

The background is a vibrant blue with a complex, abstract pattern. It features multiple layers of wavy, ribbon-like structures that resemble binary code (0s and 1s) or digital data streams. These ribbons curve and flow across the frame, creating a sense of depth and movement. The overall effect is a futuristic, high-tech aesthetic.

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

Compressive Covariance Sensing

[Structure-based compressive sensing
beyond sparsity]

Digital Object Identifier 10.1109/MSP.2015.2486805
Date of publication: 29 December 2015

Compressed sensing deals with the reconstruction of signals from sub-Nyquist samples by exploiting the sparsity of their projections onto known subspaces. In contrast, this 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.

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 allow 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 (WSS) 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, for example [9]. Although reconstruction of second-order statistics from compressed observations dates back several decades (see, e.g., [6] and the references therein), the recent interest in compressive sensing and reconstruction has propelled numerous advances in this context.

The purpose of this article is to provide a fresh look at the recent contributions in this exciting area, which is referred to as *compressive covariance sensing* (CCS). Admittedly, a straightforward approach to reconstruct second-order statistics is to apply an estimation method over a waveform uncompressed via a non-CCS procedure. However, it is not difficult to see that such a 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.

SAMPLING SECOND-ORDER STATISTICS

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

A WARM-UP EXAMPLE

Suppose that a uniform linear array (ULA) with L antennas, such as the one in Figure 1(a), observes T snapshots of a zero-mean spatial signal whose complex baseband representation is given by $\mathbf{x}_\tau \in \mathbb{C}^L$, $\tau = 0, 1, \dots, T-1$. Many array processing algorithms rely on estimates of the so-called spatial covariance matrix $\Sigma_x := E\{\mathbf{x}_\tau \mathbf{x}_\tau^H\}$ to form images or to obtain information such as the bearing of certain sources [6], [9]. A straightforward estimate of Σ_x is the sample covariance matrix, given by

$$\hat{\Sigma}_x = \frac{1}{T} \sum_{\tau=0}^{T-1} \mathbf{x}_\tau \mathbf{x}_\tau^H. \quad (1)$$

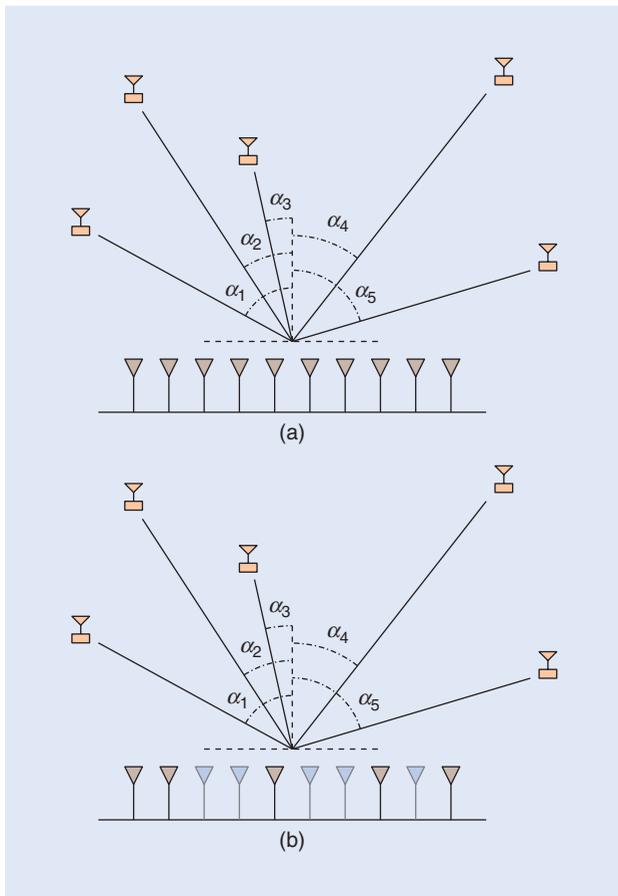
If the impinging signals are generated by uncorrelated point sources in the far field [see Figure 1(a)], the matrix Σ_x exhibits a Toeplitz structure (see the section “Modal Analysis”), meaning that the coefficients are constant along the diagonals. Thus, one may represent the (m, n) th entry of Σ_x by

$$\sigma[m-n] = E\{x_\tau[m]x_\tau^*[n]\}, \quad (2)$$

where $x_\tau[m]$ represents the m th entry of \mathbf{x}_τ . Noting that Σ_x is also Hermitian reveals that all the information is contained in the coefficients $\sigma[l]$, $l = 0, \dots, L-1$. These observations suggest the possibility of constructing estimators with improved performance [10], [11]; simply consider replacing the elements on each diagonal of $\hat{\Sigma}_x$ 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 to improve the estimation performance, the described structure can also be exploited to reduce the number of antennas required to estimate Σ_x (see, e.g., [6]–[8]). Suppose, in particular, that only a subset of the antennas in the ULA is used to sample the spatial field of interest, the others being disconnected [see Figure 1(b)].

Let the set $\mathcal{K} := \{k_0, \dots, k_{K-1}\}$ 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 $\mathbf{y}_\tau = [x_\tau[k_0], \dots, x_\tau[k_{K-1}]]^T$. The (i, j) th entry (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) of the covariance matrix $\Sigma_y := E\{\mathbf{y}_\tau \mathbf{y}_\tau^H\}$ is, therefore,



[FIG1] (a) An uncompressed ULA with ten antennas receiving the signals from five sources in the far field. (b) A compressed array with five antennas. The five antennas marked in light gray were removed, but the achievable spatial resolution remains the same.

$$E\{y_\tau[i]y_\tau^*[j]\} = E\{x_\tau[k_i]x_\tau^*[k_j]\} = \sigma[k_i - k_j]. \quad (3)$$

Thus, Σ_y is made up of a subset of the entries of Σ_x . It is clear that Σ_x can be reconstructed from a sample estimate of Σ_y if all the entries of the former show up at least once in the latter.

LINEAR SPARSE RULERS

A set $\mathcal{K} \subset \{0, \dots, L-1\}$ is a length- $(L-1)$ (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 its marks, a sparse ruler is still able to measure any integer distance between zero and $L-1$. In Figure S1, we observe that the ruler $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is capable of measuring any object of length five by using the marks three and eight.

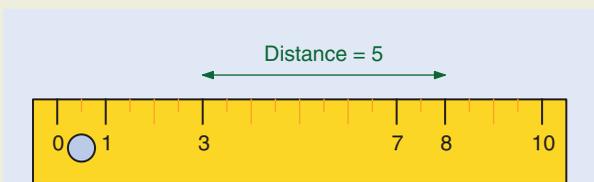
A length- $(L-1)$ minimal sparse ruler is a length- $(L-1)$ sparse ruler \mathcal{K} with minimum number of elements $|\mathcal{K}|$. The set $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is a length-10 minimal linear sparse ruler

From (3), this means that, for every $l = 0, \dots, L-1$, 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, if they contain a minimum number of elements, they are termed *minimal sparse rulers*, as explained in “Linear Sparse Rulers.” In Figure 1(b), 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 Figure 1(a).

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

An intimately related concept is the 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 allows cost savings beyond the costs associated with the antennas themselves: radio-frequency (RF) equipment, such as filters, mixers, and analog-to-digital converters (ADCs), needs to be deployed only for the active antennas. Moreover, 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 subarray 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 the $|\mathcal{K}|$, the larger the number of snapshots T required to attain a target performance. Hence, for signals defined in 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.



[FIGS1] A sparse ruler can be thought of as a ruler with a part of its marks erased, but the remaining length marks allow all integer distances between zero and its length to be measured.

since it has six elements and there exists no length-10 sparse ruler with five or fewer elements.

IMPORTANCE OF COVARIANCE STRUCTURES

In the previous example, the Hermitian Toeplitz structure of Σ_x allowed us to recover the second-order statistics of x_τ from those of its compressed version y_τ . More generally, it is expected that our ability to compress a signal while preserving the second-order statistical information depends on the structure of Σ_x . In other words, we expect that the more structured Σ_x is, the stronger the compression on x_τ it may induce.

In certain applications, such as power-spectrum estimation for 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 requirements on \mathcal{K} , which is no longer required to be a linear sparse ruler but a circular one; see “Circular Sparse Rulers” for a definition.

Because of 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 to be a length- $(L - 1)$ circular sparse ruler than 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 suboptimal designs. Immediate choices are length- $(L - 1)$ or length- $\lfloor L/2 \rfloor$ minimal linear sparse rulers [19]. In fact, the latter provide optimal solutions for most values of L below 60 [20].

Aside from Toeplitz and circulant, another common structure is the one present in those applications where the covariance matrix is known to be banded [19]. A type of Toeplitz matrix, d -banded matrices satisfy $\sigma[l] = 0$ for all $l > d$ and arise in those cases where we sample a WSS time signal whose autocorrelation sequence $\sigma[l]$ vanishes after d lags. Sampling patterns for

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.

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].

These typical covariance structures, including Toeplitz, circulant, and banded, are illustrated in Figure 2, along with their most popular applications. Generally speaking, in many cases including the

previous ones, prior knowledge constrains covariance matrices to be linear combinations of certain known matrices, say $\{\Sigma_i\}_i$. In other words, there must exist coefficients α_i such that

$$\Sigma_x = \sum_{i=0}^{S-1} \alpha_i \Sigma_i. \quad (4)$$

Without any loss of generality, we may assume that the scalars α_i are real [20] and the matrices Σ_i are linearly independent. Thus, $S = \{\Sigma_0, \dots, \Sigma_{S-1}\}$ is a basis, and S represents the dimension of the model. This expansion encompasses all the previous examples as particular cases as long as the right set of matrices Σ_i is chosen. 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 Figure 2). As we will see in the section “Optimal Designs,” S is related to how compressible Σ_x is.

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 work flourished before the surge of compressed sensing in signal processing, when the main goal was to design robust and performance-enhanced estimators with a small sample size. CCS offers a new way to exploit covariance structures for joint signal acquisition and compression.

COMPRESSION

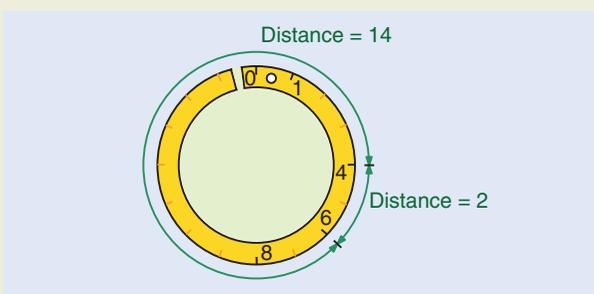
The previous array processing example describes how compression can be exerted for signals acquired in the spatial

CIRCULAR SPARSE RULERS

A set $\mathcal{K} \subset \{0, \dots, L - 1\}$ is a length- $(L - 1)$ 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 $(k - k') \bmod L = l$. 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 Figure S2).

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



[FIGS2] Generally, in a circular sparse ruler, each pair of marks allows two distances to be measured.

domain—only a subset \mathcal{K} of antennas was used to estimate Σ_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 applications 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 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

$$x[l] = x(lT_s), \quad l = 0, \dots, L-1, \quad (5)$$

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.

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 *nonuniform*

sampling, this approach advocates the acquisition of a small number of samples indexed by a subset of the Nyquist grid:

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

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.

As we will soon see, this average rate reduction has led to the technology of compressive ADCs (C-ADCs), conceived to circumvent the aforementioned hardware tradeoffs. Before exploring this topic, let us expand the families of samplers we are about to consider.

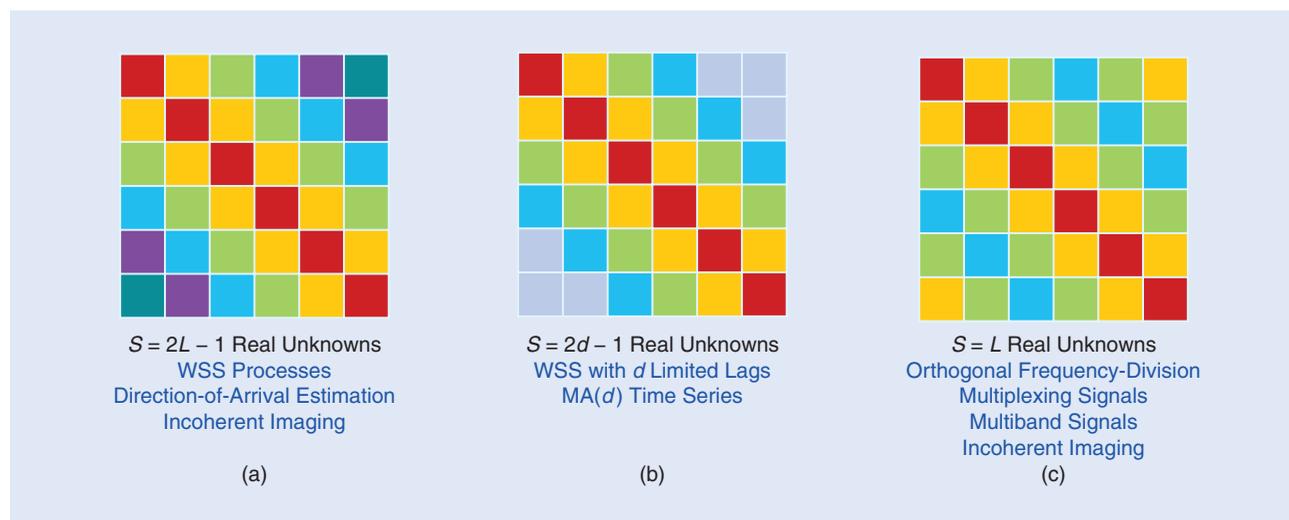
By forming $x = [x[0], \dots, x[L-1]]^T$ and $y = [y[0], \dots, y[K-1]]^T$, the operation in (6)

can be equivalently represented as a row-selection operation

$$y = \tilde{\Phi}x. \quad (7)$$

The matrix $\tilde{\Phi} \in \mathbb{C}^{K \times L}$, 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] and [5]) and in the spatial domain (see, e.g., [23]). In correspondence with this terminology, we talk about dense samplers when $\tilde{\Phi}$ is dense and about sparse samplers when $\tilde{\Phi}$ 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 why we dropped the subscript τ from x and y in (7) when compared to x_τ and y_τ in the previous section. For simplicity, we omit this subscript throughout when possible, keeping in mind that several snapshots may be available.



[FIG2] Some common covariance structures, along with their main applications: (a) Toeplitz, (b) d -banded, and (c) circulant.

[TABLE 1] OPTIMAL DESIGNS AND COMPRESSION RATIOS.

		SPARSE SAMPLERS		DENSE SAMPLERS	
	DESIGN		RATIO	DESIGN	RATIO
TOEPLITZ	\mathcal{K} IS A LENGTH- $(N-1)$ LINEAR SPARSE RULER		$\frac{1}{N}\sqrt{2.435(N-1)}$ $\leq \eta_{\min} \leq$ $\frac{1}{N}\left\lceil\sqrt{3(N-1)}\right\rceil$	$\Phi \in \mathbb{C}^{M \times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \geq \sqrt{\frac{2NB-1}{2B-1}}$	$\sqrt{\frac{\eta_{\min} \approx}{(2B-1)N^2}}$ $\frac{2NB-1}{(2B-1)N^2}$
CIRCULANT	\mathcal{K} IS A LENGTH- $(N-1)$ CIRCULAR SPARSE RULER		$\frac{2+\sqrt{4N-3}}{2N}$ $\leq \eta_{\min} \leq$ $\frac{1}{N}\left\lceil\sqrt{3\left\lfloor\frac{N}{2}\right\rfloor}\right\rceil$	$\Phi \in \mathbb{C}^{M \times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \geq \sqrt{\frac{NB}{2B-1}}$	$\sqrt{\frac{\eta_{\min} \approx}{(2B-1)N}}$ $\frac{B}{(2B-1)N}$
d -BANDED		CASE $N \leq d \leq N(B-1)$:		$\Phi \in \mathbb{C}^{M \times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \geq \sqrt{\frac{2d+1}{2B-1}}$	$\sqrt{\frac{\eta_{\min} \approx}{(2B-1)N^2}}$ $\frac{2d+1}{(2B-1)N^2}$
	\mathcal{K} IS A LENGTH- $(N-1)$ CIRCULAR SPARSE RULER		$\frac{2+\sqrt{4N-3}}{2N}$ $\leq \eta_{\min} \leq$ $\frac{1}{N}\left\lceil\sqrt{3\left\lfloor\frac{N}{2}\right\rfloor}\right\rceil$		

When observation windows for time signals are long, hardware design considerations make it convenient to split a sampling pattern into shorter pieces that 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 x are divided into B blocks of size $N = L/B$ as $x = [x^T[0], \dots, x^T[B-1]]^T$, and each block is compressed individually to produce an output block of size M :

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

It is clear that one can assemble the vector of compressed observations as $y = [y^T[0], \dots, y^T[B-1]]^T$ and the matrix $\tilde{\Phi}$ from (7) as $\tilde{\Phi} = I_B \otimes \Phi$, where \otimes represents the Kronecker product.

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

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

where $\mathcal{M} \subset \{0, \dots, N-1\}$ 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 *multicoreset samplers* [3], the matrix Φ is the result of selecting the rows of I_N indexed by \mathcal{M} .

OPTIMAL DESIGNS

One critical problem in CCS is to design a sampler $\tilde{\Phi}$ that preserves the second-order statistical information, in the sense that it allows reconstruction of the uncompressed covariance matrix from the compressed observations.

Design techniques for sparse and dense samplers hinge on different basic principles. Whereas sparse samplers are designed

based on discrete mathematics (as explained previously), 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.

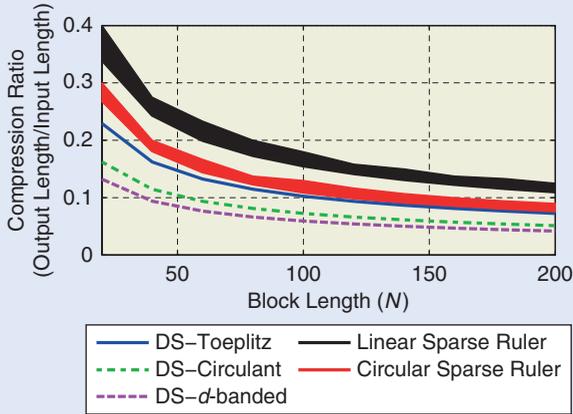
Optimal rates for dense samplers are known in closed form for linear covariance parameterizations such as Toeplitz, circulant, and banded [20]. On the other hand, their evaluation for sparse samplers requires solving combinatorial problems such as the minimal sparse ruler problem. Table 1 summarizes 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 .

The designs from Table 1 are compared in Figure 3. The vertical axis depicts the reduction in the sampling rate that can be achieved in applications such as compressive wideband spectrum sensing [24]–[26], where the spectrum occupancy over a very wide band is decided via power-spectrum estimation. Note that the sampling rate can be reduced considerably, even in the absence of sparsity. 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. Asymptotically for increasing N , sampling rate savings are proportional to $1/\sqrt{N}$ (cf. Table 1).

The superior efficiency of dense samplers over their sparse counterpart also manifests itself in Figure 3. In fact, it can be



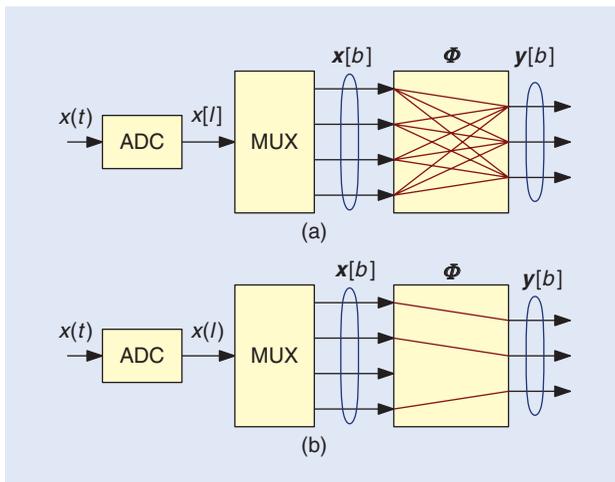
[FIG3] The optimum compression ratios when $B = 10$ and $d = BN/3$ using dense samplers (DSs) for Toeplitz, circulant, and d -banded matrices, and using a linear sparse ruler and a circular sparse ruler. Moderate block lengths yield strong compression.

shown that certain random designs for dense samplers achieve optimal compression ratios with probability one, in the sense that no other sampler (either dense or sparse) can achieve a lower ratio.

TECHNOLOGIES

The acquisition systems that can be used to implement the previously described sampling schemes, which are essentially the same as those used by many other sub-Nyquist acquisition techniques, 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 x contains the Nyquist samples of $x(t)$, their operation can be described by (7) (see Figure 4). 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.



[FIG4] The mathematical model for the operation of a C-ADC: (a) dense sampler and (b) sparse sampler.

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

MAIN APPLICATIONS

The problems that can be formulated in CCS terms are those relying exclusively on the second-order moments of a certain signal 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 $\mathcal{S} = \{\Sigma_0, \dots, \Sigma_{S-1}\}$ to be used [see (4)].

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 the section “Technologies,” this difficulty motivates the usage of C-ADCs, whose 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 Σ_x from y subject to the constraint that Σ_x is 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 Σ_x can be ignored to obtain simple estimators, or it can be enforced 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 $x = \sum_i \sqrt{\alpha_i} x^{(i)}$, where the component $\sqrt{\alpha_i} x^{(i)}$ contains the Nyquist samples of the signal received from the i th transmitter. If $x^{(i)}$ is power normalized, then α_i is the power received from the i th transmitter. Since the second-order statistics of $x^{(i)}$, collected in $\Sigma_i = E\{x^{(i)}(x^{(i)})^H\}$, 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 $x(t)$ 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 $a^{(i)}[l] = e^{j\omega_i l}$ 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 $\Sigma_x = E\{xx^H\}$, which is known to be Hermitian Toeplitz and positive semidefinite [11]. From the observations provided by a C-ADC, one can first reconstruct Σ_x and subsequently apply one of the existing techniques that take Σ_x as the input. To accomplish such reconstruction, one can use (4), with \mathcal{S} being a set spanning the subspace of Hermitian Toeplitz matrices.

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], and [33]), only the antennas corresponding to the nonnull columns of Φ 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, therefore produces a Toeplitz Σ_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 previously for compressive power-spectrum estimation, since the only structure present is that Σ_x 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 exists to this end (see, e.g., [8], [13], [33], and [34]), most of them following the same principles as described previously 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, i.e., sparse and dense sampling, either periodic or nonperiodic. For time signals, periodicity typically arises because of the block-by-block operation of C-ADCs (see, e.g., [19], [30], and [35]); for spatial signals, by consideration of periodic arrays [16], [17].

ESTIMATION AND DETECTION

Having described the model and compression schemes for CCS, we turn our attention to the reconstruction problem. For estimation, it boils down to recovering Σ_x in (4) from the compressive measurements y .

Since $y = \Phi x$, it follows that $\Sigma_y = \Phi \Sigma_x \Phi^H$. If Σ_x is given by (4), Σ_y can be similarly represented as

$$\Sigma_y = \sum_{i=0}^{S-1} \alpha_i \bar{\Sigma}_i, \quad \alpha_i \in \mathbb{R}, \quad (11)$$

where $\bar{\Sigma}_i = \Phi \Sigma_x \Phi^H$. This means that Σ_y and Σ_x share the 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 y .

MAXIMUM LIKELIHOOD

If the probability distribution of the observations is known, one may resort to a maximum-likelihood estimate of Σ_y . For example, if y is zero-mean Gaussian and

$$\hat{\Sigma}_y = \frac{1}{T} \sum_{\tau=0}^{T-1} y_\tau y_\tau^H, \quad (12)$$

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

$$\underset{\{\alpha_i\}}{\text{minimize}} \quad \log |\Sigma_y| + \text{Tr}(\Sigma_y^{-1} \hat{\Sigma}_y) \quad (13)$$

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

LEAST SQUARES

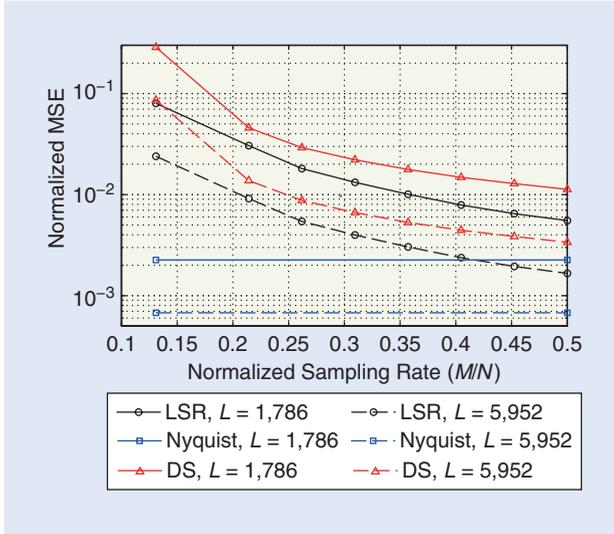
The maximum-likelihood approach involves high computational costs and requires an accurate statistical characterization of the observations. For these reasons, it is customary to rely on geometrical considerations and project the sample covariance matrix onto the span of \mathcal{S} .

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

$$\hat{\Sigma}_x^{\text{LS}} = \text{vec}^{-1} \{ S [(\Phi^* \otimes \Phi) S]^\dagger \hat{\sigma}_y \}. \quad (14)$$

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

Figure 5 illustrates the performance of this technique when Σ_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 to least-squares estimation—it negatively affects any 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 ultimate purpose of



[FIG5] The mean-squared error of the estimate of the least-squares algorithm when Σ_x is 168-banded. (Figure adapted from [19].)

compression, which is to reduce the average sampling rate—a parameter that critically affects the hardware cost.

However, note that this approach does not exploit the fact that Σ_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 $\|\hat{\sigma}_y - (\Phi^* \otimes \Phi)S\hat{\sigma}\|^2$ subject to the constraint $\Sigma_x \geq 0$, which is a convex problem. Other constraints can also be imposed if more prior information is available. For instance, the elements of $\hat{\alpha}$ might be nonnegative [30], in which case one would introduce the constraint $\hat{\alpha} \geq 0$. It can be known that $\hat{\alpha}$ is sparse either by itself or on a linearly transformed domain, in which case one may impose the constraint $\|F_s \hat{\alpha}\|_0 \leq S_0$, where S_0 is the number of nonzero entries and F_s takes $\hat{\alpha}$ to the domain where it is sparse. For instance, the elements of $F_s \hat{\alpha}$ may be samples of the power spectrum [37]. Since the zero-norm in this constraint is not convex, it is typically relaxed to an ℓ_1 -norm. For example, an ℓ_1 -norm regularized least-squares formulation can be adopted as follows:

$$\underset{\hat{\alpha}}{\text{minimize}} \quad \|\hat{\sigma}_y - (\Phi^* \otimes \Phi)S\hat{\alpha}\|^2 + \lambda \|F_s \hat{\alpha}\|_1. \quad (15)$$

In (15), signal compression is induced by the statistical structure of Σ_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).

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, 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:

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

where Σ_r and Σ_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 $y = \Phi x$, whose covariance matrix Σ_y is given by $\Phi \Sigma_w \Phi^H$ under \mathcal{H}_0 and by $\Phi (\Sigma_r + \Sigma_w) \Phi^H$ 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 $p(y; \mathcal{H}_i)$ denotes the density under hypothesis \mathcal{H}_i , this rule decides \mathcal{H}_1 when the ratio $p(y; \mathcal{H}_1)/p(y; \mathcal{H}_0)$ 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 , say α_0 , is positive or zero, while the others are unknown and treated as nuisance parameters [30]. Since in these cases no uniformly most-powerful test exists, one must resort to other classes of detectors, such as the generalized likelihood ratio test, which makes a decision by comparing $p(y; \hat{\alpha}_{\mathcal{H}_1})/p(y; \hat{\alpha}_{\mathcal{H}_0})$ against a threshold, where $\hat{\alpha}_{\mathcal{H}_i}$ is the maximum-likelihood estimate of α under hypothesis \mathcal{H}_i [30].

MODAL ANALYSIS

As mentioned in the section “Main Applications,” 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 the class of sparse spectrum estimation problems, which allow a common formulation as modal analysis [11].

Suppose that the observations are given by

$$x = \sum_{i=0}^{R-1} s_i a^{(i)} + w = As + w, \quad (17)$$

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

The covariance matrix of x is given by

$$\Sigma_x = A \Sigma_s A^H + \sigma_w^2 I_L, \quad (18)$$

where σ_w^2 is the power of the noise process, assumed white for simplicity, and Σ_s is the covariance matrix of s , which is diagonal since the sources are uncorrelated. Note that these assumptions result in Σ_x having a Toeplitz structure.

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

$$\Sigma_y = \bar{A}\Sigma_s\bar{A}^H + \sigma_w^2\bar{\Phi}\bar{\Phi}^H. \quad (19)$$

The parameters ω_i can be estimated from Σ_y using adaptations of traditional techniques such as multiple signal classification [35] and the minimum variance distortionless response algorithm [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 columnwise application of the Kronecker product, as

$$\text{vec}(\Sigma_y) = (\bar{A}^* \circ \bar{A}) \text{diag}\{\Sigma_s\} + \sigma_w^2 \text{vec}(\bar{\Phi}\bar{\Phi}^H). \quad (20)$$

The matrix $\bar{A}^* \circ \bar{A}$ 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 $\bar{A}^* \circ \bar{A}$ contains all the rows in the manifold matrix of a ULA [40]. To obtain this structure, 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{A} using a fine grid of angles ω_i and then estimate s from $y = \bar{A}s$ exploiting the idea that most of its components will be zero since, for a grid fine enough, most of the columns of \bar{A} will correspond to angles where there are no sources. In other words, s is sparse, which means that the techniques from [5] and [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 that operate on (20) instead [42].

PREPROCESSING

Most of the methods described in this article make use of the sample covariance matrix of y , defined in (12). Under general conditions, the average $T^{-1} \sum_{\tau} y_{\tau} y_{\tau}^H$ converges to the true Σ_y as T becomes large. If compression does not destroy relevant second-order statistical information, the matrix Σ_y contains all the information required to identify all entries of Σ_x , but a considerably large number T of snapshots will be required for $\hat{\Sigma}_y$ to be close to Σ_y , which is necessary to obtain a reasonable estimate of Σ_x .

Typically, in those applications involving spatial signals, the outputs of all antennas are synchronously sampled. If y_{τ} collects the samples acquired at time instant τ , it is clear that multiple observations of y can be obtained by considering successive snapshots $\tau = 0, 1, \dots, T-1$. This means that, whereas y contains samples across space, the different snapshots are acquired along the time dimension. Conversely, in applications involving time-domain signals, y contains samples acquired over time. A possible means to observe multiple realizations is by considering the vectors y_{τ} observed at different locations $\tau = 0, 1, \dots, T-1$. In this case, while 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.

THE MEAN SQUARED ERROR OF THE ESTIMATE IS LARGER WHEN COMPRESSION IS INTRODUCED SINCE IT REDUCES THE TOTAL NUMBER OF SAMPLES.

Unfortunately, many applications do not allow averaging over the dual domain, and one must cope with a single observation, say y_0 , producing the estimate $\hat{\Sigma}_y = y_0 y_0^H$. This matrix is not a satisfactory estimate of Σ_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 Σ_x to obtain a more convenient estimate [30]. In particular, because of the block-by-block operation described by (8), the fact that Σ_x is Toeplitz means that Σ_y is block Toeplitz; that is, it can be written as

$$\Sigma_y = \begin{bmatrix} \Sigma_y[0] & \Sigma_y[-1] & \cdots & \Sigma_y[-B+1] \\ \Sigma_y[1] & \Sigma_y[0] & \cdots & \Sigma_y[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_y[B-1] & \Sigma_y[B-2] & \cdots & \Sigma_y[0] \end{bmatrix}, \quad (21)$$

where the (nonnecessarily Toeplitz) $M \times M$ blocks $\Sigma_y[k]$ are given by

$$\Sigma_y[k] = E\{y[b]y^H[b-k]\}, \quad \forall b. \quad (22)$$

This suggests the estimate

$$\hat{\Sigma}_y[k] = \frac{1}{\text{number of terms}} \sum_b y[b]y^H[b-k]. \quad (23)$$

Moreover, since Σ_y is Hermitian, this computation needs only to be carried out for $k = 0, \dots, B-1$. More sophisticated estimates exhibiting different properties were analyzed in [30].

Another observation is that the smaller k , the higher the quality of the estimates of $\Sigma_y[k]$. 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

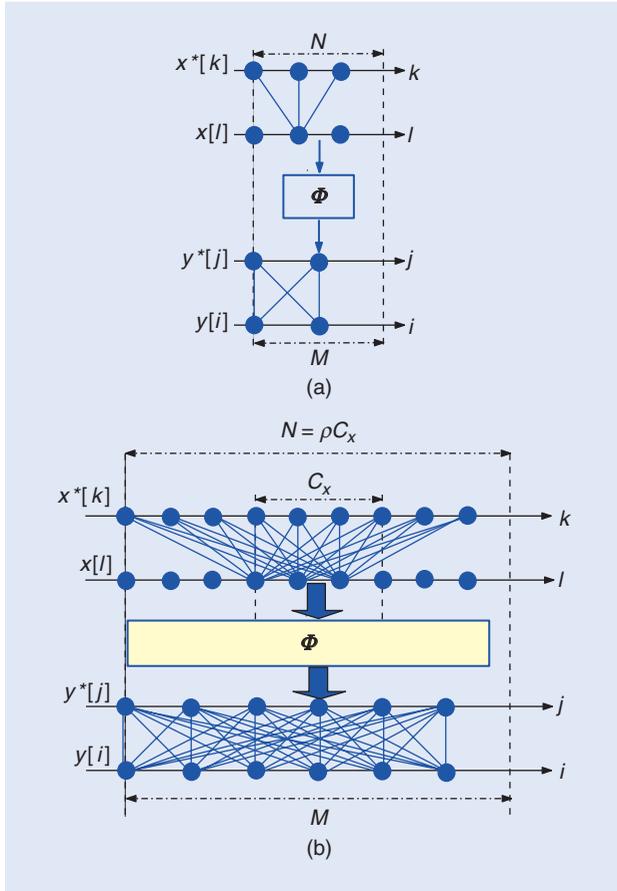
$$\Sigma_y = \begin{bmatrix} \Sigma_y[0] & \Sigma_y[-1] & \cdots & \Sigma_y[-\tilde{B}+1] \\ \Sigma_y[1] & \Sigma_y[0] & \cdots & \Sigma_y[-\tilde{B}+2] \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_y[\tilde{B}-1] & \Sigma_y[\tilde{B}-2] & \cdots & \Sigma_y[0] \end{bmatrix}, \quad (24)$$

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

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.



[FIG6] (a) A compression of a block of $N = 3$ samples of a WSS signal using a 2×3 matrix Φ , which produces compressed blocks of $M = 2$ samples. (b) A compression of a block of $N = 9$ samples of a cyclostationary signal, which produces compressed blocks of $M = 6$ samples.

CYCLOSTATIONARITY

Cyclostationarity is exhibited in many man-made signals with inherent periodicity, which is a useful feature for estimation, detection, and classification of digital communication signals [26]. Although there are several methods to reconstruct the second-order statistics of a cyclostationary signal from compressed observations (see, e.g., [26], [43], and [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 $\sigma[l, k] = E\{x[l]x^*[l-k]\}$, and it is said to be periodic when there exists an integer C_x , called the *cyclic period*, such that $\sigma[l+n_c C_x, k] = \sigma[l, k]$ 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 $x[b]$ can be divided into ρ subblocks of length C_x as

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

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

$$\Sigma_x = \begin{bmatrix} \Sigma_x[0] & \Sigma_x[-1] & \dots & \Sigma_x[-B+1] \\ \Sigma_x[1] & \Sigma_x[0] & \dots & \Sigma_x[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_x[B-1] & \Sigma_x[B-2] & \dots & \Sigma_x[0] \end{bmatrix}, \quad (26)$$

where the blocks $\Sigma_x[b]$ also have a block Toeplitz structure with blocks $\Sigma_{\tilde{x}}[\rho] = E\{\tilde{x}[\rho']\tilde{x}^H[\rho'-\rho]\}$:

$$\Sigma_x[b] = \begin{bmatrix} \Sigma_{\tilde{x}}[b\rho] & \dots & \Sigma_{\tilde{x}}[b\rho-\rho+1] \\ \Sigma_{\tilde{x}}[b\rho+1] & \dots & \Sigma_{\tilde{x}}[b\rho-\rho+2] \\ \vdots & \ddots & \vdots \\ \Sigma_{\tilde{x}}[b\rho+\rho-1] & \dots & \Sigma_{\tilde{x}}[b\rho] \end{bmatrix}. \quad (27)$$

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

Figure 6(a) 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 k -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 Figure 6(a) for the compressed sequence $y[i]$, 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 three to two after compression, the number of different correlation values has increased from three to four. This means that, whereas one cannot reconstruct the samples of $x[l]$ from $y[i]$ without further assumptions, there is a chance of reconstructing the second-order statistics of $x[l]$ from those of $y[i]$. In fact, if Φ satisfies certain conditions, one can, for instance, estimate Σ_y from $y[i]$ using sample statistics and

obtain an estimate of Σ_x via least squares, as described in the section “Least Squares.”

Now assume that $x[l]$ is a cyclostationary signal of cyclic period $C_x = 3$ and that $\sigma[l, k]$ is such that each subblock of C_x samples is only correlated with the neighboring subblocks. Figure 6(b) 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 Σ_y [c.f. (21)] can be written as

$$\Sigma_y[b] = \Phi \Sigma_x[b] \Phi^H. \quad (28)$$

To exploit the block Toeplitz structure of $\Sigma_x[b]$ [see (27)], we first vectorize both sides of (28) and apply the properties of the Kronecker product to obtain

$$\text{vec}(\Sigma_y[b]) = (\Phi^* \otimes \Phi) \text{vec}(\Sigma_x[b]). \quad (29)$$

Now we rewrite the rightmost vector of this expression as

$$\text{vec}(\Sigma_x[b]) = \mathbf{T} \boldsymbol{\beta}_x[b], \quad (30)$$

where

$$\begin{aligned} \boldsymbol{\beta}_x[b] = & [\text{vec}^T(\Sigma_x[b\rho]), \text{vec}^T(\Sigma_x[b\rho+1]), \dots, \\ & \text{vec}^T(\Sigma_x[b\rho+\rho-1]), \text{vec}^T(\Sigma_x[b\rho-\rho+1]), \\ & \dots, \text{vec}^T(\Sigma_x[b\rho-1])]^T \end{aligned} \quad (31)$$

is a $(2\rho-1)C_x^2 \times 1$ vector containing all the possibly distinct entries of $\Sigma_x[b]$, 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 $\text{vec}(\Sigma_x[b])$. Substituting (30) in (29) yields

$$\text{vec}(\Sigma_y[b]) = (\Phi^* \otimes \Phi) \mathbf{T} \boldsymbol{\beta}_x[b]. \quad (32)$$

Note that, while Φ generally has more columns than rows (as $M < N$), the $M^2 \times (2\rho-1)C_x^2$ matrix $(\Phi^* \otimes \Phi) \mathbf{T}$ 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 $\Sigma_y[b]$ 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 Σ_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 Φ with a block diagonal structure.

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 in the section “A Warm-Up Example,” where the Toeplitz structure of Σ_x allowed us to estimate Σ_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 Σ_x is Hermitian and positive semidefinite, which means that Σ_x cannot be estimated with fewer than N antennas.

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 need to be deployed. This is illustrated in Figure 7, 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.

To estimate Σ_x , the least-squares method from previous sections can be used. Let $\bar{\Phi}_g$ 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

$$\Sigma_{y_g} = \bar{\Phi}_g \Sigma_x \bar{\Phi}_g^H. \quad (33)$$

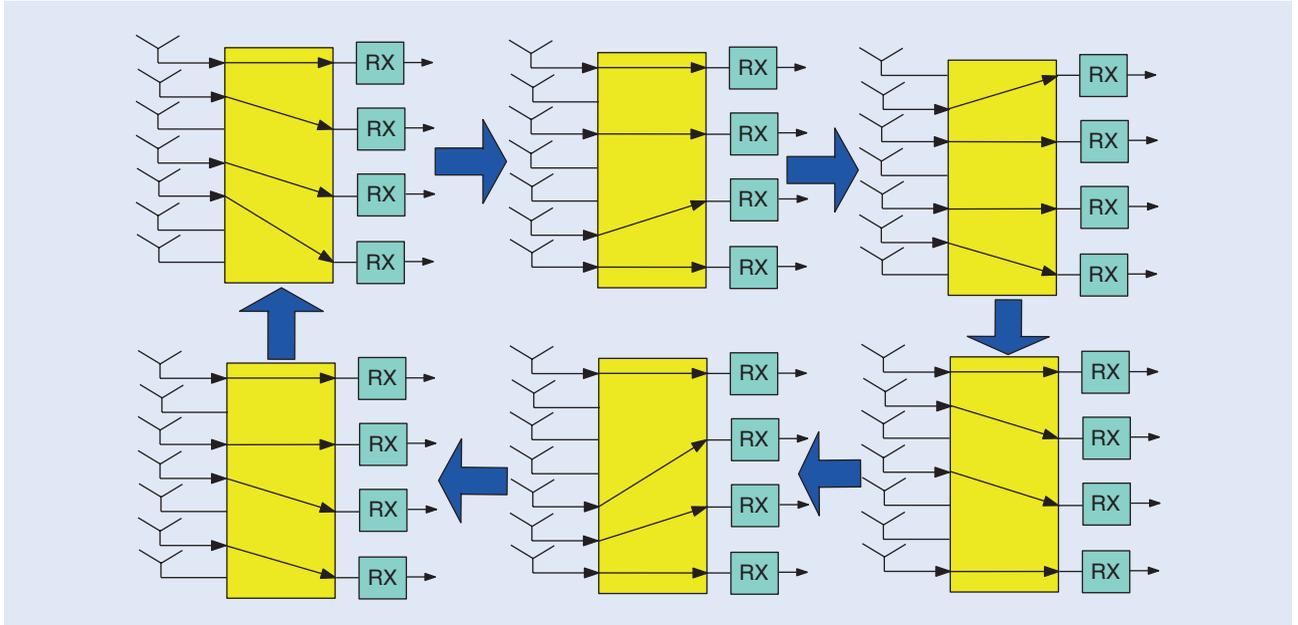
Vectorizing both sides and combining the result for the G time slots in each period yields

$$\begin{bmatrix} \text{vec}(\Sigma_{y_0}) \\ \text{vec}(\Sigma_{y_1}) \\ \vdots \\ \text{vec}(\Sigma_{y_{G-1}}) \end{bmatrix} = \begin{bmatrix} \bar{\Phi}_0^* \otimes \bar{\Phi}_0 \\ \bar{\Phi}_1^* \otimes \bar{\Phi}_1 \\ \vdots \\ \bar{\Phi}_{G-1}^* \otimes \bar{\Phi}_{G-1} \end{bmatrix} \text{vec}(\Sigma_x) = \Psi \text{vec}(\Sigma_x). \quad (34)$$

If the $GK^2 \times L^2$ matrix Ψ has full column rank, then it is possible to estimate Σ_{y_g} , $g = 0, \dots, G-1$ using sample statistics and then obtain an estimate of Σ_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]. To estimate Σ_{y_g} via sample statistics, one may simply average over the observations in the g th time slot of each period.

COMPRESSIVE COVARIANCE ESTIMATION OF MULTIBAND SIGNALS

When uncorrelated signal sources are concerned, a multiband 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



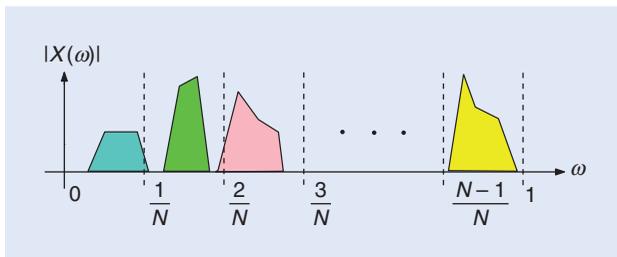
[FIG7] Implementation of dynamic spatial sampling using antenna switching.

has a multiband structure in the frequency (angular) domain (see Figure 8) [14], [35], [46]. For simplicity, consider a time-domain signal $x(t)$, although the discussion immediately carries over to the spatial domain [46]. We show how this problem can be cast as the problem of compressing a circulant covariance matrix (see the section “Importance of Covariance Structures”).

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 $x[l]$, $l = 0, \dots, L - 1$. Let us also split the frequency axis $\omega \in [0, 1)$ into N bins of size $1/N$ (see Figure 8) and introduce, for $\omega \in [0, 1/N)$, the $N \times 1$ vector $\bar{x}(\omega) = [X(\omega), X(\omega + 1/N), \dots, X(\omega + N - 1/N)]^T$.

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

$$Y = \bar{\Phi}X. \quad (35)$$



[FIG8] An example of a signal with a multiband structure. Here, the digital frequency axis ω is split into N uniform bins.

Let us form the $N \times 1$ vector $\bar{x}(\omega)$, whose n th entry contains the DTFT of the n th row of X . Note that the collection of samples in each row of 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 $\bar{x}(\omega)$ in the frequency interval $\omega \in [0, 1/N)$. Likewise, define the $M \times 1$ vector $\bar{y}(\omega)$, $\omega \in [0, 1/N)$, as the vector containing the DTFTs of the rows of Y . Clearly, (35) can then be expressed in the frequency domain using these vectors:

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

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

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

where F_N is the $N \times N$ discrete Fourier transform (DFT) matrix. From (36) and (37), it follows that

$$\Sigma_{\bar{y}}(\omega) = E[\bar{y}(\omega)\bar{y}^H(\omega)] = \Phi \Sigma_{\bar{x}}(\omega) \Phi^H, \quad (38)$$

and

$$\Sigma_{\bar{x}}(\omega) = E[\bar{x}(\omega)\bar{x}^H(\omega)] = \frac{1}{N^2} F_N^H \Sigma_x(\omega) F_N, \quad (39)$$

where $\Sigma_x(\omega) = E[x(\omega)x^H(\omega)]$. 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 $\Sigma_x(\omega)$ 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 F_N is a DFT matrix, it implies a circulant structure in $\Sigma_x(\omega)$.

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

COOPERATIVE CCS

As mentioned in the section “Pre-processing,” 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 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 Σ_x .

We now describe a particularly interesting case where the sensors use multicoset sampling. To do so, recall from the section “A Warm-Up Example” that, in the single-sensor case, Σ_x can be reconstructed from the covariance matrix of the compressed observations Σ_y if all the entries of Σ_x show up at least once in Σ_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 multicoset sampling pattern. The sought condition can be given in terms of the matrices $\Sigma_{y,z}$, $z = 0, \dots, Z-1$, where $\Sigma_{y,z}$ represents the covariance matrix of the compressed observations at the sensors within the z th group. The requirement now is that, to reconstruct Σ_x , every entry of Σ_x is only required to show up in at least one of the matrices $\{\Sigma_{y,z}\}_{z=0}^{Z-1}$. This observation yields great compression improvements per sensor, as the sampling burden is now distributed across sensors.

To illustrate this effect, suppose that Σ_x is such that, in the noncooperative scenario, the optimum compression pattern \mathcal{M} for each block is a circular sparse ruler (see Table 1). In the cooperative setting, let \mathcal{M}_z denote the multicoset 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 :

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

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

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

Clearly, for $Z = 1$, this condition reduces to the noncooperative condition, which requires \mathcal{M}_0 to be a circular sparse ruler. Each

\mathcal{M}_z , $z = 0, \dots, Z-1$, is called an *incomplete circular sparse ruler* since it does not contain all possible differences between 0 and $N-1$ (see “Circular Sparse Rulers”). However, (41) clearly implies that, for every given integer modular distance $n \in \{0, 1, \dots, N-1\}$, 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 Figure 9. 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 to minimize the number of marks in each ruler while satisfying (41). This task is intimately connected to the so-called nonoverlapping circular Golomb rulers [21].

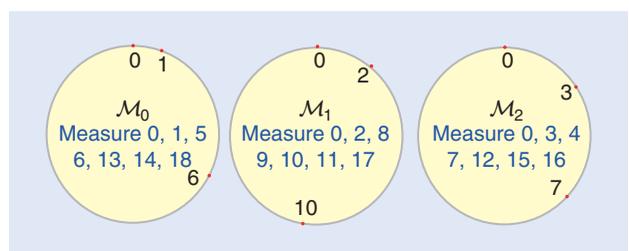
Alternative schemes for cooperative CCS include [48], which exploits the cross-correlation between observations at different sensors, and [37], where the observations are not only linearly compressed but also quantized to a single bit.

OPEN QUESTIONS

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

As for sampler design, its most existing schemes rely on identifiability criteria [8], [20], but other criteria are yet to be explored. For instance, it is important to find sampler designs minimizing the Cramér–Rao bound for unbiased estimation of the parameters of interest. Of special relevance are deterministic schemes maximizing

**CCS IS ESPECIALLY CONVENIENT
TO ACQUIRE WIDEBAND SIGNALS,
WHOSE RAPID VARIATIONS
CANNOT BE EASILY CAPTURED
BY CONVENTIONAL ADC.**



[FIG9] 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.

compression for a target estimation performance. Other sampling schemes, like gridless or continuous irregular sampling, are yet to be investigated from a CCS perspective. The problem becomes reconstructing second-order statistics when the sample locations are not a subset of a regularly-spaced grid. This problem differs from the existing literature on gridless or continuous sparse reconstruction, which aims to accurately recover sparse input signals.

Cooperative schemes also deserve extensive research. For instance, distributed implementations and data fusion techniques for CCS with affordable communication overhead need to be revisited [37]. This includes 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. The latter 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 because of 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 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 covariance estimation problems for data-starved inference networks. Such problems arise under the umbrella of probabilistic analysis for high-dimensional data sets with many variables and few samples.

As a precursor, sparse (inverse) covariance estimation has already become a popular 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 nonstationary process analysis, for instance, exploiting the framework of underspread processes [50]. Future research may also consider nonlinear parameterizations as well as nonlinear compression.

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, which can be costly in terms of both

computational and sensing resources. This standpoint entails multiple benefits, such as the possibility of introducing strong compression without need for sparsity, as required by compressed sensing.

ACKNOWLEDGMENTS

This work was supported by the Spanish Government and the European Regional Development Fund (ERDF) (project TEC2013-47020-C2-1-R COMPASS), the Galician Regional Government and ERDF (projects GRC2013/009, R2014/037, and AtlantTIC), the U.S. Army Research Office under grant W911NF-15-1-0492, 2015-2016, the National Science Foundation under grant EARS 1343248, and NWO-STW under the VICI program (10382).

AUTHORS

Daniel Romero (dromero@umn.edu) received his M.S. and Ph.D. degrees from the University of Vigo, Spain, in 2011 and 2015, respectively. He recently joined the Department of Electrical and Computer Engineering, University of Minnesota. His research interests are in the areas of statistical signal processing, communications, and big-data analytics.

Dyonisius Dony Ariananda (D.A.Dyonisius@tudelft.nl) received his bachelor's degree (2000) from Gadjah Mada University, Yogyakarta, Indonesia, and his M.S. (2009) and Ph.D. (2015) degrees from the Delft University of Technology (TU Delft), The Netherlands, all in electrical engineering. He was also granted the Australian Development Scholarship to complete his M.S. degree in inter-networking at the University of Technology, Sydney, New South Wales, Australia, in 2005. He was an academic staff member with the

Informatics Engineering Department, Sanata Dharma University, Yogyakarta, Indonesia, from 2000 to 2003 and with the Electrical Engineering Department, Atma Jaya University, Jakarta, Indonesia, from 2006 to 2007. He is now a postdoctoral researcher with TU Delft, working on compressive sampling.

Zhi Tian (ztian1@gmu.edu) is a professor with the Electrical and Computer Engineering Department, George Mason University, Virginia. She was previously on the faculty of Michigan Technological University and worked with the U.S. National Science Foundation. Her research interests lie in statistical signal processing theory and its broad applications. Her current research focuses on compressed sensing for random processes, statistical inference of large-scale data, cognitive radio networks, and distributed wireless sensor networks. She is a Fellow of the IEEE.

Geert Leus (G.J.T.Leus@tudelft.nl) is a full professor at the Delft University of Technology, The Netherlands. He received a 2002 Young Author Best Paper Award and a 2005 Best Paper Award from the IEEE Signal Processing Society (SPS). He was the chair of the IEEE Signal Processing for Communications and Networking Technical Committee and an associate editor of *IEEE Transactions on Signal Processing*, *IEEE Transactions on*

CYCLOSTATIONARITY PROVIDES AN ALTERNATIVE PERSPECTIVE TO UNDERSTAND COMPRESSION OF SECOND-ORDER STATISTICS, EVEN FOR WSS SEQUENCES.

Wireless Communications, *IEEE Signal Processing Letters*, and *EURASIP Journal on Advances in Signal Processing (JASP)*. Currently, he is a member-at-large of the Board of Governors of the IEEE SPS and a member of the IEEE Sensor Array and Multichannel Technical Committee. He serves as the editor-in-chief of *EURASIP JASP*. He is a Fellow of the IEEE.

REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. 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. Inform. 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. Inform. Theory*, vol. 46, no. 6, pp. 2173–2183, Sept. 2000.
- [4] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Top. Signal Processing*, 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. Inform. 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*. Hoboken, NJ: Wiley, 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, Sept. 1982.
- [11] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, vol. 1. Reading, MA: Addison-Wesley, 1991.
- [12] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, "Difference bases and sparse sensor arrays," *IEEE Trans. Inform. 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. Signal Processing*, 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. Signal Processing*, 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. Acoustics, Speech, and Signal Processing*, May 2011, pp. 3884–3887.
- [16] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Design and analysis of multicoset arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2013, pp. 3781–3785.
- [17] D. D. Ariananda, D. Romero, and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, San Martin, France, Dec. 2013, pp. 440–443.
- [18] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math. Soc.*, vol. 43, no. 3, pp. 377–385, 1938.
- [19] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Processing*, 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. Inform. Theory*, vol. 61, no. 3, pp. 1410–1425, Mar. 2015.
- [21] D. D. Ariananda, D. Romero, and G. Leus, "Cooperative compressive power spectrum estimation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, A Coruna, Spain, June 2014, pp. 97–100.
- [22] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Processing*, 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. Signal Processing*, vol. 58, no. 8, pp. 4131–4143, 2010.
- [24] G. Leus and D. D. Ariananda, "Power spectrum blind sampling," *IEEE Signal Processing 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. Acoustics, Speech, and Signal Processing*, Apr. 2007, vol. 4, pp. IV. 1357–1360.
- [26] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Top. Signal Processing*, 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. dissertation, California Institute of Technology, 2011.
- [29] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing 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. Signal Processing*, 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. Signal Processing*, 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. Signal Processing*, 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. Signal Processing*, vol. 46, no. 9, pp. 2458–2471, 1998.
- [34] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Int. Workshop on Signal Processing Advances in Wireless Communications*, June 2012, pp. 525–529.
- [35] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Processing*, 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 Processing*, 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. Signal Processing*, 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 Information Processing*, June 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. Acoustics, Speech, and Signal Processing*, 2009, pp. 2165–2168.
- [40] D. D. Ariananda and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Signal 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. Signal Processing*, 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. Signal Processing*, 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 on Computational Advances in Multi-Sensor Adaptive Processing*, Dec. 2011, pp. 337–340.
- [44] D. D. Ariananda and G. Leus, "Non-uniform sampling for compressive cyclic spectrum reconstruction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 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. Adv. Signal Processing*, vol. 2014, no. 1, p. 168, 2014.
- [46] D. D. Ariananda, D. Romero, and G. Leus, "Compressive periodogram reconstruction using uniform binning," *IEEE Trans. Signal Processing*, 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. Signal Processing*, vol. 52, no. 3, pp. 601–616, Mar. 2004.
- [48] D. D. Ariananda and G. Leus, "Cooperative compressive wideband power spectrum sensing," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 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. Acoustics, Speech, and Signal Processing*, 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. Inform. Theory*, vol. 52, no. 3, pp. 1067–1086, 2006.