Compressive Covariance Sampling

Daniel Romero
Dept. of Signal Theory and Communications
University of Vigo
Campus A Lagoa, 36310 Vigo, Spain
Email: dromero@gts.uvigo.es

Geert Leus
Faculty of Electrical Engineering and Mathematics
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
Email: g.j.t.leus@tudelft.nl

***Abstract*— Most research efforts in the field of compressed sensing have been pointed towards analyzing sampling and reconstruction techniques for sparse signals, where sampling rates below the Nyquist rate can be reached. When only second-order statistics or, equivalently, covariance information is of interest, perfect signal reconstruction is not required and rate reductions can be achieved even for non-sparse signals. This is what we will refer to as compressive covariance sampling. In this paper, we will study minimum-rate compressive covariance sampling designs within the class of non-uniform samplers. Necessary and sufficient conditions for perfect covariance reconstruction will be provided and connections to the well-known sparse ruler problem will be highlighted.**

I. Problem Statement

Consider the problem of estimating the covariance matrix $\mathbf{\Sigma}$ of a random vector $\boldsymbol{x}$ with components $x[n]$ when it is known that $\mathbf{\Sigma}$ is a linear combination of the matrices in the set $\mathcal{S}$. This problem has a long history and a wide range of applications [1], [2]. Now consider a modification of the same problem where only a subset of the samples $x[n]$ is available, i.e., when the observations are $y[m] = x[n_m]$ for some $\mathcal{I} = \{n_0, n_1, \dots\} \subset \mathbb{Z}$. Intuitively, if the grid defined by $\mathcal{I}$ is dense enough, then this estimation problem, which we label as *compressive covariance sampling*, can still be solved.

More specifically, the above sample selection produces a transformation of the problem: the vector $\boldsymbol{y}$ collecting the *compressed* observations $y[m]$, has a covariance matrix $\bar{\mathbf{\Sigma}}$, which is a linear combination of the matrices in $\bar{\mathcal{S}}$. The coefficients in the linear combination are in both problems the same so that solving the latter amounts to solving the former. In this paper we address the optimal design of the index set $\mathcal{I}$ so that the coefficients can be estimated, i.e., we look for sets of indices with a minimum number of elements guaranteeing that these parameters remain statistically identifiable.

A. Relation to Compressive Sampling

Recent interest in sampling signals below their Nyquist rate owes to compressive sampling (or compressed sensing) [3], which has motivated a great deal of research efforts in the last few years. In compressive sampling we are interested in reducing the number of samples by focusing on a special family of signals referred to as *sparse*. These signals are intrinsically redundant, so that this reduction does not entail any information loss if properly done.

Non-uniform sampling [4] is a particular case of compressive sampling. Mathematically, this process can be described as first acquiring a continuous-index signal $x(t)$ at rate $1/T$, resulting in the sequence $x[n] = x(nT)$, and then downsampling $x[n]$ to obtain $y[m]$ according to a set of indices $\mathcal{I}$, in a periodic or non-periodic fashion (periodic non-uniform sampling is also known as *multi-coset sampling*).

Either in the context of compressive sampling (the non-uniform sampling version) or in the context of what we call compressive covariance sampling, the design of the set $\mathcal{I}$ has been addressed in the past. The common goal in both cases is to minimize the number of elements in $\mathcal{I}$, since this critically determines the cost of the system. However, they differ as to which index sets are regarded feasible:

- In compressive sampling, the *perfect reconstruction criterion* is adopted, i.e., we must be able to exactly reconstruct $x[n]$ from the compressed observations $y[m]$. Discarding the samples whose indices are not present in $\mathcal{I}$ must entail no loss of information, at least in the noiseless case. This is the philosophy used in [5] and [6] for sampling of multi-band signals.
- In other applications, such as those described in the next section, a *covariance sampling criterion* is employed. The interest is no longer in *sparse* signals/vectors but in *stationary* stochastic processes, and the goal is to estimate the second-order statistics, not to reconstruct the signal itself. No redundancy is required: we are only interested in part of the information, thus allowing a reduction in the number of samples. Designs of this kind include nested arrays [7], coprime sampling [8], minimum-redundancy linear arrays [9] and sparse ruler samplers [10], [11].

The purpose of this paper is to formalize the covariance sampling criterion and to develop the theory of optimal designs, in the sense that the number of elements in $\mathcal{I}$ is minimized. Previous works address the compressive covariance sampling problem from completely different perspectives ranging from the number of *degrees of freedom* [7], [8], to the *difference/sum coarray* [12] and the conditions for least squares reconstruction of the second-order statistics [10]. We aim to unify the treatment of this problem under the idea of statistical identifiability of the unknown parameters, providing a framework that encompasses most compressive covariance sampling problems.

B. Applications

The formulation mentioned above accommodates several problems in signal processing and array processing. The differences lie in the particular choice of the set $\mathcal{S}$.

1) Direction of Arrival Estimation: Assume that a uniform linear array with N antennas receives $I-1$ narrowband signals impinging from $I-1$ locations, each one forming a different angle with the array. If $\boldsymbol{x} \in \mathbb{C}^N$ represents the signal received at the N antennas at a particular time, we can write

$$\boldsymbol{x} = \sum_{i=1}^{I-1} \boldsymbol{a}_i s_i + \boldsymbol{w},$$

where $\boldsymbol{a}_i \in \mathbb{C}^N$ is the channel from the i-th transmitter to the array, s_i is the i-th transmitted signal and $\boldsymbol{w} \in \mathbb{C}^N$ is spatially-white noise with power σ_0^2. Then, by considering that s_i is random with variance σ_i^2, uncorrelated with s_j for $j \neq i$, and $\boldsymbol{a}_i$ as a (typically unknown) deterministic constant, we obtain

$$\boldsymbol{\Sigma} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\} = \sum_{i=1}^{I-1} \sigma_i^2 \boldsymbol{a}_i \boldsymbol{a}_i^H + \sigma_0^2 \boldsymbol{I}_N$$

or, more compactly,

$$\boldsymbol{\Sigma} = \sum_{i=0}^{I-1} \sigma_i^2 \boldsymbol{\Sigma}_i \tag{1}$$

with $\boldsymbol{\Sigma}_i = \boldsymbol{a}_i \boldsymbol{a}_i^H, \quad i = 1, \ldots, I-1$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_N$. Because of the arrangement of the antennas in the array, the channel responses have, up to some constant, the form $\boldsymbol{a}_i = [1, e^{j\theta_i}, e^{j2\theta_i}, \ldots, e^{j(N-1)\theta_i}]^T$, where θ_i is related to the direction of arrival (DoA). The problem of DoA estimation is thus that of estimating θ_i. A closely related problem is *incoherent imaging*, where the goal is to estimate σ_i^2 [12].

In this example, the problem of compressive covariance sampling is that of retaining only the antennas with indices in $\mathcal{I}$, removing the others, while taking into account that the smaller the number of antennas, the lower the cost of the system. The set $\mathcal{S}$ is composed of the $I-1$ matrices $\boldsymbol{a}_i \boldsymbol{a}_i^H$ together with the identity matrix $\boldsymbol{I}_N$ to model noise.

2) Power Spectrum Estimation: Assume that an analog waveform $x(t)$ is to be acquired and its power spectrum, or equivalently, its auto-correlation sequence, is to be estimated. If the frequency content of this signal is limited to a finite interval, we can sample it at the Nyquist rate $1/T$ Hz obtaining the samples $x[n] = x(nT)$. If this signal is stationary, the second-order statistics can be collected in the Toeplitz covariance matrix $\boldsymbol{\Sigma} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}$, where $\boldsymbol{x} = [x[0], x[1], \ldots, x[K-1]]^T$ for some K. Thus, we can write

$$\boldsymbol{\Sigma} = \sum_{i=0}^{I-1} \alpha_i \boldsymbol{\Sigma}_i$$

where $I = 2K-1$, $\alpha_i \in \mathbb{R}$ and $\boldsymbol{\Sigma}_i \in \mathcal{B}$, with $\mathcal{B}$ given by

$$\mathcal{B} = \{\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{K-1}\} \cup \{\tilde{\boldsymbol{B}}_1, \ldots, \tilde{\boldsymbol{B}}_{K-1}\},$$

where $\boldsymbol{B}_i$ denotes the (Hermitian Toeplitz) matrix with all zeros but the diagonals $+i$ and $-i$ that have ones, $i = 0$ being used to represent the main diagonal, and $\tilde{\boldsymbol{B}}_i$ denotes the matrix with all zeros but the coefficients in the diagonals $+i$ and $-i$, that take on, respectively, the values j and $-j$. For example, $\boldsymbol{B}_1$ and $\tilde{\boldsymbol{B}}_1$ respectively look like:

$$\boldsymbol{B}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

$$\tilde{\boldsymbol{B}}_1 = \begin{bmatrix} 0 & j & 0 & 0 & \ldots & 0 & 0 \\ -j & 0 & j & 0 & \ldots & 0 & 0 \\ 0 & -j & 0 & j & \ldots & 0 & 0 \\ 0 & 0 & -j & 0 & \ldots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \ldots & 0 & j \\ 0 & 0 & 0 & 0 & \ldots & -j & 0 \end{bmatrix}$$

In this case, the problem of compressive covariance sampling is to design a pattern where we retain the samples with indices in $\mathcal{I}$ whereas we discard the others (see [10] and references therein). We thus obtain a sequence $y[m] = x[n_m], \ n_m \in \mathcal{I} = \{n_0, n_1, \ldots\}$, which, in general, is no longer stationary, but must preserve the information about the second-order statistics of $\boldsymbol{x}$, i.e., we should be able to estimate α_i if $\mathcal{I}$ is properly selected. Clearly, in this case, $\mathcal{S} = \mathcal{B}$.

3) Line Spectrum Estimation: Assume that the n-th sample of the signal under analysis is of the form

$$x[n] = \sum_{i=1}^{I-1} \sigma_i e^{j\omega_i n + j\phi_i} + w[n]$$

where $\sigma_i \in \mathbb{R}$, $w[n]$ is white noise with variance σ_0^2 and the ϕ_i's are independent and identically distributed (i.i.d.) uniform $\mathcal{U}(0, 2\pi)$ random variables. By arranging these samples as $\boldsymbol{x} = [x[0], x[1], \ldots, x[K-1]]^T$ we obtain that

$$\boldsymbol{\Sigma} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\} = \sum_{i=1}^{I-1} \sigma_i^2 \boldsymbol{e}_i \boldsymbol{e}_i^H + \sigma_0^2 \boldsymbol{I}_K$$

where $\boldsymbol{e}_i = [1, e^{j\omega_i}, e^{j2\omega_i}, \ldots, e^{j(K-1)\omega_i}]^T$. Expression (1) can be applied again with $\boldsymbol{\Sigma}_i = \boldsymbol{e}_i \boldsymbol{e}_i^H, \quad i = 1, \ldots, I-1$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_N$. The problem of compressive covariance sampling is to retain the samples $x[n]$ with index n in $\mathcal{I}$ in such a way that the coefficients σ_i^2 can still be estimated. As before, the lower the number of elements in $\mathcal{I}$, the lower the cost of the system.

4) Wideband Spectrum Sensing: Assume that a wide frequency band is sampled resulting in the sequence $x[n]$. A certain number of users are transmitting in this band, typically in different frequency channels. The sequence $x[n]$ can thus be seen as a sum of independent unit-power signals $x_i[n]$ as

$$x[n] = \sum_{i=0}^{I-1} \sigma_i x_i[n]$$

for some values of $\sigma_i \in \mathbb{R}$. By vectorizing these signals and computing their covariance, expression (1) comes up again, where $x_0[n]$ may account for noise and $\mathbf{\Sigma}_i$ contains the second-order statistics of $x_i[n]$, which are typically known since channelization information is publicly available [13], [14]. As in most communications applications, the signals $x_i[n]$ are assumed wide-sense stationary so that the associated covariance matrices $\mathbf{\Sigma}_i$ will be Toeplitz.

In wideband spectrum sensing, the set $\mathcal{S}$ is given by $\mathcal{S} = \{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_{I-1}\}$ and we wish to estimate the coefficients σ_i^2, which actually represent the power received from the i-th user since $\mathbf{\Sigma}_i$ is normalized. This is of application in the context of Dynamic Spectrum Sharing [15]. The problem of compressive covariance sampling is that of selecting the smallest subset of samples that allows estimation of the coefficients σ_i^2.

C. Paper Structure

We start by introducing periodic non-uniform sampling, also known as multi-coset sampling, in Sec. II. From this perspective, a design principle termed the *covariance sampling criterion*, which was implicitly used in some works in the literature, is formalized and generalized in Sec. III. Universal covariance samplers are defined as those satisfying the covariance sampling criterion for any set $\mathcal{S}$, a concept which is the dual of the universal sampler concept for perfect reconstruction of multi-band signals in [5]. Necessary and sufficient conditions for universality are derived, leading to the well-known sparse ruler problem [16]. Later, it is shown in Sec. V that the problem of periodic sampling simplifies under mild conditions, resulting in quasi-universal covariance samplers. A mathematical tool referred to as a *circular sparse ruler* is proposed to design optimal quasi-universal covariance samplers.

II. Multi-coset Sampling

Let $x[n]$ represent a discrete-index sequence of complex numbers which is to be processed. Multi-coset sampling considers independently each group of N samples, which can be denoted as $\boldsymbol{x}[l] = [x[lN], x[lN+1], \ldots x[lN+(N-1)]]^T \in \mathbb{C}^N$. From each period, only M samples are retained according to the pattern $\mathcal{P} = \{p_0, p_1, \ldots, p_{M-1}\}$, forming the vector $\boldsymbol{y}[l] = [x[lN+p_0], x[lN+p_1], \ldots x[lN+p_{M-1}]]^T \in \mathbb{C}^M$. If a total of L periods are acquired, it is possible to arrange these vectors as $\boldsymbol{x} = [\boldsymbol{x}^T[0], \boldsymbol{x}^T[1], \ldots, \boldsymbol{x}^T[L-1]]^T \in \mathbb{C}^{NL}$ and $\boldsymbol{y} = [\boldsymbol{y}^T[0], \boldsymbol{y}^T[1], \ldots, \boldsymbol{y}^T[L-1]]^T \in \mathbb{C}^{ML}$. The set $\mathcal{I}$ containing the indices of the samples selected from all periods can be expressed as

$$\mathcal{I} = \{lN + p : \quad l = 0, 1, \ldots, L-1, \; p \in \mathcal{P}\}. \qquad (2)$$

In view of (2), we can say that the set $\mathcal{I}$ is *periodic* with period defined by the set $\mathcal{P}$. To remove ambiguities, $\mathcal{P}$ is the smallest set satisfying (2) or, alternatively, N is the smallest integer satisfying the same equation. This is so since replacing N by any of its integer multiples may define the same set $\mathcal{I}$ by extending $\mathcal{P}$ accordingly.

This sample selection described above can also be expressed in matrix form by writing $\boldsymbol{y} = \bar{\mathbf{\Phi}}\boldsymbol{x}$, where $\bar{\mathbf{\Phi}} \in \mathbb{C}^{ML \times NL}$ is a row selection matrix where each row contains a single one and $NL-1$ zeros. In other words, $\bar{\mathbf{\Phi}}$ is a submatrix of the identity $\boldsymbol{I}_{NL}$ including only the rows indexed by the set $\mathcal{I}$. Alternatively, due to the periodicity in $\mathcal{I}$, it is clear that $\bar{\mathbf{\Phi}}$ can be written as $\bar{\mathbf{\Phi}} = \boldsymbol{I}_L \otimes \mathbf{\Phi}$ where $\otimes$ denotes Kronecker product and $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is a submatrix of $\boldsymbol{I}_N$ where only the rows indexed by $\mathcal{P}$ are retained. It is also convenient to define ϕ as the linear mapping associated with $\mathcal{I}$ that takes the covariance matrix of $\boldsymbol{x}$ and returns the covariance matrix of $\boldsymbol{y}$, i.e., $\phi(\mathbf{\Sigma}) = \bar{\mathbf{\Sigma}} = \bar{\mathbf{\Phi}}\mathbf{\Sigma}\bar{\mathbf{\Phi}}^H$. This map transforms $\mathcal{S}$ into $\bar{\mathcal{S}}$ as $\bar{\mathbf{\Sigma}}_i = \phi(\mathbf{\Sigma}_i) = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_i\bar{\mathbf{\Phi}}^H$. Thus, we may also write $\bar{\mathcal{S}} = \phi(\mathcal{S})$.

The non-periodic version of non-uniform sampling can be viewed as a particular case of multi-coset sampling where only one period is considered ($L = 1$). Typically we consider periodic non-uniform sampling when the acquisition is done in the time domain, i.e., for power spectrum estimation, line spectrum estimation and wideband spectrum sensing, whereas we consider non-periodic non-uniform sampling when the acquisition is carried out in the space domain, i.e., for direction of arrival estimation. The reason is that acquisition in the time domain is typically accomplished by means of analog to information converters (AICs) [17], which work in a periodic fashion, whereas in the space domain we are free to place the antennas using any non-periodic pattern.

III. The Covariance Sampling Criterion

Although the covariance sampling criterion can be stated in different ways, the conclusions they result in are the same. Because it is quite intuitive, we have chosen an approach based on second-order identifiability, which is described next.

Assume that the second-order statistics of the sampled signal $x[n]$, arranged in the covariance matrix $\mathbf{\Sigma} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}$, can be written as a linear combination of the covariance matrices in $\mathcal{S} = \{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_{I-1}\}$ as

$$\mathbf{\Sigma} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\} = \sum_{i=0}^{I-1} \alpha_i \mathbf{\Sigma}_i, \qquad (3)$$

where $\alpha_i \in \mathbb{R}$. After multi-coset sampling, the statistical behavior of the resulting sequence in $y[m]$ can be characterized in terms of the matrix $\bar{\mathbf{\Sigma}}$ as:

$$\bar{\mathbf{\Sigma}} = \mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^H\right\} = \bar{\mathbf{\Phi}}\mathbf{\Sigma}\bar{\mathbf{\Phi}}^H = \sum_{i=0}^{I-1} \alpha_i \bar{\mathbf{\Sigma}}_i \qquad (4)$$

where $\bar{\mathbf{\Sigma}}_i = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_i\bar{\mathbf{\Phi}}^H$. Clearly, the coefficients α_i in (4) equal those in (3) so that estimating $\bar{\mathbf{\Sigma}}$ amounts to estimating $\mathbf{\Sigma}$, but it should be taken into account that $\bar{\mathbf{\Sigma}}$ is now a linear combination of the matrices in the transformed set $\bar{\mathcal{S}} = \{\bar{\mathbf{\Sigma}}_0, \bar{\mathbf{\Sigma}}_1, \ldots, \bar{\mathbf{\Sigma}}_{I-1}\}$ determined by $\mathcal{I}$. The covariance sampling criterion can be stated as follows:

Definition 1: A sampler defined by $\mathcal{I}$ satisfies the covariance sampling criterion if the associated transformation ϕ preserves the identifiability of the coefficients α_i.

In other words, if $\mathcal{S}$ is such that no two different sets of coefficients α_i can result in the same $\boldsymbol{\Sigma}$ in (3), then $\bar{\mathcal{S}} = \phi(\mathcal{S})$ must be such that no two different sets of coefficients α_i can result in the same $\bar{\boldsymbol{\Sigma}}$ in (4). This is just the classical definition of statistical identifiability [18] applied to the covariance sampling problem. The importance of this idea is in the fact that no parameter can be estimated consistently if it is not identifiable [18].

IV. Universal Covariance Samplers

The concept of a universal covariance sampler is the dual of a *universal sampler for perfect-reconstruction* of multi-band signals defined in [5], but in the context of the covariance sampling criterion rather than in compressive sampling. Specifically, it is clear that a particular sampler defined by a given $\mathcal{I}$ may satisfy the covariance sampling criterion for certain choices of $\mathcal{S}$ but not for all of them. In cases where knowledge of $\mathcal{S}$ is available at the moment of designing the acquisition system, a set $\mathcal{I}$ may be tailored for that specific application, potentially obtaining optimal designs. However, in other cases we may be interested in general designs being able to work under any choice of $\mathcal{S}$.

Definition 2: A universal covariance sampler is a sampler $\mathcal{I}$ satisfying the covariance sampling criterion for any choice of the set $\mathcal{S}$.

Note that, in view of the covariance sampling criterion, we can confine our attention to linearly independent sets $\mathcal{S}$. Thus, a universal covariance sampler can be defined as a sampler that preserves linear independence[1]. The subsequent sections aim to formalize this notion and to provide simpler conditions that may assist in the design of universal covariance samplers.

A. Design of Universal Covariance Samplers

The results in this section require the definition of $\mathbb{S}$, which is the span, with real (not necessarily non-negative) coefficients of all possible Toeplitz covariance matrices. Clearly $\mathbb{S}$ is an $\mathbb{R}$-subspace and, in particular, the smallest subspace containing the cone of all Toeplitz covariance matrices.

Lemma 1: Let $\mathcal{B}$ be a basis for $\mathbb{S}$. Then, a necessary and sufficient condition for a sampler to be a universal covariance sampler is that $\phi(\mathcal{B})$ is an independent set of matrices.

Proof: For any $\mathcal{S}$ satisfying that

$$\sum_i \alpha_i \boldsymbol{\Sigma}_i = \sum_i \beta_i \boldsymbol{\Sigma}_i \quad \Rightarrow \alpha_i = \beta_i \forall i,$$

the set $\bar{\mathcal{S}} = \phi(\mathcal{S})$, where ϕ represents a universal covariance sampler, must also satisfy

$$\sum_i \alpha'_i \bar{\boldsymbol{\Sigma}}_i = \sum_i \beta'_i \bar{\boldsymbol{\Sigma}}_i \quad \Rightarrow \alpha'_i = \beta'_i \forall i.$$

[1]Note that in some applications of Sec. I-B the coefficients α_i are constrained to be non-negative. However, it can be readily seen that this fact does not have any influence on the aforementioned condition.

In other words, the matrices in $\bar{\mathcal{S}}$ have to be linearly independent for any choice of $\mathcal{S}$. This means that the mapping ϕ has to be injective when defined over $\mathbb{S}$, and this is so if and only if the image of a basis is an independent set. ■

Lemma 1 simplifies the task of looking for a universal sampler that preserves identifiability of any set of covariance matrices $\mathcal{S}$ to that of verifying that the image of a basis of $\mathbb{S}$ is a set of independent matrices. This result could be easily generalized for architectures other than multi-coset ones. The next Theorem, however, exploits the peculiarities of multi-coset sampling to simplify the search even more. Before moving to state this result, the following definition is necessary:

Definition 3: Given a finite or countable set $\mathcal{A} = \{a_0, a_1, \ldots\}$, the difference set Δ of $\mathcal{A}$, denoted as $\Delta(\mathcal{A})$, is defined as

$$\Delta(\mathcal{A}) = \{d \geq 0 : \exists a_i, a_j \in \mathcal{A} \, \text{s.t.} \;\, d = a_i - a_j\}.$$

In other words, it is the set of all possible non-negative differences of the elements of $\mathcal{A}$. Note that every d in $\Delta(\mathcal{A})$ is present once, i.e., we do not consider repetition. Now we are ready to state the following theorem:

Theorem 1: A multi-coset covariance sampler defined by the set $\mathcal{I}$ is universal if and only if it contains all the relative time lags at least once, i.e., iff $\Delta(\mathcal{I}) = \{0, 1, \ldots, NL-1\}$.

Proof: Let us start by taking a basis for $\mathbb{S}$, recalling that it is an $\mathbb{R}$-subspace and not a $\mathbb{C}$-subspace. This basis is given by

$$\mathcal{B} = \{\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{NL-1}\} \cup \{\tilde{\boldsymbol{B}}_1, \ldots, \tilde{\boldsymbol{B}}_{NL-1}\}$$

where the matrices $\boldsymbol{B}_i$ and $\tilde{\boldsymbol{B}}_i$ have been defined in Sec. I-B. According to Lemma 1, we must ensure that $\phi(\mathcal{B})$ is an independent set of matrices. As explained in Sec. II, the sampling matrix $\bar{\boldsymbol{\Phi}}$ is a row selection operator, which means that $\phi(\boldsymbol{\Sigma}) = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma}\bar{\boldsymbol{\Phi}}$ selects the rows and columns with indices in $\mathcal{I}$ and discards the others. It is easy to see that deleting rows and columns from a matrix in $\mathcal{B}$ preserves linear independence provided that the resulting matrices have some non-null element. This is because of two reasons. First, if matrices $\boldsymbol{B}_i$ and $\boldsymbol{B}_j$, with $i \neq j$, have their non-null elements at different positions, so do $\bar{\boldsymbol{\Phi}}\boldsymbol{B}_i\bar{\boldsymbol{\Phi}}^H$ and $\bar{\boldsymbol{\Phi}}\boldsymbol{B}_j\bar{\boldsymbol{\Phi}}^H$. Second, even though the non-null positions of $\bar{\boldsymbol{\Phi}}\boldsymbol{B}_i\bar{\boldsymbol{\Phi}}^H$ and $\bar{\boldsymbol{\Phi}}\tilde{\boldsymbol{B}}_i\bar{\boldsymbol{\Phi}}^H$ are the same, the fact that the base field is $\mathbb{R}$ makes these two matrices independent.

The condition above states that the set $\mathcal{I}$ has to be such that no matrix in $\mathcal{B}$ has a null image, that is, $\bar{\boldsymbol{\Phi}}\boldsymbol{B}_i\bar{\boldsymbol{\Phi}}^H \neq \mathbf{0}$ and $\bar{\boldsymbol{\Phi}}\tilde{\boldsymbol{B}}_i\bar{\boldsymbol{\Phi}}^H \neq \mathbf{0}$ for all i. Since the positions of the non-null elements of $\boldsymbol{B}_i$ and $\tilde{\boldsymbol{B}}_i$ are the same we may disregard all $\tilde{\boldsymbol{B}}_i$ matrices. Moreover, since the matrices $\boldsymbol{B}_i$ have their non-null coefficients at symmetric positions with respect to the main diagonal and the row/column selection is also done symmetrically, we may concentrate on the lower triangular part of these matrices (including the main diagonal).

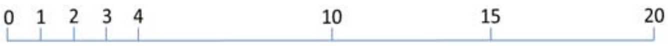

Fig. 1: Example of minimal sparse ruler of length 20.

If $\Gamma(\boldsymbol{B})$ denotes the set of indices of the non-null coefficients in the lower triangular part of matrix $\boldsymbol{B}$, we have that

$$\begin{aligned}\Gamma(\boldsymbol{B}_i) &= \Big\{(i,0),(i+1,1),\dots,(NL-1,NL-1-i)\Big\}\\ &= \Big\{(k,l):\quad 0\le k,l<NL,\quad k-l=i\Big\}\\ &= \Big\{(l+i,l):\quad 0\le l<NL-i\Big\}\end{aligned}$$

Now define $\Lambda(\mathcal{I})$ as the set of indices of the elements in the matrices in $\mathcal{B}$ that are selected (not deleted) using the index set $\mathcal{I}$, i.e.,

$$\Lambda(\mathcal{I}) = \Big\{(k,l):\quad k\in\mathcal{I}\ \text{ and }\ l\in\mathcal{I}\Big\} = \mathcal{I}\times\mathcal{I}.$$

Hence, for all $i=0,1,\dots,NL-1$ we have to ensure that $\Gamma(\boldsymbol{B}_i)\cap\Lambda(\mathcal{I})$ is non-empty, and this is achieved if and only if there exist $k,l\in\mathcal{I}$ such that $k-l=i$. ∎

Theorem 1 provides a simple means to verify whether a set $\mathcal{I}$ defines a universal multi-coset covariance sampler or not. However, in order to design a sampler $\mathcal{I}$ from scratch, the periodicity of this set must be included in the picture. Before we accomplish that task, let us devote a few words to the non-periodic case, i.e., the case where $L=1$, which results in the classical *sparse ruler problem.*

Definition 4: A (linear) sparse ruler of length $N-1$ is a set $\mathcal{I}\subset\{0,1,\dots,N-1\}$ such that $\Delta(\mathcal{I})=\{0,1,\dots,N-1\}$, and it is called minimal if no other sparse ruler of length $N-1$ exists with less elements.

Intuitively, we can say that a sparse ruler is a ruler with some marks erased but which is still able to measure all integer distances between 0 and its length. Fig. 1 shows an example of sparse ruler (actually a minimal one) when $N=21$.

Clearly, minimal sparse rulers exist for all values of N although they need not be unique. For example, a minimal sparse ruler of length 10 is given by the set $\{0,1,2,3,6,10\}$, but also by the set $\{0,1,2,5,7,10\}$. Unfortunately, there exists no quick procedure to find minimal sparse rulers: one must perform a brute-force search over the space of length-$(N-1)$ sparse rulers to find a minimal one. However, this procedure is carried out in the design stage so that this does not constitute a big problem. In view of Theorem 1 it is clear that when $L=1$ (i.e., in the non-periodic case) $\mathcal{I}$ has to be a length-$(N-1)$ sparse ruler in order to define a universal covariance sampler. Moreover, it will be optimal in the sense of minimizing the number of elements iff it is a minimal sparse ruler.

However, as we said before, the periodicity of $\mathcal{I}$ must be imposed whenever $L>1$. Therefore, one should look for a set which is not only a sparse ruler but it is also periodic in the sense of what was explained in Sec. II.

Definition 5: A periodic sparse ruler of length $NL-1$ and period N is a set $\mathcal{I}$ of indices between 0 and $NL-1$ that satisfies these two properties:

1) *if $i\in\mathcal{I}$, then $i+kN\in\mathcal{I}$ for all integers k such that $0\le i+kN<NL$*
2) $\Delta(\mathcal{I})=\{0,1,\dots,NL-1\}$

and it is called minimal if there does not exist any periodic sparse ruler with the same length and period but smaller cardinality.

Note that, because of the periodicity, we can find a set of indices $\mathcal{P}$ called period such that $\mathcal{I}=\{p+lN,\ \ p\in\mathcal{P},\ \ l=0,1,\dots L-1\}$. Clearly, if this set has M elements, then the cardinal of $\mathcal{I}$ is ML. Note also that it is possible to reformulate Theorem 1 to say that a set $\mathcal{I}$ defines a universal sampler iff it is a periodic sparse ruler.

The following result reduces the problem of finding periodic sparse rulers of length $NL-1$ to that of finding a (conventional) sparse ruler of length $N-1$.

Theorem 2: A periodic sparse ruler $\mathcal{I}$ of length $NL-1$ and period N is the result of concatenating L sparse rulers of length $N-1$, i.e., there exists a sparse ruler $\mathcal{P}$ of length $N-1$ such that

$$\mathcal{I}=\{p+lN:\quad p\in\mathcal{P},\quad l=0,1,\dots,L-1\}$$

Proof: Any element of $\Delta(\mathcal{I})$ can be written as $d+lN$, where $0\le d<N$ and $0\le l<L$. Clearly the value of d divides $\Delta(\mathcal{I})$ in N classes by gathering those elements with the same d and different l. All these classes must be present and all of them must have L elements in order for $\mathcal{I}$ to be a sparse ruler.

Since $0\le d<N$, it must be possible to obtain all d by considering two adjacent periods of the ruler. In particular, there are two kinds of distances, those of the form $i-j$ where $i\ge j$ and both i and j are in the first period; and those of the form $i-j$ where i is in the second period and j is in the first one. The distances of the second kind can also be written as $d=N+i-j$ by considering i and j in the first period, with $i<j$.

It is easy to verify that if a value of d can only be obtained as a distance of the second kind, then $d+LN\notin\Delta(\mathcal{I})$ so that that particular class will only have $L-1$ elements. On the other hand, if d can be obtained as a distance of the first kind, it is clear that $d+kN\in\Delta(\mathcal{I})$ for all $0\le k<L$. Thus, we can only rely on distances of the first kind, i.e., we need $\Delta(\mathcal{P})=\{0,1,\dots N-1\}$, where $\mathcal{P}=\mathcal{I}\cap\{0,1,\dots,N-1\}$ is the first period. In other words, the first period is a (conventional) sparse ruler of length $N-1$. ∎

Equivalently, the (fundamental) period of a periodic sparse ruler $\mathcal{I}$ is a sparse ruler itself. In other words, concatenating sparse rulers does not relax the requirements of each one. Note that this is not the case if the periodicity condition is released: concatenating two minimal sparse rulers of length $N-1$ and M elements gives a sparse ruler of length $2N-1$ and $2M$ elements, but there may exist sparse rulers of length $2N-1$ with less elements. For example, a minimal sparse ruler of

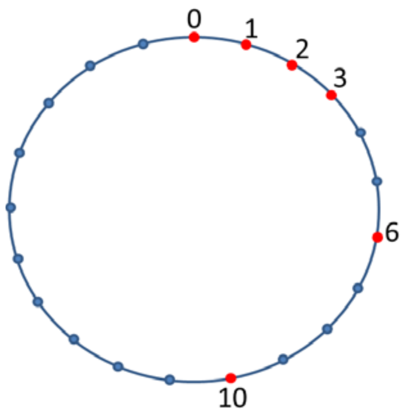

Fig. 2: Example of circular sparse ruler of length 20.

length 10 has 6 elements whereas the minimal sparse ruler of length 21 has $8 < 6 \times 2$ elements.

As a corollary of Theorem 2 we obtain that a *minimal* periodic sparse ruler of length $NL - 1$ and period N is the concatenation of *minimal* sparse rulers of length $N - 1$. This solves the problem of designing optimal universal multi-coset samplers by reducing it to finding a conventional minimal sparse ruler L times smaller.

V. Circular Sparse Rulers

Interestingly, the implications of Theorem 2 relax if we consider that the number of periods is arbitrarily high. This observation leads to a drastic reduction in the minimum M needed for obtaining universal samplers in certain situations like those where the acquisition never ends or when the covariance matrices in $\mathcal{S}$ are banded, as we will see in the next section.

A. Steady Acquisition

When the acquisition is performed over an infinite number of periods, i.e., when L approaches infinity, the requirements on every period of the periodic sparse ruler $\mathcal{I}$ relax. In particular, an infinite version of Theorem 1 would require that $\Delta(\mathcal{I}) = \{0, 1, 2, \ldots\}$, thus prompting us to introduce the concept of infinite periodic sparse ruler.

Definition 6: An infinite periodic sparse ruler of period N is a set[2] $\mathcal{J}$ *of indices satisfying these two properties:*

1) *if* $j \in \mathcal{J}$*, then* $j + lN \in \mathcal{J}$ *for all integers* l *such that* $j + lN \geq 0$
2) $\Delta(\mathcal{J}) = \{0, 1, 2 \ldots\}$

and it is called minimal if no other infinite periodic sparse ruler of period N exists whose period contains a lower number of elements.

We say that the set $\mathcal{P}$ generates an infinite periodic sparse ruler $\mathcal{J}$ of period N iff $\mathcal{J} = \{p + lN,\ \ p \in \mathcal{P},\ l \in \mathbb{N}_+\}$, where $\mathbb{N}_+$ denotes the set of non-negative integers. For instance, a generator set for an infinite periodic sparse ruler $\mathcal{J}$ is the first period $\mathcal{P} = \mathcal{J} \cap \{0, 1, \ldots, N-1\}$. In order to design this kind of rulers we shall first introduce a couple of definitions.

Definition 7: Given a finite or countable set $\mathcal{A} = \{a_0, a_1, \ldots\}$*, the N-modular difference set* Δ_N *of* $\mathcal{A}$*, denoted as* $\Delta_N(\mathcal{A})$*, is defined as*

$$\Delta_N(\mathcal{A}) = \{d \geq 0 : \exists a_i, a_j \in \mathcal{A}\, s.t.\ \ d = (a_i - a_j)_N\}$$

where $(x)_N$ *denotes the remainder of the integer division* $(x + lN)/N$*, where* l *is any integer such that* $x + lN \geq 0$*.*

In view of this definition it is clear that if d is in $\Delta_N(\mathcal{A})$, then also is $N - d$. Thus, it is clear that $\Delta(\mathcal{A}) \subset \Delta_N(\mathcal{A})$ and, consequently, that the number of elements in $\Delta_N(\mathcal{A})$ is, at least, the number of elements in $\Delta(\mathcal{A})$. For example, if $\mathcal{A} = \{0, 1, 5\}$ and $N = 10$, we have that $\Delta(\mathcal{A}) = \{0, 1, 4, 5\}$ whereas $\Delta_N(\mathcal{A}) = \{0, 1, 4, 5, 6, 9, 10\}$. The modular difference set naturally leads to the following definition:

Definition 8: A circular (or modular) sparse ruler of length $N - 1$ *is defined to be a set* $\mathcal{P} \subset \{0, 1, \ldots, N-1\}$ *such that* $\Delta_N(\mathcal{P}) = \{0, 1, \ldots, N-1\}$*, and it is called minimal if no other circular sparse ruler of length* $N - 1$ *exists with less elements.*

A circular sparse ruler can be thought of as a linear sparse ruler that wraps around. We bend the ruler until the two endpoints are at unit distance, thus making a circular ruler. In this way, given two points in the ruler, which divide the circle into two circular segments, we take into account the length of both segments. A circular sparse ruler of length 20 is depicted in Fig. 2.

Clearly, the number of elements of a minimal circular sparse ruler of length $N - 1$ is never greater than the number of elements of a minimal (linear) sparse ruler. This is so since any linear sparse ruler is also a circular sparse ruler, but the opposite needs not be true. Unfortunately like for linear sparse rulers, there is no standard means, to the best of the authors' knowledge, to design minimal circular sparse rulers: a brute-force search over the set of circular sparse rulers is, in principle, required.

The definition above allows us to state the following Theorem, which provides a means to design infinite periodic sparse rulers.

Theorem 3: A set $\mathcal{J}$ *is an infinite periodic sparse ruler* $\mathcal{J}$ *of period* N *if and only if every period is (up to an additive constant) a circular sparse ruler of length* $N - 1$*.*

Proof: Take $\mathcal{P}_0 = \mathcal{J} \cap \{0, 1, \ldots, N-1\}$ and $\mathcal{P}_1 = \mathcal{J} \cap \{N, N+1, \ldots, 2N-1\}$. Since $\Delta(\mathcal{J}) = \mathbb{N}_+$, for any integer d between 0 and $N-1$ we can find two elements $i, j \in \mathcal{J}$, $i \geq j$ such that $i - j = d$. Now take k such that $j' = j - kN \in \mathcal{P}_0$ and define $i' = i - kN$. If $i' \in \mathcal{P}_0$ we have that $d = i' - j' \in \Delta_N(\mathcal{P}_0)$. On the other hand, if $i' \in \mathcal{P}_1$, take $i'' = i' - N$, which belongs to $\mathcal{P}_0$. Then we have that $(i'' - j)_N \in \Delta_N(\mathcal{P}_0)$, but $(i'' - j)_N = (i' - N - j')_N = (d - N)_N = (d)_N = d$. Therefore $\Delta_N(\mathcal{P}_0)$ contains all integers between 0 and $N - 1$ so that $\mathcal{P}_0$ is a circular sparse ruler.

In order to show the converse theorem, take an integer $d \in \mathbb{N}_+$ and decompose it as $d = p + lN$ with $0 \leq p < N$ and l

[2] we keep the notation $\mathcal{I}$ for finite rulers whereas the infinite ones will be denoted as $\mathcal{J}$. This will be important in Sec. V-B.

an integer. If $\mathcal{P}_0$ is a circular sparse ruler, there exist $i, j \in \mathcal{P}_0$ such that $p = (i-j)_N$. If $i > j$ then $p = i - j$ so that taking $i' = i + lN$, which is also in $\mathcal{J}$ since it is periodic, together with j provides two points in $\mathcal{J}$ whose difference is d. On the contrary, if $i < j$, then $p = i - j + N$, so that considering $i' = i + (l-1)N$ together with j provides two points in $\mathcal{J}$ whose difference is d. Therefore $d \in \Delta(\mathcal{J})$ and $\mathcal{J}$ is an infinite sparse ruler. ■

An immediate consequence of this result is that the period of a minimal infinite periodic sparse ruler is a minimal circular sparse ruler. This enables us to optimally design this class of infinite rulers and, therefore, universal covariance samplers for steady acquisition.

B. Banded Covariance Matrices

A p-banded matrix is a matrix where all elements above the diagonal $+p$ and below the diagonal $-p$, these diagonals noninclusive, are zero. The fact that a covariance matrix $\boldsymbol{\Sigma}$ is p-banded means that the auto-correlation of $x[n]$ vanishes for lags greater than p. If all covariance matrices in $\mathcal{S}$ satisfy this property for p small enough, the problem of designing universal samplers boils down again to that of looking for minimal *circular* sparse rulers, thus leading to more efficient solutions. Note that in this case, the samplers are not fully universal since they are only universal for the class of sets containing p-banded matrices. For this reason we may prefer using the term quasi-universal instead.

Theorem 4: If the covariance matrices in the set $\mathcal{S}$ are p-banded with $N \leq p \leq N(L-1)$, then an N-periodic set $\mathcal{I}$ defines a quasi-universal sampler if and only if it generates an infinite periodic sparse ruler of period N.

Proof: The smallest subspace containing the p-banded covariance matrices is generated by the following basis:

$$\mathcal{B}' = \{\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_p\} \cup \{\tilde{\boldsymbol{B}}_1, \tilde{\boldsymbol{B}}_2 \ldots, \tilde{\boldsymbol{B}}_p\}$$

Thus, following similar arguments as those used to prove Theorem 1, it is possible to conclude that $\mathcal{I}$ defines a quasi-universal sampler iff $\Delta(\mathcal{I})$ contains all integers from 0 to p. In particular $\Delta(\mathcal{I})$ should include all distances from 0 to $N-1$. Since the set $\mathcal{J} = \{i + kN \geq 0,\ i \in \mathcal{I},\ k \in \mathbb{Z}\}$ is periodic and contains all integer distances, $\mathcal{J}$ is an infinite periodic sparse ruler generated by $\mathcal{I}$.

To prove the converse statement, we shall show that if $\mathcal{I}$ is periodic and generates an infinite periodic sparse ruler $\mathcal{J}$, then the difference set $\Delta(\mathcal{I})$ contains all distances from 0 to p or, in general, from 0 to $N(L-1)$. Clearly, all the distances in $\Delta(\mathcal{J})$ can be obtained as $d = i - j + lN$, where $i, j \in \mathcal{I}$ and $l \in \mathbb{N}_+$. According to this, we can classify them into two classes. The first one includes all distances that can be obtained by taking $i, j < N$. Since $i, j \in \mathcal{I}$, it is clear that $i - j + lN \in \Delta(\mathcal{I})$ for $l = 0, 1, \ldots L-1$. The second class comprises the distances not included in the first class. They can be obtained by taking $j < N$ and $N \leq i < 2N$ so that $\Delta(\mathcal{I})$ will also include the distances of the form $i - j + lN$ with $l = 0, 1, \ldots, L-2$ but not those with $l = L-1$. Thus, $\Delta(\mathcal{I})$ contains all distances between 0 and $NL-1$ except those of the form $i - j + N(L-1)$ with $j < N$ and $N \leq i < 2N$. Since in this case $i - j + N(L-1) > N(L-1)$, we have that $\Delta(\mathcal{I})$ contains all integers from 0 to $N(L-1)$ so that the sampler defined by $\mathcal{I}$ is quasi-universal. ■

This Theorem shows that the design of quasi-universal samplers for banded covariance matrices is in fact the same problem as that in Sec. V-A, i.e., to find a quasi-universal covariance sampler of length $NL-1$ amounts to finding a circular sparse ruler of length $N-1$. Observe that the condition that $p \leq N(L-1)$ is a weak assumption, since it only affects the last $N-1$ coefficients of the auto-correlation.

VI. Relation to Previous Work

All the theory in the previous sections helps us to better understand the design of universal covariance samplers, showing the ability of the covariance sampling criterion to unify the treatment of a wide variety of problems in signal processing.

In the context of power spectrum estimation, the conclusions of the present paper agree with those in [10] in the sense that a $\lfloor N/2 \rfloor$-length sparse ruler is universal for covariance sampling. This is so since they assume the covariance matrices to be banded and since any $\lfloor N/2 \rfloor$-length sparse ruler is also an $(N-1)$-length circular sparse ruler. However, this solution is not optimal since an $(N-1)$-length circular sparse ruler may have less elements than a $\lfloor N/2 \rfloor$-length linear minimal sparse ruler.

In the context of DoA estimation, the authors of [8] (see also references there) have proposed the idea of coprime sampling, where a sampling set $\mathcal{C}$ is constructed as

$$\begin{aligned}\mathcal{C} = \{&Qn,\ n = 0, 1, \ldots, P-1\} \\ &\cup \{Pm,\ m = 0, 1, \ldots, 2Q-2\},\end{aligned}$$

with P and Q two coprime integers satisfying $Q < P$; and they show that $\{0, 1, \ldots, QP\} \subset \Delta(\mathcal{C})$. Observe that this scheme spans from $\inf \mathcal{C} = 0$ to $\sup \mathcal{C} = P(2Q-2)$, achieving $\sup \Delta(\mathcal{C}) = QP$ different consecutive lags with $P + 2Q - 1$ elements. However, using a $P(2Q-2)$-length sparse ruler it is possible to attain $\sup \Delta(\mathcal{C}) = P(2Q-2)$ lags. To see the reduction in the number of elements (antennas, in this case) consider the example in [8, Sec. 5], where $Q = 5$ and $P = 7$: in that case they take $P + 2Q - 1 = 16$ elements and obtain consecutive lags up to $QP = 35$. However, with a $P(2Q-2) = 56$-length sparse ruler it is possible to achieve lags up to 56 with only 13 elements, which is the number of elements of the 56-length minimal sparse ruler. A similar remark can be done about the nested arrays of [7], where the cardinality of $\Delta(\mathcal{I})$ is known in advance, but present the disadvantage that $\Delta(\mathcal{I})$ is not a filled set, in the sense that not all integers between the minimum and the maximum are present. Both these schemes, along with minimum redundancy arrays [9], are not, in general, universal covariance samplers. The interest of these schemes is that in DoA estimation we are typically interested in resolving a large number of sources, but sometimes we do not care about the length of the array. On the contrary, the covariance sampling criterion leads to arrays that resolve the maximum number of sources for a given array length[3]. On the other hand, the scheme in [11] actually defines a universal covariance sampler.

VII. Conclusions

We have unified the treatment of several problems included in what we called compressive covariance sampling under the same framework, which is based on the covariance sampling criterion. An intuitive theory was seen to arise from that idea, resulting in simple design rules that eventually lead to the minimal sparse ruler problem. Under mild conditions, it was also seen that the problem simplifies and a novel mathematical object, namely a circular sparse ruler, was proposed to accomplish the design. Future work will be pointed to analyze the properties of this structure.

VIII. Acknowledgements

This work was supported by the the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM), COSIMA (TEC2010-19545-C04-03) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), FPU grant AP2010-0149, and by the Galician Regional Government under projects "Consolidation of Research Units" 2009/62 and 2010/85.

References

[1] J.P. Burg, D.G. Luenberger, and D.L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963 – 974, sept. 1982.

[2] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[3] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[4] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-nyquist nonuniform sampling of multiband signals," *Information Theory, IEEE Transactions on*, vol. 46, no. 6, pp. 2173 –2183, sep 2000.

[5] M. Mishali and Y.C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 993 –1009, march 2009.

[6] M.E. Dominguez-Jimenez, N. Gonzalez-Prelcic, G. Vazquez-Vilar, and R. Lopez-Valcarce, "Design of universal multicoset sampling patterns for compressed sensing of multiband sparse signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 3337 –3340.

[7] P. Pal and P.P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4167 –4181, aug. 2010.

[8] P. Pal and P.P. Vaidyanathan, "Coprime sampling and the music algorithm," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, jan. 2011, pp. 289 –294.

[9] A. Moffet, "Minimum-redundancy linear arrays," *Antennas and Propagation, IEEE Transactions on*, vol. 16, no. 2, pp. 172 – 175, mar 1968.

[10] D.D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4775–4789, 2012.

[3] Note that we are assuming uniform linear arrays (ULAs).

[11] S. Shakeri, D.D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, june 2012, pp. 525 –529.

[12] R.T. Hoctor and S.A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 735 –752, apr 1990.

[13] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *Submitted to IEEE Transactions on Signal Processing*.

[14] G. Vazquez-Vilar and R. Lopez-Valcarce, "Spectrum sensing exploiting guard bands and weak channels," *Signal Processing, IEEE Transactions on*, , no. 99, pp. 1–1, 2011.

[15] Q. Zhao and B.M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, 2007.

[16] J. Leech, "On the representation of 1, 2,. . . , n by differences," *Journal of the London Mathematical Society*, vol. 1, no. 2, pp. 160–169, 1956.

[17] J.N. Laska, S. Kirolos, M.F. Duarte, T.S. Ragheb, R.G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, may 2007, pp. 1959 –1962.

[18] E.L. Lehmann and G. Casella, *Theory of point estimation*, vol. 31, Springer, 1998.