Compressive Covariance Sensing

– Structure-Based Compressive Sensing Beyond Sparsity –

Daniel Romero, Dyonisius Dony Ariananda, Zhi Tian and Geert Leus

Compressed sensing deals with the reconstruction of signals from sub-Nyquist samples by exploiting the sparsity of their projections onto known subspaces. In contrast, the present article is concerned with the reconstruction of second-order statistics, such as covariance and power spectrum, even in the absence of sparsity priors. The framework described here leverages the statistical structure of random processes to enable signal compression and offers an alternative perspective at sparsity-agnostic inference. Capitalizing on parsimonious representations, we illustrate how compression and reconstruction tasks can be addressed in popular applications such as power spectrum estimation, incoherent imaging, direction of arrival estimation, frequency estimation, and wideband spectrum sensing.

I. Introduction

The incessantly growing size of sensing problems has spurred an increasing interest in simultaneous data acquisition and compression techniques that limit sensing, storage, and communication costs. Notable examples include compressed sensing [1], support recovery [2], sub-Nyquist sampling of multiband or multitone signals [3]–[5], and array design for aperture synthesis imaging [6]–[8]. The overarching paradigm of sub-Nyquist sampling can impact a broad swath of resource-constrained applications arising in data sciences, broadband communications, large-scale sensor networks, bioinformatics, and medical imaging, to name a few.

The aforementioned techniques rely on *parsimonious* models that capture relevant information and enable compression. In compressed sensing, for example, signals can be reconstructed from sub-Nyquist samples provided that they admit a sparse representation in a known transformed domain. Whereas this form of structure arises naturally in many applications, it is often the case that either the underlying signal is not sparse or the sparsifying transformation is difficult to model or manipulate. Those scenarios call for alternative approaches to effect compression by capturing other forms of structure.

A prominent example is the family of methods exploiting structural information in the *statistical domain*, which includes those intended to reconstruct the second-order statistics of wide-sense stationary signals, such as power, autocorrelation, or power spectral density. It is widely accepted that statistics of this class play a central role in a multitude of applications comprising audio and voice processing, communications, passive sonar, passive radar, radioastronomy, and seismology, to name a few [9]. Although reconstruction of second-order statistics from compressed observations dates back several decades (see e.g. [6] and references therein), the recent interest in compressive sensing and reconstruction has propelled numerous advances in this context.

It is the purpose of the present article to provide a refreshing look at the recent contributions in this vibrating area, which will be globally referred to as compressive covariance sensing (CCS). Admittedly, a straightforward approach to reconstruct second-order statistics is to apply an estimation method over the uncompressed waveform obtained via a non-CCS procedure. However, it is not difficult to see that this two-step approach incurs large computational complexity and heavily limits the compression ratio. CCS methods, on the other hand, proceed in a single step by directly recovering relevant second-order statistics from the compressed samples, thus allowing a more efficient exploitation of the statistical structure.

After an illustrative example providing intuition and illustrating the basic notions that underlie CCS, we delineate how the degrees of freedom of the statistical model reflect in fundamental compression limits and how efficient compression structures can be designed to attain these bounds. Focusing on the applications, we also illustrate how major savings in sampling costs can be exerted upon judiciously exploiting the statistical structure via CCS. We next provide a tutorial introduction to inference in CCS, along with advanced techniques for cyclic feature detection, cooperative distributed sensing in wireless sensor networks and power spectrum estimation for multiband signals. The article concludes with an outlook on new opportunities for efficient handling of data-intensive sensing applications in which inference from random processes is of foremost importance.

II. "Sampling" Second-Order Statistics

To introduce the basic notions of CCS, consider the problem of measuring the fine variations of a spatial field, for instance to achieve a high angular resolution in source localization. Since the large sensor arrays required in the absence of compression incur prohibitive hardware costs, many acquisition schemes have been devised to reduce the number of sensors without sacrificing resolution.

A. A Warm-up Example

Suppose that a uniform linear array (ULA) with L antennas, such as the one in Fig. 1a, observes T snapshots of a zero-

Supported by the Spanish Government and the European Regional Development Fund (ERDF) (project TEC2013-47020- C2-1-R COMPASS) and by the Galician Regional Government and ERDF (projects GRC2013/009, R2014/037 and AtlantTIC).

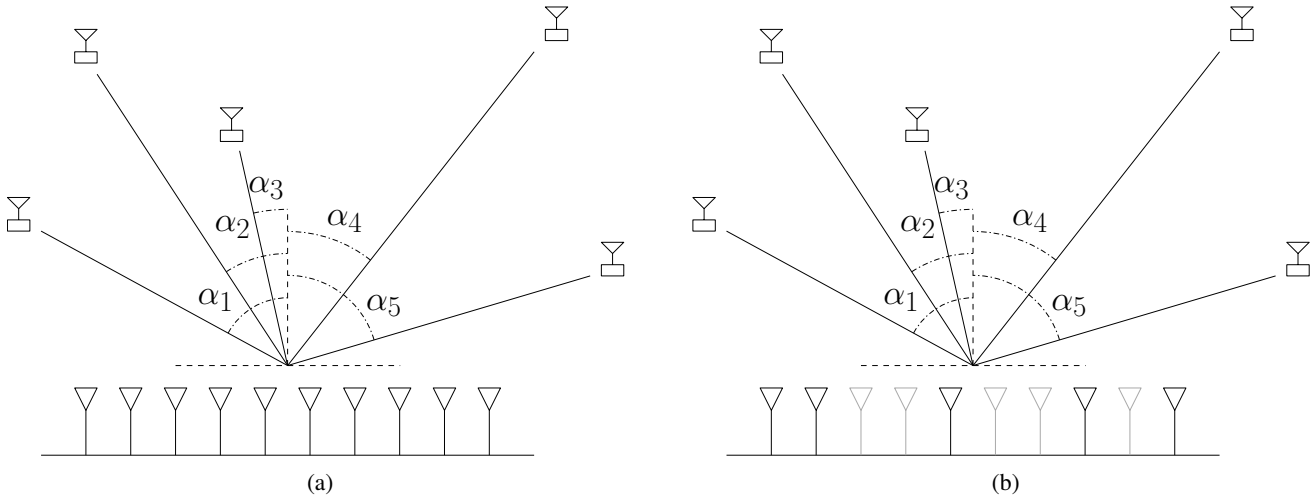


Fig. 1: (a) *Uncompressed* uniform linear array with 10 antennas receiving the signals from five sources in the far field. (b) *Compressed* array with 5 antennas. The five antennas marked in gray were removed, but the achievable spatial resolution remains the same.

mean spatial signal whose complex baseband representation is given by $\boldsymbol{x}_\tau \in \mathbb{C}^L$, $\tau = 0, 1, ...T-1$. Many array processing algorithms rely on estimates of the so-called spatial covariance matrix $\boldsymbol{\Sigma_x} := \mathrm{E}\left\{\boldsymbol{x}_\tau \boldsymbol{x}_\tau^H\right\}$ to form images or to obtain information such as the bearing of certain sources [6], [9]. A straightforward estimate of $\boldsymbol{\Sigma_x}$ is the sample covariance matrix, given by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{1}{T}\sum_{\tau=0}^{T-1} \boldsymbol{x}_\tau \boldsymbol{x}_\tau^H. \tag{1}$$

If the impinging signals are generated by uncorrelated point sources in the far field (see Fig. 1a), the matrix $\boldsymbol{\Sigma_x}$ exhibits a Toeplitz structure (see Sec. VI), meaning that its coefficients are constant along the diagonals. Thus, one may represent the (m,n)-th entry of $\boldsymbol{\Sigma_x}$ by

$$\sigma[m-n] = \mathrm{E}\left\{x_\tau[m] x_\tau^*[n]\right\}, \tag{2}$$

where $x_\tau[m]$ represents the m-th entry of $\boldsymbol{x}_\tau$. Noting that $\boldsymbol{\Sigma_x}$ is also Hermitian reveals that all its information is contained in the coefficients $\sigma[l]$, $l = 0, \ldots, L-1$. These observations suggest the possibility of constructing estimators with improved performance [10], [11]: simply consider replacing the elements on each diagonal of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ with their arithmetic mean. This operation renders a more satisfactory estimate than the sample covariance matrix in (1) because it utilizes the underlying Toeplitz structure.

Let us now adopt a different standpoint. Instead of attempting an improvement in the estimation performance, the structure described above can also be exploited to reduce the number of antennas required to estimate $\boldsymbol{\Sigma_x}$ (see e.g. [6]–[8]). Suppose, in particular, that only a subset of the antennas in the ULA are used to sample the spatial field of interest, the others being disconnected (see Fig. 1b).

Let the set $\mathcal{K} := \{k_0, \ldots, k_{K-1}\}$ collect the indices of the K active antennas. The vector signal received by this subarray, which can be thought of as a *compressed* observation, is given by $\boldsymbol{y}_\tau = [x_\tau[k_0], \ldots, x_\tau[k_{K-1}]]^T$. The (i,j)-th entry[1] of the covariance matrix $\boldsymbol{\Sigma_y} := \mathrm{E}\left\{\boldsymbol{y}_\tau \boldsymbol{y}_\tau^H\right\}$ is therefore

$$\mathrm{E}\left\{y_\tau[i] y_\tau^*[j]\right\} = \mathrm{E}\left\{x_\tau[k_i] x_\tau^*[k_j]\right\} = \sigma[k_i - k_j]. \tag{3}$$

Thus, $\boldsymbol{\Sigma_y}$ is made up of a subset of the entries of $\boldsymbol{\Sigma_x}$. It is clear therefore that $\boldsymbol{\Sigma_x}$ can be *reconstructed* from a sample estimate of $\boldsymbol{\Sigma_y}$ if all the entries of the former show up at least once in the latter. From (3), this means that, for every $l = 0, \ldots, L-1$, there must exist at least one pair of elements k, k' in $\mathcal{K}$ satisfying $k - k' = l$. Sets $\mathcal{K}$ of this nature are called *sparse rulers*, and such sets with the minimum number of elements are termed *minimal sparse rulers*, as explained in Box 1. In Fig. 1b for example, only the antennas at positions $\mathcal{K} = \{0, 1, 4, 7, 9\}$ are operative, but the array can reconstruct the same spatial covariance matrix as the array in Fig. 1a.

Mathematically, the problem of constructing sparse rulers is interesting on its own and has been extensively analyzed (see [12] and references therein). Since finding *minimal* sparse rulers is a combinatorial problem with no closed-form solution, devising structured yet sub-optimal designs has received great attention (see e.g. [12], [13]).

An intimately related concept is that of *minimum redundancy array* [7], [9], well known within the array processing community. A minimum redundancy array is a minimal linear sparse ruler whose length is maximum given its number of marks. For example, $\mathcal{K}_1 = \{0, 1, 2, 3, 7\}$, $\mathcal{K}_2 = \{0, 1, 2, 5, 8\}$ and $\mathcal{K}_3 = \{0, 1, 2, 6, 9\}$ are minimal sparse rulers of length 7, 8 and 9, respectively. However, $\mathcal{K}_1$ and $\mathcal{K}_2$ are not minimum redundancy arrays, since a minimal sparse ruler of greater length can be found with the same number of marks, an example being $\mathcal{K}_3$.

Deploying a smaller number of antennas effects cost savings beyond that associated with the antennas themselves: radio frequency (RF) equipment, such as filters, mixers, and ADCs, needs solely to be deployed for the active antennas. Moreover,

[1] We adopt the convention that the first row/column of any vector/matrix is associated with the index 0, the second with the index 1, and so on.

Box 1: **Linear Sparse Rulers**

A set $\mathcal{K} \subset \{0, \dots, L-1\}$ is a lengh-$(L-1)$ (linear) sparse ruler if for every $l = 0, \dots, L-1$, there exists at least one pair of elements k, k' in $\mathcal{K}$ satisfying $k - k' = l$. Two examples of length 10 are $\mathcal{K} = \{0, 1, 2, 5, 7, 10\}$ and $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$.

The name *sparse ruler* stems from the geometric interpretation of $\mathcal{K}$ as a *physical* ruler where all but the marks with indices in $\mathcal{K}$ have been erased. Despite lacking part of the marks, a sparse ruler is still able to measure any integer distance between 0 and $L-1$. In the example of Fig. 2, we observe that the ruler $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is capable of measuring any object of length 5 by using the marks 3 and 8.

A length-$(L-1)$ *minimal* sparse ruler is a length-$(L-1)$ sparse ruler $\mathcal{K}$ with minimum number of elements $|\mathcal{K}|$. The set $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is a length-10 minimal linear sparse ruler since it has 6 elements and there exists no length-10 sparse ruler with 5 or fewer elements.

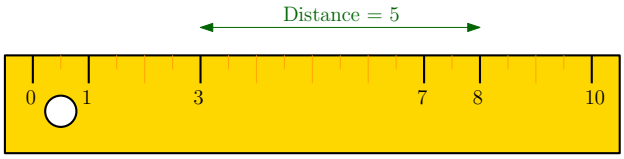


Fig. 2: A sparse ruler can be thought of as a ruler where a part of its marks have been erased, but the remaining marks allow to measure all integer distances between 0 and its length.

the fact that the endpoints 0 and $L-1$ are always in $\mathcal{K}$, for any length-$(L-1)$ linear sparse ruler $\mathcal{K}$, means that the aperture of the sub-array equals the aperture of the uncompressed array. Therefore, this antenna reduction comes at no cost in angular resolution. The price to be paid is, however, slower convergence of the estimates: generally, the smaller $|\mathcal{K}|$, the larger the amount of averaging required to attain a target performance. Hence, when sampling signals defined on the spatial domain, this kind of compression is convenient when hardware savings make up for an increase in the acquisition time, as is usually the case in array processing.

B. Importance of Covariance Structures

In the previous example, the Hermitian Toeplitz structure of $\mathbf{\Sigma}_{\boldsymbol{x}}$ allowed us to recover the second-order statistics of $\boldsymbol{x}_\tau$ from those of its compressed version $\boldsymbol{y}_\tau$. More generally, it is expected that our ability to compress a signal while preserving the second-order statistical information depends on the structure of $\mathbf{\Sigma}_{\boldsymbol{x}}$. In other words, we expect that the more structured $\mathbf{\Sigma}_{\boldsymbol{x}}$ is, the stronger compression on $\boldsymbol{x}_\tau$ it may induce.

In certain applications such as power spectrum estimation of communication signals, the covariance matrix is known to be circulant [14]–[17]. Recall that a circulant matrix is a special type of Toeplitz matrix where each row is the result of applying a circular shift to the previous one. For this reason, it can be seen that $\sigma[l] = \sigma[l - L]$. This increased structure relaxes the

Box 2: **Circular Sparse Rulers**

A set $\mathcal{K} \subset \{0, \dots, L-1\}$ is a lengh-$(L-1)$ *circular sparse ruler* if for every $l = 0, \dots, L-1$, there exists at least one pair of elements $k, k' \in \mathcal{K}$ satisfying $(k - k') \mod L = l$. An example of a length-15 circular sparse ruler is $\mathcal{K} = \{0, 1, 4, 6, 8\}$. It can be seen that any length-l *linear* sparse ruler, with $L/2 \leq l \leq L-1$, is also an example of a length-L *circular* sparse ruler.

A circular sparse ruler can be thought of as the result of wrapping around a linear ruler. This operation allows us to measure two different distances using each pair of marks (see Fig. 3).

A length-$(L-1)$ circular sparse ruler is *minimal* if there exists no length-$(L-1)$ circular sparse ruler with fewer elements.

requirements on $\mathcal{K}$, which is no longer required to be a *linear* sparse ruler but a *circular* one. See Box 2 for a definition.

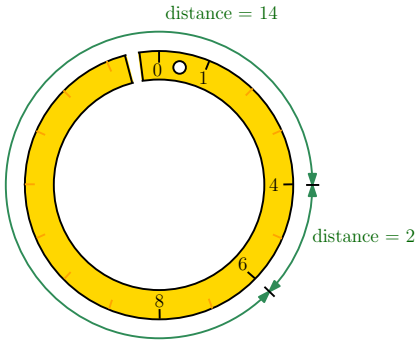
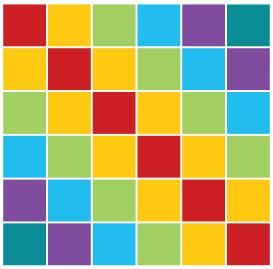


Fig. 3: Generally, in a circular sparse ruler, each pair of marks allows to measure two distances.

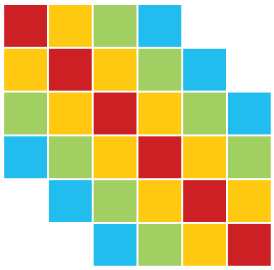
Due to their ability to measure two different distances using each pair of marks, circular sparse rulers lead to a greater compression than their linear counterparts. In other words, $\mathcal{K}$ needs fewer elements in order to be a length-$(L-1)$ *circular* sparse ruler than in order to be a length-$(L-1)$ *linear* sparse ruler.

Circular sparse rulers can be designed in several ways. For certain values of L, minimal rulers can be obtained in closed form [18]. Other cases may require exhaustive search, which motivates sub-optimal designs. Immediate choices are length-$(L-1)$ or length-$\lfloor \frac{L}{2} \rfloor$ minimal *linear* sparse rulers [19]. In fact, the latter provides optimal solutions for most values of L below 60 [20].

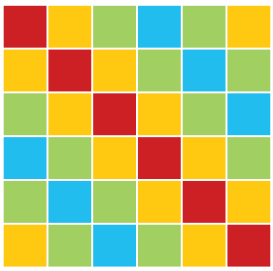
Another common structure besides Toeplitz and circulant is that present in those applications where the covariance matrix is known to be banded [19]. d-Banded matrices are Toeplitz matrices satisfying $\sigma[l] = 0$ for all $l > d$, and arise in those cases where we sample a wide-sense stationary (WSS) time signal whose autocorrelation sequence $\sigma[l]$ vanishes after d lags. Sampling patterns for banded matrices are discussed in [20], which suggests that the achievable compression is dependent on the parameter d. These designs also hold for certain situations where we are only interested in the first d correlation lags [21].



Toeplitz
$S = 2L - 1$ real unknowns
WSS processes
DOA estimation
Incoherent imaging



d-banded
$S = 2L - 1$ real unknowns
WSS with d limited lags
MA(d) time series



circular
$S = L$ real unknowns
OFDM signals
Multiband signals
Incoherent imaging

Fig. 4: Some common covariance structures, along with their main applications.

These typical covariance structures, including Toeplitz, Circulant and banded, are illustrated in Fig. 4 along with their most popular applications. Generally speaking, in many cases besides the above, prior knowledge constrains covariance matrices to be linear combinations of certain known matrices, say $\{\mathbf{\Sigma}_i\}_i$. In other words, there must exist coefficients α_i such that

$$\mathbf{\Sigma}_{\boldsymbol{x}} = \sum_{i=0}^{S-1} \alpha_i \mathbf{\Sigma}_i. \tag{4}$$

Without any loss of generality, we may assume that the scalars α_i are real [20] and the matrices $\mathbf{\Sigma}_i$ linearly independent. Thus, $\mathcal{S} = \{\mathbf{\Sigma}_0, \ldots, \mathbf{\Sigma}_{S-1}\}$ is a basis and S represents the dimension of the model. As mentioned above, this expansion encompasses all the previous examples as particular cases — suffice it to choose the right set of matrices $\mathbf{\Sigma}_i$. It can be seen that $S = 2L - 1$ for Toeplitz matrices, $S = L$ for circulant matrices, and $S = 2d - 1$ for d-banded matrices (see Fig. 4). The problem of estimating the coefficients α_i is known as *structured covariance estimation* or *covariance matching* [10], [22] and has a strong connection with CCS. Nonetheless, this line of works flourished prior to the surge of compressed sensing in signal processing, when the main goal was to design robust and performance-enhanced estimators given a small sample size. CCS offers a new look of exploiting covariance structures for joint signal acquisition and compression. It seems natural to conjure that the dimension S of the parsimonious model in (4) is related to how compressible $\mathbf{\Sigma}_{\boldsymbol{x}}$ is (see Sec. III-A below).

III. Compression

The array processing example presented above describes how compression can be accomplished for signals acquired in the spatial domain — only a subset $\mathcal{K}$ of antennas was used to estimate $\mathbf{\Sigma}_{\boldsymbol{x}}$. The remaining antennas can be disconnected or, more simply, they need not be deployed. Broadly, acquisition hardware represents the bottleneck of many current signal processing systems, whose designs aim at meeting an ever increasing demand for processing rapidly-changing signals. In practice, Nyquist acquisition of wideband signals becomes prohibitive in many circumstances since the sampling rate drastically affects power consumption and hardware complexity. The ambition to break this *bandwidth barrier* has prompted a growing interest in innovative acquisition hardware architectures that replace traditional equipment, such as the slow and power-hungry analog-to-digital converters (ADCs). In this section, we delve into compression methods that can be applied not only for compressive acquisition of spatial signals, but also for time signals and more general classes of signals.

In particular, suppose that we are interested in estimating the second-order statistics of $x(t)$, indexed by the continuous time index t. A traditional ADC ideally produces the sequence

$$x[l] = x(lT_s), \quad l = 0, \ldots, L-1, \tag{5}$$

where $1/T_s$ is the sampling rate, a number that must exceed the Nyquist rate of $x(t)$ to avoid *aliasing*. Unfortunately, power consumption, amplitude resolution, and other parameters dictated by the application establish stringent upper bounds on the values that the sampling rate can take on. These limitations conflict with the constantly increasing need for larger bandwidths and hence higher Nyquist rates.

One can think that a compression approach similar to the one described for the spatial domain may potentially alleviate these limitations by reducing the *average* sampling rate. Generally known as *non-uniform sampling*, this approach advocates the acquisition of a small number of samples indexed by a subset of the *Nyquist grid*:

$$y[i] = x(k_i T_s), \quad \mathcal{K} = \{k_0, \ldots, k_{K-1}\}. \tag{6}$$

As we will soon see, this average rate reduction has lead to the technology of *compressive ADCs* (C-ADCs), conceived to circumvent the aforementioned hardware trade-offs. Before delving into this topic, let us expand a little bit on the families of samplers we are about to consider.

By forming $\boldsymbol{x} = [x[0], \ldots, x[L-1]]^T$ and $\boldsymbol{y} = [y[0], \ldots, y[K-1]]^T$, the operation in (6) can be equivalently represented as a row-selection operation:

$$\boldsymbol{y} = \bar{\mathbf{\Phi}} \boldsymbol{x}. \tag{7}$$

The matrix $\bar{\mathbf{\Phi}} \in \mathbb{C}^{K \times L}$, which contains ones at the positions (i, k_i) and zeros elsewhere, is therefore a *sparse* matrix with

at most one nonzero entry at each row or column. Rather than restricting ourselves to matrices of this form, there are certain applications where the usage of *dense* compression matrices has proven to be successful, both in the time domain (see e.g. [4], [5]) and in the spatial domain (see e.g. [23]). In correspondence with this terminology, we will talk about *dense samplers* when $\bar{\mathbf{\Phi}}$ is dense, and about *sparse samplers* when $\bar{\mathbf{\Phi}}$ is sparse.

As opposed to most applications in array processing, it is common in time-domain applications to observe just a single realization of the signal of interest, i.e. , $T = 1$. This is the reason why we dropped the subscript τ from $\boldsymbol{x}$ and $\boldsymbol{y}$ in (7) as compared to $\boldsymbol{x}_\tau$ and $\boldsymbol{y}_\tau$ in the previous section. For simplicity, we will omit this subscript throughout when possible, keeping in mind that several realizations may be available.

When observation windows for time signals are long, hardware design considerations make it convenient to split a sampling pattern into shorter pieces which are repeated periodically. This amounts to grouping data samples in blocks that are acquired using the same pattern. Likewise, the usage of periodic arrays in the spatial domain may also present advantages [16]. In these cases, the uncompressed observations $\boldsymbol{x}$ are divided into B blocks of size $N = L/B$ as $\boldsymbol{x} = [\boldsymbol{x}^T[0], \ldots, \boldsymbol{x}^T[B-1]]^T$, and each block is compressed individually to produce an output block of size M:

$$\boldsymbol{y}[b] = \mathbf{\Phi}\boldsymbol{x}[b]. \tag{8}$$

Thus, it is clear that one can assemble the vector of compressed observations as $\boldsymbol{y} = [\boldsymbol{y}^T[0], \ldots, \boldsymbol{y}^T[B-1]]^T$, and the matrix $\bar{\mathbf{\Phi}}$ from (7) as $\bar{\mathbf{\Phi}} = \boldsymbol{I}_B \otimes \mathbf{\Phi}$, where $\otimes$ represents the Kronecker product.

In the case of sparse samplers, the above block-by-block operation means that the pattern $\mathcal{K}$ can be written as

$$\mathcal{K} = \{m + bN \ : \ m \in \mathcal{M}, \ b = 0, \ldots, B-1\}, \tag{9}$$

where $\mathcal{M} \subset \{0, \ldots, N-1\}$ is the sampling pattern used at each block. For example, $\mathcal{M}$ can be a length-$(N-1)$ linear sparse ruler. Thus, $\mathcal{M}$ can be thought of as the *period* of $\mathcal{K}$, or we may alternatively say that $\mathcal{K}$ is the result of a B-fold *concatenation* of $\mathcal{M}$. In sparse sampling schemes of this form, known in the literature as multi-coset samplers [3], the matrix $\mathbf{\Phi}$ is the result of selecting the rows of $\boldsymbol{I}_N$ indexed by $\mathcal{M}$.

A. Optimal Designs

One critical problem in CCS is to design a sampler $\bar{\mathbf{\Phi}}$ that *preserves* the second-order statistical information by allowing reconstruction of the uncompressed covariance matrix from the compressed observations. This boils down to the identifiability issue for statistical analysis.

Design techniques for sparse and dense samplers hinge on different basic principles. Whereas sparse samplers are designed based on discrete mathematics considerations (as explained earlier), existing designs for dense samplers rely on probabilistic arguments. Inspired by compressed sensing techniques, these designs generate sampling matrices at random and provide probabilistic guarantees on their admissibility.

Optimal rates for dense samplers are known in closed form for linear covariance parameterizations such as Toeplitz, circulant and banded [20]. On the other hand, their evaluation for sparse samplers requires solving combinatorial problems such as the minimal sparse ruler problem. Table I summarizes the optimum designs, along with the maximum compression ratios, for the aforementioned parameterizations [20]. The compression ratio is defined as

$$\eta := \frac{|\mathcal{K}|}{L} = \frac{|\mathcal{M}|}{N} \tag{10}$$

and satisfies $0 \leq \eta \leq 1$. Note that the stronger the compression, the smaller η. It can also be interpreted as the reduction in the *average* sampling rate: if $x[l]$ represents the sample sequence acquired at the Nyquist rate $1/T_s$, for instance, then the compressed sequence $y[k]$ corresponds to an average sampling rate of η/T_s.

With Fig. 5, the merits of CCS are readily illustrated through a popular application of compressive wideband spectrum sensing [24]–[26], where the spectrum occupancy over a very wide band is decided via power spectrum estimation. It is observed that the permissible sampling rate can be reduced considerably, even when the sparsity structure is not present. For instance, even for a moderate block length of $N = 50$, the minimum sampling rate is less than one fourth of the Nyquist rate in all cases. These cases include those deterministically designed sparse samplers that are amendable to practical hardware implementation. Asymptotically for increasing N, Table I alludes that the saving in sampling rates exhibits at a rate proportional to $1/\sqrt{N}$.

Furthermore, the superior efficiency of dense samplers also manifests itself in Fig. 5. In fact, it can be shown that certain random designs for dense samplers achieve optimum compression ratios with probability one, in the sense that no other sampler (either dense or sparse) can achieve a lower ratio.

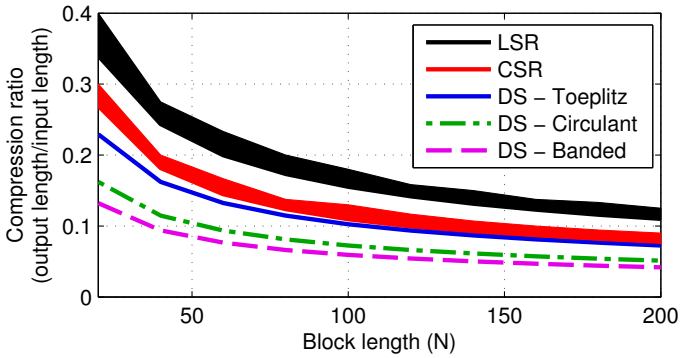


Fig. 5: Optimum compression ratios when $B = 10$ and $d = BN/3$ using dense samplers (DS) for Toeplitz, Circulant and d-banded matrices, and using a linear sparse ruler (LSR) and a circular sparse ruler (CSR). Moderate block lengths are seen to yield strong compression.

B. Technologies

The acquisition systems that can be used to implement the above sampling schemes are essentially the same as those used

	Sparse Samplers		Dense Samplers	
	Design	Ratio	Design	Ratio
Toeplitz	$\mathcal{K}$ is a length-$(N-1)$ *linear* sparse ruler	$\frac{1}{N}\sqrt{2.435(N-1)} \leq \eta_{\min} \leq \frac{1}{N}\left\lceil \sqrt{3(N-1)} \right\rceil$	$\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is drawn from a continuous distribution with $M \geq \sqrt{\frac{2NB-1}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{2NB-1}{(2B-1)N^2}}$
Circulant	$\mathcal{K}$ is a length-$(N-1)$ *circular* sparse ruler	$\frac{2+\sqrt{4N-3}}{2N} \leq \eta_{\min} \leq \frac{1}{N}\left\lceil \sqrt{3\left\lfloor \frac{N}{2} \right\rfloor} \right\rceil$	$\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is drawn from a continuous distribution with $M \geq \sqrt{\frac{NB}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{B}{(2B-1)N}}$
d-banded	$N \leq d \leq N(B-1)$		$\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is drawn from a continuous distribution with $M \geq \sqrt{\frac{2d+1}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{2d+1}{(2B-1)N^2}}$
	$\mathcal{K}$ is a length-$(N-1)$ *circular* sparse ruler	$\frac{2+\sqrt{4N-3}}{2N} \leq \eta_{\min} \leq \frac{1}{N}\left\lceil \sqrt{3\left\lfloor \frac{N}{2} \right\rfloor} \right\rceil$		

TABLE I: Optimal designs and compression ratios.

in many other sub-Nyquist acquisition techniques, which have recently experienced an intense development.

For example, time signals can be compressively acquired using C-ADCs such as *interleaved ADCs* [27], *random demodulators* [5], *modulated wideband converters* [4], and *random modulators pre-integrators* [28]. If $\boldsymbol{x}$ contains the Nyquist samples of $x(t)$, their operation can be described by (7) (see also Fig. 6). Note, however, that no C-ADC internally acquires Nyquist samples since this would entail precisely the disadvantages of conventional ADCs that they attempt to avoid. Nonetheless, they represent a convenient mathematical abstraction.

As for spatial signals, sparse samplers can be easily implemented by removing unused antennas, whereas dense samplers require analog combining (see e.g. [23]).

IV. Main Applications

The problems that can be formulated in CCS terms are those relying exclusively on the second-order moments of a certain signal $\boldsymbol{x}$. In this section, we elaborate on the mathematical formulation of the signal processing problems involved in some of the main applications. In each case we indicate the set of basis matrices $\mathcal{S} = \{\boldsymbol{\Sigma}_0, \ldots, \boldsymbol{\Sigma}_{S-1}\}$ to be used (see (4)).

A. Applications in the Time Domain

CCS is especially convenient to acquire wideband signals, whose rapid variations cannot be easily captured by conventional ADCs. As described in Sec. III-B, this difficulty motivates the usage of C-ADCs, whose linear operation can be described by (7). Their usage in CCS has been considered in a number of applications where acquisition designs and reconstruction algorithms have been proposed. Some of them are detailed next.

- **Compressive Power Spectrum Estimation:** The goal is to estimate $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ from $\boldsymbol{y}$ with the only constraint that $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ must be Hermitian Toeplitz and positive semidefinite. This means that the matrices in $\mathcal{S}$ span the subspace of Hermitian Toeplitz matrices. If the length L (in samples) of the acquisition window is greater than the length of the autocorrelation sequence $\sigma[m]$, then $\mathcal{S}$ can be set to a basis of the subspace of d-banded matrices [19]. Other approaches in the literature follow from the consideration of bases for the subspace of circulant matrices, which arise by stating the problem in the frequency domain [14], [15]. The positive (semi)definiteness of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ can be ignored in order to obtain simple estimators, or it can be exploited using methods like those in [10].
- **Wideband Spectrum Sensing:** Applications such as dynamic spectrum sharing in cognitive radio networks [29] require monitoring the power of different transmitters operating on wide frequency bands. Suppose that a spectrum sensor is receiving the signal $\boldsymbol{x} = \sum_i \sqrt{\alpha_i}\boldsymbol{x}^{(i)}$, where the component $\sqrt{\alpha_i}\boldsymbol{x}^{(i)}$ contains the Nyquist samples of the signal received from the i-th transmitter. If $\boldsymbol{x}^{(i)}$ is power normalized, then α_i is the power received from the i-th transmitter. Since the second-order statistics of $\boldsymbol{x}^{(i)}$, collected in $\boldsymbol{\Sigma}_i = \mathrm{E}\left\{\boldsymbol{x}^{(i)}(\boldsymbol{x}^{(i)})^H\right\}$, are typically known [25], [30], [31], estimating the power of each transmitter amounts to estimating the α_i's in the expansion (4).
  CCS is of special relevance in this application since the typically large number of transmitters means that $\boldsymbol{x}$ is wideband, which motivates the usage of C-ADCs. Various estimation algorithms have been proposed on these grounds in [30].
- **Frequency estimation:** C-ADCs can be used to identify sinusoids in wideband signals [32]. If R denotes the number of sinusoids, the uncompressed signal samples can be modeled as $x[l] = \sum_{i=0}^{R-1} s_i a^{(i)}[l] + w[l]$, where $s_i \in \mathbb{C}$ is random, $w[l]$ is noise, and $a^{(i)}[l] = e^{\jmath\omega_i l}$ is a complex exponential whose frequency ω_i is to be estimated, possibly along with the variance of s_i. This is the problem of estimating a *sparse* power spectrum [5]. Many algorithms for estimating these frequencies rely on estimates of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}$, which is known to be Hermitian Toeplitz and positive semidefinite [11]. From the observations provided by a C-ADC, whose usage is especially convenient if $\boldsymbol{x}$ is wideband, one can first reconstruct $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ and subsequently apply one of the existing techniques, which take $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ as the input parameter. To accomplish the reconstruction, one can use (4) with $\mathcal{S}$ being a set spanning the subspace

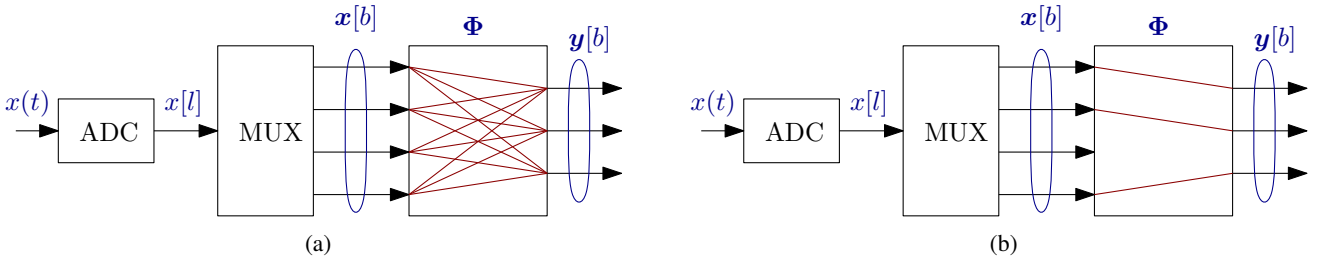


Fig. 6: Mathematical model for the operation of a C-ADC. (a) dense sampler (b) sparse sampler.

of Hermitian Toeplitz matrices.

B. Applications in the Spatial Domain

In applications requiring estimating the so-called *angular spectrum* (e.g. sonar, radar, astronomy, localization), introducing compression may considerably decrease hardware costs. In schemes using sparse sampling (see e.g. [6]–[8], [13], [33]), only the antennas corresponding to the non-null columns of $\bar{\mathbf{\Phi}}$ need to be physically deployed, whereas in schemes employing dense sampling [23], the number of antennas is preserved after introducing compression, but the number of RF chains is reduced.

In applications employing CCS, the received signal is typically modeled as a sum of incoherent planar waves emitted by a collection of sources in the far field. The spatial field produced by each source results in a Toeplitz spatial covariance matrix which depends on the angle of arrival of that source. The sum of all contributions and noise, assumed white for simplicity, produces therefore a Toeplitz $\mathbf{\Sigma}_{\boldsymbol{x}}$.

Two problems are usually considered:

- **Incoherent Imaging:** If a continuous source distribution is assumed, then the angular spectrum is *dense*. The problem can be formulated as described above for compressive power spectrum estimation, since the only structure present in $\mathbf{\Sigma}_{\boldsymbol{x}}$ is that it is Hermitian Toeplitz and positive semidefinite [8]. However, recent works show that the problem can also be stated using circulant covariance matrices [16], [17].
- **Direction of Arrival Estimation:** The goal is to estimate the angles of arrival of a finite number of sources. A broad family of methods exist to this end (see e.g. [8], [13], [33], [34]), most of them following the same principles as described above for frequency estimation, since both problems admit the formulation of sparse power spectrum estimation.

Most applications listed in this section have been covered with the two compression methods introduced in previous sections, namely sparse and dense sampling, either in a periodic or non-periodic fashion. For time signals, periodicity typically arises due to the block-by-block operation of C-ADCs (see e.g. [19], [30], [35]); for spatial signals, by consideration of periodic arrays [16], [17].

V. Estimation and Detection

Having described the modeling and compression schemes for CCS, we turn attention to the reconstruction problems in

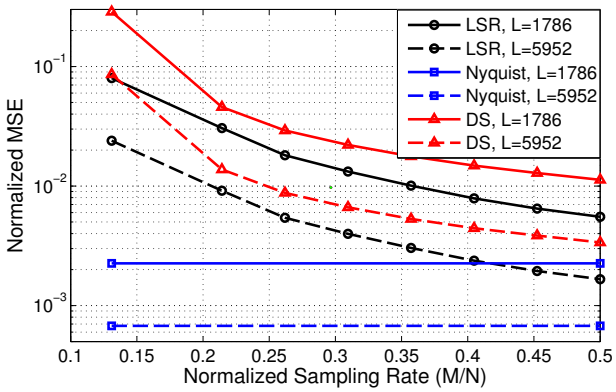


Fig. 7: Mean squared error of the estimate of the least squares algorithm when $\mathbf{\Sigma}_{\boldsymbol{x}}$ is 168-banded (figure obtained from [19]).

CCS. For estimation, it boils down to reconstructing $\mathbf{\Sigma}_{\boldsymbol{x}}$ in (4) from the compressive measurements $\boldsymbol{y}$.

Since $\boldsymbol{y} = \bar{\mathbf{\Phi}}\boldsymbol{x}$, it follows that $\mathbf{\Sigma}_{\boldsymbol{y}} = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_{\boldsymbol{x}}\bar{\mathbf{\Phi}}^H$. When $\mathbf{\Sigma}_{\boldsymbol{x}}$ is given by (4), $\mathbf{\Sigma}_{\boldsymbol{y}}$ can be similarly represented as

$$\mathbf{\Sigma}_{\boldsymbol{y}} = \sum_{i=0}^{S-1} \alpha_i \bar{\mathbf{\Sigma}}_i, \quad \alpha_i \in \mathbb{R}, \tag{11}$$

where $\bar{\mathbf{\Sigma}}_i = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_i\bar{\mathbf{\Phi}}^H$. This means that $\mathbf{\Sigma}_{\boldsymbol{y}}$ and $\mathbf{\Sigma}_{\boldsymbol{x}}$ share the same coordinates α_i. If the compression is accomplished properly, for example using the designs discussed in previous sections, these coordinates are identifiable and can be estimated from the observations of $\boldsymbol{y}$.

A. Maximum Likelihood

If the probability distribution of the observations is known, one may resort to a maximum likelihood estimate of $\mathbf{\Sigma}_{\boldsymbol{y}}$. For example, if $\boldsymbol{y}$ is zero-mean Gaussian and

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{y}} = \frac{1}{T}\sum_{\tau=0}^{T-1} \boldsymbol{y}_\tau \boldsymbol{y}_\tau^H, \tag{12}$$

is the sample covariance matrix of the compressed observations, the maximization of the log-likelihood leads to the following problem:

$$\underset{\{\alpha_i\}_i}{\text{minimize}} \quad \log|\mathbf{\Sigma}_{\boldsymbol{y}}| + \text{Tr}\left(\mathbf{\Sigma}_{\boldsymbol{y}}^{-1}\hat{\mathbf{\Sigma}}_{\boldsymbol{y}}\right) \tag{13}$$

subject to (11). Numerous algorithms have been proposed to solve this non-convex problem (see e.g. [10], [30], [36]).

B. Least Squares

The maximum likelihood approach involves high computational costs and requires an accurate statistical characterization of the observations. For these reasons, it is customary to rely on geometrical considerations and project the sample covariance matrix onto the span of $\mathcal{S}$.

From $\boldsymbol{\Sigma_y} = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma_x}\bar{\boldsymbol{\Phi}}^H$, it follows that $\boldsymbol{\sigma_y} = (\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{\sigma_x}$, where $\boldsymbol{\sigma_y}$ and $\boldsymbol{\sigma_x}$ are, respectively, the vectorizations of $\boldsymbol{\Sigma_y}$ and $\boldsymbol{\Sigma_x}$. Vectorizing (4) yields $\boldsymbol{\sigma_x} = \sum_{i=0}^{S-1} \alpha_i \boldsymbol{\sigma}_i$ or, in matrix form, $\boldsymbol{\sigma_x} = \boldsymbol{S\alpha}$, where we have arranged the vectors $\boldsymbol{\sigma}_i$ as columns of the matrix $\boldsymbol{S}$ and the coordinates α_i as elements of the vector $\boldsymbol{\alpha}$. This results in the relation $\boldsymbol{\sigma_y} = (\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S\alpha}$. If the $M^2B^2 \times S$ matrix $(\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S} \in \mathbb{C}$ is full column rank, substituting $\boldsymbol{\sigma_y}$ by a sample estimate $\hat{\boldsymbol{\sigma}}_{\boldsymbol{y}}$ produces an overdetermined system $\hat{\boldsymbol{\sigma}}_{\boldsymbol{y}} = (\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S}\hat{\boldsymbol{\alpha}}$, whose solution via least squares yields the desired estimate in closed form [14], [19], [30]:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{\text{LS}} = \text{vec}^{-1}\left\{\boldsymbol{S}[(\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S}]^{\dagger}\hat{\boldsymbol{\sigma}}_{\boldsymbol{y}}\right\}. \tag{14}$$

Here, the operator $\text{vec}^{-1}\{\cdot\}$ re-stacks a vector into a square matrix.

Fig. 7 illustrates the performance of this technique when $\boldsymbol{\Sigma_x}$ is 168-banded (see [19] for more details) and several sampling designs are used. Clearly, the mean squared error of the estimate is larger when compression is introduced since it reduces the total number of samples. This effect is not exclusive of least squares estimation – it negatively affects any reasonable estimator. For this reason, including compression usually requires longer observation time if a certain target performance metric is to be achieved. This does not conflict with the true purpose of compression, which is to reduce the *average* sampling rate – a parameter that affects most decisively to the cost of the hardware.

However, note that this approach does not exploit the fact that $\boldsymbol{\Sigma_x}$ is positive semidefinite. This constraint can be enforced to improve the estimation performance at the expense of greater complexity. For instance, one may attempt to minimize the least squares cost $||\hat{\boldsymbol{\sigma}}_{\boldsymbol{y}} - (\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S}\hat{\boldsymbol{\alpha}}||^2$ subject to the constraint $\boldsymbol{\Sigma_x} \geq \boldsymbol{0}$, which is a convex problem. Other constraints can also be imposed if more prior information is known. For instance, the elements of $\hat{\boldsymbol{\alpha}}$ might be non-negative [30], in which case one would introduce the constraint $\hat{\boldsymbol{\alpha}} \geq 0$; or it can be known that $\hat{\boldsymbol{\alpha}}$ is sparse either by itself or on a linearly transformed domain, in which case one may impose the constraint $||\boldsymbol{F}_s\hat{\boldsymbol{\alpha}}||_0 \leq S_0$, where S_0 is the number of non-zero entries and $\boldsymbol{F}_s$ takes $\hat{\boldsymbol{\alpha}}$ to the domain where it is sparse. For instance, the elements of $\boldsymbol{F}_s\hat{\boldsymbol{\alpha}}$ may be samples of the power spectrum [37]. Since the zero-norm in this constraint is not convex, it is typically relaxed to a ℓ_1-norm. For example, an ℓ_1-norm regularized least squares formulation can be adopted as follows:

$$\underset{\hat{\boldsymbol{\alpha}}}{\text{minimize}} \quad \|\hat{\boldsymbol{\sigma}}_{\boldsymbol{y}} - (\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}})\boldsymbol{S}\hat{\boldsymbol{\alpha}}\|^2 + \lambda\|\boldsymbol{F}_s\hat{\boldsymbol{\alpha}}\|_1 \tag{15}$$

In (15), signal compression is induced by the statistical structure of $\boldsymbol{\Sigma_x}$ beyond sparsity, while the additional sparsity structure can lead to stronger compression at the expense of increased computational complexity compared to the closed-form solution in (14).

C. Detection

In detection theory, we are interested in deciding whether a signal of interest is present or not. This operation is typically hindered by the presence of noise and other waveforms, such as *clutter* in radar or interference in communications.

In many cases of interest, this problem can be stated in terms of the second-order statistics of the signals involved, so the goal is to decide one of the following hypotheses:

$$\begin{aligned} \mathcal{H}_0 &: \quad \boldsymbol{\Sigma_x} = \boldsymbol{\Sigma_w} \\ \mathcal{H}_1 &: \quad \boldsymbol{\Sigma_x} = \boldsymbol{\Sigma_r} + \boldsymbol{\Sigma_w}, \end{aligned} \tag{16}$$

where $\boldsymbol{\Sigma_r}$ and $\boldsymbol{\Sigma_w}$ respectively collect the second-order statistics of the signal of interest and noise/interference. Our decision must be based on the observation of the compressed samples $\boldsymbol{y} = \bar{\boldsymbol{\Phi}}\boldsymbol{x}$, whose covariance matrix $\boldsymbol{\Sigma_y}$ is given by $\bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma_w}\bar{\boldsymbol{\Phi}}^H$ under $\mathcal{H}_0$ and $\bar{\boldsymbol{\Phi}}(\boldsymbol{\Sigma_r} + \boldsymbol{\Sigma_w})\bar{\boldsymbol{\Phi}}^H$ under $\mathcal{H}_1$. A *most powerful* detection rule exists for this simple setting and can be found using the Neyman-Pearson lemma [11]. If $p(\boldsymbol{y}; \mathcal{H}_i)$ denotes the density under hypothesis $\mathcal{H}_i$, this rule decides $\mathcal{H}_1$ when the ratio $p(\boldsymbol{y}; \mathcal{H}_1)/p(\boldsymbol{y}; \mathcal{H}_0)$ exceeds a certain threshold set to achieve a target probability of false alarm [11].

More general problems arise by considering basis expansions like the one in (4). In this case, the goal may be to decide whether one of the α_i's, say α_0, is positive or zero; while the others are unknown and treated as nuisance parameters [30]. In these cases, no uniformly most powerful test exists and one must resort to other classes of detectors such as the *generalized likelihood ratio test*, which makes a decision by comparing $p(\boldsymbol{y}; \hat{\boldsymbol{\alpha}}_{\mathcal{H}_1})/p(\boldsymbol{y}; \hat{\boldsymbol{\alpha}}_{\mathcal{H}_0})$ against a threshold, using $\hat{\boldsymbol{\alpha}}_{\mathcal{H}_i}$ as the maximum likelihood estimate of $\boldsymbol{\alpha}$ under hypothesis $\mathcal{H}_i$ [30].

VI. Modal Analysis

As mentioned in Sec. IV, the problem of estimating the frequency of a number of noise-corrupted sinusoids and the problem of estimating the direction of arrival of a number of sources in the far field are instances of a class of sparse spectrum estimation problems, which allow a common formulation as modal analysis [11].

Suppose that the observations are given by

$$\boldsymbol{x} = \sum_{i=0}^{R-1} s_i \boldsymbol{a}^{(i)} + \boldsymbol{w} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{w}, \tag{17}$$

where $\boldsymbol{a}^{(i)} = [1, e^{\jmath\omega_i}, \ldots, e^{\jmath\omega_i(L-1)}]^T$ are the so-called *steering vectors*, $\boldsymbol{A} = [\boldsymbol{a}^{(0)}, \ldots, \boldsymbol{a}^{(R-1)}]$ is the *manifold* matrix, $\boldsymbol{w}$ is noise and the coefficients s_i, collected in the vector $\boldsymbol{s}$, are uncorrelated random variables. The structure of $\boldsymbol{a}^{(i)}$ stems from the fact that each antenna receives the signal s_i with a different phase shift. Because in a ULA the antennas are uniformly spaced, the relative phase shift between each pair of antennas is an integer multiple of a normalized frequency quantity ω_i, which is a function of the angle of arrival.

The covariance matrix of $\boldsymbol{x}$ is given by

$$\boldsymbol{\Sigma_x} = \boldsymbol{A}\boldsymbol{\Sigma_s}\boldsymbol{A}^H + \sigma_{\boldsymbol{w}}^2\boldsymbol{I}_L, \tag{18}$$

where $\sigma_{\boldsymbol{w}}^2$ is the power of the noise process, assumed white for simplicity, and $\boldsymbol{\Sigma}_{\boldsymbol{s}}$ is the covariance matrix of $\boldsymbol{s}$, which is diagonal since the sources are uncorrelated. Note that these assumptions result in $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ having a Toeplitz structure.

The compressed observations can be written as $\boldsymbol{y} = \bar{\boldsymbol{\Phi}}\boldsymbol{x} = \bar{\boldsymbol{A}}\boldsymbol{s}$, where $\bar{\boldsymbol{A}} = \bar{\boldsymbol{\Phi}}\boldsymbol{A}$, and have covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{y}} = \bar{\boldsymbol{A}}\boldsymbol{\Sigma}_{\boldsymbol{s}}\bar{\boldsymbol{A}}^H + \sigma_{\boldsymbol{w}}^2\bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}^H. \tag{19}$$

The parameters ω_i can be estimated from $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ using adaptations of traditional techniques such as MUSIC [35] and MVDR [38].

Alternative approaches are based on the observation that the vectorization of (19) can be written in terms of the Khatri-Rao product, defined as the column-wise application of the Kronecker product, as

$$\operatorname{vec}(\boldsymbol{\Sigma}_{\boldsymbol{y}}) = (\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}})\operatorname{diag}\{\boldsymbol{\Sigma}_{\boldsymbol{s}}\} + \sigma_{\boldsymbol{w}}^2\operatorname{vec}(\bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}^H). \tag{20}$$

The matrix $\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}}$ can be thought of as a *virtual manifold matrix*, since this expression has the same structure as (17) [13], [39]. An especially convenient structure is when $\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}}$ contains all the rows in the manifold matrix of a ULA [40]. To this end, array geometries like two-level nested arrays [13], coprime arrays [32] and linear sparse rulers [34] can be used.

Other approaches stem from the idea of *gridding*. One can construct the matrix $\bar{\boldsymbol{A}}$ using a fine grid of angles ω_i and then estimate $\boldsymbol{s}$ from $\boldsymbol{y} = \bar{\boldsymbol{A}}\boldsymbol{s}$ exploiting the idea that most of its components will be zero since, for a grid fine enough, most of the columns of $\bar{\boldsymbol{A}}$ will correspond to angles where there are no sources. In other words, $\boldsymbol{s}$ is *sparse*, which means that the techniques from [5], [41] can be applied to recover this vector. This technique does not have to rely on second-order statistics, but similar grid-based approaches can be devised which operate on (20) instead [42].

VII. Preprocessing

Most of the methods described in this article make use of the sample covariance matrix of $\boldsymbol{y}$, defined in (12). Under general conditions, the average $T^{-1}\sum_\tau \boldsymbol{y}_\tau\boldsymbol{y}_\tau^H$ converges to the true $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ as T becomes large. If the compression does not destroy relevant second-order statistical information, the matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ contains all the information required to solve the problem with identifiability, whereas the accuracy performance of these methods strongly depends on the number T of available realizations.

Typically, in those applications involving spatial signals, the output of a sensor array is synchronously sampled. If $\boldsymbol{y}_\tau$ collects the samples acquired at time instant τ, it is clear that multiple observations of $\boldsymbol{y}$ can be obtained by considering successive snapshots $\tau = 0, 1, \ldots, T-1$. This means that, whereas $\boldsymbol{y}$ contains samples across space, the different snapshots are acquired along the time dimension. Conversely, in applications involving time-domain signals, $\boldsymbol{y}$ contains samples acquired over time. A possible means to observe multiple realizations is by considering the vectors $\boldsymbol{y}_\tau$ observed at different locations $\tau = 0, 1, \ldots, T-1$. In this case, while $\boldsymbol{y}$ contains time samples, τ ranges across space. This establishes a duality relation between the space and the time domains: when the observed signals are defined on one domain, multiple observations can be acquired over the other.

Unfortunately, many applications do not allow averaging over the dual domain and one must cope with a single observation, say $\boldsymbol{y}_0$, producing the estimate $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}} = \boldsymbol{y}_0\boldsymbol{y}_0^H$. In fact, this matrix is not a satisfactory estimate of $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ since it is always rank-one and is not Toeplitz. For this reason, an estimation/detection method working on this kind of estimate may exhibit a poor performance.

The key observation in this case is that, although multiple realizations cannot be acquired, sometimes it is possible to gather a large number of samples in the domain where the signal is defined. One can therefore exploit the Toeplitz structure of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ to obtain a more convenient estimate [30]. In particular, due to the block-by-block operation described by (8), the fact that $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is Toeplitz means that $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ is *block Toeplitz*; that is, it can be written as

$$\boldsymbol{\Sigma}_{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-1] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-B+1] \\ \boldsymbol{\Sigma}_{\boldsymbol{y}}[1] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{\boldsymbol{y}}[B-1] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[B-2] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] \end{bmatrix}, \tag{21}$$

where the (non-necessarily Toeplitz) $M \times M$ blocks $\boldsymbol{\Sigma}_{\boldsymbol{y}}[k]$ are given by

$$\boldsymbol{\Sigma}_{\boldsymbol{y}}[k] = \mathrm{E}\left\{\boldsymbol{y}[b]\boldsymbol{y}^H[b-k]\right\}, \quad \forall b. \tag{22}$$

This suggests the estimate

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{y}}[k] = \frac{1}{\text{no. of terms}}\sum_b \boldsymbol{y}[b]\boldsymbol{y}^H[b-k]. \tag{23}$$

Moreover, since $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ is Hermitian, this computation needs only to be carried out for $k = 0, \ldots, B-1$. More sophisticated estimates exhibiting different properties were analyzed in [30].

Another observation worth mentioning is that the smaller k, the higher the quality of the estimates of $\boldsymbol{\Sigma}_{\boldsymbol{y}}[k]$. The reason is that the number of averaging terms in (23) is larger for blocks lying close to the main diagonal than for distant ones. Thus, it seems reasonable to operate on a *cropped* covariance matrix:

$$\boldsymbol{\Sigma}_{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-1] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-\tilde{B}+1] \\ \boldsymbol{\Sigma}_{\boldsymbol{y}}[1] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[-\tilde{B}+2] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{\boldsymbol{y}}[\tilde{B}-1] & \boldsymbol{\Sigma}_{\boldsymbol{y}}[\tilde{B}-2] & \ldots & \boldsymbol{\Sigma}_{\boldsymbol{y}}[0] \end{bmatrix}, \tag{24}$$

where $\tilde{B} < B$. Note that in this case, the dimension of the cropped matrix is less than the length of the observation vector $\boldsymbol{y}$.

In certain cases, this technique leads to important computational savings at a small performance loss since the terms being retained are those of the *highest quality* [30].

VIII. Advanced Techniques

Having explained the basic principles of CCS, we now illustrate the broad applications of CCS by considering other forms of second-order statistics as well as implementation issues in practical systems.

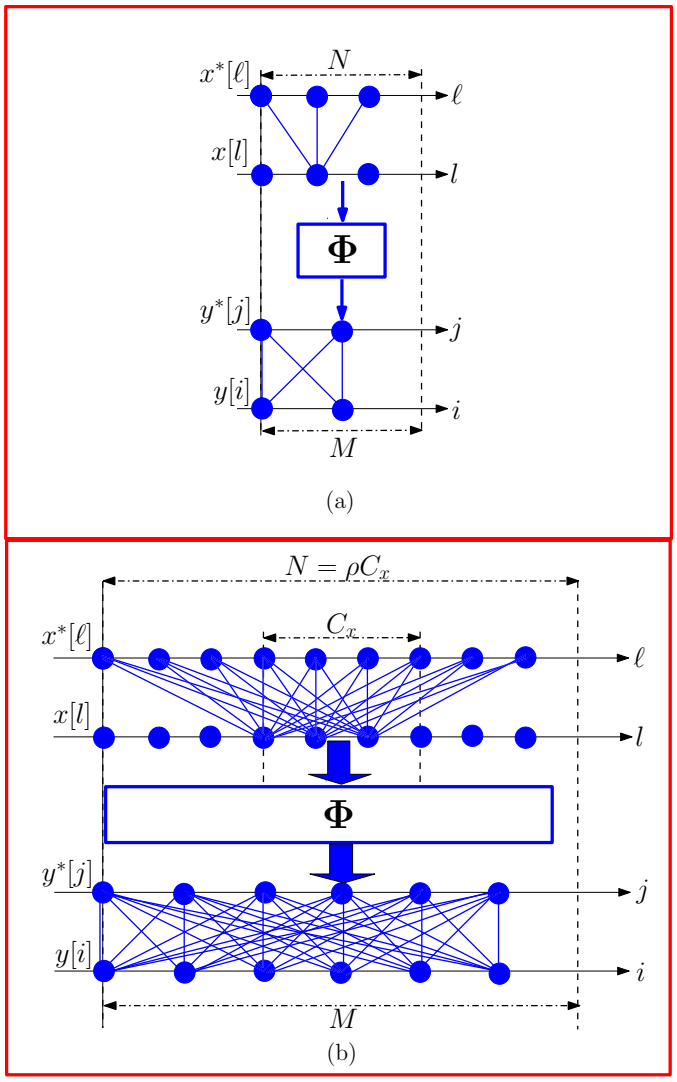


Fig. 8: (a) Compression of a block of $N = 3$ samples of a WSS signal using a 2×3 matrix $\mathbf{\Phi}$, which produces compressed blocks of $M = 2$ samples. (b) Compression of a block of $N = 9$ samples of a cyclostationary signal, which produces compressed blocks of $M = 6$ samples.

A. Cyclostationarity

Cyclostationarity exhibits in many man-made signals with inherent periodicity, which is a useful feature for estimation, detection and classification of digital communication signals [26]. While there exist several methods to reconstruct the second-order statistics of a cyclostationary signal from compressed observations (see e.g. [26], [43], [44]), in this section, we only illustrate the main principles underlying these techniques using a simple model.

We say that a signal is cyclostationary if its *time-varying covariance function* is periodic. Formally, the time-varying covariance function of a zero-mean process $x[l]$ is defined as $\sigma[l, \ell] = \mathrm{E}\{x[l]x^*[l-\ell]\}$, and it is said to be periodic when there exists an integer C_x, called *cyclic period*, such that $\sigma[l + n_c C_x, \ell] = \sigma[l, \ell]$ for any integer n_c [26], [43]. Although other forms of cyclostationarity exist, we confine ourselves to this one for simplicity. Note that cyclostationary signals generalize WSS signals, since the latter may be viewed as a particular case of the former with $C_x = 1$.

Suppose that the length of the sampling block is an integer multiple of the cyclic period, that is, $N = \rho C_x$ for some integer ρ. Then, the vector $\boldsymbol{x}[b]$ can be divided into ρ sub-blocks of length C_x as

$$\boldsymbol{x}[b] = [\tilde{\boldsymbol{x}}^T[b\rho], \tilde{\boldsymbol{x}}^T[b\rho+1], \ldots, \tilde{\boldsymbol{x}}^T[b\rho+\rho-1]]^T. \quad (25)$$

The fact that $\sigma[l, \ell]$ is periodic along l means that $\mathbf{\Sigma}_{\boldsymbol{x}}$ is block Toeplitz with $C_x \times C_x$ blocks. By defining an $N \times N$ matrix $\mathbf{\Sigma}_{\boldsymbol{x}}[b] = \mathrm{E}\{\boldsymbol{x}[b']\boldsymbol{x}^H[b'-b]\}$, we can write

$$\mathbf{\Sigma}_{\boldsymbol{x}} = \begin{bmatrix} \mathbf{\Sigma}_{\boldsymbol{x}}[0] & \mathbf{\Sigma}_{\boldsymbol{x}}[-1] & \ldots & \mathbf{\Sigma}_{\boldsymbol{x}}[-B+1] \\ \mathbf{\Sigma}_{\boldsymbol{x}}[1] & \mathbf{\Sigma}_{\boldsymbol{x}}[0] & \ldots & \mathbf{\Sigma}_{\boldsymbol{x}}[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{\boldsymbol{x}}[B-1] & \mathbf{\Sigma}_{\boldsymbol{x}}[B-2] & \ldots & \mathbf{\Sigma}_{\boldsymbol{x}}[0] \end{bmatrix}, \quad (26)$$

where the blocks $\mathbf{\Sigma}_{\boldsymbol{x}}[b]$ also have a block Toeplitz structure with blocks $\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[\varrho] = \mathrm{E}\{\tilde{\boldsymbol{x}}[\varrho']\tilde{\boldsymbol{x}}^H[\varrho'-\varrho]\}$:

$$\mathbf{\Sigma}_{\boldsymbol{x}}[b] = \begin{bmatrix} \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho] & \ldots & \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho-\rho+1] \\ \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho+1] & \ldots & \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho-\rho+2] \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho+\rho-1] & \ldots & \mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho] \end{bmatrix}. \quad (27)$$

Cyclostationarity provides an alternative perspective to understand compression of second-order statistics, even for WSS sequences. The main idea is that the resulting sequence $y[k]$ of compressed observations is cyclostationary with cyclic period M, which is larger than that of the original signal C_x.

Fig. 8a intuitively explains this effect for a WSS signal ($C_x = 1$) satisfying $\sigma[l] = 0$ for $|l| > 1$. In that figure, the dots on the l-axis represent a block of $N = 3$ samples of the WSS sequence $x[l]$ and the dots on the ℓ-axis represent their complex conjugates. The three lines connecting the dots in both axes represent the (possibly) different values of correlation between samples. Note that no extra lines need to be drawn since only $\sigma[-1]$, $\sigma[0]$, and $\sigma[1]$ are allowed to be different from zero. Since the correlation of a WSS signal is determined by the time-lags independent of the time origin, only one representative dot along $x[l]$ is chosen as the time origin. A similar representation is provided at the bottom of Fig. 8a for the compressed sequence $y[k]$, which can be seen to be cyclostationary with cyclic period $C_y = M$ (just apply the above considerations to (21)). Note that the four line segments effectively capture all the different correlation values between samples of $y[i]$. Here, $y[i]$ is no longer WSS due to the compression process, and hence all time origins along $y[i]$ within a block are selected to depict the correlations.

Observe that, although the number of samples in each block was reduced from 3 to 2 after compression, the number of different correlation values has increased from 3 to 4 !! This means that, whereas one cannot reconstruct the samples of $x[l]$ from $y[k]$ without further assumptions, there is a chance of reconstructing the second-order statistics of $x[l]$ from those of $y[k]$. In fact, if $\mathbf{\Phi}$ satisfies certain conditions, one can, for instance, estimate $\mathbf{\Sigma}_{\boldsymbol{y}}$ from $y[k]$ using sample statistics and

obtain an estimate of $\mathbf{\Sigma}_{\boldsymbol{x}}$ via least squares as described in Sec. V-B.

Now assume that $x[l]$ is a cyclostationary signal of cyclic period $C_x = 3$ and assume that $\sigma[l, \ell]$ is such that each sub-block of C_x samples is only correlated with the neighboring sub-blocks. Fig. 8b illustrates a case where a block of $N = \rho C_x = 9$ samples is compressed to produce a block of $M = 6$ samples. As before, all (possibly) distinct correlation values have been represented with the corresponding line segments. Observe that, although the number of output samples is lower than the number of input samples, it may be possible to use the $M^2 = 36$ correlation values at the output to reconstruct the $\rho C_x^2 = 27$ correlation values at the input.

We next describe a reconstruction method based on least squares. Note from (8) and (22) that the $M \times M$ blocks of $\mathbf{\Sigma}_{\boldsymbol{y}}$ (c.f. (21)) can be written as

$$\mathbf{\Sigma}_{\boldsymbol{y}}[b] = \mathbf{\Phi}\mathbf{\Sigma}_{\boldsymbol{x}}[b]\mathbf{\Phi}^H. \tag{28}$$

In order to exploit the block Toeplitz structure of $\mathbf{\Sigma}_{\boldsymbol{x}}[b]$ (see (27)) we first vectorize both sides of (28) and apply the properties of the Kronecker product to obtain

$$\text{vec}(\mathbf{\Sigma}_{\boldsymbol{y}}[b]) = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\text{vec}(\mathbf{\Sigma}_{\boldsymbol{x}}[b]). \tag{29}$$

Now we rewrite the rightmost vector of this expression as

$$\text{vec}(\mathbf{\Sigma}_{\boldsymbol{x}}[b]) = \mathbf{T}\boldsymbol{\beta}_x[b], \tag{30}$$

where

$$\begin{aligned}\boldsymbol{\beta}_x[b] =&[\text{vec}^T(\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho]), \text{vec}^T(\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho+1]), \dots,\\ &\text{vec}^T(\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho+\rho-1]), \text{vec}^T(\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho-\rho+1]),\\ &\dots, \text{vec}^T(\mathbf{\Sigma}_{\tilde{\boldsymbol{x}}}[b\rho-1])]^T\end{aligned} \tag{31}$$

is a $(2\rho-1)C_x^2 \times 1$ vector containing all the possibly distinct entries of $\mathbf{\Sigma}_{\boldsymbol{x}}[b]$ and where $\mathbf{T}$ is the $N^2 \times (2\rho-1)C_x^2$ repetition matrix which maps (and repeats) the elements of $\boldsymbol{\beta}_x[b]$ into $\text{vec}(\mathbf{\Sigma}_{\boldsymbol{x}}[b])$. Substituting (30) in (29) yields

$$\text{vec}(\mathbf{\Sigma}_{\boldsymbol{y}}[b]) = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\mathbf{T}\boldsymbol{\beta}_x[b]. \tag{32}$$

Note that, while $\mathbf{\Phi}$ generally has more columns than rows (as $M < N$), the $M^2 \times (2\rho-1)C_x^2$ matrix $(\mathbf{\Phi}^* \otimes \mathbf{\Phi})\mathbf{T}$ can have more rows than columns. Hence, under certain conditions, (32) is an overdetermined system for each $b = -B+1, \dots, 0, 1, \dots, B-1$. Substituting $\mathbf{\Sigma}_{\boldsymbol{y}}[b]$ by a sample estimate one can obtain an estimate of $\boldsymbol{\beta}_x[b]$ as the least squares solution of that system and obtain an estimate of $\mathbf{\Sigma}_{\boldsymbol{x}}$ by plugging the result in (30).

This approach has been proposed in [43] using dense samplers. A more specific case is discussed in [44], which specifically proposes the usage of a sparse matrix $\mathbf{\Phi}$ with a block diagonal structure.

B. Dynamic Sampling

There are situations where the signal itself does not possess evident covariance structure, but we can effect compression by means of dynamic sampling.

Let us go back to the array processing example of Section II-A, where the Toeplitz structure of $\mathbf{\Sigma}_{\boldsymbol{x}}$ allowed us to estimate $\mathbf{\Sigma}_{\boldsymbol{x}}$ using $M < N$ antennas. This structure relies on the assumption that the sources are uncorrelated. If this is not the case, then the only structure present in $\mathbf{\Sigma}_{\boldsymbol{x}}$ is that it is Hermitian and positive semidefinite, which means that $\mathbf{\Sigma}_{\boldsymbol{x}}$ cannot be estimated with less than N antennas.

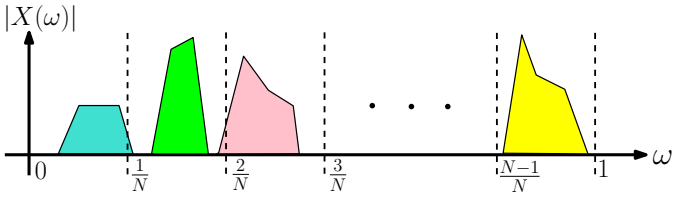


Fig. 9: Example of a signal with a multi-band structure. Here, the digital frequency axis ω is splitted into N uniform bins.

A possible way to circumvent this problem is to adopt a dynamic scheme where a *full* array of N antennas (the uncompressed array) is deployed but only a certain subset of antennas is activated at each time slot [40]. The activation pattern may change periodically over time, which allows computing sample statistics for every activation pattern. With this technique, only a small number of RF chains needs to be deployed. This is illustrated in Fig. 10, where only $K = 4$ out of the $L = 7$ physical antennas are active at each time slot. The antenna selection may be implemented using analog circuitry. Note that a similar scheme could be used relying on dense samplers. Alternative settings include [45], where different arrays are obtained by sampling different frequencies.

In order to estimate $\mathbf{\Sigma}_{\boldsymbol{x}}$, the least squares method from previous sections can be used. Let $\bar{\mathbf{\Phi}}_g$ denote the $K \times L$ compression matrix used during the g-th time slot. The covariance matrix of the compressed observations at time slot g is given by

$$\mathbf{\Sigma}_{\boldsymbol{y}_g} = \bar{\mathbf{\Phi}}_g \mathbf{\Sigma}_{\boldsymbol{x}} \bar{\mathbf{\Phi}}_g^H. \tag{33}$$

Vectorizing both sides and combining the result for the G time slots in each period yields

$$\begin{bmatrix} \text{vec}(\mathbf{\Sigma}_{\boldsymbol{y}_0}) \\ \text{vec}(\mathbf{\Sigma}_{\boldsymbol{y}_1}) \\ \vdots \\ \text{vec}(\mathbf{\Sigma}_{\boldsymbol{y}_{G-1}}) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{\Phi}}_0^* \otimes \bar{\mathbf{\Phi}}_0 \\ \bar{\mathbf{\Phi}}_1^* \otimes \bar{\mathbf{\Phi}}_1 \\ \vdots \\ \bar{\mathbf{\Phi}}_{G-1}^* \otimes \bar{\mathbf{\Phi}}_{G-1} \end{bmatrix} \text{vec}(\mathbf{\Sigma}_{\boldsymbol{x}}) = \mathbf{\Psi}\text{vec}(\mathbf{\Sigma}_{\boldsymbol{x}}). \tag{34}$$

If the $GK^2 \times L^2$ matrix $\mathbf{\Psi}$ has full column rank, then it is possible to estimate $\mathbf{\Sigma}_{\boldsymbol{y}_g}$, $g = 0, \dots, G-1$ using sample statistics and then obtain an estimate of $\mathbf{\Sigma}_{\boldsymbol{x}}$ as the least squares solution of (34). It can be shown that this full rank condition is satisfied if every pair of antennas is simultaneously active in at least one time slot per scanning period [40]. In order to estimate $\mathbf{\Sigma}_{\boldsymbol{y}_g}$ via sample statistics, one may simply average over the observations in the g-th time slot of each period.

C. Compressive Covariance Estimation of Multi-band Signals

When uncorrelated signal sources are concerned, a multi-band signal structure arises in many applications [14], [35], [46]. Suppose that our goal is to estimate the second-order

statistics, e.g. the power spectrum, of a time-domain (spatial-domain) signal which has a multiband structure in the frequency (angular) domain (see Fig. 9) [14], [35], [46]. For simplicity, consider a time-domain signal $x(t)$, although the discussion immediately carries over to the spatial domain [46]. We will show how this problem can be cast as the problem of compressing a circulant covariance matrix (see Sec. II-B).

The *trick* is to reformulate the problem in the frequency domain. Let $X(\omega)$ denote the discrete-time Fourier transform (DTFT) at digital frequency $\omega \in [0,1)$ of the sequence $x[l]$, $l = 0, \ldots, L-1$. Let us also split the frequency axis $\omega \in [0,1)$ into N bins of size $\frac{1}{N}$ (see Fig. 9) and introduce, for $\omega \in [0, \frac{1}{N})$, the $N \times 1$ vector $\boldsymbol{x}(\omega) = [X(\omega), X(\omega+\frac{1}{N}), \ldots, X(\omega+\frac{N-1}{N})]^T$.

Now, suppose that instead of concatenating the vectors $\boldsymbol{x}[b]$ vertically to form $\boldsymbol{x}$ (see Sec. III), we arrange them as columns of the $N \times B$ matrix $\boldsymbol{X}$. Repeating the same operation for the compressed samples in $\boldsymbol{y}$ produces the $M \times B$ matrix $\boldsymbol{Y}$. Clearly, since $\bar{\boldsymbol{\Phi}} = \mathbf{I}_B \otimes \boldsymbol{\Phi}$, it follows that the compression model of (7) can be rewritten as

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}. \tag{35}$$

Let us form the $N \times 1$ vector $\bar{\boldsymbol{x}}(\omega)$, whose n-th entry contains the DTFT of the n-th row of $\boldsymbol{X}$. Note that the collection of samples in each row of $\boldsymbol{X}$ is the result of downsampling $x[l]$ by a factor of N. This operation produces N aliases in the frequency domain, which means that the spectrum has period $1/N$. Thus, it suffices to consider $\bar{\boldsymbol{x}}(\omega)$ in the frequency interval $\omega \in [0, \frac{1}{N})$. Likewise, define the $M \times 1$ vector $\bar{\boldsymbol{y}}(\omega)$, $\omega \in [0, \frac{1}{N})$, as the vector containing the DTFTs of the rows of $\boldsymbol{Y}$. Clearly, (35) can then be expressed in the frequency domain using these vectors:

$$\bar{\boldsymbol{y}}(\omega) = \boldsymbol{\Phi}\bar{\boldsymbol{x}}(\omega). \tag{36}$$

The relationship between $\boldsymbol{x}(\omega)$ and $\bar{\boldsymbol{x}}(\omega)$ can be shown to be given by [14], [35], [46]:

$$\bar{\boldsymbol{x}}(\omega) = \frac{1}{N}\boldsymbol{F}_N^H \boldsymbol{x}(\omega), \quad \omega \in [0, 1/N), \tag{37}$$

where $\boldsymbol{F}_N$ is the $N \times N$ discrete Fourier transform (DFT) matrix. From (36) and (37) it follows that

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{y}}}(\omega) = E[\bar{\boldsymbol{y}}(\omega)\bar{\boldsymbol{y}}^H(\omega)] = \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\bar{\boldsymbol{x}}}(\omega)\boldsymbol{\Phi}^H \tag{38}$$

and

$$\boldsymbol{\Sigma}_{\bar{\boldsymbol{x}}}(\omega) = E[\bar{\boldsymbol{x}}(\omega)\bar{\boldsymbol{x}}^H(\omega)] = \frac{1}{N^2}\boldsymbol{F}_N^H \boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega)\boldsymbol{F}_N, \tag{39}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega) = E[\boldsymbol{x}(\omega)\boldsymbol{x}^H(\omega)]$. If the frequency bands are uncorrelated, for instance because they were produced by different sources, and if the width of each band is less than $1/N$, which is the width of the bin, then $\boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega)$ in (39) is a diagonal matrix for all $\omega \in [0, 1/N)$ [46]. Such a diagonal structure is characteristic of multiband signals, which enables compression beyond sparsity. Likewise, since $\boldsymbol{F}_N$ is a DFT matrix, it implies a circulant structure in $\boldsymbol{\Sigma}_{\bar{\boldsymbol{x}}}(\omega)$.

Compare (38) with the expression $\boldsymbol{\Sigma}_{\boldsymbol{y}} = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma}_{\boldsymbol{x}}\bar{\boldsymbol{\Phi}}^H$ from previous sections. We observe that $\boldsymbol{\Sigma}_{\bar{\boldsymbol{y}}}(\omega)$ is the result of compressing the circulant matrix $\boldsymbol{\Sigma}_{\bar{\boldsymbol{x}}}(\omega)$. A possible means of estimating the second-order statistics of $x[l]$ is, for example, by using sample statistics to estimate $\boldsymbol{\Sigma}_{\bar{\boldsymbol{y}}}(\omega)$, reconstruct $\boldsymbol{\Sigma}_{\bar{\boldsymbol{x}}}(\omega)$ using least squares, and finally recover $\boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega)$ from (39) [46].

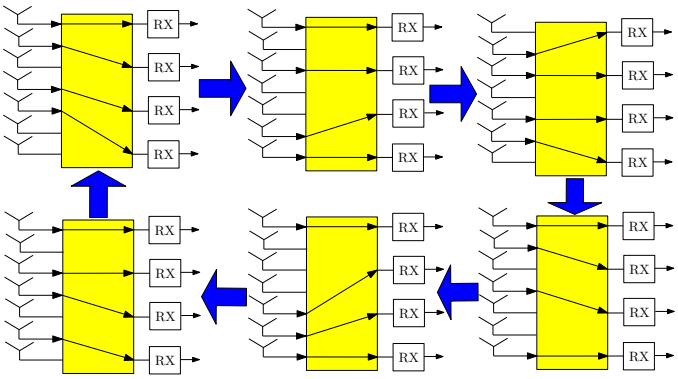


Fig. 10: Implementation of dynamic spatial sampling using antenna switching.

D. Cooperative CCS

As mentioned in Sec. VII, in certain cases multiple sensors are used to observe a time signal in multiple spatial locations, which can result in improved convergence of the sample statistics [47]. Here we show that this setting can also be used to introduce strong compression.

Suppose that a collection of sensors are deployed across a certain area in order to estimate the second-order statistics of a certain WSS time signal $x(t)$. Although different sensors observe different signal values, we can assume that the second-order statistics of the received signals are approximately the same for all sensors. This is the case, for example, if the channels from each signal source to all sensors (possibly after passing through an automatic gain control) have approximately the same statistics [21]. As before, let us collect those statistics in the Toeplitz covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$.

We now describe a particularly interesting case where the sensors use multi-coset sampling. To do so, recall from Sec. II-A that, in the single sensor case, $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ can be reconstructed from the covariance matrix of the compressed observations $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ if all the entries of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ show up at least once in $\boldsymbol{\Sigma}_{\boldsymbol{y}}$. In the cooperative scenario, a milder condition may be imposed by capitalizing on the availability of multiple sensors.

Let us form Z groups of sensors by arranging together all the sensors that share the same multi-coset sampling pattern. The sought condition can be given in terms of the matrices $\boldsymbol{\Sigma}_{\boldsymbol{y},z}$, $z = 0, \ldots, Z-1$, where $\boldsymbol{\Sigma}_{\boldsymbol{y},z}$ represents the covariance matrix of the compressed observations at the sensors within the z-th group. The requirement now is that, in order to reconstruct $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, every entry of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is only required to show up in at least one of the matrices $\{\boldsymbol{\Sigma}_{\boldsymbol{y},z}\}_{z=0}^{Z-1}$. This observation yields great compression improvements per sensor, as the sampling burden is now distributed across sensors.

To illustrate this effect, suppose that $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is such that, in the non-cooperative scenario, the optimum compression pattern $\mathcal{M}$ for each block is a circular sparse ruler (see Table I). In the

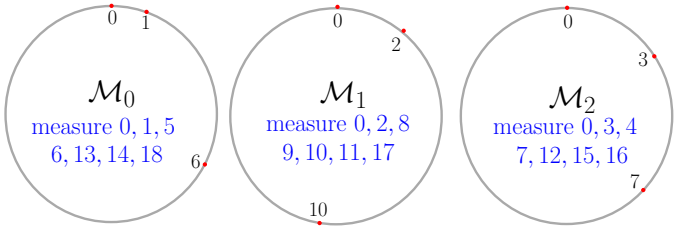


Fig. 11: Incomplete circular sparse rulers used in a setting with $Z = 3$ groups of sensors. The correlation lags that the sensors in each group measure are listed inside each circumference.

cooperative setting, let $\mathcal{M}_z$ denote the multi-coset sampling pattern used by all sensors in group z and let $\Omega(\mathcal{M}_z)$ represent the set containing all modular differences between elements of $\mathcal{M}_z$:

$$\Omega(\mathcal{M}_z) = \{(m - m') \bmod N \ : \ m, m' \in \mathcal{M}_z\} \tag{40}$$

It can be shown that a collection of sampling patterns $\{\mathcal{M}_z\}_{z=0}^{Z-1}$ ensures the identifiability of $\mathbf{\Sigma}_{\boldsymbol{x}}$ if and only if [21]

$$\bigcup_{z=0}^{Z-1} \Omega(\mathcal{M}_z) = \{0, 1, \ldots, N-1\}. \tag{41}$$

Clearly, for $Z = 1$ this condition reduces to the non-cooperative condition, which requires $\mathcal{M}_0$ to be a circular sparse ruler. Each $\mathcal{M}_z$, $z = 0, \ldots, Z-1$, is called an *incomplete* circular sparse ruler since it does not contain all possible differences between 0 and $N-1$ (see Box 2). However, (41) clearly implies that, for every given integer modular distance $n \in \{0, 1, \ldots, N-1\}$, *at least one* of those incomplete circular sparse rulers can measure n. An example of collection of incomplete circular sparse rulers is the one composed of the sets $\mathcal{M}_0 = \{0, 1, 6\}$, $\mathcal{M}_1 = \{0, 2, 10\}$, and $\mathcal{M}_2 = \{0, 3, 7\}$, represented geometrically in Fig. 11. Observe that, as in the case of circular sparse rulers, each mark provides two distances, one clockwise and the other counterclockwise.

The next question is how to minimize the overall compression ratio. The idea is therefore to minimize the number of marks in each ruler while satisfying (41). This task is intimately connected to the so-called non-overlapping circular Golomb rulers [21].

Alternative schemes for cooperative CCS include [48], which exploits the cross-correlation between observations at different sensors, and [37], where the observations are not only linearly compressed but also quantized to a single bit.

IX. Open Lines

Despite the long history of structured covariance estimation and recent excitement on compressed sensing of sparse signals, the research on CCS is still at an early stage. A great deal of research is required to improve its applicability and theoretical understanding. Some possible future directions are listed in this section.

As for sampler design, most existing schemes rely on identifiability criteria [8], [20], but other criteria are yet to be explored. For instance, it would be important to find sampler designs minimizing the Cramér-Rao bound for unbiased estimation of the parameters of interest. Of special relevance would be deterministic schemes with the strongest possible compression yet capable of reaching a target performance. Other sampling schemes, say gridless or continuous irregular sampling, are yet to be investigated from a CCS perspective. Here, we envision a gridless or continuous observation space where samples can be drawn from, and the focus of CCS is on the reconstruction of second-order statistics regardless of sparsity. This problem differs from the existing literature on gridless or continuous sparse reconstruction, which aims to accurately recover sparse input signals with nonzeros that lie anywhere in the continuous input domain.

Cooperative schemes deserve more extensive research. For instance, distributed implementations and data fusion techniques need to be revisited for CCS with affordable communication overhead [37]. The latter include schemes where sensors quantize their observations before reporting them to the fusion center. In this context, either the correlations or the raw data can be quantized before taking the correlations. This latter technique is possible since under some conditions the correlation function of the original raw data can be computed from the correlation function of the quantized data.

CCS may also be of critical relevance in big data analytics due to its ability to meaningfully reduce the dimension of the data set. In this context, online, adaptive and distributed implementations are yet to be devised. Moreover, as more and more big-data applications employ a network of high-dimensional signals for data mining and exploration, it is an interesting new direction to see how the CCS framework benefits the covariance estimation problems for data-starved inference networks. Such problems arise under the umbrella of probabilistic analysis for high dimensional datasets with many variables and few samples. As a precursor, sparse (inverse) covariance estimation has already become a widely pursued topic in statistical inference for analysis on graphs, where the sparsity of the (inverse) covariance matrix is exploited, in the context of correlation mining. When high-dimensional or wideband random processes are concerned, CCS has been applied for covariance estimation based on the exploitation of various structures in the data: Gaussianity, stationarity and compression [49]. Fruitful exploration along this direction may lead to CCS for inference networks, which will find broad applications in analyzing astronomical data, network data, biomedical diagnostics and video imaging, to name a few.

Finally, we highlight the relevance of extending the reviewed techniques to non-stationary process analysis, for instance exploiting the framework of underspread processes [50]. Future research may also consider non-linear parameterizations as well as non-linear compression.

X. Conclusions

This article presented a renewed perspective on a traditional topic in signal processing, which we dubbed CCS. We introduced a joint signal acquisition and compression framework for a number of applications and problems that deal with second-order statistics. The basic principle underlying CCS is that the desired signal statistics can be reconstructed directly

from properly compressed observations, without having to recover the original signal itself that can be costly in terms of both computational and sensing resources. This standpoint entails multiple benefits such as the possibility of introducing stronger compression beyond sparsity, compared to traditional compressed sensing.

REFERENCES

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

[2] Y. Jin and B. D. Rao, "Support recovery of sparse signals in the presence of multiple measurement vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3139–3157, May 2013.

[3] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep 2000.

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

[5] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

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

[7] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[8] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proc. IEEE*, vol. 73, no. 10, pp. 1522–1524, 1985.

[9] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2004.

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

[11] L. L. Scharf, *Statistical signal processing: detection, estimation, and time series analysis*, vol. 1, Addison-Wesley, 1991.

[12] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, "Difference bases and sparse sensor arrays," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 716–721, 1993.

[13] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Sig. Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[14] C. P. Yen, Y. Tsai, and X. Wang, "Wideband spectrum sensing based on sub-Nyquist sampling," *IEEE Trans. Sig. Process.*, vol. 61, no. 12, pp. 3028–3040, 2013.

[15] M.A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, May 2011, pp. 3884–3887.

[16] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Design and analysis of multi-coset arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, 2013.

[17] D. D. Arianananda, D. Romero, and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, San Martin, France, Dec 2013, pp. 440–443.

[18] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. American Math. Soc.*, vol. 43, no. 3, pp. 377–385, 1938.

[19] D. D. Arianananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Sig. Process.*, vol. 60, no. 9, pp. 4775–4789, 2012.

[20] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1410–1425, Mar. 2015.

[21] D. D. Arianananda, D. Romero, and G. Leus, "Cooperative compressive power spectrum estimation," in *Proc. of IEEE Sensor Array Multichannel Sig. Process. Workshop*, A Corunha, Spain, Jun. 2014, pp. 97–100.

[22] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Sig. Process.*, vol. 8, no. 3, pp. 185–210, 1998.

[23] V. Venkateswaran and A. J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Sig. Process.*, vol. 58, no. 8, pp. 4131–4143, 2010.

[24] G. Leus and D. D. Arianananda, "Power spectrum blind sampling," *IEEE Sig. Process. Lett.*, vol. 18, no. 8, pp. 443–446, 2011.

[25] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, Apr. 2007, vol. 4, pp. IV–1357–IV–1360.

[26] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Topics Sig. Process.*, vol. 6, no. 1, pp. 58–69, Feb. 2012.

[27] W. C. Black and D. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.

[28] S. Becker, *Practical Compressed Sensing: Modern Data Acquisition and Signal Processing*, Ph.D. thesis, California Institute of Technology, 2011.

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

[30] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Sig. Process.*, vol. 61, no. 24, pp. 6232–6246, 2013.

[31] G. Vázquez-Vilar and R. López-Valcarce, "Spectrum sensing exploiting guard bands and weak channels," *IEEE Trans. Sig. Process.*, vol. 59, no. 12, pp. 6045–6057, 2011.

[32] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Sig. Process.*, vol. 59, no. 2, pp. 573–586, 2011.

[33] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. Part I: Fully augmentable arrays," *IEEE Trans. Sig. Process.*, vol. 46, no. 9, pp. 2458–2471, 1998.

[34] S. Shakeri, D. D. Arianananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Int. Workshop Sig. Process. Advances Wireless Commun.*, Jun. 2012, pp. 525–529.

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

[36] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Process.*, vol. 92, no. 7, pp. 1580–1590, 2012.

[37] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *IEEE Trans. Sig. Process.*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[38] Y. Wang, A. Pandharipande, and G. Leus, "Compressive sampling based MVDR spectrum sensing," in *Proc. Cognitive Inf. Process.*, Jun. 2010, pp. 333–337.

[39] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, "DOA estimation of quasi-stationary signals via Khatri-Rao subspace," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, 2009, pp. 2165–2168.

[40] D. D. Arianananda and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Sig. Processing*, vol. 93, no. 12, pp. 3435–3448, 2013.

[41] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Sig. Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[42] P. Pal and P.P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Sig. Process.*, vol. 63, no. 3, pp. 711–726, Feb 2015.

[43] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, Dec. 2011, pp. 337–340.

[44] D. D. Arianananda and G. Leus, "Non-uniform sampling for compressive cyclic spectrum reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.*, May 2014, pp. 41–45.

[45] E. BouDaher, F. Ahmad, and M. G. Amin, "Sparse reconstruction for direction-of-arrival estimation using multi-frequency co-prime arrays," *EURASIP J. Advances Sig. Process.*, vol. 2014, no. 1, pp. 168, 2014.

[46] D. D. Arianananda, D. Romero, and G. Leus, "Compressive periodogram reconstruction using uniform binning," *IEEE Trans. Sig. Process.*, vol. 63, no. 16, pp. 4149–4164, Aug. 2015.

[47] R. J. Kozick and B. M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," *IEEE Trans. Sig. Process.*, vol. 52, no. 3, pp. 601–616, March 2004.

[48] D. D. Arianananda and G. Leus, "Cooperative compressive wideband power spectrum sensing," in *Proc. Asilomar Conf. Sig., Syst., Comput.*, Nov. 2012, pp. 303–307.

[49] A. O. Nasif, Z. Tian, and Q. Ling, "High-dimensional sparse covariance estimation for random signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Process.* IEEE, 2013, pp. 4658–4662.

[50] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1067–1086, 2006.