compressed sensing), then, in a fixed number of iterations, a relatively simple iterative projection algorithm can compute near optimal estimates of signals that lie on, or close to, a given UoS model. These results are similar to those derived for K-sparse signals in [22] and for structured sparse models in [8]. Our contribution here is to show that these results extend to more general UoS models (whether finite or infinite) as long as the bi-Lipschitz embedding property holds.

A. Sampling and the union of subspaces models

Union of subspaces models have been considered in [3], [4] and [5]. To formally define the UoS model in a general Hilbert space $\mathcal{H}$, consider a set of arbitrary subspaces $\mathcal{A}_i \subset \mathcal{H}$. We then define the UoS as the set

$$\mathcal{A} = \bigcup \mathcal{A}_i. \tag{4}$$

In analogy with compressed sensing, sampling of an element $\mathbf{x} \in \mathcal{H}$ is done using a linear operator $\mathbf{\Phi} : \mathcal{H} \to \mathcal{L}$, where $\mathcal{L}$ is some Hilbert space. We then write the observations as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}, \tag{5}$$

where $\mathbf{e} \in \mathcal{L}$ is again an error term.

B. The bi-Lipschitz condition

In order to guarantee stability, it is necessary to impose a bi-Lipschitz condition on $\mathbf{\Phi}$ as a map from $\mathcal{A}$ to $\mathcal{L}$.

Definition 1. *We say that $\mathbf{\Phi}$ is bi-Lipschitz on a set $\mathcal{A}$, if there exist constants $0 < \alpha \le \beta$, such that for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$*

$$\alpha\|\mathbf{x}_1 + \mathbf{x}_2\|^2 \le \|\mathbf{\Phi}(\mathbf{x}_1 + \mathbf{x}_2)\|^2 \le \beta\|\mathbf{x}_1 + \mathbf{x}_2\|^2. \tag{6}$$

The bi-Lipschitz constants of $\mathbf{\Phi}$ on $\mathcal{A}$ are the largest α and smallest β for which the above inequalities hold for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$.

Whilst β is the square of the Lipschitz constant of the map $\mathbf{\Phi}$ (as a map from $\mathcal{A}$ to $\mathcal{L}$), $1/\alpha$ is the square of the Lipschitz constant of the inverse of $\mathbf{\Phi}$ defined as a map from $\mathbf{\Phi}\mathcal{A} \subset \mathcal{L}$ to $\mathcal{A}$. Note that the requirement $\alpha > 0$ is equivalent to the requirement that $\mathbf{\Phi}$ is one to one as a map from $\mathcal{A}$ to $\mathcal{L}$. Therefore, the inverse of P is well defined as a function from $\mathbf{\Phi}\mathcal{A}$ to the set $\mathcal{A}$ whenever $\alpha > 0$.

C. The separation property

When dealing with infinite dimensions and infinite unions, extra care has to be taken. We therefore here define an additional property that allows us to avoid problems that might arise in this setting. Note that this is not the only way to deal with these problems, but to avoid unnecessary complication of the discussion, we leave details of an alternative approach for a future publication.

Both, our definition of an optimal solution and of our proposed algorithmic strategy rely on the definition of a projection operator, which projects a given point onto the closest subspace. In the infinite dimensional setting, such projections are not guaranteed to exist in general. We will therefore restrict our discussion here to those UoS models, for which such a projection makes sense.

In order to define projections onto a general set, we introduce the following property.

Definition 2. *Let $\mathcal{A}_i$ be subspaces of $\mathcal{H}$. We say that a set of subspaces $\{\mathcal{A}_i\}$ has the separation property, if for all $\mathbf{x} \in \mathcal{H}, \mathbf{x} \ne \mathbf{0}$ there is a $\delta > 0$, such that for all $\delta_0 < \delta$ the closed balls $\mathcal{B}(\mathbf{x}, \delta_0)$ do not intersect $\mathcal{A}$, whilst the closed ball $\mathcal{B}(\mathbf{x}, \delta)$ intersects at least one subspaces $\mathcal{A}_i$.*

This somewhat technically looking *separation property* is equivalent to the requirement that for each $\mathbf{x}$ in $\mathcal{H}$ there exists a (possibly non-unique) $\mathbf{x}_{\mathcal{A}}$ that satisfies

$$\|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| = \inf_{\tilde{\mathbf{x}} \in \mathcal{A}} \|\mathbf{x} - \tilde{\mathbf{x}}\|, \tag{7}$$

where $\mathbf{x}_{\mathcal{A}}$ is the orthogonal projection of $\mathbf{x}$ onto one of the subspaces $\mathcal{A}_i$.

The *separation property* thus guarantees that we can define a projection operator onto $\mathcal{A}$. In particular, we can define the projection of a point $\mathbf{x}$ onto the set $\mathcal{A}$ as

$$P_{\mathcal{A}}(\mathbf{x}) = \arg\min_{\tilde{\mathbf{x}} \in \mathcal{A}} \|\tilde{\mathbf{x}} - \mathbf{x}\|, \tag{8}$$

where $\arg\min_{\tilde{\mathbf{x}} \in \mathcal{A}}$ returns one of the elements in A that achieves the infimum $\inf \tilde{\mathbf{x}} \in \mathcal{A}\|\mathbf{x} - \tilde{\mathbf{x}}\|$. Note that in general, the elements in $\mathcal{A}$ that achieve the infimum might not be defined uniquely, however, we assume that the operator $P_{\mathcal{A}}(\mathbf{x})$ returns any one such element. How this element is chosen in practice does not influence the theoretical results derived here so that we do not specify any particular approach in this paper.

II. THE OPTIMAL SOLUTION

The bi-Lipschitz condition guarantees that $\mathbf{\Phi}$ is one to one as a function from $\mathcal{A}$ to $\mathcal{L}$, that is, it maps distinct points form $\mathcal{A}$ into distinct points in $\mathcal{L}$. We are therefore able, at least in theory, to invert $\mathbf{\Phi}$ on $\mathcal{A}$. The condition also guarantees stability in that, for any $\mathbf{x} \in \mathcal{A}$, if we are given an observation $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{\Phi}\hat{\mathbf{x}} + \mathbf{e}$, where $\mathbf{e} \in \mathcal{L}$ and $\hat{\mathbf{x}} \in \mathcal{H}$ are general errors, then we could, at least in theory, recover a good approximation of $\mathbf{x}$ as follows. We let $\hat{\mathbf{y}}$ be the projection of $\mathbf{y}$ onto the closest[1] element in $\mathbf{\Phi}\mathcal{A}$. We then look for the unique $\mathbf{x} \in \mathcal{A}$ for which $\hat{\mathbf{y}} = \mathbf{\Phi}\hat{\mathbf{x}}$. As will be shown more rigorous below, the bi-Lipschitz property of $\mathbf{\Phi}$ then guarantees that $\hat{\mathbf{x}}$ is close to $\mathbf{x}$.

More formally, consider

$$\inf_{\tilde{\mathbf{x}} \in \mathcal{A}} \|\mathbf{y} - \mathbf{\Phi}\tilde{\mathbf{x}}\|. \tag{9}$$

If $\mathbf{\Phi}\mathcal{A}$ has the *separation property*, then we can define optimal solutions as those elements $\mathbf{x}_{opt} \in \mathcal{A}$ for which

$$\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}_{opt}\| = \inf_{\tilde{\mathbf{x}} \in \mathcal{A}} \|\mathbf{y} - \mathbf{\Phi}\tilde{\mathbf{x}}\|. \tag{10}$$

We now show that all $\mathbf{x}_{opt}$ are basically optimal if the bi-Lipschitz property holds, that is, we can't define an estimate that performs substantially better.

Let us first derive an upper bound for the error. Note that by definition of $\mathbf{x}_{opt}$, $\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}_{opt}\| \leq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}_{\mathcal{A}}\|$, where we define $\mathbf{x}_{\mathcal{A}} = P_{\mathcal{A}}(\mathbf{x})$. Defining $\mathbf{e}_{opt} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}_{opt}$ and $\mathbf{e}_{\mathcal{A}} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}_{\mathcal{A}}$ we thus have

$$
\begin{aligned}
\|\mathbf{x} - \mathbf{x}_{opt}\| &\leq \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt}\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| \\
&\leq \frac{1}{\sqrt{\alpha}} \|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt})\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| \\
&= \frac{1}{\sqrt{\alpha}} \|\mathbf{e}_{\mathcal{A}} - \hat{\mathbf{e}}\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| \\
&= \frac{1}{\sqrt{\alpha}} \|\mathbf{e}_{\mathcal{A}}\| + \frac{1}{\sqrt{\alpha}} \|\hat{\mathbf{e}}\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| \\
&\leq \frac{2}{\sqrt{\alpha}} \|\mathbf{e}_{\mathcal{A}}\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|,
\end{aligned}
$$

where the second inequality is due to the Lipschitz property and the last inequality due to the fact that $\|\mathbf{e}_{opt}\| \leq \|\mathbf{e}_{\mathcal{A}}\|$.

We furthermore have the following 'worst case' lower bound

Theorem 1. *For each* $\mathbf{x}$ *there exists an* $\mathbf{e}$, *such that*

$$\|\mathbf{x} - \mathbf{x}_{opt}\| \geq \sqrt{\frac{0.5}{\beta}} \|\mathbf{e}_{\mathcal{A}}\| + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|$$

Proof: We have the lower bound

$$
\begin{aligned}
\|\mathbf{x} - \mathbf{x}_{opt}\|^2 &= \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt}\|^2 + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|^2 - 2\langle (\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt}), (\mathbf{x} - \mathbf{x}_{\mathcal{A}}) \rangle \\
&\geq \frac{1}{\beta} \|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt})\|^2 + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|^2 - 2\langle (\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt}), (\mathbf{x} - \mathbf{x}_{\mathcal{A}}) \rangle,
\end{aligned}
$$

where from now on we simplify the notation and write $\langle \cdot, \cdot \rangle$ for the real part of the inner product $Re\langle \cdot, \cdot \rangle$.

---

[1] Again, there might not exist a closest $\mathbf{y} \in \mathbf{\Phi}\mathcal{A}$ in general. In this case, we could instead take an $\mathbf{y} \in \mathbf{\Phi}\mathcal{A}$ that is within ϵ of the $\inf_{\mathbf{x} \in \mathcal{A}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2$. To avoid carrying around these ϵ terms, we here assume that the subspaces $\mathbf{\Phi}\mathcal{A}$ also have the *separation property*, so that this complication is not required.

Let $\mathcal{C}_i$ be the cone of elements $\mathbf{y} \in \mathcal{L}$ for which $\mathbf{x}_{opt} \in \mathcal{A}_i$. Because $\mathbf{x}_{\mathcal{A}}$ is the orthogonal projection of $\mathbf{x}$ onto the closest subspace, if $\mathbf{x} \in \mathcal{A}_i$, then $\mathbf{x} - \mathbf{x}_{\mathcal{A}}$ is orthogonal to $\mathcal{A}_i$. Thus, if $\mathbf{x}_{\mathcal{A}} \in \mathcal{A}_i$ and if $\mathbf{y} \in \mathcal{C}_i$, then

$$\langle (\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt}), (\mathbf{x} - \mathbf{x}_{\mathcal{A}}) \rangle = 0. \tag{11}$$

Also, for all $\mathbf{y} \in \mathcal{C}_i$, because $\mathbf{e}_{opt} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}_{opt}$ is orthogonal to $\mathbf{\Phi}\mathbf{x}_{\mathcal{A}} - \mathbf{\Phi}\mathbf{x}_{opt}$,

$$\frac{1}{\beta}\|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}_{opt})\|^2 + \frac{1}{\beta}\|\mathbf{e}_{opt}\|^2 = \frac{1}{\beta}\|\mathbf{e}_{\mathcal{A}}\|^2, \tag{12}$$

so that for all $\mathbf{y} \in \mathcal{C}_i$

$$\|\mathbf{x} - \mathbf{x}_{opt}\|^2 \geq \frac{1}{\beta}\|\mathbf{e}_{\mathcal{A}}\|^2 - \frac{1}{\beta}\|\mathbf{e}_{opt}\|^2 + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|^2.$$

We can now choose $\mathbf{e} = c\mathbf{\Phi}\mathbf{x}_{\mathcal{A}}$, where $c > 1$ is chosen large enough for $\mathbf{y}$ to be in $\mathcal{C}_i$. Because $\mathbf{e}_{opt}$ is orthogonal to $\mathbf{\Phi}\mathcal{A}_i$, $\|\mathbf{e}_{opt}\|$ is constant as a function of c, whilst $\|\mathbf{e}_{\mathcal{A}}\|$ increases for $c > 1$. We can thus choose c (and thus $\mathbf{e}$) such that $\mathbf{y} \in \mathcal{C}_i$ and

$$-\|\mathbf{e}_{opt}\|^2 > -0.5\|\mathbf{e}_{\mathcal{A}}\|^2 + \sqrt{2\beta}\|\mathbf{e}_{\mathcal{A}}\|\|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|, \tag{13}$$

so that for all $\mathbf{x}$ there is an $\mathbf{e}$ such that

$$\|\mathbf{x} - \mathbf{x}_{opt}\|^2 \geq 2\frac{0.5}{\beta}\|\mathbf{e}_{\mathcal{A}}\|^2 + \|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|^2 + \sqrt{\frac{0.5}{\beta}}\|\mathbf{e}_{\mathcal{A}}\|\|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\|,$$

from which the theorem follows.   ∎

III. THE ITERATIVE PROJECTION ALGORITHM

As stated above, $\mathbf{x}_{opt}$ is not always defined and even in cases in which $\mathbf{\Phi}\mathcal{A}$ has the *separation property* so that $\mathbf{x}_{opt}$ is defined, calculating $\mathbf{x}_{opt}$ is highly non-trivial for most $\mathbf{\Phi}$ and $\mathcal{A}$. We therefore propose an iterative algorithm and show that under certain conditions on α and β we can efficiently calculate solutions whose error is of the same order as that achieved by $\mathbf{x}_{opt}$. In order for our algorithm to be applicable, we require that we are able to efficiently calculate the projection of any $\mathbf{x} \in \mathbb{R}^N$ onto the closest $\mathcal{A}_i$ (which therefore has to be well defined).

The Iterative Projection Algorithm is a generalization of the Iterative Hard Thresholding algorithm of [23], [24] and [22] to general UoS models.

Assume $\mathcal{A}$ has the *separation property*. Given $\mathbf{y}$ and $\mathbf{\Phi}$, let $\mathbf{x}^0 = \mathbf{0}$. The Iterative Projection Algorithm is the iterative procedure defined by the recursion

$$\mathbf{x}^{n+1} = P_{\mathcal{A}}(\mathbf{x}^n + \mu\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)), \tag{14}$$

where the non-linear operator $P_{\mathcal{A}}(\mathbf{a})$ is defined in subsection I-C.

In many problems, calculation of $P_{\mathcal{A}}(\mathbf{a})$ is much easier than a brute force search for $\mathbf{x}_{opt}$. For example, in the K-sparse model, $P_{\mathcal{A}}(\mathbf{a})$ simply keeps the largest (in magnitude) K elements of $\mathbf{a}$ and sets the other elements to zero, whilst in the low rank matrix approximation problem, different efficient projections have been defined in [20]. Furthermore, the above algorithm only requires the application of $\mathbf{\Phi}$ and its adjoint, which can often be computed efficiently. Importantly, the next result shows that under certain conditions, not only does the algorithm calculate near optimal solutions, it does so in a fixed number of iterations (depending only on a form of signal to noise ratio)!

We have the following main result.

Theorem 2. *Let $\mathcal{A}$ have the* separation property. *Given $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}$ where $\mathbf{x}$ is arbitrary. Assume $\mathbf{\Phi}$ is bi-Lipschitz as a map from $\mathcal{A}$ to L with constants α and β. If $\beta \leq \frac{1}{\mu} < 1.5\alpha$, then, after*

$$n^{\star} = \left\lceil 2\frac{\ln(\delta\frac{\|\mathbf{e}_{\mathcal{A}}\|}{\|\mathbf{x}_{\mathcal{A}}\|})}{\ln(2/(\mu\alpha) - 2)} \right\rceil \tag{15}$$

iterations, the Iterative Projection Algorithm calculates a solution $\mathbf{x}^{n^{\star}}$ satisfying

$$\|\mathbf{x} - \mathbf{x}^{n^{\star}}\| \leq (c^{0.5} + \delta)\|\mathbf{e}_{\mathcal{A}}\| + \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\|, \tag{16}$$

where $c \leq \frac{4}{3\alpha - 2\mu}$ and $\tilde{\mathbf{e}} = \mathbf{\Phi}(\mathbf{x} - \mathbf{x}_{\mathcal{A}}) + e$.

Note that this bound is of the same order as that derived for $\mathbf{x}_{opt}$.

The above theorem has been proved for the K-sparse model in [22] and for constraint sparse models in [8]. Our main contribution is to show that it holds for general UoS constrained inverse problems[2], as long as the bi-Lipschitz property holds with appropriate constants.

---

[2]It might also be worth noting that the proof of Theorem 2 is not only valid for union of subspaces, but holds for arbitrary subsets of $\mathcal{A} \subset \mathcal{H}$ for which $\mathbf{\Phi}$ satisfies the bi-Lipschitz requirement. A more detailed discussion of this fact is left for an upcoming publication.

To derive the result, we pursue a slightly different approach to that in [22] and [8] and instead follow the ideas of [25]. The proof is based on the following lemma.

Lemma 3. *If $\frac{1}{\mu} \geq \beta$ then, using $\mathbf{x}^{n+1} = P_{\mathcal{A}}(\mathbf{x}^n + \mu \mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n))$, we have*

$$\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 - \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n\|^2$$
$$\leq -\langle (\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2, \tag{17}$$

where $\mathbf{g} = 2\mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)$.

Proof: The left hand side in the equality of the lemma can be bounded by

$$\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 - \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n\|^2$$
$$= -\langle (\mathbf{x}^{n+1} - \mathbf{x}^n), \mathbf{g} \rangle + \|\mathbf{\Phi}(\mathbf{x}^{n+1} - \mathbf{x}^n)\|^2$$
$$\leq -\langle (\mathbf{x}^{n+1} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\mathbf{x}^{n+1} - \mathbf{x}^n)\|^2$$

We will now show that $\mathbf{x}^{n+1} = H_{\mathcal{A}}(\mathbf{x}^n + \frac{\mu}{2}\mathbf{g})$ minimizes $-\langle (\tilde{\mathbf{x}} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\tilde{\mathbf{x}} - \mathbf{x}^n)\|^2$ over all $\tilde{\mathbf{x}} \in \mathcal{A}$ so that $\mathbf{x}_{\mathcal{A}} \in \mathcal{A}$ implies that

$$-\langle (\mathbf{x}^{n+1} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\mathbf{x}^{n+1} - \mathbf{x}^n)\|^2 \leq -\langle (\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\|^2, \tag{18}$$

from which the lemma will follow.

We write the infimum of $-\langle (\tilde{\mathbf{x}} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\tilde{\mathbf{x}} - \mathbf{x}^n)\|^2$ as

$$\inf_{\mathbf{x} \in \mathcal{A}} \left(-\langle \mathbf{x}, \mathbf{g} \rangle + \langle \mathbf{x}^n, \mathbf{g} \rangle + \frac{1}{\mu}\|(\mathbf{x} - \mathbf{x}^n)\|^2 \right)$$
$$\propto \inf_{\mathbf{x} \in \mathcal{A}} \left(-\mu\langle \mathbf{x}, \mathbf{g} \rangle + \langle \mathbf{x}, \mathbf{x} \rangle + \|\mathbf{x}^n\|^2 - 2\langle \mathbf{x}, \mathbf{x}^n \rangle \right)$$
$$\propto \inf_{\mathbf{x} \in \mathcal{A}} \left(-\mu\langle \mathbf{x}, \mathbf{g} \rangle + \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{x}^n \rangle \right)$$
$$\propto \inf_{\mathbf{x} \in \mathcal{A}} \|\mathbf{x} - \mathbf{x}^n - \frac{\mu}{2}\mathbf{g}\|^2$$
$$= \|\mathbf{x}^{n+1} - \mathbf{x}^n - \mu\mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)\|^2,$$

where the last equality comes from the definition of $\mathbf{x}^{n+1} = P_{\mathcal{A}}(\mathbf{x}^n + \mu\mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n))$. Thus, the infimum of $-\langle (\tilde{\mathbf{x}} - \mathbf{x}^n), \mathbf{g} \rangle + \frac{1}{\mu}\|(\tilde{\mathbf{x}} - \mathbf{x}^n)\|^2$ is proportional to $\inf_{\mathbf{x} \in \mathcal{A}} \|\mathbf{x} - \mathbf{x}^n - \frac{\mu}{2}\mathbf{g}\|^2$ so that $\mathbf{x}^{n+1}$ simultaneously minimises both quantities. ∎

Proof of Theorem 2: Let $\mathbf{x}_{\mathcal{A}} = P_{\mathcal{A}}(\mathbf{x})$, so that the triangle inequality implies that.

$$\|\mathbf{x} - \mathbf{x}^{n+1}\| \leq \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^{n+1}\| + \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\|. \tag{19}$$

The square of the first term on the right is bounded using the bi-Lipschitz property of $\mathbf{\Phi}$

$$\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^{n+1}\|^2 \leq \frac{1}{\alpha}\|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^{n+1})\|^2 \tag{20}$$

We expand this, so that

$$\|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^{n+1})\|^2 = \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1} - \mathbf{e}_{\mathcal{A}}\|^2$$
$$\leq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 + \|\mathbf{e}_{\mathcal{A}}\|^2 - 2\langle \mathbf{e}_{\mathcal{A}}, (\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}) \rangle$$
$$\leq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 + \|\mathbf{e}_{\mathcal{A}}\|^2 + \|\mathbf{e}_{\mathcal{A}}\|^2 + \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2$$
$$= 2\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 + 2\|\mathbf{e}_{\mathcal{A}}\|^2, \tag{21}$$

where the last inequality follows from $-2\langle \mathbf{e}_{\mathcal{A}}, (\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}) \rangle \leq \|\mathbf{e}_{\mathcal{A}}\|\|(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1})\| \leq 0.5(\|\mathbf{e}_{\mathcal{A}}\|^2 + \|(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1})\|^2)$.

We will now show that under the Lipschitz assumption of the theorem, the first term on the right is bounded by

$$\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 \leq (\mu - \alpha)\|(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\|^2 + \|\mathbf{e}_{\mathcal{A}}\|^2. \tag{22}$$

To show this, we write

$$
\begin{aligned}
& \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{n+1}\|^2 - \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n\|^2 \\
\leq\ & -2\langle(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n), \mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)\rangle + \frac{1}{\mu}\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2 \\
=\ & -2\langle(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n), \mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)\rangle + \alpha\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2 + (\frac{1}{\mu} - \alpha)\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2 \\
\leq\ & -2\langle(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n), \mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n)\rangle + \|\mathbf{\Phi}(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\|^2 + (\frac{1}{\mu} - \alpha)\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2 \\
=\ & \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}_{\mathcal{A}}\|^2 - \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n\|^2 + (\frac{1}{\mu} - \alpha)\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n\|^2 \\
=\ & \|\mathbf{e}_{\mathcal{A}}\|^2 - \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n\|^2 + (\frac{1}{\mu} - \alpha)\|(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\|^2
\end{aligned}
\tag{23}
$$

where the first inequality is due to Lemma 3.

We have thus shown that

$$
\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^{n+1}\|^2 \leq 2\left(\frac{1}{\mu\alpha} - 1\right)\|(\mathbf{x}_{\mathcal{A}} - \mathbf{x}^n)\|^2 + \frac{4}{\alpha}\|\mathbf{e}_{\mathcal{A}}\|^2.
\tag{24}
$$

Under the condition of the Theorem, $2(\frac{1}{\mu\alpha} - 1) < 1$, so that we can iterate the above expression

$$
\|\mathbf{x}_{\mathcal{A}} - \mathbf{x}^k\|^2 \leq \left(2\left(\frac{1}{\mu\alpha} - 1\right)\right)^k \|\mathbf{x}_{\mathcal{A}}\|^2 + c\|\mathbf{e}_{\mathcal{A}}\|^2,
\tag{25}
$$

where $c \leq \frac{4}{3\alpha - 2\frac{1}{\mu}}$.

In conclusion, we have

$$
\begin{aligned}
\|\mathbf{x} - \mathbf{x}^k\| &\leq \sqrt{\left(2\frac{1}{\mu\alpha} - 2\right)^k \|\mathbf{x}_{\mathcal{A}}\|^2 + c\|\mathbf{e}_{\mathcal{A}}\|^2} + \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\| \\
&\leq \left(2\frac{1}{\mu\alpha} - 2\right)^{k/2} \|\mathbf{x}_{\mathcal{A}}\| + c^{0.5}\|\mathbf{e}_{\mathcal{A}}\| + \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\|,
\end{aligned}
\tag{26}
$$

which means that after $k^{\star} = \left\lceil 2\frac{\ln(\delta\frac{\|\mathbf{e}_{\mathcal{A}}\|}{\|\mathbf{x}_{\mathcal{A}}\|})}{\ln(2/(\mu\alpha) - 2)} \right\rceil$ iterations we have

$$
\|\mathbf{x} - \mathbf{x}^{k^{\star}}\| \leq (c^{0.5} + \delta)\|\mathbf{e}_{\mathcal{A}}\| + \|\mathbf{x}_{\mathcal{A}} - \mathbf{x}\|.
\tag{27}
$$

∎

A. A remark on $\mathbf{e}_{\mathcal{A}}$

For readers familiar with the literature on compressed sensing a remark is in order. In our general result, we have written the bound on the result in terms of $\|\mathbf{e}_{\mathcal{A}}\| = \|\mathbf{\Phi}(\mathbf{x} - \mathbf{x}_{\mathcal{A}})\mathbf{e}\|$. This is the most general statement in which we do not assume additional structure on $x - x_{\mathcal{A}}$ and $\mathbf{\Phi}$. This differs from results in sparse inverse problems, where, under the bi-Lipschitz property, $\mathbf{e}_{\mathcal{A}}$ is proportional to $\|\mathbf{x} - \mathbf{x}_K\| + \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{K}$. Here x_K is the best K-term approximation to $\mathbf{x}$. It also differs from results derived in [8] where $\mathcal{A}$ satisfies certain nesting properties and where a Restricted Amplification property is used to bound $\|\mathbf{e}_{\mathcal{A}}\|$ by a function of $\mathbf{x} - \mathbf{x}_{\mathcal{A}}$. Unfortunately, in the general setting of this paper, such a bound is not possible without additional assumptions and the best one could hope for would be to bound $\|\mathbf{e}_{\mathcal{A}}\|$ by $\|\mathbf{\Phi}\|\|\mathbf{x} - \mathbf{x}_{\mathcal{A}}\| + \|\mathbf{e}\|$.

IV. EXAMPLES OF BI-LIPSCHITZ EMBEDDINGS

The bi-Lipschitz property depends on both, $\mathbf{\Phi}$ and $\mathcal{A}$. In this section we will study three particular sets cases found in the literature. For the first two cases, bi-Lipschitz maps have already been studied and we here review the main results before deriving a new result that demonstrates how such properties can be proved even in an infinite dimensional setting.

A. Finite Unions of Finite dimensional Subspaces

We start with the finite dimensional setting and with unions of finite dimensional subspaces. In particular, let $\mathcal{A}$ be the union of $L < \infty$ subspaces each of dimension no more than K and let $\mathbf{\Phi}\mathcal{A} \subset \mathbb{R}^M$. This is an important special case of UoS models which covers many of the problems studied in practice, such as the K-sparse models used in compressed sensing [1], [2], block sparse signal models [8], [9], [10], the simultaneous sparse approximation problem [11], [12], [13], [14], [15], signals sparse in an over-complete dictionary [16], [10], the union of statistically independent subspaces as considered by Fletcher et. [17] and signals sparse in an analysis frame [18]. Finite unions of finite dimensional subspaces have therefore be studied in, for example, [4] where the following result was derived.

Theorem 4. *For any $t > 0$, let*

$$M \geq \frac{2}{c\delta_{\mathcal{A}}}\left(\ln(2L) + 2K\ln\left(\frac{12}{\delta_{\mathcal{A}}}\right) + t\right),\tag{28}$$

then there exist a $\mathbf{\Phi}$ and a constant $c > 0$ such

$$(1 - \delta_{\mathcal{A}}(\mathbf{\Phi}))\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2 \leq \|\mathbf{\Phi}(\mathbf{y}_1 - \mathbf{y}_2)\|_2^2 \leq (1 + \delta_{\mathcal{A}}(\mathbf{\Phi}))\|\mathbf{y}_1 - \mathbf{y}_2\|_2^2\tag{29}$$

holds for all $\mathbf{y}_1, \mathbf{y}_2$ from the union of L arbitrary K dimensional subspaces $\mathcal{A}$. What is more, if $\mathbf{\Phi}$ is an $M \times N$ matrix generated by randomly drawing i.i.d. entries from an appropriately scaled subgaussian distribution[3], then this matrix satisfies equation (29) with probability at least

$$1 - e^{-t}.\tag{30}$$

The constant c then only depends on the distribution of the entries in $\mathbf{\Phi}$ and is $c = \frac{7}{18}$ if the entries of $\mathbf{\Phi}$ are i.i.d. normal.

B. Infinite Unions of Finite dimensional Subspaces in $\mathbb{R}^N$

Recently, similar results could also be derived for a union of infinitely many subspaces. In [20] minimum rank constrained linear matrix valued inverse problems are studied. These problems are another instance of the linear inverse problem studied in this paper and can be stated as follows: Find a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ with rank no more than K, such that $\mathbf{y} = P(\mathbf{X})$, where $P(\cdot)$ is a linear function that maps $\mathbb{R}^{m \times n}$ into $\mathbb{R}^M$. Vectorising $\mathbf{X}$ as an element $\mathbf{x} \in \mathbb{R}^N$ and by writing P in matrix form, we have the linear inverse problem where $\mathcal{A}$ is the set of vectorised matrices with rank at most K. This problem was solved with the Iterative Projection Algorithm in [27] where it was also shown that

Theorem 5. *If P is a random nearly isometrically distributed linear map[4], then with probability $1 - e^{(-c_1 N)}$,*

$$(1 - \delta)\|\mathbf{X}_1 - \mathbf{X}_2\|_F \leq \|P(\mathbf{X}_1 - \mathbf{X}_2)\| \leq (1 + \delta)\|\mathbf{X}_1 - \mathbf{X}_2\|_F\tag{31}$$

for all rank K matrices $\mathbf{X}_1 \in \mathbb{R}^m \times n$ and $\mathbf{X}_2 \in \mathbb{R}^m \times n$, whenever $N \geq c_0 K(m+n)\log(mn)$, where c_1 and c_0 are constants depending on δ only.

C. Infinite Unions of Infinite dimensional Subspaces

We now show that non-trivial bi-Lipschitz embeddings also exist between infinite dimensional spaces $\mathcal{H}$ and $\mathcal{L}$, where $\mathcal{A}$ is an infinite union of infinite dimensional subspaces in $\mathcal{H}$. We here consider the example from [29]. A continuous real valued time series $x(t)$ is assumed to be band-limited, that is, its Fourier transform $\mathcal{X}(f)$ is assumed to be zero apart from the set $S \subset [-B_N \ B_N]$. Furthermore, the support of $\mathcal{X}(f)$ is assumed to be 'sparse' in the sense that we can write S as the union of K intervals of 'small' bandwidth B_K, i.e. $S \subset \bigcup_{k=1}^{K}[d_k \ d_k + B_K]$, where the d_k are arbitrary scalars from the interval $[0 \ B_N - B_K]$. Note, due to symmetry, we only consider the support in the positive interval $[0 \ B_N]$. Crucially, we assume that $KB_K < B_N$, so that $\mathcal{X}(f)$ is zero for most (in terms of Lebesgue measure) f in $[0 \ B_N]$. Fixing the support S, $\mathcal{X}(f)$ and therefore $x(t)$ lie on a subspace of the space of all square integrable functions with bandwidth B_N. If $KB_K < B_N$, then there are infinitely many distinct sets S satisfying this definition, so that $x(t)$ lies in the union of infinitely many infinite dimensional subspaces.

Classical sampling theory tells us that there exists sampling operators that map band-limited functions into ℓ_2. What is more, these sampling operator are not only one to one, but also isometric, that is, bi-Lipschitz embeddings with $\alpha = 1$ and $\beta = 1$. These sampling operators are given by the Nyquist sampling theorem, which only takes account of the bandwidth B_N, but does not consider additional structure in $\mathcal{A}$. To improve on the classical theory, we are thus interested in sampling schemes with a sampling rate that is less than the Shannon rate.

To this end, we show that there exist bi-Lipschitz embeddings of functions from $\mathcal{A}$ into the space of band-limited signals with bandwidth B_M, where $B_M < B_N$. Combining this embedding with the standard (isometric) Nyquist sampling kernel

---

[3]Examples of these distributions include the Gaussian distribution and random variables that are $\pm\frac{1}{\sqrt{N}}$ with equal probability [26] [16].

[4]See [27] for an exact definition of nearly isometrically distributed linear maps. An example would again be if the matrix $\mathbf{\Phi}$ has appropriately scaled i.i.d. Gaussian entries.

for functions with bandwidth B_M, gives a stable sampling scheme where the sampling rate is $2B_M$ instead of $2B_N$. The iterative projection algorithm will therefore also be applicable to this sampling problem. It is worth noting that the bi-Lipschitz embedding property shown here not only guarantees invertability of the sampling process, which was demonstrated for the problem under consideration in [29], but also guarantees stability of this inverse.

Our treatment here is theoretical in nature and is meant as an example to show how bi-Lipschitz embeddings can be constructed in the infinite dimensional setting, it is not meant as a fully fledged practical sampling method and many practical issues remain to be addressed.

Compressed Sensing theory has shown that there is a constant c such that there are matrices $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M \leq cKln(N/k)$ which are bi-Lipschitz embeddings from the set of all K sparse vectors in $\mathbb{R}^N$ to $\mathbb{R}^M$ [28]. Therefore, assume $\mathbf{\Phi}$ satisfies

$$\alpha \|\mathbf{x}\|^2 \leq \|\mathbf{\Phi}\mathbf{x}\|^2 \leq \beta \|\mathbf{x}\|^2 \tag{32}$$

for all vectors $\mathbf{x} \in \mathbb{C}^N$ with no-more than $2K$ non-zero elements.

The following sampling approach is basically that proposed in [21] and is based on mixing of the spectrum of $x(t)$. It uses a matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ to define this mixing procedure. Our contribution is to show that if the matrix $\mathbf{\Phi}$ has the bi-Lipschitz property with constants α and β, then so will this sampling operator.

Let $\mathcal{A} \subset L^2_{\mathbb{C}}([0\ B_N])$ be the subset of the set of square integrable real valued functions whose Fourier transform has positive support $S \subset [0\ B_N]$, where S is the union of no more than K intervals of width no more than B_K. Let $M = \lceil B_N/B_K \rceil$ and let $B_M = MB_K$. We then split the interval $[0\ B_N]$ into M blocks of length B_K as follows. Let S_j be the interval $[(j-1)B_K\ jB_K)$ for integers $1 \leq j \leq N-1$ and $S_N = [(j-1)B_K\ B_N]$. Similarly, let $\tilde{S}_i$ be the interval $[(i-1)B_K\ iB_K)$ for integers $1 \leq i \leq M-1$ and $S_M = [(i-1)B_K\ iB_K]$. We can then define a linear map from $x(t)$ to $y(t)$ by mapping the Fourier transform of $x(t)$ into the Fourier transform of $y(t)$ as follows

$$\mathcal{Y}(S_i) = \sum_{j=1}^{N} [\mathbf{\Phi}]_{i,j} \mathcal{X}(S_j), \tag{33}$$

where we use the convention that $\mathcal{X}(f) = 0$ for $f > B_N$. In words, the new function has the Fourier transform $\mathcal{Y}$ (defined by symmetry also for $f < 0$) which is constructed by concatenating M functions of length B_K. Each of these blocks is a weighted sum of the N blocks of $\mathcal{X}$, where the weights are the entries of the matrix $\mathbf{\Phi}$.

We have the following result

Theorem 6. *Let $\mathcal{A} \subset L^2_{\mathbb{C}}([0\ B_N])$ be the subset of the set of square integrable real valued functions whose Fourier transform has positive support $S \subset [0\ B_N]$, where S is the union of no more than K intervals of width no more than B_K. If the matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is bi-Lipschitz as a map from the set of all K-sparse vectors in $\mathbb{R}^N$ into $\mathbb{R}^M$, with bi-Lipschitz constants α and β, then the map defined by equation 33 is a bi-Lipschitz map from $\mathcal{A}$ to $L^2_{\mathbb{C}}([0\ B_M])$ such that*

$$\alpha \|\mathcal{X}_1 - \mathcal{X}_2\|^2 \leq \|\mathcal{Y}_1 - \mathcal{Y}_2\|_2^2 \leq \beta \|\mathcal{X}_1 - \mathcal{X}_2\|_2^2, \tag{34}$$

for all $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{A}$.

Proof: To see that this map is bi-Lipschitz from $\mathcal{A}$ to $L^2([0\ B_M])$, consider stacking up the blocks $\mathcal{Y}(\tilde{S}_i)$ and $\mathcal{X}(S_j)$ in two vectors. For $f \in [0\ B_K)$ we use $f_i = (i-1)*B_K + f$ and $\tilde{f}_j = (j-1)*B_K + f$ and define the vectors

$$\mathbf{y}(f) = \begin{bmatrix} \mathcal{Y}(\tilde{f}_1) \\ \mathcal{Y}(\tilde{f}_2) \\ \cdots \\ \mathcal{Y}(\tilde{f}_M) \end{bmatrix} = \mathbf{\Phi} \begin{bmatrix} \mathcal{X}(f_1) \\ \mathcal{X}(f_2) \\ \cdots \\ \mathcal{X}(f_N) \end{bmatrix} = \mathbf{\Phi}\mathbf{x}(f). \tag{35}$$

This model is known as an infinite measurement vector model [21].

Using the norm of L^2, we can write

$$\|\mathcal{Y}_1 - \mathcal{Y}_2\|^2 = \int_0^{B_M} (\mathcal{Y}_1(f) - \mathcal{Y}_2(f))^2\ df$$

$$= \int_0^{B_K} \sum_i (\mathcal{Y}_1((i-1)B_K + f) - \mathcal{Y}_2((i-1)B_K + f))^2\ df$$

$$= \int_0^{B_K} \|\mathbf{y}_1(f) - \mathbf{y}_2(f)\|_2^2\ df$$

$$= \int_0^{B_K} \|\mathbf{\Phi}\mathbf{x}_1(f) - \mathbf{\Phi}\mathbf{x}_2(f)\|_2^2\ df. \tag{36}$$

Noting that for fixed f, the vectors $\mathbf{x}_1(f)$ and $\mathbf{x}_2(f)$ are K-sparse, the bi-Lipschitz property of $\mathbf{\Phi}$ leads to the inequalities

$$\int_0^{B_K} \alpha \|\mathbf{x}_1(f) - \mathbf{x}_2(f)\|_2^2 \, df \le \int_0^{B_K} \|\mathbf{\Phi}\mathbf{x}_1(f) - \mathbf{\Phi}\mathbf{x}_2(f)\|_2^2 \, df \le \int_0^{B_K} \beta \|\mathbf{x}_1(f) - \mathbf{x}_2(f)\|_2^2 \, df, \tag{37}$$

so that

$$\alpha \|\mathcal{X}_1 - \mathcal{X}_2\|^2 \le \|\mathcal{Y}_1 - \mathcal{Y}_2\|_2^2 \le \beta \|\mathcal{X}_1 - \mathcal{X}_2\|_2^2, \tag{38}$$

i.e the mapping defined above satisfies the bi-Lipschitz condition with constants α and β defined by the bi-Lipschitz constants of the matrix $\mathbf{\Phi}$. ∎

If we consider signals whose Fourier transform has support S and if we let $|S|$ be the size of the support, then, if we assume that the support is the union of finitely many intervals of length B_K, then we have $|S| = KB_K$ for some K. If we then use $N = B_N/B_K$ and select M and B_M such that $B_M = MB_K$, then the fact that there are bi-Lipschitz matrices with $M = cK \ln N/K$ together with the above theorem implies the following corollary

Corollary 7. *Let* $\mathcal{A} \subset L_{\mathbb{C}}^2([0\ B_N])$ *be the subset of the set of square integrable real valued functions whose Fourier transform has positive support* $S \subset [0\ B_N]$, *where* $|S|$ *is bounded and where* S *is the union of finitely many intervals of finite width. There exist bi-Lipschitz embeddings from* $\mathcal{A}$ *to* $L_{\mathbb{C}}^2([0\ B_M])$ *whenever*

$$B_M \ge c|S| \ln \left(\frac{B_N}{|S|} \right), \tag{39}$$

where c is some constant.

V. CONCLUSION

We have here presented a unified framework that allows us to sample and reconstruct signals that lie on or close to the union of subspaces. The bi-Lipschitz property is necessary to guarantee stable reconstruction. We have shown that bounds on the bi-Lipschitz constants α and β are sufficient for the near optimal reconstruction with the iterative projection algorithm. Whilst we have here concentrated on the general theory for arbitrary union of subspaces models, we have highlighted several more concrete examples from the literature. We could also show that bandlimited signals with 'sparse' frequency support admit sub-Nyquist sampling methods that are bi-Lipschitz.

We hope that this note offers the basis for the development of novel sampling approaches to several problems that fit into the union of subspaces framework. On the one hand, we have shown on several examples, how bi-Lipschitz sampling operators can be constructed. On the other hand, we have suggested an algorithmic framework which can reconstruct signals with near optimal accuracy. Whilst our contribution was theoretical in nature, our results point the way toward practical strategies that can be developed further in order to tackle a given sampling problem. To achieve this, four problems need that captures relevant prior knowledge, 2) designing realisable sampling operators that satisfy the bi-Lipschitz property, 3) implementing efficient ways to store and manipulate the signals on a computer and 4) developing efficient algorithms to project onto the constraint set.

———————————————————————————-

REFERENCES

[1] E. Candès and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Conf, Wavelet Applications in Signal and Image Processing XI*, Jan. 2005.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Y. Lu and M. Do, "A theory for sampling signals from a union of subspaces," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, 2008.

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

[5] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[6] J. M. Shapiro, "Embedded image coding using the zero-trees of wavelet coefficients," *IEEE Transactions on Image Processing*, vol. 41, no. 12, pp. 3445–3462, Dec 1993.

[7] C. La and M. Do, "Signal reconstruction using sparse tree representations," in *Proc. SPIE Conf, Wavelet Applications in Signal and Image Processing XI*, San Diego, California, Sep 2005.

[8] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *preprint*, 2008.

[9] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *submitted*, 2009.

[10] T. Blumensath and M. Davies, "Compressed sensing and source separation," in *International Conference on Independent Component Analysis and Blind Source Separation*, 2007.

[11] S. F. Cotter, B. D. Rao, K. Engan, and K. K-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[12] Chen J and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[13] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, 2006.

[14] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, 2006.

[15] K. Schnass R. Gribonval, H. Rauhut and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier analysis and applications*, vol. Published online, no. DOI:10.1007/s00041-008-9044-y, October 2008.

[16] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, May 2007.

[17] A. K. Fletcher, S. Rangan, and V. K. Goyal, "The rate-distortion performance of compressed sensing," in *Proc. IEEE conf. Acoustics, Speech and Signal Processing*, 2007.

[18] P Milanfar M. Elad and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947968, 2007.

[19] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[20] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *arXiv*, , no. arXiv:0706.4138v1, 2008.

[21] Y. C. Eldar and M. Mishali, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," Tech. Rep., arXiv:0902.4291v2, 2009.

[22] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3,. pp. 265–274 2009.

[23] N. G. Kingsbury and T. H. Reeves, "Iterative image coding with overcomplete complex wavelet transforms," in *Proc. Conf. on Visual Communications and Image Processing*, 2003.

[24] T. Blumensath and M.E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.

[25] R. Garg and R. Khandekar, "Gradient descend with sparsification: An iterative algorithm for sparse recovery with restricted isometry property.," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 337–344.

[26] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, 2007.

[27] D. Goldfarb and S. Ma, "Convergence of fixed point continuation algorithms for materix rank minimisation," *arXiv*, , no. arXiv:0906.3499v2, 2009.

[28] Emmanuel Candès and Terence Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406 – 5425, 2006.

[29] M. Mishali and Y. C. Eldar, "Blind Multi-Band Signal Reconstruction: Compressed Sensing for Analog Signals," *IEEE Trans. on on Signal Processing*, vol. 57, no. 3, pp. 993-1009, 2009