Sparse Signal Recovery in Hilbert Spaces

Graeme Pope and Helmut Bölcskei

Dept. of IT & EE, ETH Zurich, 8092 Zurich, Switzerland
Email: {gpope, boelcskei}@nari.ee.ethz.ch

Abstract—This paper reports an effort to consolidate numerous coherence-based sparse signal recovery results available in the literature. We present a single theory that applies to general Hilbert spaces with the sparsity of a signal defined as the number of (possibly infinite-dimensional) subspaces participating in the signal's representation. Our general results recover uncertainty relations and coherence-based recovery thresholds for sparse signals, block-sparse signals, multi-band signals, signals in shift-invariant spaces, and signals in finite unions of (possibly infinite-dimensional) subspaces. Moreover, we improve upon and generalize several of the existing results and, in many cases, we find shortened and simplified proofs.

I. INTRODUCTION

The sparse signal recovery literature is vast and has evolved along several threads with recent focus mostly on probabilistic results. This paper constitutes an attempt to consolidate the numerous coherence-based recovery results available in the literature. More specifically, we formulate a single theory that applies to finite- and infinite-dimensional Hilbert spaces, in combination with sparsity defined as the (finite) number of (possibly infinite-dimensional) subspaces participating in a signal's representation. The general coherence-based recovery thresholds we find contain the known thresholds in the following settings as special cases: (i) sparse signals in finite-dimensional spaces [1]–[4], (ii) block-sparse signals [5], [6], (iii) multi-band signals [7]–[9], (iv) signals in shift-invariant spaces [10], and (v) signals in finite unions of finite or infinite-dimensional subspaces [11]–[13]. In addition, we improve upon the thresholds in [5] and we generalize the uncertainty relation in [10]. We introduce suitable generalizations of P0-minimization [2], basis pursuit [2], and orthogonal matching pursuit [14]. Finally, we indicate how the results on signal separation reported in [15], [16] can be extended to the general Hilbert space setting considered here.

Key to our results are definitions of coherence [2] and mutual coherence [3], [16] that work for our general setting. Based on these definitions, we obtain a general kernel uncertainty relation which is then used to establish general recovery thresholds. Similarly our definition of mutual coherence paves the way to a general uncertainty relation that yields fundamental limits on how sparse a signal in a general Hilbert space can be under two different representations. All theorems in this paper are given without proof.

Notation: Lowercase boldface letters stand for column vectors and uppercase boldface letters designate matrices. For a vector $\boldsymbol{a}$, the kth element is written a_k. For the matrix $\mathbf{A}$,

$\mathbf{A}^H$ is its conjugate transpose, its kth column is written $\boldsymbol{a}_k$, and the entry in the kth row and ℓth column is denoted by $A_{k,\ell}$. The spectral norm of $\mathbf{A}$ is $\|\mathbf{A}\|_{2\to 2}$, $\sigma_{\min}(\mathbf{A})$ and $\sigma_{\max}(\mathbf{A})$ are the minimum and maximum singular value of $\mathbf{A}$, respectively.

$\mathscr{H}$ and $\mathscr{G}$ are Hilbert spaces equipped with the norm $\|\cdot\|_{\mathscr{H}}$ and $\|\cdot\|_{\mathscr{G}}$, respectively, and $\mathscr{H}$ has direct sum decomposition [17, Ch. 5.20] $\mathscr{H} = \bigoplus_{i=1}^n \mathscr{H}^{(i)}$ where $n < \infty$. We define $v^{(i)}$ to be the canonical projection of v onto $\mathscr{H}^{(i)}$. For $v \in \mathscr{H}$, $\|v\|_{\mathscr{H},0} \triangleq \left| \{i : \|v^{(i)}\|_{\mathscr{H}} > 0\} \right|$ and $\|v\|_{\mathscr{H},1} \triangleq \sum_{i=1}^n \|v^{(i)}\|_{\mathscr{H}}$. We define $\mathscr{H}^{(\mathcal{S})} \triangleq \bigoplus_{s\in\mathcal{S}} \mathscr{H}^{(s)}$ and $v^{(\mathcal{S})}$ to be the projection of v onto $\mathscr{H}^{(\mathcal{S})}$. We say that a signal $v \in \mathscr{H}$ is $\varepsilon_{\mathcal{S}}$-concentrated to the set $\mathcal{S}$ if $\|v^{(\mathcal{S})}\|_{\mathscr{H},1} \geqslant (1 - \varepsilon_{\mathcal{S}}) \|v\|_{\mathscr{H},1}$, where $0 \leqslant \varepsilon_{\mathcal{S}} \leqslant 1$. We define $\boldsymbol{e}_i \in \mathbb{C}^N$ to be the all zero vector with a one in the ith position. For an operator $\varphi \colon \mathscr{H} \to \mathscr{G}$ with adjoint φ^H, $\omega_{\min}(\varphi) \triangleq \inf_{v\in\mathscr{H}} \|\varphi(v)\|_{\mathscr{G}} / \|v\|_{\mathscr{H}}$, $\omega_{\max}(\varphi) \triangleq \sup_{v\in\mathscr{H}} \|\varphi(v)\|_{\mathscr{G}} / \|v\|_{\mathscr{H}}$, and $\ker(\varphi) \triangleq \{v \in \mathscr{H} : \varphi(v) = 0\}$. For $\alpha \in \mathbb{R}$, we set $[\alpha]^+ \triangleq \max\{0, \alpha\}$. The cardinality of a set $\mathcal{S}$ is denoted as $|\mathcal{S}|$. The Fourier transform operator is written $\mathcal{F}$.

II. SIGNAL AND SAMPLING MODEL

Let $\mathscr{H}$ and $\mathscr{G}$ be Hilbert spaces, with dimensions N and M, respectively, possibly infinite. Assume that $\mathscr{H} = \bigoplus_{i=1}^n \mathscr{H}^{(i)}$, $n < \infty$, and set $d_i = \dim(\mathscr{H}^{(i)})$. We describe the sampling of signals in $\mathscr{H}$ through the application of a bounded linear operator $\Phi \colon \mathscr{H} \to \mathscr{G}$, which we call a *sampling* operator. With Φ we associate the operators $\varphi_i \colon \mathscr{H}^{(i)} \to \mathscr{G}$, for $i = 1, ..., n$, obtained by restricting the action of Φ to the subspace $\mathscr{H}^{(i)}$. It follows from the linearity of Φ that $\Phi(v) = \sum_{i=1}^n \varphi_i(v^{(i)})$. We require that each φ_i be injective.

For $N = \sum_{i=1}^n d_i < \infty$, the action of Φ can be represented through a matrix $\mathbf{D} \in \mathbb{C}^{M\times N}$ according to $\Phi(v) = \mathbf{D}v$, $v \in \mathbb{C}^N$. Taking $\mathbf{D}[i] = [\, \boldsymbol{d}_{i_1} \cdots \boldsymbol{d}_{i_{d_i}} \,]$ to be the set of columns of $\mathbf{D}$ that correspond to $\mathscr{H}^{(i)}$ we have $\varphi_i(\boldsymbol{v}^{(i)}) = \mathbf{D}[i]\, \boldsymbol{v}^{(i)}$, for $i = 1, ..., n$.

III. DEFINITIONS OF COHERENCE

Key to our results are definitions of coherence, mutual coherence, and spark for general sampling operators.

Definition 1 (Hilbert space coherence): Let $\mathscr{H}$ and $\mathscr{G}$ be Hilbert spaces and let $\Phi \colon \mathscr{H} \to \mathscr{G}$ be a sampling operator. We define the *Hilbert space coherence of* Φ as[1]

$$\mu_{\mathscr{H}} = \mu_{\mathscr{H}}(\Phi) \triangleq \max_{i,j,i\neq j} \frac{\omega_{\max}\big(\varphi_i^H \varphi_j\big)}{\omega_{\min}^2(\varphi_i)}. \tag{1}$$

---
[1] By assumption the operators φ_i are injective, hence $\omega_{\min}(\varphi_i) > 0$.

We can interpret $\mu_{\mathscr{H}}(\Phi)$ as a measure of closeness of the subspaces $\mathscr{H}^{(i)}$ under the action of Φ.

Definition 2 (Mutual Hilbert space coherence): Let $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{G}$ be Hilbert spaces and let $\Phi\colon \mathscr{H}_1 \to \mathscr{G}$ and $\Psi\colon \mathscr{H}_2 \to \mathscr{G}$ be sampling operators. We define the *mutual Hilbert space coherence of Φ and Ψ* as

$$\mu_{\mathscr{H}}(\Phi, \Psi) \triangleq \max_{i,j} \frac{\omega_{\max}(\varphi_i^H \psi_j)}{\omega_{\min}(\varphi_i)\,\omega_{\min}(\psi_j)}. \tag{2}$$

The mutual Hilbert space coherence extends the definition of mutual coherence in [3], [16]. The setting of [3], [16] is recovered as follows. Let $\mathscr{H}_1 = \mathbb{C}^{N_1}$, $\mathscr{H}_2 = \mathbb{C}^{N_2}$, and $\mathscr{G} = \mathbb{C}^{M}$. Represent the sampling operators $\Phi\colon \mathscr{H}_1 \to \mathscr{G}$ and $\Psi\colon \mathscr{H}_2 \to \mathscr{G}$ by the matrices $\mathbf{A}$ and $\mathbf{B}$, respectively, so that $\Phi(\boldsymbol{v}) = \mathbf{A}\boldsymbol{v}$ and $\Psi(\boldsymbol{u}) = \mathbf{B}\boldsymbol{u}$. Then, we have

$$\mu_{\mathscr{H}}(\Phi, \Psi) = \max_{i,j} \frac{\left\|\boldsymbol{a}_i^H \boldsymbol{b}_j\right\|_2}{\left\|\boldsymbol{a}_i\right\|_2 \left\|\boldsymbol{b}_j\right\|_2} \stackrel{(a)}{=} \max_{i,j}|\langle \boldsymbol{a}_i, \boldsymbol{b}_j\rangle| = \mu_m,$$

where μ_m is the mutual coherence as specified in [16], and (a) follows since in [16] $\mathbf{A}$ and $\mathbf{B}$ are assumed to have columns with unit ℓ_2-norm.

We will also need a general definition of spark [4], [18].

Definition 3 (Hilbert space spark): Let $\mathscr{H}$ and $\mathscr{G}$ be Hilbert spaces and let $\Phi\colon \mathscr{H} \to \mathscr{G}$ be a sampling operator. Then

$$\mathrm{spark}(\Phi) \triangleq \min_{v\in\ker(\Phi)\setminus\{0\}} \|v\|_{\mathscr{H},0}. \tag{3}$$

The spark of a sampling operator is the smallest number of subspaces that a non-zero signal $v \in \mathscr{H}$ in $\ker(\Phi)$ can occupy.

IV. RECOVERY THRESHOLDS

With our general definitions of coherence and spark, the general recovery thresholds below follow without difficulties. We start with a general kernel uncertainty relation.

Theorem 1 (Kernel uncertainty relation): Let $\Phi\colon \mathscr{H} \to \mathscr{G}$ be a sampling operator with Hilbert space coherence $\mu_{\mathscr{H}}(\Phi)$. Let $v \in \mathscr{H}$ be $\varepsilon_{\mathcal{S}}$-concentrated to $\mathcal{S}$. If $\Phi(v) = 0$, then

$$|\mathcal{S}| \geqslant (1 - \varepsilon_{\mathcal{S}}) \left(1 + (\mu_{\mathscr{H}}(\Phi))^{-1}\right). \tag{4}$$

We next define two optimization problems for the recovery of a signal $v \in \mathscr{H}$ from its measurements $z = \Phi(v) \in \mathscr{G}$. The first one, $\mathscr{H}$-P0, aims to find the signal that explains the given measurements while occupying the fewest subspaces:

$$(\mathscr{H}\text{-P0}) \quad \underset{\hat{v}\in\mathscr{H}}{\mathrm{minimize}} \ \|\hat{v}\|_{\mathscr{H},0} \ \text{subject to} \ \Phi(\hat{v}) = z. \tag{5}$$

Furthermore, we consider a modified version of basis pursuit:

$$(\mathscr{H}\text{-BP}) \quad \underset{\hat{v}\in\mathscr{H}}{\mathrm{minimize}} \ \|\hat{v}\|_{\mathscr{H},1} \ \text{subject to} \ \Phi(\hat{v}) = z. \tag{6}$$

Recovery thresholds for $\mathscr{H}$-P0 and $\mathscr{H}$-BP can now be derived from the kernel uncertainty relation in Theorem 1.

Theorem 2: If $v \in \mathscr{H}$ satisfies $\Phi(v) = z$ and

$$\|v\|_{\mathscr{H},0} < \mathrm{spark}(\Phi)/2, \tag{7}$$

then v is the unique minimizer of $\mathscr{H}$-P0 applied to z.

In addition, we have the following bound, $\mathrm{spark}(\Phi) \geqslant 1 + (\mu_{\mathscr{H}}(\Phi))^{-1}$, which combined with Theorem 2 allows us to conclude that $\mathscr{H}$-P0 returns the correct solution if

$$\|v\|_{\mathscr{H},0} < \left(1 + (\mu_{\mathscr{H}}(\Phi))^{-1}\right)/2. \tag{8}$$

We next provide a recovery condition for $\mathscr{H}$-BP.

Theorem 3: If $v \in \mathscr{H}$ satisfies $\Phi(v) = z$ and (8) holds, then $\mathscr{H}$-BP applied to z returns the correct solution v.

A commonly used alternative to BP is orthogonal matching pursuit (OMP) [14], [19]. We next present a Hilbert-space version of OMP, which we call $\mathscr{H}$-OMP. This algorithm works by iteratively identifying the subspaces $\mathscr{H}^{(i)}$ participating in the representation of v and computes an approximation to v, denoted as v_i, in the ith iteration. The corresponding residual in the ith iteration is given by $r_i \triangleq z - \Phi(v_i)$. The algorithm is initialized with $r_0 \leftarrow z$ and $i \leftarrow 1$, and performs the following steps until $\|r_i\|_{\mathscr{G}} = 0$:

1) Find

$$\ell = \arg \max_{\hat{\ell}} \left\|\varphi_{\hat{\ell}}^H(r_{i-1})\right\|_{\mathscr{H}} / \omega_{\min}(\varphi_{\hat{\ell}}).$$

2) Update the list of participating subspaces: $\mathcal{S}_i \leftarrow \mathcal{S}_{i-1} \cup \{\ell\}$.

3) Find the best approximation to v with support $\mathcal{S}_i$:

$$v_i \leftarrow \arg \min_{u\in\mathscr{H}^{(\mathcal{S}_i)}} \|z - \Phi_{\mathcal{S}_i}(u)\|_{\mathscr{G}}.$$

4) Update the residual and i: $r_i \leftarrow z - \Phi(v_i)$, $i \leftarrow i + 1$.

Theorem 4: Let $\Phi\colon \mathscr{H} \to \mathscr{G}$ be a sampling operator. Then $\mathscr{H}$-OMP applied to $z = \Phi(v)$ returns the correct solution v if (8) is satisfied and will require exactly $\|v\|_{\mathscr{H},0}$ iterations.

Note that implementing the algorithms mentioned above, when $\mathscr{H}$ is infinite-dimensional, is non-trivial. Some alternatives to $\mathscr{H}$-BP and $\mathscr{H}$-OMP, such as SBR2/4, have been proposed for blind multi-band sampling [9], which is a special case of our setup. It is an interesting open problem to extend these algorithms to the general framework in this paper.

V. DISCUSSION OF RECOVERY THRESHOLDS

We next show how the recovery thresholds in [1]–[5], [7]–[9], [11] follow from the general recovery threshold (8). The results in [6], which pertain to a generalization of [5] allowing for different subspace dimensions, can be recovered following the same methodology, but this will not be detailed here due to space constraints.

A. Sparse signal recovery

The (coherence-based) thresholds in [1]–[4] are recovered as follows. Set $\mathscr{H} = \mathbb{C}^N$ and $\mathscr{G} = \mathbb{C}^M$. Take the sampling operator Φ to be represented by the matrix $\mathbf{D} \in \mathbb{C}^{M\times N}$, with unit ℓ_2-norm columns $\boldsymbol{d}_i$. Take $\mathscr{H}^{(i)}$ to be the 1-dimensional subspace spanned by $\boldsymbol{e}_i \in \mathbb{C}^N$, so that $N = n$. The action of $\varphi_i\colon \mathscr{H}^{(i)} \to \mathscr{G}$ is represented by $\varphi_i(\boldsymbol{v}^{(i)}) = \boldsymbol{d}_i v^{(i)} = \boldsymbol{d}_i v_i$. Since $\omega_{\min}(\varphi_i) = \|\boldsymbol{d}_i\|_2 = 1$ and $\omega_{\max}(\varphi_i^H \varphi_j) = |\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|$, we get $\mu_{\mathscr{H}} = \max_{i\neq j}|\langle \boldsymbol{d}_i, \boldsymbol{d}_j\rangle|$, which is exactly the definition of coherence as introduced in [2]–[4]. The recovery threshold (8) for $\mathscr{H}$-P0, $\mathscr{H}$-BP, and $\mathscr{H}$-OMP (which then reduce to P0, BP, and OMP, respectively) is thus equal to the

corresponding thresholds in [2]–[4]. As an aside the general result (8) shows how dictionaries with unnormalized columns should be treated, specifically what the appropriate measure of coherence is, and what the selection criterion in Step 1 of ($\mathscr{H}$-)OMP should be.

B. Block-sparsity

The results for the block-sparse setting considered in [5] are recovered as follows. Set $\mathscr{H} = \mathbb{C}^N$, $\mathscr{G} = \mathbb{C}^M$, and $N = nd$, where d is the block size and n is the number of blocks (and hence the number of subspaces $\mathscr{H}^{(i)}$). As before, the sampling operator Φ is represented by the matrix $\mathbf{D} \in \mathbb{C}^{M \times N}$ with unit ℓ_2-norm columns. Let $\mathscr{H}^{(i)}$ be the subspace spanned by $\{e_{(i-1)d+1}, ..., e_{id}\}$, and set $\mathbf{D}[i] = [d_{(i-1)d+1} \cdots d_{id}]$, so that $\varphi_i(v^{(i)}) = \mathbf{D}[i]\,v^{(i)}$. From (1) the Hilbert space coherence is

$$\mu_{\mathscr{H}}(\Phi) = \max_{i,j,j \neq i} \frac{\sigma_{\max}\left((\mathbf{D}[i])^H \mathbf{D}[j]\right)}{\sigma^2_{\min}(\mathbf{D}[i])}. \qquad (9)$$

We next show how the recovery threshold (8) improves upon that reported in [5, Thms. 2 and 3], which states that recovery using (L-OPT) [5, Eq. 32] and BOMP [5, Sec. IV-A] (our $\mathscr{H}$-BP and $\mathscr{H}$-OMP, respectively), is successful if $\|v\|_{\mathscr{H},0} < \left(1 + \hat{\mu}^{-1}\right)/2$. Here

$$\hat{\mu} \triangleq \frac{d\mu_{\mathrm{B}}(\mathbf{D})}{1 - (d-1)\nu}$$

$$\mu_{\mathrm{B}} = \mu_{\mathrm{B}}(\mathbf{D}) \triangleq \max_{i,j,j \neq i} \frac{\sigma_{\max}\left((\mathbf{D}[i])^H \mathbf{D}[j]\right)}{d}$$

$$\nu = \nu(\mathbf{D}) \triangleq \max_{\ell=1,\dots,n} \max_{i,j,j \neq i} \left|(\mathbf{D}[\ell]_i)^H \mathbf{D}[\ell]_j\right|,$$

and $\mathbf{D}[\ell]_i$ is the ith column of $\mathbf{D}[\ell]$. The following steps establish that $\mu_{\mathscr{H}} \leqslant \hat{\mu}$, thereby proving our claim[2]

$$\mu_{\mathscr{H}}(\Phi) \leqslant \frac{\max_{j \neq i} \sigma_{\max}\left((\mathbf{D}[i])^H \mathbf{D}[j]\right)}{\min_k \sigma_{\min}\left((\mathbf{D}[k])^H \mathbf{D}[k]\right)}$$

$$\overset{(a)}{\leqslant} \max_{i,j,j \neq i} \frac{\sigma_{\max}\left((\mathbf{D}[i])^H \mathbf{D}[j]\right)}{[1 - (d-1)\nu]^+} = \hat{\mu},$$

where we applied the Geršgorin disc theorem [20, Th. 6.1.1] in (a). When $(\mathbf{D}[i])^H \mathbf{D}[i] = \mathbf{I}_d$, for all i, we have $\mu_{\mathscr{H}} = \hat{\mu}$, but one can easily find examples where the strict inequality $\mu_{\mathscr{H}} < \hat{\mu}$ holds.

C. Multi-band signals

We next show how our results apply to sparse multi-band signals as considered in [8], [9], [21], [22]. Let $\mathscr{H}$ be the space of functions band-limited to the interval $[0, 1/T)$ and for a signal $v \in \mathscr{H}$, let V be its Fourier transform. For

---

[2]It is possible that $\hat{\mu} < 0$ and since $\mu_{\mathscr{H}}(\Phi)$ is always non-negative, we do not have $\mu_{\mathscr{H}}(\Phi) \leqslant \hat{\mu}$ in this case. However, in this instance [5, Thms. 2 and 3] say that we cannot guarantee the recovery of any signal, but the right-hand side of (8) is positive, thus trivially improving upon the recovery thresholds in [5, Thms. 2 and 3].

---

simplicity of exposition, assume that the interval $[0, 1/T)$, is divided into n disjoint intervals $\mathcal{I}_1, ..., \mathcal{I}_n$, with $\mathcal{I}_i = [(i-1)/(nT), i/(nT))$, $i = 1, ..., n$. Define the subspaces $\mathscr{H}^{(i)} = \{v \in \mathrm{L}_2(\mathbb{R}) : V(f) = 0, \text{ for all } f \notin \mathcal{I}_i\}$. Thus, for a signal $v \in \mathscr{H}$, the sparsity level $\|v\|_{\mathscr{H},0}$ is the number of frequency bands $\mathcal{I}_i$ occupied by V.

We next demonstrate how the multi-coset sampling scheme of [7], [8] can be analyzed in our framework. Multi-coset sampling maps the signal v to $m \leqslant n$ sequences $z^{(k)}$ as follows:

$$z^{(k)}_\ell = v(\ell nT + kT), \quad k = 1, \dots, m, \ \ell \in \mathbb{Z}.$$

To obtain an explicit characterization of the corresponding sampling operator Φ we will work in the frequency domain. The Fourier transform of $z^{(k)}$ is given by

$$Z^{(k)}(f) = \frac{1}{nT} \sum_{\ell=1}^{n} V\left(f + \frac{\ell}{nT}\right) e^{2\pi i k\ell/n}$$

$$= \frac{1}{nT} \sum_{\ell=1}^{n} V^{(\ell)}(f)\, e^{2\pi i k\ell/n} = \sum_{\ell=1}^{n} \lambda_{k,\ell} V^{(\ell)}(f),$$

where $\lambda_{k,\ell} = (nT)^{-1} \exp(2\pi i k\ell/n)$ and $V^{(\ell)}(f) = V(f + \ell/(nT))$. Then, the action of the sampling operator, $\Phi : \mathscr{H} \to \mathscr{G}$, can be represented in terms of the continuously parametrized linear system of equations

$$\begin{pmatrix} Z^{(1)}(f) \\ \vdots \\ Z^{(m)}(f) \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m,1} & \lambda_{m,2} & \cdots & \lambda_{m,n} \end{pmatrix}}_{\triangleq \mathbf{\Lambda}} \underbrace{\begin{pmatrix} V^{(1)}(f) \\ \vdots \\ V^{(n)}(f) \end{pmatrix}}_{\triangleq \widetilde{V}(f)},$$

(10)

for $f \in [0, 1/(nT))$. We have thus established a finite-dimensional continuously indexed matrix representation of Φ [17]. Based on this insight, we next show that

$$\mathrm{spark}(\Phi) = \mathrm{spark}(\mathbf{\Lambda}) = m, \qquad (11)$$

and

$$\mu_{\mathscr{H}}(\Phi) = \mu_{\mathscr{H}}(\mathbf{\Lambda}), \qquad (12)$$

which means that we can reduce the computation of Hilbert space spark and Hilbert space coherence of an infinite-dimensional operator to that of a finite matrix that does not depend on f. Since (10) holds for all $f \in [0, 1/(nT))$, for v to lie in the kernel of Φ, $\widetilde{V}(f)$ must be in $\ker(\mathbf{\Lambda})$ for each $f \in [0, 1/(nT))$. One can then show that this implies that $\mathrm{spark}(\Phi) = \mathrm{spark}(\mathbf{\Lambda})$. The second equality in (11) follows since $\mathbf{\Lambda}$ consists of the first m rows of the $n \times n$ DFT matrix and hence $\mathrm{spark}(\mathbf{\Lambda}) = m$ [21].

To prove (12), note that for $u \in \mathscr{H}^{(i)}$ with Fourier transform U, $\varphi_i : \mathscr{H}^{(i)} \to \mathscr{G}$ is given by the matrix representation

$$\varphi_i(U)(f) \triangleq U(f) \begin{pmatrix} \lambda_{1,i} \\ \vdots \\ \lambda_{m,i} \end{pmatrix},$$

and has adjoint

$$\varphi_i^H(X)(f) = \varphi_i^H\left(\begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(m)} \end{pmatrix}\right)(f) \triangleq \left(\sum_{k=1}^m \lambda_{k,i}^* X^{(k)}\right)(f),$$

where

$$X = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(m)} \end{pmatrix} \in \mathscr{G}.$$

Hence, for $u \in \mathscr{H}^{(\ell)}$ with Fourier transform U, we have

$$\left\|\varphi_j^H \varphi_\ell(U)\right\|_{\mathscr{H}} = \left\|\sum_{i=1}^m \lambda_{i,j}^* \lambda_{i,\ell} U\right\|_{\mathscr{H}} = \left|\sum_{i=1}^m \lambda_{i,j}^* \lambda_{i,\ell}\right| \|U\|_{\mathscr{H}}$$
$$= \left|\boldsymbol{\lambda}_j^H \boldsymbol{\lambda}_\ell\right| \|U\|_{\mathscr{H}}, \tag{13}$$

where $\boldsymbol{\lambda}_j$ is the jth column of $\boldsymbol{\Lambda}$. Since (13) holds for all U, it follows that

$$\omega_{\min}(\varphi_j^H \varphi_\ell) = \omega_{\max}(\varphi_j^H \varphi_\ell) = \left|\boldsymbol{\lambda}_j^H \boldsymbol{\lambda}_\ell\right|,$$

and hence

$$\mu_{\mathscr{H}}(\Phi) = \max_{j,\ell,j\neq\ell} \frac{\omega_{\max}(\varphi_j^H \varphi_\ell)}{\omega_{\min}(\varphi_\ell^H \varphi_\ell)} = \max_{j,\ell,j\neq\ell} \frac{\left|\boldsymbol{\lambda}_j^H \boldsymbol{\lambda}_\ell\right|}{\|\boldsymbol{\lambda}_\ell\|_2^2} = \mu_{\mathscr{H}}(\boldsymbol{\Lambda}).$$

From [7], [9] we know that to recover a multi-band signal with bandwidth $s/(nT)$ (and with unknown spectral occupancy), it is necessary to sample at a rate $f_s = m/(nT) \geqslant 2s/(nT)$. Theorem 2 implies that uniqueness of $\mathscr{H}$-P0 recovery is guaranteed for multi-band coset sampling if $\mathrm{spark}(\Phi)/2 = m/2 > s$. Hence, sampling at rate at least $2s/(nT)$ is also sufficient to recover an s-sparse signal and recovery of the (multi-coset sampled) signal can be achieved through $\mathscr{H}$-P0.

D. Relation to further results

Theorem 2 in this paper implies [11, Prop. 4] and [11, Eq. (23)] with the observation that the generalized Gram matrix in [11, Eq. (17)] plays the role of the sampling operator Φ in our framework. Our Theorem 2 also implies [12, Th. 2.2].

VI. Uncertainty Relations and Signal Separation

Another thrust in the sparse signal recovery literature deals with the recovery of sparsely corrupted signals [16]. The main tool underlying this line of work is an uncertainty relation that sets a limit on how sparsely a given signal can be represented concurrently in two different dictionaries [1], [15], [16]. We next formulate a Hilbert space version of this uncertainty relation, which is then used to recover and generalize results in [10] and [16].

Theorem 5 (Uncertainty relation): Let $\mathscr{H}_1, \mathscr{H}_2$, and $\mathscr{G}$ be Hilbert spaces and let $\Phi: \mathscr{H}_1 \to \mathscr{G}$ and $\Psi: \mathscr{H}_2 \to \mathscr{G}$ be sampling operators. Let $u \in \mathscr{H}_1$ and $v \in \mathscr{H}_2$ be signals that

are $\varepsilon_{\mathcal{U}}$- and $\varepsilon_{\mathcal{V}}$-concentrated to the sets $\mathcal{U}$ and $\mathcal{V}$, respectively, and assume that $\Phi(u) = \Psi(v)$. Then, we have

$$|\mathcal{U}||\mathcal{V}| \geqslant \frac{1}{\mu_{\mathscr{H}}^2(\Phi, \Psi)} [(1 - \varepsilon_{\mathcal{U}})(1 + \mu_{\mathscr{H}}(\Phi)) - |\mathcal{U}|\,\mu_{\mathscr{H}}(\Phi)]^+$$
$$\times [(1 - \varepsilon_{\mathcal{V}})(1 + \mu_{\mathscr{H}}(\Psi)) - |\mathcal{V}|\,\mu_{\mathscr{H}}(\Psi)]^+. \tag{14}$$

Remark: [16, Th. 1] can be recovered from Theorem 5 by noting that Φ and Ψ play the role of the dictionaries $\mathbf{A}$ and $\mathbf{B}$, respectively, as used in [16]. Then $\Phi(\boldsymbol{u}) = \Psi(\boldsymbol{v})$ becomes $\mathbf{A}\boldsymbol{u} = \mathbf{B}\boldsymbol{v}$ and [16, Th. 1] follows since $\mu_{\mathscr{H}}(\Phi, \Psi) = \mu_m$, $\mu_{\mathscr{H}}(\Phi) = \mu_a$, and $\mu_{\mathscr{H}}(\Psi) = \mu_b$, with μ_m, μ_a, and μ_b as defined in [16].

A. Shift-invariant spaces

We next show how Theorem 5 can be used to recover [10, Th. 1]. Consider the shift-invariant space

$$\mathfrak{S}_\phi \triangleq \left\{ z : z(t) = \sum_{\substack{i=1,\ldots,n_1 \\ k \in \mathbb{Z}}} v_k^{(i)} \phi_i(t - kT), v^{(i)} \in \ell_2, \forall i \right\}, \tag{15}$$

with n_1 generators $\phi_i \in \mathrm{L}_2(\mathbb{R})$ and $\|\phi_i\|_2 = 1$, for all i. Set $\mathscr{H}_1$ to be the space of vector sequences

$$v = \begin{pmatrix} v^{(1)} \\ \vdots \\ v^{(n_1)} \end{pmatrix}, \tag{16}$$

with $v^{(i)} \in \ell_2$, for all i. Define the operator $\varphi_i : \ell_2 \to \mathfrak{S}_\phi$ by

$$\varphi_i\left(v^{(i)}\right) \triangleq \sum_{k \in \mathbb{Z}} v_k^{(i)} \phi_i(\cdot - kT), \tag{17}$$

with adjoint $\varphi_i^H : \mathfrak{S}_\phi \to \ell_2$ given by $\varphi_i^H(z) = \{\langle z(\cdot), \phi_i(\cdot - \ell T)\rangle\}_{\ell \in \mathbb{Z}}$. The sampling operator[3] $\Phi : \mathscr{H}_1 \to \mathfrak{S}_\phi$ is then given by

$$\Phi(v) \triangleq \sum_{i=1}^{n_1} \sum_{k \in \mathbb{Z}} v_k^{(i)} \phi_i(\cdot - kT) = \sum_{i=1}^{n_1} \varphi_i\left(v^{(i)}\right). \tag{18}$$

A signal $v \in \mathscr{H}$ is s-sparse if at most s of the sequences $v^{(i)}$ in (18) are non-zero, i.e., if $\|v\|_{\mathscr{H},0} \leqslant s$, and in the terminology of [10], $\|v\|_{\mathscr{H},0}$ is the number of active generators.

Now let us consider a set of n_2 generators $\theta_i \in \mathrm{L}_2(\mathbb{R})$ where $\|\theta_i\|_2 = 1$, for all i, and the space

$$\mathfrak{S}_\theta \triangleq \left\{ z : z(t) = \sum_{\substack{i=1,\ldots,n_2 \\ k \in \mathbb{Z}}} v_k^{(i)} \theta_i(t - kT), v^{(i)} \in \ell_2, \forall i \right\}.$$

Let $\mathscr{H}_2$ be the space of vector sequences, as in (16), but with n_1 replaced by n_2, and define the operators $\vartheta_i : \ell_2 \to \mathfrak{S}_\theta$ and $\Theta : \mathscr{H}_2 \to \mathfrak{S}_\theta$ as in (17) and (18), respectively, with ϕ_i replaced by θ_i. Suppose that $z = \Phi(v) = \Theta(u)$. We now establish a limit on the sparsity of u and v.

---

[3] In this case, the sampling operator rather behaves like an interpolation operator as it maps a sequence to a continuous-time signal, but to maintain consistency with the rest of the paper we still refer to it as a sampling operator.

Following [10] the generators will be assumed to satisfy:

$$\langle \phi_i(\cdot - kT), \phi_j(\cdot - \ell T) \rangle = \langle \theta_i(\cdot - kT), \theta_j(\cdot - \ell T) \rangle$$

$$= \begin{cases} 1 & \text{if } i = j \text{ and } k = \ell \\ 0 & \text{otherwise.} \end{cases}$$

Then $\left\| \varphi_i(v^{(i)}) \right\|_2 = \left\| v^{(i)} \right\|_2$, for all i, and for all $v^{(i)} \in \ell_2$, hence $\omega_{\min}(\varphi_i) = 1$, for all i, and similarly $\omega_{\min}(\vartheta_\ell) = 1$, for all ℓ. For $i \neq j$ and $v^{(j)} \in \ell_2$, we have

$$\left\| \varphi_i^H \varphi_j \left(v^{(j)} \right) \right\|_2 = \sum_{k \in \mathbb{Z}} \left| \left\langle \varphi_j \left(v^{(j)} \right), \phi_i(\cdot - kT) \right\rangle \right|^2$$

$$= \sum_{k \in \mathbb{Z}} \left| \left\langle \sum_{\ell \in \mathbb{Z}} v_\ell^{(j)} \phi_j(\cdot - \ell T), \phi_i(\cdot - kT) \right\rangle \right|^2$$

$$= \sum_{\substack{k \in \mathbb{Z} \\ \ell \in \mathbb{Z}}} \left| v_\ell^{(j)} \right|^2 \underbrace{\left| \langle \phi_j(\cdot - \ell T), \phi_i(\cdot - kT) \rangle \right|^2}_{=0} = 0,$$

and similarly $\left\| \vartheta_i^H \vartheta_j(u^{(j)}) \right\|_2 = 0$, for all $u^{(j)} \in \ell_2$. Therefore, $\mu_{\mathscr{H}}(\Phi) = \max_{i \neq j} \omega_{\max}(\varphi_i^H \varphi_j) = 0$ and similarly $\mu_{\mathscr{H}}(\Theta) = 0$. This gives

$$\mu_{\mathscr{H}}(\Phi, \Theta) = \max_{i,j} \frac{\omega_{\max}(\varphi_i^H \vartheta_j)}{\omega_{\min}(\varphi_i) \, \omega_{\min}(\vartheta_i)} = \max_{i,j} \omega_{\max}(\varphi_i^H \vartheta_j).$$

The uncertainty relation (14) hence reduces to

$$\|u\|_{\mathscr{H},0} \, \|v\|_{\mathscr{H},0} \, \mu_{\mathscr{H}}^2(\Phi, \Theta) \geqslant 1, \tag{19}$$

where we assume perfect concentration (since this is the case considered in [10]), i.e., $\varepsilon_{\mathcal{U}} = \varepsilon_{\mathcal{V}} = 0$. We now show that (19) is the uncertainty relation in [10, Th. 1], which, in our notation, is given by (19) but with $\mu_{\mathscr{H}}(\Phi, \Theta)$ replaced by

$$\mu(\Phi, \Theta) = \max_{\ell, r} \, \underset{\xi \in [0, 2\pi)}{\text{ess sup}} \left| R_{\phi_\ell, \theta_r}(e^{i\xi}) \right|,$$

where

$$R_{\phi_\ell, \theta_r}(e^{i\xi}) \triangleq \mathcal{F}\left\{ \left\{ \langle \theta_r(\cdot), \phi_\ell(\cdot - kT) \rangle \right\}_{k \in \mathbb{Z}} \right\}(e^{i\xi}).$$

It therefore suffices to prove that $\omega_{\max}(\varphi_\ell^H \vartheta_r) = \text{ess sup}_{\xi \in [0, 2\pi)} \left| R_{\phi_\ell, \theta_r}(e^{i\xi}) \right|$. For $v^{(r)} \in \ell_2$, we have

$$\varphi_\ell^H \vartheta_r \left(v^{(r)} \right) = \left\{ \sum_{k \in \mathbb{Z}} v_k^{(r)} \langle \theta_r(\cdot - kT), \phi_\ell(\cdot - mT) \rangle \right\}_{m \in \mathbb{Z}}$$

$$= \left\{ \sum_{k \in \mathbb{Z}} v_k^{(r)} \Lambda_{m,k} \right\}_{m \in \mathbb{Z}},$$

where $\Lambda_{m,k} \triangleq \langle \theta_r(\cdot - kT), \phi_\ell(\cdot - mT) \rangle$. Then

$$\omega_{\max}(\varphi_\ell^H \vartheta_r) = \sup_{\|v^{(r)}\|_2 = 1} \left\| \varphi_\ell^H \vartheta_r(v^{(r)}) \right\|_2$$

$$= \sup_{\|v^{(r)}\|_2 = 1} \left\| \mathbf{\Lambda} v^{(r)} \right\|_2 = \|\mathbf{\Lambda}\|_{2 \to 2}.$$

Since $\mathbf{\Lambda}$ is a doubly infinite Toeplitz matrix, its operator norm $\|\mathbf{\Lambda}\|_{2 \to 2}$ is given by the essential supremum of the Fourier transform of a row of $\mathbf{\Lambda}$ [23, p. 62]. We therefore

have $\omega_{\max}(\varphi_\ell^H \vartheta_r) = \text{ess sup}_{\xi \in [0, 2\pi)} \left| R_{\phi_\ell, \theta_r}(e^{i\xi}) \right|$, which concludes the proof.

We finally note that our Theorem 5 also applies to nonorthogonal generator sets $\{\phi_i\}$ and $\{\theta_j\}$ with potentially different shift parameters, thereby extending the uncertainty relation in [10].

REFERENCES

[1] D. L. Donoho and P. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, pp. 906–931, Jun. 1989.

[2] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[3] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[4] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[5] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Sig. Proc.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[6] P. T. Boufounos, G. Kutyniok, and H. Rauhut, "Sparse recovery from combined fusion frame measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3864–3876, Jun. 2011.

[7] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," *Proc. IEEE Int. Conf. Acoustics, Speech, and Sig. Proc.*, vol. 3, pp. 1688–1691, Apr. 1996.

[8] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," *Proc. Inf. Theory and Appl. Workshop*, pp. 547–554, Jan. 2008.

[9] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Sig. Proc.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[10] Y. C. Eldar, "Uncertainty relations for shift-invariant analog signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5742–5757, Dec. 2009.

[11] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Sig. Proc.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.

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

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

[14] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[15] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[16] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[17] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York, NY: Springer, 2000.

[18] R. Gribonval and M. Nielsen, "Sparse decompositions in "incoherent" dictionaries," *Proc. IEEE Int. Conf. on Image Proc.*, vol. 1, pp. 33–36, Sep. 2003.

[19] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, pp. 57–98, 1997.

[20] R. A. Horn and C. Johnson, *Matrix Analysis*. New York, NY: Cambridge Univ. Press, 1990.

[21] R. Venkataramani and Y. Bresler, "Sub-Nyquist sampling of multiband signals: Perfect reconstruction and bounds on aliasing error," *Proc. IEEE Int. Conf. Acoustics, Speech, and Sig. Proc.*, vol. 3, pp. 1633–1636, May 1998.

[22] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Sel. Topics in Sig. Proc.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[23] U. Grenander and G. Szegő, *Toeplitz forms and their applications*. New York, NY: AMS Chelsea, 1984.