UNCERTAINTY RELATIONS AND SPARSE DECOMPOSITIONS OF ANALOG SIGNALS

Yonina C. Eldar

Technion—Israel Institute of Technology, Haifa, Israel
yonina@ee.technion.ac.il

ABSTRACT

We consider uncertainty principles for analog signals that lie in a finitely-generated shift-invariant (SI) space. By adapting the notion of coherence defined for finite dictionaries to infinite SI representations, we develop an uncertainty principle similar in spirit to its finite counterpart. Building upon these results and similar work in the finite setting, we show how to find a sparse decomposition of an analog signal in an overcomplete dictionary by solving a convex optimization problem. The distinguishing feature of our approach is the fact that even though the problem is defined over an infinite domain with infinitely many variables and constraints, under certain conditions on the dictionary spectrum our algorithm can find the sparsest representation by solving a finite dimensional problem.

Index Terms— Sparse decompositions, uncertainty principle.

1. INTRODUCTION

In recent years there has been a surge of research into discrete uncertainty relations in finite-dimensional bases [1, 2]. This work has been spurred in part by the relationship between sparse representations and compressed sensing [3, 4]. In particular, several works have shown that discrete uncertainty relations can be used to establish uniqueness of sparse decompositions. Furthermore, there is an intimate connection between uncertainty principles and the ability to recover sparse expansions via convex programming [1, 2, 5].

Using redundant dictionaries can lead to sparse decompositions that use only few dictionary elements. However, finding a sparse expansion in practice is generally a difficult combinatorial problem. Two fundamental questions are how sparse a given signal can be represented, and whether this sparse expansion can be found efficiently. In recent years, several key papers have addressed both of these questions in a discrete setting, in which the signals to be represented are finite-length vectors [1, 2, 5, 6, 7, 8, 4].

Our goal here is to extend these results to the analog domain by first deriving uncertainty relations for decompositions of analog signals in pairs of bases, and then suggesting concrete algorithms to decompose a continuous-time signal into a sparse expansion. In our development, we consider signals that lie in shift-invariant (SI) subspaces $\mathcal{A}$ of L_2 [9, 10]. Such signals can be expressed in terms of linear combinations of shifts of a finite set of generators:

$$\mathcal{A}=\left\{x(t)=\sum_{\ell=1}^{N}\sum_{n\in\mathbb{Z}}a_\ell[n]\phi_\ell(t-nT):a_\ell[n]\in\ell_2\right\},\qquad(1)$$

where $\phi_\ell(t), 1\leq\ell\leq N$ are the SI generators, $a_\ell[n]$ are the expansions coefficients, and T is the period. Clearly, $x(t)$ is characterized by infinitely many coefficients $a_\ell[n]$. Therefore, the finite results which provide bounds on the number of non-zero expansion coefficients in pairs of bases are not immediately relevant here. Instead, we characterize analog sparsity as the number of active generators that comprise a given representation, where the ℓth generator is said to be active if $a_\ell[n]$ is not identically zero.

Considering expansions in two orthonormal bases, we show that the number of active generators in each representation obeys an uncertainty principle similar in spirt to that of finite decompositions. The key to establishing this relation is in defining the analog coherence between the two bases, by replacing the inner product in the finite setting by the largest spectral value of the sampled cross-correlation between basis elements in the analog case. We next address the problem of sparse decomposition in an overcomplete dictionary consisting of two bases. In the finite setting, it can be shown that under certain conditions on the dictionary, a sparse decomposition can be found using computationally efficient algorithms. However, directly generalizing these results to the analog setting is challenging. By exploiting recent results on analog compressed sensing [11, 12, 13] we show that in certain cases an analog sparse decomposition can be found by solving a finite-dimensional convex problem.

In Section 2 we review the discrete uncertainty principle and introduce the class of analog signals we focus on. The analog uncertainty principle is formulated in Section 3. Sparse decompositions in two orthonormal analog bases are discussed in Section 4.

2. PROBLEM FORMULATION

2.1. Discrete Uncertainty Principles

The generalized uncertainty principle is concerned with pairs of representations of a vector $\mathbf{x}\in\mathbb{R}^N$ in two different orthonormal bases [1, 2]: $\{\boldsymbol{\phi}_\ell, 1\leq\ell\leq N\}$ and $\{\boldsymbol{\psi}_\ell, 1\leq\ell\leq N\}$. Any $\mathbf{x}$ can then be decomposed uniquely in terms of each one of these vector sets:

$$\mathbf{x}=\sum_{\ell=1}^{N}a_\ell\boldsymbol{\phi}_\ell=\sum_{\ell=1}^{N}b_\ell\boldsymbol{\psi}_\ell.\qquad(2)$$

The uncertainty relation sets limits on the sparsity of the decomposition for any $\mathbf{x}$. Specifically, let $A=\|\mathbf{a}\|_0$ and $B=\|\mathbf{b}\|_0$ denote the number of non-zero elements in both expansions. Then [2, 1]

$$\frac{1}{2}(A+B)\geq\sqrt{AB}\geq\frac{1}{\mu(\mathbf{\Phi},\mathbf{\Psi})},\qquad(3)$$

where $\mu(\mathbf{\Phi},\mathbf{\Psi})$ is the coherence between the bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$:

$$\mu(\mathbf{\Phi},\mathbf{\Psi})=\max_{\ell,r}|\boldsymbol{\phi}_\ell^H\boldsymbol{\psi}_r|.\qquad(4)$$

It can easily be shown that $1/\sqrt{N}\leq\mu(\mathbf{\Phi},\mathbf{\Psi})\leq 1$ [1].

As we discuss in Section 4, the uncertainty principle provides insight into how sparse a signal $\mathbf{x}$ can be represented in an overcomplete dictionary consisting of $\mathbf{\Phi}$ and $\mathbf{\Psi}$. It also sheds light on

This work was supported by the European Commission FP7 Network of Excellence in Wireless COMmunications NEWCOM++.

the ability to compute such decompositions efficiently. Most of the research on sparse expansions has focused on representing finite-length vectors. First steps towards extending the notions underlying sparse representations and compressed sensing to the analog domain are considered in [11, 13]. Here we continue this development by extending the discrete uncertainty relations to the analog domain.

2.2. Analog Problem Formulation

In our development, we focus on signals of the form (1). In the Fourier domain, we can represent any $x(t) \in \mathcal{A}$ as

$$X(\omega) = \sum_{\ell=1}^{N} A_\ell(e^{j\omega T})\Phi_\ell(\omega), \tag{5}$$

where $X(\omega)$, $\Phi_\ell(\omega)$ are the Fourier transforms of $x(t)$, $\phi_\ell(t)$, and

$$A_\ell(e^{j\omega T}) = \sum_{n\in\mathbb{Z}} a_\ell[n]e^{j\omega nT} \tag{6}$$

is the discrete-time Fourier transform (DTFT) of $a_\ell[n]$.

In order to guarantee a unique stable representation of any signal in $\mathcal{A}$, the generators $\phi_\ell(t)$ are chosen to form a Riesz basis for L_2. This means that there exists constants $\alpha > 0$ and $\beta < \infty$ such that

$$\alpha\mathbf{I} \preceq \mathbf{M}_{\phi\phi}(e^{j\omega}) \preceq \beta\mathbf{I}, \quad \text{a.e. } \omega, \tag{7}$$

where $\mathbf{M}_{\phi\phi}(e^{j\omega})$ is the $N \times N$ matrix with elements $R_{\phi_\ell\phi_r}$. For any two functions $\phi(t), \psi(t)$ with Fourier transforms $\Phi(\omega), \Psi(\omega)$,

$$R_{\phi\psi}(e^{j\omega}) = \frac{1}{T}\sum_{k\in\mathbb{Z}} \overline{\Phi}\left(\frac{\omega}{T} - \frac{2\pi}{T}k\right) \Psi\left(\frac{\omega}{T} - \frac{2\pi}{T}k\right), \tag{8}$$

where $\overline{(\cdot)}$ denotes the complex conjugate. Note that $R_{\phi\psi}(e^{j\omega})$ is the DTFT of the cross-correlation sequence $r_{\phi\psi}[n] = \langle\phi(t - nT), \psi(t)\rangle$, where

$$\langle s(t), x(t)\rangle = \int_{t=-\infty}^{\infty} \overline{s}(t)x(t)dt. \tag{9}$$

Our main interest is in expansions of a signal $x(t)$ in a SI subspace $\mathcal{A}$ of L_2 in terms of orthonormal bases for $\mathcal{A}$. The generators $\{\phi_\ell(t)\}$ of $\mathcal{A}$ form an orthonormal basis if

$$\langle\phi_\ell(t - nT), \phi_r(t - mT)\rangle = \delta_{nm}\delta_{\ell r}, \tag{10}$$

for all ℓ, r and n, m, where $\delta_{nm} = 1$ if $n = m$ and 0 otherwise. This condition can be expressed in the Fourier domain as

$$R_{\phi_\ell\phi_r}(e^{j\omega}) = \delta_{\ell r}. \tag{11}$$

Given an orthonormal basis $\{\phi_\ell(t - nT)\}$ for $\mathcal{A}$, the unique representation coefficients $a_\ell[n]$ in (1) can be computed as $a_\ell[n] = \langle\phi_\ell(t - nT), x(t)\rangle$. This can be seen by taking the inner product of $x(t)$ in (1) with $\phi_\ell(t - mT)$ and using (10).

In the finite-dimensional setting, sparsity is defined as the number of non-zero coefficients in a given basis. However, in an analog decomposition of the form (1), there are in general infinitely many coefficients. Instead, we measure analog sparsity by the number of generators needed to represent $x(t)$. In other words, some of the sequences $a_\ell[n]$ in (1) may be identically zero, in which case

$$x(t) = \sum_{|\ell|=A}\sum_{n\in\mathbb{Z}} a_\ell[n]\phi_\ell(t - nT), \tag{12}$$

where the notation $|\ell| = A$ means a sum over at most A elements. Evidently, sparsity is defined by the energy of the entire sequence $a_\ell[n]$ and not be the values of the individual elements.

In general, the number of zero sequences depends on the choice of basis. Suppose we have an alternative representation

$$x(t) = \sum_{|\ell|=B}\sum_{n\in\mathbb{Z}} b_\ell[n]\psi_\ell(t - nT), \tag{13}$$

where $\{\psi_\ell(t)\}$ also generate an orthonormal basis for $\mathcal{A}$. An interesting question is whether there are limitations on A and B. This question is addressed in the next section and leads to an analog uncertainty principle, similar to (3).

As in the discrete setting we expect to be able to use fewer generators in a SI expansion by allowing for an overcomplete dictionary. In particular, if we expand $x(t)$ using both sets of orthonormal bases we may be able to reduce the number of sequences in the decomposition. The problem is how to find such a sparse representation in practice. In Section 4 we show that despite the combinatorial complexity and infinite dimensions of the problem, under certain conditions on the bases functions, we can recover a sparse decomposition by solving a finite-dimensional convex optimization problem.

3. UNCERTAINTY RELATIONS IN SI SPACES

We begin by developing an analog uncertainty principle for signals in SI subspaces. Specifically, we show that the minimal number of sequences required to express $x(t)$ in terms of any two orthonormal bases has to satisfy the same inequality (3) as in the discrete setting, with an appropriate modification of the coherence [16].

Theorem 1 *Suppose we have a signal $x(t) \in \mathcal{A}$ where $\mathcal{A}$ is a subspace of L_2. Let $\{\phi_\ell(t), 1 \le \ell \le N\}$ and $\{\psi_\ell(t), 1 \le \ell \le N\}$ denote two orthonormal generators of $\mathcal{A}$ so that $x(t)$ can be expressed in both bases with coefficient sequences $a_\ell[n], b_\ell[n]$:*

$$x(t) = \sum_{|\ell|=A}\sum_{n\in\mathbb{Z}} a_\ell[n]\phi_\ell(t - nT) = \sum_{|\ell|=B}\sum_{n\in\mathbb{Z}} b_\ell[n]\psi_\ell(t - nT). \tag{14}$$

Then,

$$\frac{1}{2}(A + B) \ge \sqrt{AB} \ge \frac{1}{\mu(\Phi, \Psi)}, \tag{15}$$

where

$$\mu(\Phi, \Psi) = \max_{\ell,r} \operatorname*{ess\,sup}_{\omega} \left|R_{\phi_\ell\psi_r}(e^{j\omega})\right|, \tag{16}$$

and $R_{\phi\psi}(e^{j\omega})$ is defined by (8).

An interesting question is how small $\mu(\Phi, \Psi)$ can be made by appropriately choosing the bases. From Theorem 1 the smaller $\mu(\Phi, \Psi)$, the stronger the restriction on the sparsity in both decompositions. As we will see in Section 4, such a limitation is helpful in recovering the true sparse coefficients.

Theorem 2 *Let $\{\phi_\ell(t), 1 \le \ell \le N\}$ and $\{\psi_\ell(t), 1 \le \ell \le N\}$ denote two orthonormal generators of a SI subspace $\mathcal{A} \subset L_2$ and let $\mu(\Phi, \Psi) = \max_{\ell,r} \operatorname{ess\,sup} \left|R_{\phi_\ell\psi_r}(e^{j\omega})\right|$, where $R_{\phi\psi}(e^{j\omega})$ is defined by (8). Then*

$$\frac{1}{\sqrt{N}} \le \mu(\Phi, \Psi) \le 1. \tag{17}$$

4. RECOVERY OF SPARSE REPRESENTATIONS

4.1. Discrete Representations

One of the important implications of the discrete uncertainty principle is its relation with sparse approximations [1, 2, 7, 8]. Given two orthonormal bases $\mathbf{\Phi}, \mathbf{\Psi}$ for $\mathbb{R}^N$ an interesting question is whether one can reduce the number of non-zero expansion coefficients by decomposing $\mathbf{x} \in \mathbb{R}^N$ in terms of the concatenated dictionary

$$\mathbf{D} = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\Psi} \end{bmatrix}. \quad (18)$$

In many cases such a representation can be much sparser than the decomposition in either of the bases alone. Finding a sparse expansion $\mathbf{x} = \mathbf{D}\boldsymbol{\gamma}$ in which $\boldsymbol{\gamma}$ has as few non-zero components as possible can be translated into the combinatorial optimization problem

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_0 \quad \text{s.t. } \mathbf{x} = \mathbf{D}\boldsymbol{\gamma}. \quad (19)$$

Clearly, (19) is NP-complete in general and cannot be solved efficiently. The surprising result of [1, 2, 5] is that if $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ is small enough with respect to the sparsity of $\boldsymbol{\gamma}$, then the sparsest possible $\boldsymbol{\gamma}$ is unique and can be found by the basis pursuit algorithm in which the non-convex ℓ_0 norm is replaced by the convex ℓ_1 norm:

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_1 \quad \text{s.t. } \mathbf{x} = \mathbf{D}\boldsymbol{\gamma}. \quad (20)$$

Proposition 1 *Let $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$ be a dictionary consisting of two orthonormal bases with coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{\ell,r} |\phi_\ell^H \psi_r|$. If a vector $\mathbf{x}$ has a sparse decomposition in $\mathbf{D}$ such that $\mathbf{x} = \mathbf{D}\boldsymbol{\gamma}$ and $\|\boldsymbol{\gamma}\|_0 < 1/\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ then this representation is unique, namely there cannot be another $\boldsymbol{\gamma}'$ with $\|\boldsymbol{\gamma}'\|_0 < 1/\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ and $\mathbf{x} = \mathbf{D}\boldsymbol{\gamma}'$. Furthermore, if*

$$\|\boldsymbol{\gamma}\|_0 < \frac{\sqrt{2} - 0.5}{\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})}, \quad (21)$$

then the unique sparse representation can be found by solving (20).

As detailed in [1, 2], the proof of Proposition 1 follows from the generalized discrete uncertainty principle.

Another useful result on dictionaries with low coherence is that every set of $k \leq 2/\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) - 1$ columns are linearly independent [7, Theorem 6].

Proposition 2 *Let $\mathbf{D} = [\boldsymbol{\Phi}\ \boldsymbol{\Psi}]$ be a dictionary consisting of two orthonormal bases with coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$. Then $\sigma(\mathbf{D}) \geq 2/\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) - 1$ where $\sigma(\mathbf{D})$ is the Kruskal rank of $\mathbf{D}$ and is equal to the maximal number q such that every set of q columns of $\mathbf{D}$ is linearly independent.*

4.2. Analog Representations

We would now like to generalize these recovery results to the analog setup. However, it is not immediately clear how to extend the finite ℓ_1 basis pursuit algorithm of (20) to the analog domain.

To set up the analog sparse decomposition problem, suppose we have a signal $x(t)$ that lies in a space $\mathcal{A}$, and let $\{\phi_\ell(t - nT)\}, \{\psi_\ell(t - nT)\}$ be two orthonormal generators of $\mathcal{A}$ with $1 \leq \ell \leq N$. Our goal is to represent $x(t)$ in terms of the joint dictionary $\{d_\ell(t - nT), 1 \leq \ell \leq 2N\}$ with

$$d_\ell(t) = \begin{cases} \phi_\ell(t), & 1 \leq \ell \leq N; \\ \psi_{\ell-N}(t), & N+1 \leq \ell \leq 2N, \end{cases} \quad (22)$$

using as few non-zero sequences as possible. Denoting by $\boldsymbol{\gamma}[n]$ the vector at point-n whose elements are $\gamma_\ell[n]$, our problem is to choose the vector sequence $\boldsymbol{\gamma}[n]$ such that

$$x(t) = \sum_{\ell=1}^{2N} \sum_{n \in \mathbb{Z}} \gamma_\ell[n] d_\ell(t - nT), \quad (23)$$

and $\gamma_\ell[n]$ is identically zero for the largest possible number of indices ℓ. The number of non-zero sequences $\gamma_\ell[n]$ is equal to $\|\mathbf{c}\|_0$ where $c_\ell = \|\gamma_\ell[n]\|_2$. For ease of notation, we denote $\|\boldsymbol{\gamma}\|_{2,0} = \|\mathbf{c}\|_0$, and similarly $\|\boldsymbol{\gamma}\|_{2,1} = \|\mathbf{c}\|_1$. Finding the sparsest decomposition (23) can then be written as

$$\min_{\boldsymbol{\gamma}} \|\boldsymbol{\gamma}\|_{2,0} \quad \text{s.t. } x(t) = \sum_{\ell=1}^{2N} \sum_{n \in \mathbb{Z}} \gamma_\ell[n] d_\ell(t - nT). \quad (24)$$

Problem (24) is the analog version of (19). However, in addition to being combinatorial as its finite counterpart, (24) also has infinitely many variables and constraints. In order to extend the finite-dimensional decomposition results to the analog domain, we need to address whether there is a unique sparse representation for any input signal in a given dictionary, and how (24) can be solved in practice, despite the combinatorial complexity and infinite dimensions.

The uniqueness condition of Proposition 1 can be readily extended to the analog case, since its proof is based on the uncertainty relation (3) which is identical to (15) with the appropriate modification to the coherence measure.

Proposition 3 *Suppose that a signal $x(t) \in \mathcal{A}$ has a sparse representation in the joint dictionary $\{d_\ell(t - nT)\}$ of (22) which consists of two orthonormal bases $\{\phi_\ell(t-nT), \psi_\ell(t-nT)\}$ for $1 \leq \ell \leq N$. If $\|\boldsymbol{\gamma}\|_{2,0} < 1/\mu(\Psi, \Phi)$ where $\mu(\Psi, \Phi)$ is the coherence defined by (16), then this representation is unique.*

The second more fundamental question is how to find a unique sparse representation when it exists. We may attempt to develop a solution by replacing the ℓ_0 norm in (24) by an ℓ_1 norm, as in the finite-dimensional case. However, in practice, it is not clear how to solve the resulting problem since it is defined over an infinite set of variables $\gamma_\ell[n]$, and has infinitely many constraints (for all t).

In Section 4.4 we show that (24) can be converted into an equivalent finite-dimensional problem when $\mathbf{M}_{\phi\psi}(e^{j\omega})$ can be written as

$$\mathbf{M}_{\phi\psi}(e^{j\omega}) = \mathbf{A}\mathbf{Z}(e^{j\omega}). \quad (25)$$

Here $\mathbf{A}$ is a fixed matrix independent of ω, and $\mathbf{Z}(e^{j\omega})$ is an invertible diagonal matrix with diagonal elements $Z_\ell(e^{j\omega})$. The columns of $\mathbf{A}$ are normalized such that $\operatorname{ess\,sup} Z_\ell(e^{j\omega}) = 1$ for all ℓ.

Our main result is that under the condition (25) we can convert (24) into a multiple measurement vector (MMV) problem in which our goal is to represent a set of m vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ using a finite-dimensional dictionary $\mathbf{D} = [\mathbf{I}\ \mathbf{A}]$ with common sparsity, namely the non-zero elements in the expansion of the m vectors are all supported on a fixed location set. In order to develop these results, in the next section we review the MMV model and a recently developed generalization to the case in which it is desirable to jointly decompose infinitely many vectors $\mathbf{x}_i$ in terms of a given dictionary $\mathbf{D}$. This extension is referred to as the infinite measurement model (IMV) [12]. In Section 4.4 we show how these ideas can be used to find the sparsest decomposition in a SI subspace when (25) holds, by first solving a finite-dimensional convex optimization problem.

4.3. MMV and IMV Models

In an MMV problem our goal is to jointly decompose m vectors $\mathbf{x}_i, 1 \leq i \leq m$ in a dictionary $\mathbf{D}$. Denoting by $\mathbf{X}$ the matrix with columns $\mathbf{x}_i$, our goal is to seek a matrix $\boldsymbol{\Gamma}$ with columns $\boldsymbol{\gamma}_i$ such that $\mathbf{X} = \mathbf{D}\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}$ has as few non-zero rows as possible. The results in [14, 15] establish that under the same conditions as Proposition 1, the unique $\boldsymbol{\Gamma}$ can be found by solving an extension of the ℓ_1 program:

$$\min_{\boldsymbol{\Gamma}} \|\mathbf{s}(\boldsymbol{\Gamma})\|_1 \quad \text{s.t. } \mathbf{X} = \mathbf{D}\boldsymbol{\Gamma}. \quad (26)$$

Here $\mathbf{s}(\boldsymbol{\Gamma})$ is a vector whose ℓth element is equal to $\|\boldsymbol{\Gamma}^\ell\|$ where $\boldsymbol{\Gamma}^\ell$ is the ℓth row of $\boldsymbol{\Gamma}$, and the norm is an arbitrary vector norm.

Proposition 4 *Let* $\mathbf{X}$ *be an* $N\times m$ *matrix with columns* $\mathbf{x}_i, 1\leq i\leq m$ *that are jointly sparse in the dictionary* $\mathbf{D}=[\mathbf{\Phi}\ \mathbf{\Psi}]$ *consisting of two orthonormal bases, so that* $\mathbf{X}=\mathbf{D}\mathbf{\Gamma}$ *with* $\|\mathbf{s}(\mathbf{\Gamma})\|_0=k$. *If* $k<1/\mu(\mathbf{\Phi},\mathbf{\Psi})$ *where* $\mu(\mathbf{\Phi},\mathbf{\Psi})=\max_{\ell,r}|\phi_\ell^H\psi_r|$, *then this representation is unique. Furthermore, if*

$$k<\frac{\sqrt{2}-0.5}{\mu(\mathbf{\Phi},\mathbf{\Psi})}, \tag{27}$$

then the unique sparse representation can be found by solving (26) with any vector norm.

The MMV model has been recently generalized to the IMV case in which there are infinitely many vectors $\mathbf{x}$ of length N, and infinitely many coefficient vectors $\boldsymbol{\gamma}$:

$$\mathbf{x}(\lambda)=\mathbf{D}\boldsymbol{\gamma}(\lambda),\quad \lambda\in\Lambda, \tag{28}$$

where Λ is some set whose cardinality can be infinite and uncountable, such as the set of frequencies $\omega\in[-\pi,\pi)$. The k-sparse IMV model assumes that the vectors $\{\boldsymbol{\gamma}(\lambda)\}$, which we denote for brevity by $\boldsymbol{\gamma}(\Lambda)$, share a joint sparsity pattern, so that the non-zero elements are all supported on a fixed location set of size k [12].

A major difficulty with the IMV model is that it is not clear how to determine the entire solution $\boldsymbol{\gamma}(\Lambda)$ since there are infinitely many equations to solve. In [12] it was shown that (28) can be converted to a finite MMV without loosing any information by a set of operations that are referred to as the continuous-to-finite (CTF) block. The essential idea is to first recover the support of $\boldsymbol{\gamma}(\Lambda)$, namely the non-zero location set, by solving a finite MMV, and then reconstruct $\boldsymbol{\gamma}(\Lambda)$ from the data $\mathbf{x}(\Lambda)$ and the knowledge of the support, which we denote by S. The reason for this separation is that once S is known, (28) becomes invertible when the coherence is low enough.

Let $\mathbf{D}_S$ denote the matrix containing the subset of the columns of $\mathbf{D}$ whose indices belong to S. Then (28) can be written as

$$\mathbf{x}(\lambda)=\mathbf{D}_S\boldsymbol{\gamma}^S(\lambda),\quad \lambda\in\Lambda, \tag{29}$$

where the superscript $\boldsymbol{\gamma}^S(\lambda)$ is the vector that consists of the entries of $\boldsymbol{\gamma}(\lambda)$ in the locations S. Since $\boldsymbol{\gamma}(\Lambda)$ is k-sparse, $|S|\leq k$. In addition, from Proposition 2 it follows that if $\mu(\mathbf{\Phi},\mathbf{\Psi})<1/k$ then every k columns of $\mathbf{D}$ are linearly independent. Therefore $\mathbf{D}_S$ consists of linearly independent columns implying that $(\mathbf{D}_S)^\dagger\mathbf{D}_S=\mathbf{I}$, where $(\mathbf{D}_S)^\dagger=\left(\mathbf{D}_S^H\mathbf{D}_S\right)^{-1}\mathbf{D}_S^H$ is the Moore-Penrose pseudo-inverse of $\mathbf{D}_S$. Multiplying (29) by $(\mathbf{D}_S)^\dagger$ on the left gives

$$\boldsymbol{\gamma}^S(\lambda)=(\mathbf{D}_S)^\dagger\mathbf{x}(\lambda),\quad \lambda\in\Lambda. \tag{30}$$

The elements in $\boldsymbol{\gamma}(\lambda)$ not supported on S are all zero. Therefore (30) allows for exact recovery of $\boldsymbol{\gamma}(\Lambda)$ once S is correctly identified.

In order to determine S we exploit the fact that $\mathrm{span}(\mathbf{x}(\Lambda))$ is finite, since $\mathbf{x}(\lambda)$ is of length N. In addition, it is shown in [12] that if there exists a solution set $\boldsymbol{\gamma}(\Lambda)$ with sparsity k, and the matrix $\mathbf{D}$ has Kruskal rank $\sigma(\mathbf{D})\geq 2k$, then every finite collection of vectors spanning the subspace $\mathrm{span}(\mathbf{x}(\Lambda))$ contains sufficient information to recover S exactly. Therefore, to find S all we need is to construct a matrix $\mathbf{V}$ whose range space is equal to $\mathrm{span}(\mathbf{x}(\Lambda))$. We are then guaranteed that the linear system

$$\mathbf{V}=\mathbf{D}\mathbf{U} \tag{31}$$

has a unique k-sparse solution $\mathbf{U}$ whose row support is equal S. This allows to avoid the infinite structure of (28) and to concentrate on finding S by solving the single MMV system of (31). The solution can be determined using an ℓ_1 relaxation of the form (26) with $\mathbf{V}$ replacing $\mathbf{X}$, as long as the conditions of Proposition 4 hold.

In practice, an appropriate matrix $\mathbf{V}$ can be constructed by first forming the matrix $\mathbf{Q}=\int_{\lambda\in\Lambda}\mathbf{x}(\lambda)\mathbf{x}^H(\lambda)d\lambda$. Every $\mathbf{V}$ satisfying $\mathbf{Q}=\mathbf{V}\mathbf{V}^H$ will then have a column span equal to $\mathrm{span}(\mathbf{x}(\Lambda))$ [12].

4.4. Analog Dictionaries

We now return to the analog decomposition problem (24) and show how to exploit the results presented in the previous section on IMV models in order to find the sparsest SI expansion by solving a finite-dimensional problem. Our approach is comprised of three steps:

1. Convert (24) into an IMV system.
2. Use the reduction from IMV to MMV in order to find the active generators $d_\ell(t), \ell\in S$.
3. Invert the system of equations over S.

The last two steps were detailed in the previous section. It remains to show how (24) can be transformed into an IMV.

We begin by noting that since $\{\phi_\ell(t)\}$ generate an orthonormal basis for $\mathcal{A}$, $x(t)$ is uniquely determined by the N sequences

$$c_\ell[n]=\langle\phi_\ell(t-nT),x(t)\rangle=r(nT), \tag{32}$$

where $r(t)$ is the convolution $r(t)=\phi(-t)*x(t)$. Therefore, constraining $x(t)$ is equivalent restricting $c_\ell[n]$. Taking the inner products on both sides of (24) with respect to $\phi_\ell(t-nT)$ leads to

$$c_r[m]=\sum_{\ell=1}^{2N}\sum_{n\in\mathbb{Z}}\gamma_\ell[n]a_{r\ell}[m-n], \tag{33}$$

where $a_{r\ell}[n]=\langle\phi_r(t-nT),d_\ell(t)\rangle$, or, in the Fourier domain,

$$C_r(e^{j\omega})=\sum_{\ell=1}^{2N}\Gamma_\ell(e^{j\omega})A_{r\ell}(e^{j\omega}),\quad 1\leq r\leq N. \tag{34}$$

To simplify (34) we use the definition (22) of $d_\ell(t)$. Since $\langle\phi_r(t-nT),\phi_\ell(t)\rangle=\delta_{r\ell}\delta_{n0}$ and the Fourier transform of $\langle\phi_r(t-nT),\psi_\ell(t)\rangle$ is $R_{\phi_r\psi_\ell}(e^{j\omega})$, (34) can be written as

$$\mathbf{c}(e^{j\omega})=\begin{bmatrix}\mathbf{I} & \mathbf{M}_{\phi\psi}(e^{j\omega})\end{bmatrix}\boldsymbol{\gamma}(e^{j\omega}), \tag{35}$$

where $\mathbf{c}(e^{j\omega}),\boldsymbol{\gamma}(e^{j\omega})$ are the vectors with elements $C_\ell(e^{j\omega}),\Gamma_\ell(e^{j\omega})$ respectively. Our sparse recovery problem (24) is therefore equivalent to

$$\begin{array}{ll}\min_{\boldsymbol{\gamma}} & \|\boldsymbol{\gamma}(e^{j\omega})\|_{2,0}\\ \text{s.t.} & \mathbf{c}(e^{j\omega})=\begin{bmatrix}\mathbf{I} & \mathbf{M}_{\phi\psi}(e^{j\omega})\end{bmatrix}\boldsymbol{\gamma}(e^{j\omega}).\end{array} \tag{36}$$

The minimization in (36) is very similar to the IMV problem (28). Indeed, we seek a set of vectors $\boldsymbol{\gamma}$ with joint sparsity that have the smallest number of non-zero rows, and satisfy an infinite set of linear matrix relations. However, in contrast to (28), the matrix here depends on ω. If $\mathbf{M}_{\phi\psi}(e^{j\omega})$ has the form (25), then (36) can be converted to a finite MMV problem. Indeed, let the first N elements of $\boldsymbol{\gamma}(e^{j\omega})$ be denoted by $\mathbf{a}(e^{j\omega})$ and the remaining N elements by $\mathbf{b}(e^{j\omega})$. Then (36) becomes

$$\begin{array}{ll}\min_{\mathbf{a},\mathbf{d}} & \|\mathbf{a}(e^{j\omega})\|_{2,0}+\|\mathbf{d}(e^{j\omega})\|_{2,0}\\ \text{s.t.} & \mathbf{c}(e^{j\omega})=\begin{bmatrix}\mathbf{I} & \mathbf{A}\end{bmatrix}\begin{bmatrix}\mathbf{a}(e^{j\omega})\\ \mathbf{d}(e^{j\omega})\end{bmatrix},\end{array} \tag{37}$$

where $\mathbf{d}(e^{j\omega})=\mathbf{Z}(e^{j\omega})\mathbf{b}(e^{j\omega})$, and we used the fact that since $\mathbf{Z}(e^{j\omega})$ is diagonal and invertible, $\|\mathbf{b}(e^{j\omega})\|_{2,0}=\|\mathbf{d}(e^{j\omega})\|_{2,0}$ so that the two vector sequences have the same sparsity. Problem (37) has the required IMV form. It can be solved by first finding the sparsest matrix $\mathbf{U}$ that satisfies $\mathbf{C}=[\mathbf{I}\ \mathbf{A}]\mathbf{U}$ where the columns of $\mathbf{C}$ form a basis for the span of $\{\mathbf{c}(e^{j\omega}),-\pi\leq\omega\leq\pi\}$. To determine $\mathbf{U}$ we consider the convex program

$$\min_{\mathbf{U}}\|\mathbf{s}(\mathbf{U})\|_1\quad \text{s.t.}\ \mathbf{C}=\begin{bmatrix}\mathbf{I} & \mathbf{A}\end{bmatrix}\mathbf{U}. \tag{38}$$

Let S denote the rows in $\mathbf{U}$ that are not identically zero and let $\boldsymbol{\gamma}^S[n]$ be the corresponding sequences $\gamma_\ell[n], \ell \in S$. Then

$$\boldsymbol{\gamma}^S(e^{j\omega}) = \begin{bmatrix} \mathbf{I} \\ \mathbf{Z}_{S'}^{-1}(e^{j\omega}) \end{bmatrix} (\mathbf{D}_S^H \mathbf{D}_S)^{-1} \mathbf{D}_S^H \mathbf{c}(e^{j\omega}), \qquad (39)$$

where S' denotes the rows in S between $N+1$ and $2N$. The remaining sequences $\gamma_\ell, \ell \notin S$ are identically zero.

Proposition 4 provides conditions under which (38) will find the sparsest representation assuming that $\mathbf{A}$ is a unitary matrix. These conditions are stated in terms of the coherence $\mu(\mathbf{I}, \mathbf{A})$. It is easy to see that $\mu(\mathbf{I}, \mathbf{A}) = \mu(\Phi, \Psi)$. Furthermore, it can readily shown that $\mathbf{A}$ is indeed unitary [16]. We therefore have the following theorem.

Theorem 3 *Let $\{\phi_\ell(t), 1 \leq \ell \leq N\}$ and $\{\psi_\ell(t), 1 \leq \ell \leq N\}$ denote two orthonormal generators of a subspace $\mathcal{A}$ of L_2 with coherence $\mu(\Psi, \Phi)$. Let $x(t)$ be a signal in $\mathcal{A}$ and suppose there exists sequences $a_\ell[n], b_\ell[n]$ such that*

$$x(t) = \sum_{\ell=1}^{N} \sum_{n \in \mathbb{Z}} (a_\ell[n]\phi_\ell(t - nT) + b_\ell[n]\psi_\ell(t - nT)) \qquad (40)$$

with $k = \|\mathbf{a}\|_{2,0} + \|\mathbf{b}\|_{2,0}$ satisfying $k < (\sqrt{2} - 0.5)/\mu(\mathbf{\Phi}, \mathbf{\Psi})$. Let $\mathbf{M}_{\phi\psi}(e^{j\omega}) = \mathbf{A}\mathbf{Z}(e^{j\omega})$ be the cross-correlation matrix, where $\mathbf{Z}(e^{j\omega})$ is an invertible diagonal matrix with diagonal elements $Z_\ell(e^{j\omega})$ satisfying $\max_\omega Z_\ell(e^{j\omega}) = 1$. Then, the sequences $a_\ell[n]$ and $b_\ell[n]$ can be found by solving

$$\begin{array}{ll} \min_{\mathbf{\Gamma}_1, \mathbf{\Gamma}_2} & \|\mathbf{s}(\mathbf{\Gamma}_1)\|_1 + \|\mathbf{s}(\mathbf{\Gamma}_2)\|_1 \\ \text{s.t.} & \mathbf{C} = \begin{bmatrix} \mathbf{I} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \end{bmatrix}. \end{array} \qquad (41)$$

Here $\mathbf{C}$ is chosen such that its columns form a basis for the range of $\{\mathbf{c}(e^{j\omega}), \omega \in (-\pi, \pi]\}$ where the ℓth component of $\mathbf{c}(e^{j\omega})$ is the Fourier transform at frequency ω of $c_\ell[n] = \langle \phi_\ell(t - nT), x(t) \rangle$, and $\mathbf{s}(\mathbf{\Gamma}_i)$ is a vector whose ℓth element is equal to $\|\mathbf{\Gamma}_i^\ell\|$ where the norm is arbitrary. Let S_1, S_2 denote the rows of $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2$ that are not identically equal 0, and define $\mathbf{D}_S = [\mathbf{I}_{S_1} \ \mathbf{A}_{S_2}]$. Then the non-zero sequences $a_\ell[n], b_\ell[n], \ell \in S$ are given in the Fourier domain by

$$\begin{bmatrix} \mathbf{a}_S(e^{j\omega}) \\ \mathbf{b}_S(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{Z}_{S_2}^{-1}(e^{j\omega}) \end{bmatrix} (\mathbf{D}_S^H \mathbf{D}_S)^{-1} \mathbf{D}_S^H \mathbf{c}(e^{j\omega}). \quad (42)$$

5. CONCLUSION

In this paper, we extended the recent line of work on generalized uncertainty principles to the analog domain, by considering sparse representations in SI bases. We showed that there is a fundamental limit on the ability to sparsely represent an analog signal in an infinite-dimensional SI space in two orthonormal bases. The sparsity bound is similar to that obtained in the finite-dimensional discrete setting with an appropriate measure of coherence. We then treated the problem of sparsely representing an analog signal in an overcomplete dictionary. Building upon the uncertainty principle and some of the recent works in the area of compressed sensing for analog signals, we showed that under certain conditions on the Fourier domain representation of the dictionary, the sparsest representation can be found by solving a finite-dimensional convex optimization problem. This is possible by first sampling the analog signal in a lossless manner. Exploiting the Fourier domain representation of the dictionary together with recent results on infinite measurement models, the problem is converted to finite dimensions.

6. REFERENCES

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

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

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.

[4] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[5] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Transactions on Inform. Theory*, vol. 49, no. 6, pp. 1579–1581, 2003.

[6] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions Info. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

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

[9] C. de Boor, R. DeVore, and A. Ron, "The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$," *J. Funct. Anal*, vol. 119, no. 1, pp. 37–78, 1994.

[10] O. Christensen and Y. C. Eldar, "Generalized shift-invariant systems and frames for subspaces," *J. Fourier Analys. Appl.*, vol. 11, pp. 299–313, 2005.

[11] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," to appear in *IEEE Trans. Signal Process.*

[12] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," to appear in *IEEE Trans. Signal Process.*

[13] Y. C. Eldar, "Compressed sensing of analog signals," submitted to *IEEE Trans. Signal Processing*.

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

[15] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," submitted to *IEEE Trans. Inform. Theory*.

[16] Y. C. Eldar, "Uncertainty relations for analog signals," submitted to *IEEE Trans. Inform. Theory*.