# Compression Limits for Random Vectors with Linearly Parameterized Second-Order Statistics

Daniel Romero, *Student Member, IEEE,* Roberto López-Valcarce,
*Member, IEEE* and Geert Leus, *Fellow, IEEE.*

*Abstract*—The class of complex random vectors whose covariance matrix is linearly parameterized by a basis of Hermitian Toeplitz (HT) matrices is considered, and the maximum compression ratios that preserve all second-order information are derived — the statistics of the uncompressed vector must be recoverable from a set of linearly compressed observations. This kind of vectors arises naturally when sampling wide-sense stationary random processes and features a number of applications in signal and array processing.

Explicit guidelines to design optimal and nearly optimal schemes operating both in a periodic and non-periodic fashion are provided by considering two of the most common linear compression schemes, which we classify as dense or sparse. It is seen that the maximum compression ratios depend on the structure of the HT subspace containing the covariance matrix of the uncompressed observations. Compression patterns attaining these maximum ratios are found for the case without structure as well as for the cases with circulant or banded structure. Universal samplers are also proposed to compress unknown HT subspaces.

*Index Terms*—Compressive Covariance Sensing, Covariance Matching, Compression Matrix Design.

## I. Preliminaries

Consider the problem of estimating the second-order statistics of a zero-mean random vector $x \in \mathbb{C}^L$ from a set of $K$ linear observations collected in the vector $y \in \mathbb{C}^K$ given by

$$y = \bar{\Phi} x, \tag{1}$$

where $\bar{\Phi} \in \mathbb{C}^{K \times L}$ is a known matrix and several realizations of $y$ may be available. This matrix may be referred to as the *compression matrix*, *measurement matrix* or *sampler*, where compression is achieved by setting $K < L$ (typically $K \ll L$). The covariance matrix $\Sigma = \mathrm{E}\{xx^H\}$ contains the second-order statistics of $x$ and is assumed to be a linear combination of the Hermitian Toeplitz (HT) matrices in a given set $\mathcal{S} = \{\Sigma_0, \Sigma_1, \cdots, \Sigma_{S-1}\} \subset \mathbb{C}^{L \times L}$, that is, there exist some scalars $\alpha_s$ such that $\Sigma = \sum_s \alpha_s \Sigma_s$.

This problem arises in inference operations over the second-order statistics of a random vector with a Toeplitz covariance matrix. Operating on the compressed observations $y$ entails multiple advantages due to their smaller dimension. In fact, many research efforts in the last decades have been pointed towards designing compression methods and reconstruction algorithms that allow for sampling rate reductions. While most efforts have been focused on reconstructing $x$, there were also important advances when only the second-order statistics of this vector are of interest. This paper is concerned with problems of the second kind.

The *compression ratio* $\rho = L/K$ measures how much $x$ is compressed. The maximum compression ratio remains an open problem in many cases of interest; and most existing results rely on the usage of specific reconstruction algorithms (see Sec. I-D). This paper presents a general and unifying framework built on abstract criteria where the maximum compression ratio is defined and computed for most relevant settings. The proofs involved in this theory are constructive, resulting in several methods for designing optimal compression matrices.

### A. Covariance Matching Formulation

The prior information restricts the structure of $\Sigma$, thus determining how much $x$ can be compressed. When no information at all is available, $\Sigma$ is simply constrained to be Hermitian positive semidefinite and no compression is possible. However, if $x$ contains samples from a wide-sense stationary process, the fact that $\Sigma$ is HT and positive semidefinite allows for a certain degree of compression. More generally, $\Sigma$ may be assumed to lie in the intersection of the cone of positive semidefinite matrices and the subspace spanned by a set of HT matrices (not necessarily positive semidefinite) $\mathcal{S} = \{\Sigma_0, \Sigma_1, \cdots, \Sigma_{S-1}\} \subset \mathbb{C}^{L \times L}$. This subspace, throughout referred to as the *covariance subspace*, captures the prior information available and, intuitively, the smaller its dimension, the higher the compression that can be reached.

Without any loss of generality, we consider real scalars:

$$\Sigma = \sum_{s=0}^{S-1} \alpha_s \Sigma_s, \quad \text{with} \quad \alpha_s \in \mathbb{R}, \tag{2}$$

and $\mathcal{S}$ is assumed to be a linearly independent set of matrices:

$$\sum_{s=0}^{S-1} \alpha_s \Sigma_s = \sum_{s=0}^{S-1} \beta_s \Sigma_s \quad \Rightarrow \quad \alpha_s = \beta_s \; \forall s. \tag{3}$$

Thus, $\mathcal{S}$ is a basis for the covariance subspace, which means that the decomposition in (2) is unique and, consequently,

knowing the $\alpha_s$'s is equivalent to knowing $\boldsymbol{\Sigma}$. Since the coefficients are real-valued and the matrices HT, it is necessary that $S \leq 2L - 1$ in order for $\mathcal{S}$ to be linearly independent. The second-order statistics of $\boldsymbol{y}$, arranged in $\bar{\boldsymbol{\Sigma}} = \mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^H\right\}$, and those of $\boldsymbol{x}$, arranged in $\boldsymbol{\Sigma}$, are related by:

$$\bar{\boldsymbol{\Sigma}} = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma}\bar{\boldsymbol{\Phi}}^H = \sum_{s=0}^{S-1} \alpha_s \bar{\boldsymbol{\Sigma}}_s, \quad \text{where} \quad \bar{\boldsymbol{\Sigma}}_s = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma}_s\bar{\boldsymbol{\Phi}}^H. \quad (4)$$

In other words, the expansion coefficients of $\boldsymbol{\Sigma}$ with respect to $\mathcal{S}$ are those of $\bar{\boldsymbol{\Sigma}}$ with respect to $\bar{\mathcal{S}} = \{\bar{\boldsymbol{\Sigma}}_0, \bar{\boldsymbol{\Sigma}}_1, \cdots, \bar{\boldsymbol{\Sigma}}_{S-1}\} \subset \mathbb{C}^{K \times K}$. Albeit Hermitian, the matrices in $\bar{\mathcal{S}}$ are not Toeplitz in general. If the compression operation preserves all relevant information, then $\bar{\mathcal{S}}$ is linearly independent and knowing $\bar{\boldsymbol{\Sigma}}$ is equivalent to knowing the $\alpha_s$'s, which in turn amounts to knowing $\boldsymbol{\Sigma}$. Conversely, if the compression is so strong that the linear independence is lost, then some second-order information about $\boldsymbol{x}$ cannot be recovered.

This paper unifies the treatment of a number of problems arising in different applications (see Sec. I-C) by noting that they can be stated as the estimation of a linearly parameterized covariance matrix $\bar{\boldsymbol{\Sigma}}$ from the compressed observations $\boldsymbol{y}$, that is, they admit a *covariance matching* formulation [1], [2]. For simplicity, a linear parameterization such as the one in (4) is assumed, but the results still apply to certain non-linear parameterizations [1] (see the discussion around Lemma 1).

### B. Signal Acquisition

Compression is particularly convenient in the acquisition stage since otherwise part of the resources would be devoted to acquire data that is afterwards discarded. For this reason, the literature contains many compressive acquisition and reconstruction procedures. Remarkable examples are sub-Nyquist sampling of multiband/multitone [3]–[7] signals, compressed sensing [8], [9], and array design for aperture synthesis imaging [10]–[12]. They differ as to which structure is assumed for the data and which information is deemed important.

Most consider reconstructing a signal $\boldsymbol{x}$ from linearly compressed observations $\boldsymbol{y} = \bar{\boldsymbol{\Phi}}\boldsymbol{x}$. Although this procedure is, in principle, possible when the goal is to estimate the second-order statistics of $\boldsymbol{x}$, saving the intermediate step of reconstructing $\boldsymbol{x}$ may entail computational advantages and greater compression ratios. This problem will be globally referred to as *compressive covariance sampling* (CCS).

These approaches (including CCS) share similar compression structures, classified here according to the nature of $\bar{\boldsymbol{\Phi}}$:

- *Sparse samplers* are those where $\bar{\boldsymbol{\Phi}}$ is a sparse matrix. Commonly, $\bar{\boldsymbol{\Phi}}$ is composed of $K$ different rows of the identity matrix $\boldsymbol{I}_L$, thus performing a *component selection* of $\boldsymbol{x}$. If this selection is periodic, it is known as *multi-coset sampling* (see Secs. II-C and IV-A).
- *Dense samplers* are those where $\bar{\boldsymbol{\Phi}}$ is a dense matrix. Each component of $\boldsymbol{y}$ is therefore a *linear combination* of the components of $\boldsymbol{x}$. In the case of periodic samplers, $\bar{\boldsymbol{\Phi}}$ is block diagonal where all diagonal blocks are replicas of a certain dense matrix (see Sec. II-C).

The nature of the acquisition architecture depends on the domain where the signal of interest is defined:

- *Time-domain signals:* several alternatives have been proposed to replace *analog-to-digital converters* (ADCs), which are known to be slow, expensive and power-hungry. Some examples include interleaved ADCs [13], non-uniform sampling and its generalizations [4], [5], [14], the random demodulator [7], [15], the modulated wideband converter [6] and the random modulator pre-integrator [16], [17]. We will globally refer to these devices as *compressive-ADCs* (C-ADCs). Their operation is described by (1) when $\boldsymbol{x}$ contains the Nyquist samples of the signal of interest, which are not physically acquired but can be used as a convenient mathematical abstraction.
- *Space-domain signals:* Compression is accomplished using (1), where $\boldsymbol{x}$ is a snapshot of the *uncompressed array*. With sparse sampling (see e.g. [10]–[12], [18]–[23]), only the antennas corresponding to the non-null columns of $\bar{\boldsymbol{\Phi}}$ need to be physically deployed to obtain $\boldsymbol{y}$, whereas in dense sampling [24]–[26], analog combiners are used to reduce the number of radio frequency chains.

### C. Applications of CCS

We show how CCS can be applied to several problems that can be formulated using covariance matching models. These models need not be used for estimation [1], [2]; they are simply used to capture the information to be preserved. The most common covariance subspaces, defined in Sec. II-B, are the *Toeplitz subspace*, the *circulant subspace* and the *d-banded subspace*. Sparse and dense samplers have been considered in most applications, either in a periodic or non-periodic fashion.

*1) Compressive Power Spectrum Estimation:* The goal is to estimate $\boldsymbol{\Sigma}$ from $\boldsymbol{y}$ with the only constraint that it must be HT and positive semidefinite, which means that the covariance subspace is the Toeplitz subspace. One can employ any basis for this subspace, reconstruct $\boldsymbol{\Sigma}$ and apply a Fourier transform to find the power spectrum. More directly, one can consider the Fourier basis (see (12) below) where the coordinate $\alpha_s$ in (2) will represent the value of the power spectrum at frequency $2\pi s/(2L - 1)$. Assuming bounded autocorrelation supports enables $d$-banded subspaces [27], whereas a frequency domain formulation results in circulant subspaces [28], [29].

*2) Wideband Spectrum Sensing:* If $\boldsymbol{x} = \sum_s \sigma_s \boldsymbol{x}_s$, where $\boldsymbol{x}_s$ corresponds to a signal whose second-order statistics are known up to a scale, the parameters $\sigma_s$ capturing the power of each component can be estimated based on the observations provided by a C-ADC [30]–[33]. The covariance subspace is the span of the set of covariance matrices of the $\boldsymbol{x}_s$'s.

*3) Incoherent Imaging:* Arbitrary distributions of uncorrelated sources in the far field of a uniform linear array (the uncompressed array) produce HT spatial covariance matrices. The angular spectrum can be obtained as the coefficients $\alpha_s$ in the expansion (12) (see [10]), which correspond to the intensity impinging from $2L-1$ looking directions. Recent formulations have also considered circulant subspaces [34], [35].

*4) Sparse Spectrum Estimation:* modal analysis can be used to identify the components of a *sum of sinusoids* in noise (time-domain signals) [36], [37] or to estimate the *direction of arrival* (DoA) of a number of point sources

in the far field (space-domain signals) [12], [20]–[23], [38] using the compressed observations $\boldsymbol{y}$. $\boldsymbol{\Sigma}$ is expanded as $\boldsymbol{\Sigma} = \sum_{s=0}^{R-1} \alpha_s \boldsymbol{v}(\phi_s) \boldsymbol{v}^H(\phi_s)$, where $R$ is the number of sinusoids/sources and $\boldsymbol{v}(\phi_s)$ corresponds either to the sinusoid with frequency $\phi_s$ or to the source at angle $\phi_s$. If $\boldsymbol{x}$ is uniformly sampled, then $\boldsymbol{v}(\phi_s) \boldsymbol{v}^H(\phi_s)$ is Toeplitz. Since the angles $\phi_s$ are unknown, the only structure present in $\boldsymbol{\Sigma}$ is that it is HT and positive semidefinite [12]. Therefore, $\bar{\boldsymbol{\Phi}}$ must preserve the structure of *any* Toeplitz matrix. An equivalent approach uses *universal samplers* (see Sec. II).

### D. Related Work and Contributions

Most works on reconstructing second-order statistics from compressed measurements deal with estimating Toeplitz covariance matrices using non-periodic sparse samplers, where the observation is that at least a pair of samples at each possible distance is required to estimate the statistics of the uncompressed signal [10]–[12], [23], [36]. The optimal solution, termed *restricted minimum redundancy array* or *minimal sparse ruler*, was analyzed in [18], [19], [39]–[42] and shown to be optimal in direction finding [12]. Suboptimal, yet more structured, schemes were proposed in [18], [23], [41]–[44].

Periodic sampling in $d$-banded subspaces was considered in [27], where the maximum $\rho$ was bounded using the conditions for unique reconstruction of a least squares algorithm. Suboptimal compression schemes were proposed in [27] and [45]. Non-periodic sparse sampling in circulant subspaces was considered in [28] and [35], where optimal and suboptimal designs are respectively found based on specific algorithms.[1]

The sampler design criteria used in most of these works are tailored to specific reconstruction algorithms. Furthermore, their formulation is not general enough to accommodate periodic samplers, dense samplers or prior information. The contributions of this paper can be summarized as follows:

- We present a formal and general framework, irrespective of any algorithm, that establishes the conditions for a compression pattern to be *admissible* and defines the maximum compression ratio based on abstract criteria.
- Optimal sparse and dense samplers are found for most cases of interest. Novel designs include (non-)periodic sparse samplers for circulant and banded subspaces, periodic sparse samplers for Toeplitz subspaces and (non-)periodic dense samplers for Toeplitz, circulant and banded subspaces.
- The notion of *universal sampler* is proposed as the one preserving all second-order information for any HT covariance subspace.
- We provide simple tools to assess admissibility in all linear and certain non-linear cases. Particularly, we show that the positive semidefinite nature of covariance matrices does not generally allow greater compression ratios.

### E. Notation

If a set $\mathcal{A}$ is finite, then $|\mathcal{A}|$ denotes its cardinality. If $\mathbb{F}$ is a field, then the $\mathbb{F}$-span of a set of matrices $\mathcal{A}$ is defined as $\mathrm{span}_{\mathbb{F}} \mathcal{A} = \{ \boldsymbol{A} \in \mathbb{C}^{P \times P} : \boldsymbol{A} = \sum_s \alpha_s \boldsymbol{A}_s, \ \boldsymbol{A}_s \in \mathcal{A}, \ \alpha_s \in$

$\mathbb{F}\}$. The $\mathbb{F}$-dimension of a set $\mathcal{B}$, denoted as $\dim_{\mathbb{F}} \mathcal{B}$, is the smallest $n \in \mathbb{N}$ such that there exists some $\mathcal{A}$ with $|\mathcal{A}| = n$ such that $\mathcal{B} \subset \mathrm{span}_{\mathbb{F}} \mathcal{A}$. The image of a set $\mathcal{A}$ through a function $\phi$ is denoted as $\phi(\mathcal{A})$.

Lowercase is used for scalars, bold lowercase for vectors and bold capital for matrices. Superscript $^T$ stands for transpose, $^H$ for conjugate transpose and $\otimes$ represents the Kronecker product [46]. The $(i, j)$ entry of the $P \times Q$ matrix $\boldsymbol{A}$ is $a_{i,j}$, where we start with index zero (that is, the top-left entry is $a_{0,0}$). The vectorization of $\boldsymbol{A}$ is the vector $\mathrm{vec}\{\boldsymbol{A}\} = [\boldsymbol{a}_0^T, \cdots, \boldsymbol{a}_{Q-1}^T]^T$, where $\boldsymbol{a}_j = [a_{0,j}, \cdots, a_{P-1,j}]^T$. The $d$-th diagonal refers to the entries $(i, j)$ with $j - i = d$, where $d$ is a negative, null or positive integer. $\boldsymbol{E}_{i,j}$ is a matrix with all zeros except for a 1 at the position $(i, j)$ and it is represented as $\boldsymbol{e}_i$ if it has a single column.

The symbol $\jmath$ denotes the imaginary unit and $(x)_N$ is the remainder of the integer division of $x \in \mathbb{Z}$ by $N$, i.e., $(x)_N$ is the only element in the set $\{x + bN, \ b \in \mathbb{Z}\} \cap \{0, \ldots, N-1\}$.

### F. Paper Structure

The rest of the paper is structured as follows. Sec. II sets the theoretical background, where maximum compression ratios and covariance samplers are defined. Sec. III presents some results to design covariance samplers, which are applied in Secs. IV and V to design universal and non-universal covariance samplers, respectively. Asymptotic compression ratios are discussed in Sec. VI, whereas some remarks and conclusions are respectively provided in Secs. VII and VIII.

## II. THEORETICAL FRAMEWORK

The definition of the maximum compression ratio requires to first decide which samplers we are willing to accept. As explained in Sec. I, we are interested in those samplers preserving all the second-order statistical information of $\boldsymbol{x}$, i.e., those samplers that allow to recover the statistics of $\boldsymbol{x}$ from the statistics of $\boldsymbol{y}$. In order to formalize this notion, let us start by associating the compression matrix $\bar{\boldsymbol{\Phi}} \in \mathbb{C}^{K \times L}$ with a linear function that relates the covariance matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$ and which is defined as

$$\begin{array}{ccc} \mathrm{span}\,\mathcal{S} & \xrightarrow{\phi} & \mathrm{span}\,\bar{\mathcal{S}} \\ \underset{\mathbb{R}}{\phantom{x}} & & \underset{\mathbb{R}}{\phantom{x}} \\ \boldsymbol{\Sigma} & \longrightarrow & \phi(\boldsymbol{\Sigma}) = \bar{\boldsymbol{\Phi}} \boldsymbol{\Sigma} \bar{\boldsymbol{\Phi}}^H \end{array} \qquad (5)$$

where, recall, $\mathcal{S}$ is a linearly independent set of $S$ HT matrices.[2] We next specify which sampling matrices are *admissible*:

*Definition 1: A matrix $\bar{\boldsymbol{\Phi}}$ defines an $\mathcal{S}$-covariance sampler*[3] *if the associated function $\phi$, defined in* (5)*, is invertible.*

The maximum compression ratio is the largest value of $L/K$ for which a covariance sampler $\bar{\boldsymbol{\Phi}} \in \mathbb{C}^{K \times L}$ can be found. Above this value, it is not possible to consistently estimate the second-order statistics of $\boldsymbol{x}$, even from an arbitrarily large number of realizations of $\boldsymbol{y}$, since the statistical identifiability[4] of $\boldsymbol{\Sigma}$ is lost [48]. For convenience, we will regard $L$ as given and attempt to minimize $K$.

---

[1]The initial statement in [28], [35] uses periodic sampling, but their considerations in the frequency domain lead to non-periodic sampling.

[2]For mathematical convenience, $\phi$ is not only defined for positive semidefinite matrices.

[3]When the set $\mathcal{S}$ is clear from the context, we will simply say that $\bar{\boldsymbol{\Phi}}$ defines a *covariance sampler*.

[4]See [47] for a discussion on the statistical identifiability in CCS.

One may argue that the requirement in Definition 1 is too strong since it suffices to require $\phi$ to be invertible only for those matrices in $\operatorname{span}_{\mathbb{R}} \mathcal{S}$ that are positive semidefinite. More generally, the prior information may constrain $\boldsymbol{\Sigma}$ to be in a certain non-linear set $\mathcal{A}$ such as the set of positive semidefinite matrices, the set of covariance matrices of auto-regressive processes with a given order, the non-linear sets in [1], etc. In that case, we may reformulate Definition 1 to require $\phi$ to be invertible only in $\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}$. However, this is unnecessary, as shown next:

*Lemma 1: Let $\phi$ be the function defined in (5), where $\mathcal{S}$ is an independent set of $S$ HT matrices, let $\mathcal{A}$ be a set of matrices such that $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] = S$ and let $\phi_{|\mathcal{A}}$ be the restriction of $\phi$ to $\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}$, defined as:*

$$
\begin{array}{ccc}
\mathcal{A} \cap \operatorname*{span}_{\mathbb{R}} \mathcal{S} & \xrightarrow{\phi_{|\mathcal{A}}} & \phi(\mathcal{A} \cap \operatorname*{span}_{\mathbb{R}} \mathcal{S}) \\
\boldsymbol{\Sigma} & \longrightarrow & \phi_{|\mathcal{A}}(\boldsymbol{\Sigma}) = \phi(\boldsymbol{\Sigma}).
\end{array} \quad (6)
$$

*Then, $\phi$ is invertible if and only if $\phi_{|\mathcal{A}}$ is invertible.*

*Proof:* See Appendix A. ∎

Therefore, the non-linear information collected in $\mathcal{A}$ is irrelevant from the linear compression perspective whenever $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] = S$. If this condition is not satisfied, one must choose a different basis $\mathcal{S}'$ such that $\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}' = \mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}$ and $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}'] = |\mathcal{S}'|$, which is always possible. This establishes the generality of Definition 1 and enables us to work with covariance subspaces without further concerns.

If $\mathcal{A}$ is the cone of positive semidefinite matrices, then $\mathcal{S}$ satisfies $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] = S$ in most cases of interest:

*Lemma 2: Let $\mathcal{A}$ be the set of positive semidefinite matrices. Then $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] = S$ if at least one of the following conditions holds:*

1) $\boldsymbol{\Sigma} \geq \mathbf{0}$ *for all* $\boldsymbol{\Sigma} \in \mathcal{S}$
2) $\exists \boldsymbol{\Sigma} \in \operatorname{span}_{\mathbb{R}} \mathcal{S}$ *such that* $\boldsymbol{\Sigma} > \mathbf{0}$

*Proof:* 1) means that $\mathcal{S} \subset [\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}]$. Then $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] \geq \dim_{\mathbb{R}} \mathcal{S} = S$. Noting that $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] \leq S$ for any $\mathcal{S}$ shows that $\dim_{\mathbb{R}}[\mathcal{A} \cap \operatorname{span}_{\mathbb{R}} \mathcal{S}] = S$. On the other hand, if 2) holds, we can assume without any loss of generality that $\mathcal{S} = \{\boldsymbol{\Sigma}_0, \ldots, \boldsymbol{\Sigma}_{S-1}\}$ where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma} > 0$. If $\mathcal{S}' = \{\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1 + \alpha \boldsymbol{\Sigma}_0, \ldots, \boldsymbol{\Sigma}_{S-1} + \alpha \boldsymbol{\Sigma}_0\}$, then $\operatorname{span}_{\mathbb{R}} \mathcal{S} = \operatorname{span}_{\mathbb{R}} \mathcal{S}'$ for any $\alpha$. Choose $\alpha = -\min_s \lambda_{\min}(\boldsymbol{\Sigma}_s)/\lambda_{\min}(\boldsymbol{\Sigma}_0)$, with $\lambda_{\min}$ representing the minimum eigenvalue. Then $\mathcal{S}'$ satisfies 1), which concludes the proof. ∎

Since at least one of the above sufficient conditions will be satisfied in all cases considered in this paper, Lemma 2 establishes that positive semidefiniteness plays no role in the compression of Toeplitz, circulant or banded subspaces. Hence, no compression improvements are possible in those cases.

Clearly, a matrix $\bar{\boldsymbol{\Phi}}$ may define a covariance sampler for certain sets $\mathcal{S}$ but not for others. If a matrix $\bar{\boldsymbol{\Phi}}$ is a covariance sampler for any choice of $\mathcal{S}$, we call it *universal*:

*Definition 2: A sampling matrix $\bar{\boldsymbol{\Phi}} \in \mathbb{C}^{K \times L}$ defines a universal covariance sampler if it is an $\mathcal{S}$-covariance sampler for any linearly independent set $\mathcal{S}$ of $L \times L$ HT matrices.*

Knowing $\mathcal{S}$ is always beneficial since $\bar{\boldsymbol{\Phi}}$ may be tailored to obtain optimal compression ratios and estimation performance. Universal samplers are motivated by those cases where $\mathcal{S}$, or

even $S$, is unknown at the moment of designing the compression matrix. Note that other notions of universal samplers have been introduced in different contexts [9], [28], [49]–[51].

### A. Interpretation

Due to the definition of domain and codomain in (5), $\phi$ clearly represents a surjective map. Therefore, the notion of invertibility actually means that $\phi$ must be injective, that is, for any set of real coefficients $\alpha_s$ and $\beta_s$,

$$
\phi\left(\sum_s \alpha_s \boldsymbol{\Sigma}_s\right) = \phi\left(\sum_s \beta_s \boldsymbol{\Sigma}_s\right) \quad \Rightarrow \quad \alpha_s = \beta_s \ \forall s. \quad (7)
$$

This condition is, in turn, equivalent to

$$
\sum_s \alpha_s \bar{\boldsymbol{\Sigma}}_s = \sum_s \beta_s \bar{\boldsymbol{\Sigma}}_s \Rightarrow \alpha_s = \beta_s \ \forall s, \quad (8)
$$

which means that $\bar{\mathcal{S}}$ must be linearly independent. Thus, determining whether a given matrix $\bar{\boldsymbol{\Phi}}$ defines an $\mathcal{S}$-covariance sampler amounts to checking whether $\bar{\mathcal{S}} = \phi(\mathcal{S})$ is linearly independent or not. Alternatively, (8) states that no two different linear combinations of the matrices in $\bar{\mathcal{S}}$ can result in the same $\bar{\boldsymbol{\Sigma}}$, which means that covariance samplers can also be defined as those samplers preserving the identifiability of the coefficients $\alpha_s$.

To the best of our knowledge, Definition 1 is the first attempt to formalize the design of samplers for CCS problems using abstract criteria not depending on specific algorithms. In the sequel, several results will be established to determine whether a matrix defines a covariance sampler or, in some cases, even a universal covariance sampler.

### B. Notable Covariance Subspaces

The results about covariance samplers derived in this paper will be particularized in Sec. V for the most common covariance subspaces, which are defined next:

*1) Toeplitz Subspace:* A matrix is Toeplitz if it is constant along its diagonals [52]. The set of all $L \times L$ HT matrices, represented as $\mathbb{S}^L$, is a subspace of $\mathbb{C}^{L \times L}$ over the real scalar field,[5] and it is the largest subspace considered in this paper. The *standard basis* of $\mathbb{S}^L$ is defined as the set

$$
\mathcal{S}_T = \{\boldsymbol{I}_L\} \cup \{\boldsymbol{T}_1, \cdots, \boldsymbol{T}_{L-1}\} \cup \{\tilde{\boldsymbol{T}}_1, \cdots, \tilde{\boldsymbol{T}}_{L-1}\}, \quad (9)
$$

where $\boldsymbol{T}_l$ denotes the HT matrix with all zeros except for the entries on the diagonals $+l$ and $-l$, which have ones, and $\tilde{\boldsymbol{T}}_l$ represents the HT matrix with all zeros except for the entries on the diagonal $+l$, which have the imaginary unit $\jmath$, and those on the diagonal $-l$, which have $-\jmath$. Formally,

$$
\boldsymbol{T}_l = \boldsymbol{J}_L^l + (\boldsymbol{J}_L^l)^T \quad l \geq 1 \quad (10)
$$

$$
\tilde{\boldsymbol{T}}_l = \jmath \boldsymbol{J}_L^l - \jmath (\boldsymbol{J}_L^l)^T \quad l \geq 1, \quad (11)
$$

where $\boldsymbol{J}_L$ is the first linear shift of $\boldsymbol{I}_L$ to the right, i.e., the matrix whose element $(m, n)$ is one if $n - m = 1$ and

---

[5]The reason is that any linear combination with real coefficients of HT matrices is also HT. This statement is false for complex coefficients.

zero otherwise. The basis $\mathcal{S}_T$ shows that $\dim_{\mathbb{R}} \mathbb{S}^L = 2L - 1$. Another important basis for this subspace is the Fourier basis:

$$\mathcal{S}_F = \{\boldsymbol{\Sigma}_0, \cdots, \boldsymbol{\Sigma}_{2L-2}\}, \quad (\boldsymbol{\Sigma}_s)_{m,n} = \frac{e^{j\frac{2\pi}{2L-1}(m-n)s}}{2L-1}. \quad (12)$$

*2) Circulant Subspace:* A circulant matrix is a matrix whose $n$-th row equals the $n$-th circular rotation of the zeroth row[6] to the right [52]. In other words, the element $(m, n)$ equals the element $(m', n')$ if $(m-n)_L = (m'-n')_L$. In our case, the matrices in the circulant subspace must be HT and circulant simultaneously. A possible basis for $L$ odd is

$$\mathcal{S}_C = \{\boldsymbol{I}_L\} \cup \{\boldsymbol{C}_1, \cdots, \boldsymbol{C}_{\frac{L-1}{2}}\} \cup \{\tilde{\boldsymbol{C}}_1, \cdots, \tilde{\boldsymbol{C}}_{\frac{L-1}{2}}\}, \quad (13)$$

where

$$\boldsymbol{C}_l = \boldsymbol{T}_l + \boldsymbol{T}_{L-l}, \quad l = 1, \ldots, \lfloor (L-1)/2 \rfloor \quad (14)$$
$$\tilde{\boldsymbol{C}}_l = \tilde{\boldsymbol{T}}_l - \tilde{\boldsymbol{T}}_{L-l}, \quad l = 1, \ldots, \lfloor (L-1)/2 \rfloor, \quad (15)$$

and

$$\mathcal{S}_C = \{\boldsymbol{I}_L\} \cup \{\boldsymbol{C}_1, \cdots, \boldsymbol{C}_{\frac{L}{2}-1}\} \cup \{\tilde{\boldsymbol{C}}_1, \cdots, \tilde{\boldsymbol{C}}_{\frac{L}{2}-1}\} \cup \{\boldsymbol{T}_{\frac{L}{2}}\}$$

for $L$ even. Clearly, the dimension of this subspace equals $L$.

*3) d-banded Subspace:* A *d-banded matrix* is a matrix where all the elements above the diagonal $+d$ and below the diagonal $-d$ (these diagonals noninclusive) are zero. A possible basis for this subspace is given by

$$\mathcal{S}_B^d = \{\boldsymbol{I}_L\} \cup \{\boldsymbol{T}_1, \cdots, \boldsymbol{T}_d\} \cup \{\tilde{\boldsymbol{T}}_1, \cdots, \tilde{\boldsymbol{T}}_d\}, \quad (16)$$

which is a subset of $\mathcal{S}_T$. The dimension is therefore $2d + 1$.

### C. The Role of Periodicity

The fact that many sampling schemes operate repeatedly on a block-by-block basis leads to the concept of periodicity (see Sec. I-C). Note, however, that subsequent stages may process multiple blocks jointly. Assume that $\boldsymbol{x}$ is partitioned into $B$ blocks of $N = L/B$ samples as[7] $\boldsymbol{x} = [\boldsymbol{x}[0]^T, \cdots, \boldsymbol{x}[B-1]^T]^T$, with $\boldsymbol{x}[b] \in \mathbb{C}^N \ \forall b$ and that sampling a block with $N$ elements results in another block with $M$ elements:

$$\boldsymbol{y}[b] = \boldsymbol{\Phi}\boldsymbol{x}[b], \quad b = 0, 1, \ldots, B-1, \quad (17)$$

where $\boldsymbol{y}[b] \in \mathbb{C}^M$ and $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$. The use of the term *periodicity* owes to the fact that the matrix $\boldsymbol{\Phi}$ does not depend on $b$. By making $\boldsymbol{y} = [\boldsymbol{y}[0]^T, \cdots, \boldsymbol{y}[B-1]^T]^T$ and

$$\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}, \quad (18)$$

expression (17) results in (1). From (18), it also follows that the matrices in $\bar{\mathcal{S}}$ are block Toeplitz with $M \times M$ blocks.

Since $K = MB$, the compression ratio in the periodic setting takes the form

$$\rho = \frac{L}{K} = \frac{N}{M}. \quad (19)$$

Further conventions are useful when dealing with sparse sampling, in which case, as seen in Sec. I-B, $\bar{\boldsymbol{\Phi}}$ equals a

---

[6]Recall the conventions introduced in Sec. I-E.
[7]For simplicity, we assume that $L$ is an integer multiple of $B$.

---

submatrix of $\boldsymbol{I}_L$ up to row permutations. For concreteness, assume that the rows of $\bar{\boldsymbol{\Phi}}$ are ordered as they are in $\boldsymbol{I}_L$. If $\mathcal{K} = \{l_0, \cdots, l_{K-1}\}$ denotes the set containing the indices of the non-null columns of $\bar{\boldsymbol{\Phi}}$, the entries of $\boldsymbol{y} = \bar{\boldsymbol{\Phi}}\boldsymbol{x}$ are given by $y_k = x_{l_k}$, $l_k \in \mathcal{K}$, where $\boldsymbol{x} = [x_0, \cdots, x_{L-1}]^T$ and $\boldsymbol{y} = [y_0, \cdots, y_{K-1}]^T$. The set $\mathcal{M}$, which contains the indices of the non-null columns $\boldsymbol{\Phi}$, is related to $\mathcal{K}$ by

$$\mathcal{K} = \{m + bN, \ m \in \mathcal{M}, b = 0, 1, \ldots, B-1\}. \quad (20)$$

Loosely speaking, we say that $\mathcal{K}$ is periodic with period $\mathcal{M}$. These sets have $|\mathcal{K}| = K = MB$ and $|\mathcal{M}| = M$ elements.

Note that periodic sampling indeed generalizes non-periodic sampling, since the latter can be retrieved just by making $B = 1$. For this reason, most results will be presented for periodic samplers, with occasional comments on the non-periodic setting if needed.

### III. DESIGN OF COVARIANCE SAMPLERS

The results in this section allow to determine whether a matrix $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler or not, and provide useful means to design these matrices for a given $\mathcal{S}$. They are based on the following basic result from linear algebra:

*Lemma 3:* Let $\mathcal{S} = \{\boldsymbol{\Sigma}_0, \cdots, \boldsymbol{\Sigma}_{S-1}\}$ be a set of Hermitian matrices. If $\mathcal{S}$ is linearly independent when considering real coefficients, that is,

$$\sum_{s=0}^{S-1} \alpha_s \boldsymbol{\Sigma}_s = \boldsymbol{0}, \quad \alpha_s \in \mathbb{R} \ \Rightarrow \alpha_s = 0 \ \forall s, \quad (21)$$

then it is also independent when considering coefficients in $\mathbb{C}$, i.e., (21) also applies when $\alpha_s \in \mathbb{C}$.

*Proof:* It easily follows by combining expression (21) with the fact that $\boldsymbol{\Sigma}_s = \boldsymbol{\Sigma}_s^H$, $\forall s$. ∎

The importance of this basic fact is that it allows us to focus on the complex extension of $\phi$, defined as

$$\begin{array}{ccc} \operatorname*{span}_{\mathbb{C}} \mathcal{S} & \xrightarrow{\phi_{\mathbb{C}}} & \operatorname*{span}_{\mathbb{C}} \bar{\mathcal{S}} \\ \boldsymbol{\Sigma} & \longrightarrow & \phi_{\mathbb{C}}(\boldsymbol{\Sigma}) = \bar{\boldsymbol{\Phi}}\boldsymbol{\Sigma}\bar{\boldsymbol{\Phi}}^H. \end{array} \quad (22)$$

In other words, $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler iff $\phi_{\mathbb{C}}$ is an invertible function. An equivalent statement is provided by the following lemma, which is the basic tool to be used in the design of covariance samplers.

*Lemma 4:* Let $\ker \phi_{\mathbb{C}}$ denote the set of matrices $\boldsymbol{\Sigma} \in \operatorname{span}_{\mathbb{C}} \mathcal{S}$ satisfying $\phi_{\mathbb{C}}(\boldsymbol{\Sigma}) = \boldsymbol{0}$. Then, a matrix $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler if and only if $\ker \phi_{\mathbb{C}} = \{\boldsymbol{0}\}$.

*Proof:* It is an immediate consequence of Definition 1 and Lemma 3. ∎

### A. Design of Sparse Samplers

Designing sparse samplers involves manipulating *difference sets*, which contain all possible distances between elements of another set:

*Definition 3:* The difference set of $\mathcal{A} \subset \mathbb{Z}$, denoted as $\Delta(\mathcal{A})$, is defined as:

$$\Delta(\mathcal{A}) = \{\delta \geq 0 : \exists a_1, a_2 \in \mathcal{A} \text{ s.t. } \delta = a_2 - a_1\}. \quad (23)$$

Note that the difference set considers no repetition of elements, i.e., every distance shows up at most once. The cardinality of $\Delta(\mathcal{A})$ is upper bounded by one plus the number of unordered subsets of $\mathcal{A}$ with two elements:

$$|\Delta(\mathcal{A})| \leq \frac{|\mathcal{A}| \cdot (|\mathcal{A}| - 1)}{2} + 1, \qquad (24)$$

where the $+1$ term accounts for the fact that $0 \in \Delta(\mathcal{A})$ for any non-empty $\mathcal{A}$.

The correlation vector $\boldsymbol{\sigma}_s$ associated with the HT matrix $\boldsymbol{\Sigma}_s$ is defined as the the first column of $\boldsymbol{\Sigma}_s$. The following theorem is a quick method to verify whether a sparse sampler defined by a set $\mathcal{K}$ is a covariance sampler.

*Theorem 1:* Let $\mathcal{S} = \{\boldsymbol{\Sigma}_s\}_{s=0}^{S-1}$ be a linearly independent set of HT matrices, let $\{\boldsymbol{\sigma}_s\}_{s=0}^{S-1}$ be the associated set of correlation vectors, and let $\tilde{\boldsymbol{\sigma}}_s$ be the vector whose entries are the elements of $\boldsymbol{\sigma}_s$ indexed by $\Delta(\mathcal{K})$. Then, $\mathcal{K}$ defines an $\mathcal{S}$-covariance sampler if and only if $\mathrm{rank}\,\boldsymbol{R} = S$, where

$$\boldsymbol{R} = \left[ \begin{array}{cccc} \tilde{\boldsymbol{\sigma}}_0 & \tilde{\boldsymbol{\sigma}}_1 & \dots & \tilde{\boldsymbol{\sigma}}_{S-1} \\ \tilde{\boldsymbol{\sigma}}_0^* & \tilde{\boldsymbol{\sigma}}_1^* & \dots & \tilde{\boldsymbol{\sigma}}_{S-1}^* \end{array} \right]. \qquad (25)$$

*Proof:* Observe that $\bar{\boldsymbol{\Sigma}}_s$ contains an element from the $\delta$-th diagonal of $\boldsymbol{\Sigma}_s$ iff $|\delta| \in \Delta(\mathcal{K})$. Now vectorize the matrices in $\bar{\mathcal{S}}$ and arrange these vectors as columns of a matrix. By removing repeated rows and duplicating the row corresponding to the main diagonal we obtain $\boldsymbol{R}$. Therefore, the number of linearly independent columns in $\boldsymbol{R}$ equals the number of linearly independent matrices in $\bar{\mathcal{S}}$. The result follows from Lemma 4 by noting that $\ker \phi_{\mathbb{C}} = \{\boldsymbol{0}\}$ iff $\mathrm{rank}\,\boldsymbol{R} = S$. ∎

From (25), it is easy to conclude[8] that $2|\Delta(\mathcal{K})| - 1 \geq S$ in order for $\boldsymbol{R}$ to be full column rank. Combining this expression with (24) results in the following necessary condition for $\mathcal{K}$ to define a covariance sampler:

$$K \cdot (K-1) + 1 \geq S. \qquad (26)$$

### B. Design of Dense Samplers

Designing sampling matrices is oftentimes involved due to the nature of the design criteria. In many cases, it is convenient to draw $\bar{\boldsymbol{\Phi}}$ at random using a distribution that provides an admissible sampler with a certain probability [9], [51]. Following this idea, this paper employs probabilistic techniques to obtain optimal designs for dense samplers.

These techniques provide sampling matrices with an acceptable behavior without considering any structure of the covariance subspace other than its dimension. The next result establishes the minimum size of a random matrix $\boldsymbol{\Phi}$ to define a covariance sampler. The only requirement is that this matrix be drawn from a continuous probability distribution.

*Theorem 2:* Let $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$, with $M \leq N$, be a random matrix with a continuous probability distribution.[9] Then, with probability one, the matrix $\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}$ defines an $\mathcal{S}$-covariance sampler if and only if $S \leq M^2(2B-1)$, where $S$ is the cardinality of the HT basis set $\mathcal{S}$.

---

[8]Note the existence of a duplicate row in $\boldsymbol{R}$.

[9]Formally, we say that a distribution $\mu$ is continuous if it is *absolutely continuous with respect to Lebesgue measure*, that is, $\mu(B) = 0$ for all Borel sets of $\mathbb{C}^{M \times N}$ with zero Lebesgue measure [53]. Intuitively, this means that there are no probability *masses*.

*Proof:* See Appendix B. ∎

Note that the matrices in $\mathrm{span}_{\mathbb{R}}\,\bar{\mathcal{S}}$ are Hermitian and block Toeplitz with $M \times M$ blocks. It can be seen that the dimension of such a subspace is at most $M^2(2B-1)$, which is exactly the one achieved by the random design from Theorem 2 when $S = M^2(2B-1)$ (see Sec. II-A). Therefore, no other design can achieve a higher compression ratio.

## IV. UNIVERSAL COVARIANCE SAMPLERS

After having laid the mathematical framework, we are ready to provide designs that result in covariance samplers independently of which basis of HT matrices is considered. The first result of this section reduces the task of checking whether a given matrix defines a covariance sampler for all possible bases to that of checking just for one.

*Lemma 5:* Let $\mathcal{S}$ be a basis for $\mathbb{S}^L$. Then, a sampler $\bar{\boldsymbol{\Phi}}$ is universal if and only if it is an $\mathcal{S}$-covariance sampler.

*Proof:* Clearly, if $\bar{\boldsymbol{\Phi}}$ is universal, it is also an $\mathcal{S}$-covariance sampler. Conversely, if $\bar{\boldsymbol{\Phi}}$ is an $\mathcal{S}$-covariance sampler, it is also an $\mathcal{S}'$-covariance sampler for any basis $\mathcal{S}'$ of HT matrices since the restriction of an injective map is always injective. ∎

The rest of this section applies this result to obtain sparse and dense universal covariance samplers.

### A. Sparse Samplers

The next necessary and sufficient condition for a sparse sampler to be universal basically states that all autocorrelation lags must be identifiable from the compressed observations.

*Theorem 3:* The set $\mathcal{K} \subset \{0, \dots, L-1\}$ defines a universal covariance sampler if and only if $\Delta(\mathcal{K}) = \{0, \dots, L-1\}$.

*Proof:* Consider the basis $\mathcal{S}_T$ from (9). If $\Delta(\mathcal{K}) = \{0, \dots, L-1\}$, the matrix $\boldsymbol{R}$ from Theorem 1 becomes

$$\boldsymbol{R} = \left[ \begin{array}{cc} \boldsymbol{I}_L & -\jmath\tilde{\boldsymbol{I}}_L \\ \boldsymbol{I}_L & \jmath\tilde{\boldsymbol{I}}_L \end{array} \right], \qquad (27)$$

where $\tilde{\boldsymbol{I}}_L$ is the submatrix of $\boldsymbol{I}_L$ that results from removing the first column. Since $\boldsymbol{R}$ has rank $2L-1$, $\mathcal{K}$ defines an $\mathcal{S}_T$-covariance sampler and, due to Lemma 5, it is universal.

If one or more elements of $\{0, \dots, L-1\}$ are missing in $\Delta(\mathcal{K})$, at least two of the rows of $\boldsymbol{R}$ are missing, meaning that $\mathrm{rank}\,\boldsymbol{R} < 2L-1$. Then, $\mathrm{rank}\,\boldsymbol{R} = 2L-1$ iff $\{0, \dots, L-1\} \subset \Delta(\mathcal{K})$. From Theorem 1, $\mathcal{K}$ defines an $\mathcal{S}_T$-covariance sampler iff $\Delta(\mathcal{K}) = \{0, \dots, L-1\}$. Now apply Lemma 5. ∎

This theorem provides a very simple means to check whether $\mathcal{K}$ is universal or not. Interestingly, this is closely related to the classical problem in number theory known as the *sparse ruler problem*, or as the representation of integers by difference bases (see [40], [54] and references therein). Its application to array processing dates back to the 60's [11].

*Definition 4:* A length-$(L-1)$ (linear) sparse ruler is a set $\mathcal{K} \subset \{0, 1, \dots, L-1\}$ satisfying $\Delta(\mathcal{K}) = \{0, 1, \dots, L-1\}$. It is called minimal if there exists no other length-$(L-1)$ sparse ruler with smaller cardinality.

Intuitively, we may associate this set with a classical ruler (the physical object) with some marks erased, which is still capable of measuring all integer distances between 0 and its length using pairs of marks. Two examples of minimal sparse
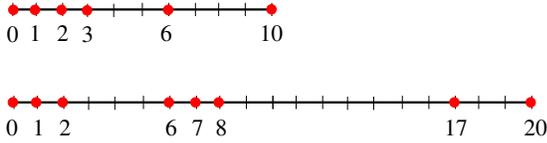
Fig. 1: Example of a length-10 minimal sparse ruler (above) and length-20 minimal sparse ruler (below).

rulers are shown in Fig. 1, where red dots correspond to the marks that have not been erased. Sparse rulers exist for all $L$, although they are not necessarily unique. For instance, two different length-10 sparse rulers are $\{0, 1, 2, 3, 6, 10\}$ and $\{0, 1, 2, 5, 7, 10\}$. The most remarkable properties of a length-$(L-1)$ sparse ruler are that the endpoints are always present, i.e., $\{0, L-1\} \subset \mathcal{K}$, and that its reflection $(L-1) - \mathcal{K} = \{(L-1) - k : k \in \mathcal{K}\}$ is also a sparse ruler. Trivially, if $\mathcal{K}$ is minimal, then $(L-1) - \mathcal{K}$ is also minimal. Therefore, (minimal) sparse rulers exist at least in pairs unless $\mathcal{K} = (L-1) - \mathcal{K}$. The cardinality $K = |\mathcal{K}|$ of a sparse ruler is lower bounded as

$$K \geq \frac{1}{2} + \sqrt{2(L-1) + \frac{1}{4}}, \qquad (28)$$

which follows directly from (24) and is only attained for $L-1 = 0, 1, 3$ and $6$ (see e.g. [42]); or as (see [39], [40]):

$$K \geq \sqrt{\tau(L-1)}, \qquad (29)$$

where $\tau = \max_\theta 2(1 - \frac{\sin\theta}{\theta}) \approx 2.4345$; and, if it is minimal it is upper bounded by [18]:

$$K \leq \left\lceil \sqrt{3(L-1)} \right\rceil, \quad L-1 \geq 3. \qquad (30)$$

Thus, in the non-periodic case ($B = 1$), Theorem 3 reduces our design problem to finding a length-$(L-1)$ sparse ruler, for which design algorithms abound. A trivial example is $\{0, \ldots, L-1\}$, which clearly represents a universal sampler since in that case $\boldsymbol{y} = \boldsymbol{x}$. More sophisticated constructions were discussed in [18], [23], [39]–[43]. However, if the compression ratio is to be maximized, then one should look for a *minimal* sparse ruler, which is an exhaustive-search problem. Fortunately, there exist tables for values of $L-1$ up to the order of 100. Although higher values of this parameter demand, in principle, intensive computation, one may resort to the designs in [18], [41], [42], which provide nearly minimal rulers despite being really simple.

On the other hand, it is not clear how to design sampling patterns in the periodic case ($B > 1$) since periodicity needs to be enforced on $\mathcal{K}$. Before that, the next definition is required.

*Definition 5:* A length-$(L-1)$ periodic sparse ruler of period $N$, where $N$ divides $L$, is a set $\mathcal{K} \subset \{0, 1, \ldots, L-1\}$ satisfying two conditions:

1) if $k \in \mathcal{K}$, then $k + bN \in \mathcal{K}$ for all $b \in \mathbb{Z}$ such that $0 \leq k + bN < L$
2) $\Delta(\mathcal{K}) = \{0, 1, \ldots, L-1\}$.

*It is called minimal if there exists no other periodic sparse ruler with the same length and period but smaller cardinality.*

Observe that any periodic sparse ruler is also a sparse ruler, whereas the converse need not be true. Clearly, Theorem 3

could be rephrased to say that $\mathcal{K}$ is universal iff it is a length-$(NB-1)$ periodic sparse ruler of period $N$. The problem of designing sparse covariance samplers becomes that of designing periodic sparse rulers. The next result simplifies this task by stating that a length-$(NB-1)$ periodic sparse ruler of period $N$ is indeed the concatenation of $B$ length-$(N-1)$ sparse rulers:

*Theorem 4:* A set $\mathcal{K}$ is a periodic sparse ruler of length $NB-1$ and period $N$ if and only if there exists a sparse ruler $\mathcal{M}$ of length $N-1$ such that

$$\mathcal{K} = \{m + bN : \quad m \in \mathcal{M}, \quad b = 0, 1, \ldots, B-1\}. \qquad (31)$$

*Proof:* See Appendix D. ∎

One of the consequences of Theorem 4 is that increasing the number of blocks in a periodic sparse sampler cannot improve the compression ratio. For example, concatenating two equal length-$(N-1)$ minimal sparse rulers with $M$ elements results in a length-$(2N-1)$ sparse ruler with $2M$ elements. Note, however, that the situation is different if the periodicity requirement is dropped. For instance, a minimal length-10 sparse ruler has 6 elements, whereas a length-21 minimal sparse ruler has $8 < 6 \times 2$ elements.

As a corollary of Theorem 4, we conclude that a minimal periodic sparse ruler is the concatenation of minimal sparse rulers. Thus, the problem of designing optimal sparse universal covariance samplers (either periodic or non-periodic) reduces to designing a minimal length-$(N-1)$ sparse ruler $\mathcal{M}$.

Table I illustrates the minimum value of $M = |\mathcal{M}|$ (labeled as $M_{\mathrm{LSR}}$) for several values of $N$, enabling us to obtain the optimum compression ratio for block lengths $N$ up to 60, which covers most practical cases. For higher $N$, one may resort to another table, to a computer program, or to the asymptotic considerations in Sec. VI. However, although there is no closed form expression for the maximum achievable compression ratio $\rho$, the bounds in (29) and (30) show that

$$\frac{N}{\left\lceil \sqrt{3(N-1)} \right\rceil} \leq \rho \leq \frac{N}{\sqrt{\tau(N-1)}}. \qquad (32)$$

### B. Dense Samplers

Deriving conditions for universality of dense samplers is simpler than for sparse samplers since most mathematical complexity has been subsumed by Theorem 2. Moreover, the results are simpler and can be expressed in closed form.

*Theorem 5:* Let $\boldsymbol{\Phi}$ be an $M \times N$ random matrix satisfying the hypotheses of Theorem 2. Then, $\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}$ defines a universal covariance sampler with probability 1 if and only if

$$M \geq \sqrt{\frac{2NB-1}{2B-1}}. \qquad (33)$$

*Proof:* If $\mathcal{S}$ is a basis for $\mathbb{S}^L$, then $|\mathcal{S}| = 2L-1 = 2NB-1$. From Theorem 2, $\bar{\boldsymbol{\Phi}}$ is an $\mathcal{S}$-covariance sampler iff $2NB-1 \leq M^2(2B-1)$, which is equivalent to (33). Universality then follows from Lemma 5. ∎

Expression (33) can be interpreted as $M^2(2B-1) \geq 2NB-1$, where $2NB-1$ is the dimension of the uncompressed subspace and $M^2(2B-1)$ is the maximum dimension

| $N$ | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $M_{\mathrm{CSR}}$ | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 |
| $M_{\mathrm{HLSR}}$ | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 |
| $M_{\mathrm{LSR}}$ | 4 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 7 | 7 |

| $N$ | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $M_{\mathrm{CSR}}$ | 5 | 5 | 5 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 6 |
| $M_{\mathrm{HLSR}}$ | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| $M_{\mathrm{LSR}}$ | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 9 | 9 | 9 |

| $N$ | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $M_{\mathrm{CSR}}$ | 6 | 7 | 7 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 8 |
| $M_{\mathrm{HLSR}}$ | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 8 | 8 | 8 |
| $M_{\mathrm{LSR}}$ | 9 | 9 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 11 |

| $N$ | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $M_{\mathrm{CSR}}$ | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| $M_{\mathrm{HLSR}}$ | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 9 | 9 |
| $M_{\mathrm{LSR}}$ | 11 | 11 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 |

| $N$ | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $M_{\mathrm{CSR}}$ | 8 | 8 | 9 | 9 | 9 | 9 | 9 | 8 | 9 | 9 | 9 |
| $M_{\mathrm{HLSR}}$ | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 10 |
| $M_{\mathrm{LSR}}$ | 12 | 12 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 14 |



(a) Example of length-20 minimal circular sparse ruler.

(b) Example of length-20 circular sparse ruler designed with a length-10 *linear* sparse ruler.

Fig. 2: Comparison of two length-20 circular sparse rulers. The ruler on the left, with 5 elements, is minimal whereas the one on the right, with 6 elements, is not.

TABLE I: Values of $M$ for a length-$(N-1)$ minimal circular sparse ruler ($M_{\mathrm{CSR}}$), length-$\lfloor \frac{N}{2} \rfloor$ minimal linear sparse ruler ($M_{\mathrm{HLSR}}$) and length-$(N-1)$ minimal linear sparse ruler ($M_{\mathrm{LSR}}$).

of a subspace of Hermitian block-Toeplitz matrices. Thus, this design provides optimal compression, which is achieved when

$$M = \left\lceil \sqrt{\frac{2NB-1}{2B-1}} \right\rceil, \tag{34}$$

and given by

$$\rho = \frac{N}{M} \approx \sqrt{\frac{(2B-1)N^2}{2NB-1}}. \tag{35}$$

## V. NON-UNIVERSAL COVARIANCE SAMPLERS

Universal samplers are used when no structure exists or when it is unknown. However, when prior information is available, the values that $\bar{\Sigma}$ can take on are restricted, allowing for larger compression ratios. This section analyzes this effect for the covariance subspaces introduced in Sec. II-B. Since the Toeplitz subspace has already been considered in Sec. IV, we proceed to analyze circulant and $d$-banded subspaces.

### A. Circulant Covariance Subspace

*1) Sparse Samplers:* Restricting $\Sigma$ to be circulant yields considerable compression gains with respect to the Toeplitz case since the requirements on every period of $\mathcal{K}$ relax. In particular, every period must be a *circular* sparse ruler, which is a much weaker requirement than that of being a *linear* sparse ruler. This concept is related to the *modular* difference set defined next. Recall from Sec. I-E that $(x)_A$ denotes the remainder of the integer division of $x$ by $A$.

*Definition 6: Let $\mathcal{A}$ be a set of integers. The $A$-modular difference set of $\mathcal{A}$, denoted as $\Delta_A(\mathcal{A})$, is defined as*

$$\Delta_A(\mathcal{A}) = \{\delta \geq 0 : \exists a_1, a_2 \in \mathcal{A} \ s.t. \ \delta = (a_2 - a_1)_A\}. \tag{36}$$

Clearly, for any $\mathcal{A} \subset \{0, 1, \ldots, A-1\}$, we have that $\Delta(\mathcal{A}) \subset \Delta_A(\mathcal{A})$, which means that $|\Delta_A(\mathcal{A})|$ is never less than $|\Delta(\mathcal{A})|$. Actually, $\Delta_A(\mathcal{A})$ will typically be larger than $\Delta(\mathcal{A})$ since the fact that $\delta$ is in $\Delta_A(\mathcal{A})$ implies that $A - \delta$ is also in that set. For example, if $\mathcal{A} = \{0, 1, 5\}$ and $A = 10$, then $\Delta(\mathcal{A}) = \{0, 1, 4, 5\} \subset \Delta_{10}(\mathcal{A}) = \{0, 1, 4, 5, 6, 9\}$. Finally, the cardinality of the modular difference set is upper bounded by noting that any pair of elements in a set $\mathcal{A}$ with cardinality $|\mathcal{A}|$ generates at most two distances in $\Delta_A(\mathcal{A})$:

$$|\Delta_A(\mathcal{A})| \leq |\mathcal{A}| \cdot (|\mathcal{A}| - 1) + 1. \tag{37}$$

Now it is possible to state the requirements to compress circulant subspaces:

*Theorem 6: Let $\mathcal{S}_C$ be given by (13). Then, the set $\mathcal{K} \subset \{0, \ldots, L-1\}$ is an $\mathcal{S}_C$-covariance sampler if and only if $\Delta_L(\mathcal{K}) = \{0, \ldots, L-1\}$.*

*Proof:* See Appendix E. ∎

Theorem 6 is therefore the analogue of Theorem 3 for circulant subspaces. However, in this case the conclusion does not lead to a *linear* sparse ruler but to a *circular* one:

*Definition 7: A length-$(L-1)$ circular (or modular) sparse ruler is a set $\mathcal{K} \subset \{0, \ldots, L-1\}$ satisfying $\Delta_L(\mathcal{K}) = \{0, \ldots, L-1\}$; and it is said to be minimal if no other length-$(L-1)$ circular sparse ruler exists with smaller cardinality.*

As with linear sparse rulers, a geometric interpretation is possible in terms of a physical ruler. Suppose that we wrap around a conventional ruler (made of some flexible material) until the first mark and the last mark lie at unit distance, thus making a *circular ruler*. Now assume that some of the marks are erased, but that it is still possible to measure all distances between 0 and the length of the original ruler using pairs of marks. The advantage with respect to a linear ruler is that any pair of marks provides, in general, *two* distances, which are the lengths of the two circular segments that they define. Two length-20 circular sparse rulers are illustrated in Fig. 2, the one on the left being minimal. Other examples of length-$(L-1)$ circular sparse rulers are $\{0, \ldots, L-1\}$ and $\{0, \ldots, \lfloor \frac{L}{2} \rfloor\}$, which are referred to as *trivial* circular sparse rulers.

Circular sparse rulers, also known as *difference cycles*, were analyzed by the mathematical community using finite group theory and additive number theory (see [54] for an overview of the main results). Among the most remarkable properties,

we mention that a reflection of a circular sparse ruler is also a circular sparse ruler (see Sec. IV-A) and that any circular rotation of a circular sparse ruler $\mathcal{K}$, defined as

$$\mathcal{K}_{(i)} = \{(k+i)_L : \quad k \in \mathcal{K}\}, \quad i \in \mathbb{Z}, \tag{38}$$

is also a circular sparse ruler. Moreover, since $\Delta(\mathcal{K}) \subset \Delta_L(\mathcal{K})$ for any $\mathcal{K} \subset \{0, \ldots, L-1\}$, any *linear* sparse ruler is also a *circular* sparse ruler. Hence, the cardinality of a *minimal* circular sparse ruler can never be greater than the cardinality of a *minimal* linear sparse ruler if both have the same length. It is possible to go even further by noting that any length-$\lfloor \frac{L}{2} \rfloor$ linear sparse ruler is also a length-$(L-1)$ circular sparse ruler. For example, Fig. 2b shows a length-20 circular sparse ruler constructed with a length-10 *linear* sparse ruler. From this observation and (30), we obtain

$$|\mathcal{K}| \leq \left\lceil \sqrt{3 \left\lfloor \frac{L}{2} \right\rfloor} \right\rceil \tag{39}$$

for any minimal circular sparse ruler. On the other hand, expression (37) yields

$$|\mathcal{K}| \geq \frac{1}{2} + \sqrt{L - \frac{3}{4}}. \tag{40}$$

A length-$(L-1)$ circular sparse ruler can be designed in several ways. For certain values of $L$, minimal rulers attaining (40) can be obtained in closed form (see [55, Sec. III-B] for an overview; also [56]). 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 [27]. In fact, the latter provides optimal solutions for most values of $L$ below 60 (see Table I). Further alternatives include [57].

Circular sparse rulers seem to have been introduced in signal/array processing in [47] and used later in [34], [35], [45]. Theorem 6 basically states that a covariance sampler for circulant subspaces is a length-$(L-1)$ circular sparse ruler, which gives a practical design criterion just for the non-periodic case. We now move on to introduce periodicity:

*Definition 8:* A length-$(L-1)$ periodic circular sparse ruler of period $N$, where $N$ divides $L$, is a set $\mathcal{K} \subset \{0, 1, \ldots, L-1\}$ satisfying:
1) if $k \in \mathcal{K}$, then $k + bN \in \mathcal{K}$ for all $b \in \mathbb{Z}$ such that $0 \leq k + bN < L$;
2) $\Delta_L(\mathcal{K}) = \{0, 1, \ldots, L-1\}$.

*It is called minimal if there is no other periodic circular sparse ruler with the same length and period but smaller cardinality.*

Hence, Theorem 6 could be rephrased to say that $\mathcal{K}$ is an $\mathcal{S}_C$-covariance sampler iff it is a length-$(NB-1)$ periodic circular sparse ruler of period $N$. Although designing these rulers may seem difficult, the next result simplifies this task by stating that every period is, indeed, a circular sparse ruler.

*Theorem 7:* A set $\mathcal{K}$ is a periodic circular sparse ruler of length $NB-1$ and period $N$ if and only if there exists a circular sparse ruler $\mathcal{M}$ of length $N-1$ such that

$$\mathcal{K} = \{m + bN : \quad m \in \mathcal{M}, \quad b = 0, 1, \ldots, B-1\} \tag{41}$$

*Proof:* See Appendix F. ∎

Table I reveals that the cardinality $M$ of a minimal circular sparse ruler is not monotone with $N$. For example, minimal length-19 circular sparse rulers have 6 elements whereas minimal length-20 circular sparse rulers have 5 elements (see [54] for a proof). Table I also illustrates the compression gain due to the knowledge that $\boldsymbol{\Sigma}$ is circulant. For example, when $N = 60$, a universal sampler has a compression ratio of $\frac{N}{M} = \frac{60}{14} \approx 4.28$, whereas a covariance sampler for circulant subspaces has a compression ratio of $\frac{N}{M} = \frac{60}{9} \approx 6.67$.

Although maximum compression ratios cannot be expressed in closed form, simple bounds follow from (39) and (40):

$$\frac{N}{\left\lceil \sqrt{3 \lfloor \frac{N}{2} \rfloor} \right\rceil} \leq \rho \leq \frac{2N}{1 + \sqrt{4N-3}}. \tag{42}$$

*2) Dense Samplers:* As in universal sampling, designing dense samplers is much easier than designing sparse samplers. The following corollary of Theorem 2 follows by noting that any basis for the circulant subspace has $L = NB$ elements.

*Corollary 1:* Let $\boldsymbol{\Phi}$ be an $M \times N$ random matrix satisfying the hypotheses of Theorem 2 and let $\mathcal{S}_C$ be given by (13). Then, with probability one, the matrix $\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}$ defines an $\mathcal{S}_C$-covariance sampler if and only if

$$M \geq \sqrt{\frac{NB}{2B-1}}. \tag{43}$$

The optimum compression ratio is, therefore,

$$\rho = \frac{N}{M} \approx \sqrt{\frac{(2B-1)N}{B}}. \tag{44}$$

For large $B$, this represents an approximate gain of $\sqrt{2}$ with respect to the universal case.

### B. *d-banded Covariance Subspace*

*1) Sparse samplers:* The prior knowledge that $\boldsymbol{\Sigma}$ is $d$-banded may also provide compression gains. In particular, we will see that, for sparse samplers, $d$-banded subspaces with $N \leq d \leq N(B-1)$ are compressed like circulant subspaces.

*Theorem 8:* Let $\mathcal{S}_B^d$ be given by (16) with $N \leq d \leq N(B-1)$. Then, the set

$$\mathcal{K} = \{m + bN, \ m \in \mathcal{M}, b = 0, 1, \ldots, B-1\}, \tag{45}$$

where $\mathcal{M} \subset \{0, \ldots, N-1\}$, defines an $\mathcal{S}_B^d$-covariance sampler iff $\mathcal{M}$ is a length-$(N-1)$ circular sparse ruler.

*Proof:* See Appendix G. ∎

Observe that the condition $d \leq N(B-1)$ is a mild assumption since we are only requiring that the last $N-1$ lags of the associated autocorrelation sequence be zero.[10] Note as well that other cases rather than $N \leq d \leq N(B-1)$ may be considered, resulting in different conclusions. For example, in the non-periodic case ($B = 1$) it can be shown from Theorem 1 that the only requirement on $\mathcal{K}$ sampler is that $\Delta(\mathcal{K}) = \{0, \ldots, d\}$.

From Theorem 8 and Theorem 7, it follows that $\mathcal{K}$ must be a length-$(NB-1)$ periodic circular sparse ruler of period

---

[10]Strictly speaking, we only need the lags $NB - N + 1$ through $NB - 1$ to be zero since the lags greater than $NB - 1$ are not relevant in the model.

$N$, which means that samplers for $d$-banded subspaces mimic those for circulant subspaces. Thus, one should apply the design and compression ratio considerations from Sec. V-A1. Interestingly, note that the latter does not depend on $d$ provided that this parameter remains within the aforementioned limits.

*2) Dense Samplers:* From Theorem 2 and noting that $d$-banded subspaces have dimension $2d + 1$ we obtain:

*Corollary 2: Let $\boldsymbol{\Phi}$ be an $M \times N$ random matrix satisfying the hypotheses of Theorem 2 and let $\mathcal{S}_B^d$ be given by* (16). *Then, with probability one, the matrix $\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}$ defines an $\mathcal{S}_B^d$-covariance sampler if and only if*

$$M \geq \sqrt{\frac{2d + 1}{2B - 1}}. \tag{46}$$

According to this result, the maximum compression ratio is:

$$\rho = \frac{N}{M} \approx \sqrt{\frac{(2B - 1)N^2}{2d + 1}}, \tag{47}$$

which clearly improves the ratio in (35) since $d \leq NB - 1$.

## VI. ASYMPTOTIC REGIME

We next provide the optimal compression ratios for universal dense samplers and bound the optimal ratios for universal sparse samplers as $M$ and $N$ become larger.

- **Dense Samplers:** The maximum compression ratio $\rho_{\mathrm{DS}}$ of universal dense samplers is given by (35). Asymptotically in $N$, we have that $\rho_{\mathrm{DS}} \to \sqrt{\frac{2B-1}{2B}N}$, which becomes $\rho_{\mathrm{DS}} \to \sqrt{\frac{N}{2}}$ in the non-periodic case and $\rho_{\mathrm{DS}} \to \sqrt{N}$ if the number of periods $B$ also becomes large. Alternatively, we observe that $M \to \sqrt{\frac{2B}{2B-1}N}$ as $N$ becomes large, which means that $M \to \sqrt{2N}$ in the non-periodic case and $M \to \sqrt{N}$ as $B \to \infty$.

- **Sparse Samplers:** In [18], [40] it is established that the quotient $M^2/N$ asymptotically converges to a constant $c$, which is between[11] $\tau \approx 2.434$ and 3, with $M$ and $N - 1$ respectively denoting the cardinality and length of a minimal linear sparse ruler. Therefore, the asymptotic optimal compression ratio is given by

$$\rho_{\mathrm{SS}} \to \sqrt{\frac{N}{c}}. \tag{48}$$

In terms of $M$, this means that $M \to \sqrt{cN}$. Interestingly, if we use nested arrays [23], [43], the maximum achievable compression we can obtain for suitable choices of the parameters is $\rho_{\mathrm{NA}} \to \sqrt{\frac{N}{4}}$, which is therefore suboptimal. However, they present the advantage of having a simple design. The scheme in [18], [41] allows the simple construction of sparse rulers satisfying $M^2/N < 3$, which entail compression ratios greater than $\sqrt{\frac{N}{3}}$ even for finite $M$ and $N$.

To sum up, dense samplers provide better asymptotic compression ratios than sparse samplers. The compression loss between both approaches is quantified by the constant $c$, which

[11]As an informal guess, consider the length-90 minimal sparse ruler, which has 16 elements. A simple approximation yields $c \approx 16^2/91 \approx 2.8132$.

means that between 36% and 42% compression may be lost for large $B$ if we use sparse sampling instead of dense sampling. Similar observations arise for non-universal samplers by using the expressions in Sec. V.

Interestingly, these expressions show that the compression ratio can be made arbitrarily large just by increasing the number of observations. This conclusion agrees with [58].

## VII. DISCUSSION

The compression ratio was defined such that it is preserved for any number of realizations of $\boldsymbol{y}$ — note that each one is compressed using that ratio. In case of an arbitrarily large number of realizations, the maximum compression ratio separates consistency from inconsistency in the estimation. However, the notion of consistency is not truly meaningful in case of just one realization. For those cases, the values presented here provide simple guidelines to select suitable compression ratios and a guess of the quality of the estimation, in the sense that a good performance is expected when the actual compression ratio is much lower than the maximum one and *vice versa*.

## VIII. CONCLUSIONS

We have derived maximum compression ratios and optimal covariance samplers for a number of cases including Toeplitz, circulant, and banded covariance subspaces. The results were derived for the general periodic case, but they can be immediately particularized to the non-periodic setting. One of the effects observed is the convenience of having long blocks.

Two common schemes were considered: sparse and dense samplers. The design of optimal sparse samplers is related to the minimal sparse ruler problem, which is an exhaustive search problem with known near-optimal simple approximations. Some cases deal with *linear* and others with *circular* sparse rulers.

For dense samplers, the proposed random design is much simpler since it solely depends on the size of the compression matrix relative to the dimension of the covariance subspace. As opposed to the designs presented for sparse samplers, which result in samplers which are optimal *only* among the family of sparse samplers, the random designs proposed here result in samplers which are optimal in general, that is, no other covariance sampler (either dense or sparse) can do better.

## APPENDIX A
### PROOF OF LEMMA 1

Clearly, if $\phi$ is invertible so is $\phi_{|\mathcal{A}}$. In order to prove the converse statement, it suffices to show that $\phi$ is injective if $\phi_{|\mathcal{A}}$ is injective. This is a simple consequence of the definition of the codomains for both functions. Therefore, we need to prove that, given any two vectors $\boldsymbol{a} = [a_0, \cdots, a_{S-1}]^T$ and $\boldsymbol{b} = [b_0, \cdots, b_{S-1}]^T$ in $\mathbb{R}^S$, the matrices

$$\boldsymbol{\Sigma_a} = \sum_s a_s \boldsymbol{\Sigma}_s \quad \text{and} \quad \boldsymbol{\Sigma_b} = \sum_s b_s \boldsymbol{\Sigma}_s \tag{49}$$

must satisfy that

$$\phi(\boldsymbol{\Sigma_a}) = \phi(\boldsymbol{\Sigma_b}) \quad \Rightarrow \quad \boldsymbol{\Sigma_a} = \boldsymbol{\Sigma_b} \tag{50}$$

or, equivalently, that

$$\phi(\Sigma_a) = \phi(\Sigma_b) \quad \Rightarrow \quad a = b, \tag{51}$$

since $\mathcal{S}$ is linearly independent. To do so, let us take $S$ linearly independent vectors $\alpha_0, \cdots, \alpha_{S-1}$, where $\alpha_i = [\alpha_{i,0} \ldots \alpha_{i,S-1}]^T$, such that the $S$ matrices

$$\Sigma_{\alpha_i} = \sum_s \alpha_{i,s} \Sigma_s, \quad i = 0, \ldots, S-1 \tag{52}$$

are in $\mathcal{A}$. This operation is possible since $\dim_\mathbb{R}[\mathcal{A} \cap \mathrm{span}_\mathbb{R} \mathcal{S}] = S$. Moreover, since $\phi_{|\mathcal{A}}$ is injective and $\{\Sigma_{\alpha_i}\}_{i=0}^{S-1}$ is a linearly independent set of matrices, it follows that the matrices

$$\bar{\Sigma}_{\alpha_i} = \phi_{|\mathcal{A}}(\Sigma_{\alpha_i}) = \phi(\Sigma_{\alpha_i}) = \sum_s \alpha_{i,s} \bar{\Sigma}_s \tag{53}$$

also form an independent set of matrices. On the other hand, since the $S$ vectors $\alpha_i$ constitute a basis for $\mathbb{R}^S$, it is possible to write $a$ and $b$ as:

$$a = \sum_i \tilde{a}_i \alpha_i \quad \text{and} \quad b = \sum_i \tilde{b}_i \alpha_i, \tag{54}$$

for some $\tilde{a}_i, \tilde{b}_i \in \mathbb{R}$, which in turn means that

$$\Sigma_a = \sum_i \tilde{a}_i \Sigma_{\alpha_i} \quad \text{and} \quad \Sigma_b = \sum_i \tilde{b}_i \Sigma_{\alpha_i} \tag{55}$$

or

$$\phi(\Sigma_a) = \sum_i \tilde{a}_i \bar{\Sigma}_{\alpha_i} \quad \text{and} \quad \phi(\Sigma_b) = \sum_i \tilde{b}_i \bar{\Sigma}_{\alpha_i}. \tag{56}$$

Noting that the matrices $\bar{\Sigma}_{\alpha_i}$ are linearly independent leads to the statement

$$\phi(\Sigma_a) = \phi(\Sigma_b) \quad \Rightarrow \quad \tilde{a}_i = \tilde{b}_i \ \forall i, \tag{57}$$

which is equivalent to (51), thus concluding the proof.

## APPENDIX B
## PROOF OF THEOREM 2

In order to show Theorem 2 we will proceed by computing the dimension of $\ker \phi_\mathbb{C}$, and deriving the conditions under which $\dim \ker \phi_\mathbb{C} = 0$, which, in virtue of Lemma 4, are the conditions determining whether $\bar{\Phi}$ defines a covariance sampler. However, since the direct computation of $\ker \phi_\mathbb{C}$ is not a simple task, we perform several intermediate steps. First, we compute $\ker \tilde{\phi}_\mathbb{C}$, where $\tilde{\phi}_\mathbb{C}$ is defined as the extension of $\phi_\mathbb{C}$ to $\mathbb{C}^{L \times L}$:

$$\begin{array}{ccc} \mathbb{C}^{NB \times NB} & \xrightarrow{\tilde{\phi}_\mathbb{C}} & \mathbb{C}^{MB \times MB} \\ \Sigma & \longrightarrow & \bar{\Sigma} = \bar{\Phi} \Sigma \bar{\Phi}^H \end{array} \tag{58}$$

We later compute $\dim \ker \phi_\mathbb{C}$ by successive intersections as

$$\ker \phi_\mathbb{C} = \mathrm{span}_\mathbb{C} \mathcal{S} \cap \left( \mathbb{T}^{NB} \cap \left( \mathbb{B}^{N,B} \cap \ker \tilde{\phi}_\mathbb{C} \right) \right), \tag{59}$$

where $\mathbb{T}^{NB}$ represents the set of (not necessarily Hermitian) $NB \times NB$ Toeplitz matrices and $\mathbb{B}^{N,B}$ represents the set of $NB \times NB$ matrices with Toeplitz $N \times N$ blocks. The matrices in $\mathbb{B}^{N,B}$ can thus be written as

$$\begin{bmatrix} A_{0,0} & , \cdots, & A_{0,B-1} \\ \vdots & & \vdots \\ A_{B-1,0} & , \cdots, & A_{B-1,B-1} \end{bmatrix} \tag{60}$$

where the blocks $A_{b,p} \in \mathbb{C}^{N \times N}$ are Toeplitz. Expression (59) results from the fact that $\ker \phi_\mathbb{C} = \mathrm{span}_\mathbb{C} \mathcal{S} \cap \ker \tilde{\phi}_\mathbb{C}$ and

$$\mathrm{span}_\mathbb{C} \mathcal{S} \subset \mathbb{T}^{NB} \subset \mathbb{B}^{N,B}. \tag{61}$$

On the other hand, the requirement that the probability measure is absolutely continuous means that the probability that any row (or column) of $\Phi$ is in a given subspace of dimension less than $N$ (resp. $M$) is zero. Another consequence is that $\mathrm{rank}\, \Phi = M \le N$ with probability one and, as a result, the (right) null space of $\Phi$ has dimension $N - M$. Let us denote by $V$ an $N \times (N - M)$ matrix whose columns span this null space. Due to the properties of $\Phi$, it is clear that the probability that the columns of $V$ are contained in a given subspace of dimension less than $N$ is zero.

We start by computing a basis for $\ker \tilde{\phi}_\mathbb{C}$ in terms of $V$.

*Lemma 6:* Let $E_{i,j} \in \mathbb{C}^{B \times B}$ be the matrix with all entries set to zero but the $(i,j)$-th entry, which is one, and let $e_k$ denote the $k$-th column of the identity matrix $I_N$. Let also $\tilde{\phi}_\mathbb{C}$ be defined as in (58), and let the columns of $V = [v_0, \cdots, v_{N-M-1}] \in \mathbb{C}^{N \times (N-M)}$ form a basis for the null space of $\Phi$. Then, a basis for $\ker \tilde{\phi}_\mathbb{C}$ is given by

$$\mathcal{W} = \bigcup_{i=0}^{B-1} \bigcup_{j=0}^{B-1} \mathcal{W}_{i,j}, \tag{62}$$

*where*

$$\mathcal{W}_{i,j} = \Big\{ E_{i,j} \otimes e_k \otimes v_l^H, \tag{63}$$

$$k = 0, 1, \ldots, N-1, \ l = 0, 1, \ldots, N-M-1 \Big\}$$

$$\cup \Big\{ E_{i,j} \otimes e_k^H \otimes v_l, \tag{64}$$

$$k = 0, 1, \ldots, M-1, \ l = 0, 1, \ldots, N-M-1 \Big\}.$$

*Proof:* See Appendix C. ∎

Now let us evaluate the intersection $\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_\mathbb{C}$, which means that we must look for the matrices in $\ker \tilde{\phi}_\mathbb{C}$ whose $N \times N$ blocks have a Toeplitz structure. For the sake of simplicity, let us proceed block-by-block by separately considering the subspaces generated by each $\mathcal{W}_{i,j}$. Clearly, the matrices in $\mathrm{span}_\mathbb{C} \mathcal{W}_{i,j}$ can have, at most, a single non-null $N \times N$ block, which is the $(i,j)$-th block. This block is in the subspace generated by the following basis:

$$\{ e_k \otimes v_l^H, \ k = 0, 1, \ldots, N-1,$$
$$l = 0, 1, \ldots, N-M-1 \}$$
$$\cup \{ e_k^H \otimes v_l, \ k = 0, 1, \ldots, M-1,$$
$$l = 0, 1, \ldots, N-M-1 \}.$$

Therefore, all blocks in this subspace can be written in terms of this basis as

$$\sum_k \sum_l \alpha_{k,l} (e_k \otimes v_l^H) + \sum_k \sum_l \beta_{k,l} (e_k^H \otimes v_l) \tag{65}$$

for some $\alpha_{k,l} \in \mathbb{C}$ and $\beta_{k,l} \in \mathbb{C}$. The blocks with Toeplitz structure must necessarily satisfy

$$\sum_{n=-N+1}^{N-1} \gamma_n \boldsymbol{P}_n = \sum_{k=0}^{N-1} \sum_{l=0}^{N-M-1} \alpha_{k,l}(\boldsymbol{e}_k \otimes \boldsymbol{v}_l^H) \qquad (66)$$
$$+ \sum_{k=0}^{M-1} \sum_{l=0}^{N-M-1} \beta_{k,l}(\boldsymbol{e}_k^H \otimes \boldsymbol{v}_l)$$

for some $\gamma_n \in \mathbb{C}$, where $\boldsymbol{P}_n$ equals $\boldsymbol{J}_N^n$ for $n \geq 0$ and $(\boldsymbol{J}_N^{-n})^T$ for $n < 0$, with $\boldsymbol{J}_N$ defined in Sec. II-B.

Expression (66) represents a system of linear equations in $\alpha_{k,l}$, $\beta_{k,l}$ and $\gamma_n$, with $N^2 - M^2 + 2N - 1$ unknowns and $N^2$ equations. On the other hand, since $\boldsymbol{\Phi}$, and consequently $\boldsymbol{V}$, follow a continuous distribution, it follows that there are $\min(N^2, N^2 - M^2 + 2N - 1)$ independent matrices in (66). Consequently, if $N^2 \geq N^2 - M^2 + 2N - 1$ the only solution is just the zero matrix, and $\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}} = \{\boldsymbol{0}\}$, which in turn means that $\ker \phi_{\mathbb{C}} = \{\boldsymbol{0}\}$. Therefore, a sufficient condition for $\bar{\boldsymbol{\Phi}}$ to define a covariance sampler (see Lemma 4) is

$$M^2 \geq 2N - 1. \qquad (67)$$

Conversely, if $N^2 < N^2 - M^2 + 2N - 1$ the subspace of solutions has dimension $N^2 - M^2 + 2N - 1 - N^2 = 2N - M^2 - 1$. Therefore, the blocks of the matrices in $\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}$ can be written as a linear combination of $2N - M^2 - 1$ Toeplitz matrices $\boldsymbol{M}_k$. By considering all blocks, it follows that $\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}$ is generated by the following basis:

$$\Big\{ \boldsymbol{E}_{i,j} \otimes \boldsymbol{M}_k, \quad i,j = 0,1,\ldots,B-1;$$
$$k = 0,1,\ldots,2N-M^2-2 \Big\}. \qquad (68)$$

Thus, any matrix in $\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}$ can be written as

$$\boldsymbol{\Sigma} = \sum_{i,j,k} \eta_{i,j}^k \boldsymbol{E}_{i,j} \otimes \boldsymbol{M}_k. \qquad (69)$$

Now we compute the dimension of $\mathbb{T}^{NB} \cap \left(\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}\right)$. First note that $\dim(\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}) = B^2(2N - M^2 - 1)$. In order for $\boldsymbol{\Sigma} \in \mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}$ to be Toeplitz, we require that $\eta_{i,j}^k$ only depend on the difference $i - j$, which reduces the dimension of this space to $2N - M^2 - 1$ times the number of block diagonals, i.e., $(2N - M^2 - 1)(2B - 1)$. Moreover, since any two adjacent block diagonals share $N - 1$ diagonals, this imposes $(2B - 2)(N - 1)$ additional equations and results in

$$\dim\left(\mathbb{T}^{NB} \cap \left(\mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}\right)\right) \qquad (70)$$
$$= (2N - M^2 - 1)(2B - 1) - (2B - 2)(N - 1).$$

At this point, note that $\mathbb{T}^{NB}$ is the smallest subspace containing of both $\mathbb{T}^{NB} \cap \mathbb{B}^{N,B} \cap \ker \tilde{\phi}_{\mathbb{C}}$ and $\mathrm{span}_{\mathbb{R}} \mathcal{S}$. Since $\boldsymbol{\Phi}$ was generated according to a continuous distribution, then with probability one these two subspaces will not overlap (except for the zero matrix) unless the sum of their dimensions exceeds the dimension of the parent subspace, which is $2NB - 1$. Therefore, $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler if and only if

$$(2N - M^2 - 1)(2B - 1) - (2B - 2)(N - 1) + S \leq 2NB - 1$$

or, equivalently

$$S \leq M^2(2B - 1). \qquad (71)$$

It remains only to show that one only needs to look at (71) in order to assess whether a matrix $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler, the condition in (67) being completely irrelevant. This follows from the fact that (67) implies (71). Indeed, if we multiply both sides of (67) by $(2B - 1)$, we obtain

$$M^2(2B - 1) \geq (2N - 1)(2B - 1) \qquad (72)$$
$$= (2NB - 1) + 2(N - 1)(B - 1) \qquad (73)$$
$$\geq (2NB - 1) \geq S \qquad (74)$$

where the second inequality follows from the fact that $(N - 1)(B - 1) \geq 0$ and the third one is a consequence of the linear independence of $\mathcal{S}$. Therefore, (67) implies (71), and $\bar{\boldsymbol{\Phi}}$ defines a covariance sampler if and only if (71) holds.

## APPENDIX C
## PROOF OF LEMMA 6

Computing $\ker \tilde{\phi}_{\mathbb{C}}$ amounts to finding a basis for the subspace of matrices $\boldsymbol{\Sigma}$ in $\mathbb{C}^{NB \times NB}$ satisfying $\bar{\boldsymbol{\Phi}} \boldsymbol{\Sigma} \bar{\boldsymbol{\Phi}}^H = \boldsymbol{0}$. Vectorizing this expression (see e.g. [46]) results in the condition $(\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}}) \boldsymbol{z} = \boldsymbol{0}$, where $\boldsymbol{z} = \mathrm{vec}\, \boldsymbol{\Sigma}$. Thus, $\ker \tilde{\phi}_{\mathbb{C}}$ is given (up to inverse vectorization) by the null space of the $(MB)^2 \times (NB)^2$ matrix $\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}}$.

Since the columns of $\boldsymbol{V}$ constitute a basis for the null space of $\boldsymbol{\Phi}$ and since $\bar{\boldsymbol{\Phi}} = \boldsymbol{I}_B \otimes \boldsymbol{\Phi}$, the columns of $\bar{\boldsymbol{V}} = \boldsymbol{I}_B \otimes \boldsymbol{V}$ constitute a basis for the null-space of $\bar{\boldsymbol{\Phi}}$. It can be shown that $\ker \tilde{\phi}_{\mathbb{C}}$ is composed of matrices of the form $\boldsymbol{\Sigma} = \bar{\boldsymbol{V}} \boldsymbol{A}^H + \boldsymbol{B} \bar{\boldsymbol{V}}^H$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are arbitrary matrices of the appropriate dimensions. It follows that the null space of $\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}}$ is spanned by the columns of the matrix

$$\bar{\boldsymbol{W}} = [\boldsymbol{I}_{NB} \otimes \bar{\boldsymbol{V}}, \quad \bar{\boldsymbol{V}}^* \otimes \boldsymbol{I}_{NB}]. \qquad (75)$$

By the properties of the Kronecker product [46], the fact that $\boldsymbol{\Phi}$ has maximum rank implies that $\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}}$ has maximum rank as well, so that its null space has dimension $(N^2 - M^2)B^2$. However, since $\bar{\boldsymbol{V}}$ is $NB \times (N - M)B$, it is clear that $\bar{\boldsymbol{W}}$ has $2(N - M)NB^2$ columns, which is greater than $(N^2 - M^2)B^2$. Thus, in order to obtain a basis for the null space of $\bar{\boldsymbol{\Phi}}^* \otimes \bar{\boldsymbol{\Phi}}$ we should remove dependent columns from $\bar{\boldsymbol{W}}$. This procedure is carried out in the following lemma:

*Lemma 7:* Let $\boldsymbol{V} \in \mathbb{C}^{N \times (N-M)}$, with $M \leq N$, be a matrix whose columns generate the null space of $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$, which follows a continuous distribution, and let $\bar{\boldsymbol{V}} = \boldsymbol{I}_B \otimes \boldsymbol{V}$. Then, the columns of $\bar{\boldsymbol{W}}$, defined by (75), span the same subspace as the columns of $\bar{\bar{\boldsymbol{W}}}$, which is defined as

$$\bar{\bar{\boldsymbol{W}}} = [\boldsymbol{I}_{NB} \otimes \bar{\boldsymbol{V}}, \quad \bar{\boldsymbol{V}}^* \otimes \boldsymbol{I}_B \otimes \boldsymbol{F}_M], \qquad (76)$$

where $\boldsymbol{F}_M = [\boldsymbol{I}_M, \; \boldsymbol{0}_{M,N-M}]^T$.

*Proof:* The procedure we follow in this proof is to remove linearly dependent columns from $\bar{\boldsymbol{W}}$. Since the case $B > 1$ is quite tedious, here we only show this result for the case $B = 1$. The proof for the general case follows the same lines and it is easily extrapolated, but it requires an overloaded notation. For $B = 1$ we have that

$$\bar{\boldsymbol{W}} = [\boldsymbol{I}_N \otimes \boldsymbol{V}, \quad \boldsymbol{V}^* \otimes \boldsymbol{I}_N]. \qquad (77)$$

Now scale the last $N(N-M)$ columns of $\bar{W}$ to obtain

$$\bar{W}' = [I_N \otimes V, \quad G \otimes I_N], \qquad (78)$$

where $G$ is the result of scaling the columns of $V^*$ such that the first row contains only ones[12]:

$$G = \begin{bmatrix} 1 & 1 & \dots & 1 \\ g_{1,0} & g_{1,1} & \cdots & g_{1,N-M-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N-1,0} & g_{N-1,1} & \cdots & g_{N-1,N-M-1} \end{bmatrix} \qquad (79)$$

Now consider a submatrix of $\bar{W}'$ obtained by retaining the first $N(N-M)$ columns and the columns with indices $N(N-M)+Ni, \dots, N(N-M)+N(i+1)-1$, i.e.,

$$\bar{W}'_i = \begin{bmatrix} V & 0 & \dots & 0 & I_N \\ 0 & V & \dots & 0 & g_{1,i}I_N \\ & & \ddots & & \\ 0 & 0 & \dots & V & g_{N-1,i}I_N \end{bmatrix}, \qquad (80)$$

where $i = 0, \dots, N-M-1$. Scaling the diagonal blocks on the left yields:

$$\bar{W}''_i = \begin{bmatrix} V & 0 & \dots & 0 & I_N \\ 0 & g_{1,i}V & \dots & 0 & g_{1,i}I_N \\ & & \ddots & & \\ 0 & 0 & \dots & g_{N-1,i}V & g_{N-1,i}I_N \end{bmatrix}. \qquad (81)$$

Now, since $\Phi$ follows a continuous distribution, the last $N-M$ columns of $[V, I_N]$ can be written as linear combinations of the first $N$ columns, which means that the last $N-M$ columns of $\bar{W}'_i$ are linearly dependent of the others. Repeating this operation for $i = 0, \dots, N-M-1$ and removing from $\bar{W}$ the columns declared as dependent at each $i$ gives

$$\bar{\bar{W}} = [I_N \otimes V, \quad V^* \otimes F_M], \qquad (82)$$

which clearly spans the same subspace as $\bar{W}$. In the general case with $B \geq 1$ we obtain

$$\bar{\bar{W}} = [I_{NB} \otimes \bar{V}, \quad \bar{V}^* \otimes I_B \otimes F_M]. \qquad (83)$$

■

Note that, indeed, the matrix defined in (76) has $(N^2 - M^2)B^2$ columns, which means that they constitute a basis for the null space of $\bar{\Phi}^* \otimes \bar{\Phi}$. Upon inverse vectorization of the columns of $\bar{\bar{W}}$ we obtain the sought basis in matrix form:

$$\mathcal{W} = \Big\{ E_{i,j} \otimes e_k \otimes v_l^H, \quad i,j = 0,1,\dots B-1, \qquad (84)$$

$$k = 0,1,\dots,N-1, \; l = 0,1,\dots,N-M-1 \Big\}$$

$$\cup \Big\{ E_{i,j} \otimes e_k^H \otimes v_l, \quad i,j = 0,1,\dots B-1, \qquad (85)$$

$$k = 0,1,\dots,M-1, \; l = 0,1,\dots,N-M-1 \Big\}.$$

[12]This is always possible whenever the elements of the first row of $V$ are all different from zero. However, it is possible with probability one to choose $V$ such that it generates the null space of $\Phi$ and satisfies this condition.

# APPENDIX D
## PROOF OF THEOREM 4

Clearly, if $\mathcal{M}$ is a length-$(N-1)$ sparse ruler, then (31) defines a periodic sparse ruler. To show the converse statement, assume that $\mathcal{K}$ is a periodic sparse ruler and take $\mathcal{M} = \mathcal{K} \cap \{0,\dots,N-1\}$. Then, $\{0,\dots,NB-1\} \subset \Delta(\mathcal{K})$ and, in particular, $\{N(B-1),\dots,NB-1\} \subset \Delta(\mathcal{K})$, meaning that

$$\forall \delta \in \{N(B-1),\dots,NB-1\}, \; \exists q,p \in \mathcal{K} \text{ s.t. } q-p = \delta.$$

Due to the periodicity of $\mathcal{K}$, any $k \in \mathcal{K}$ can be uniquely decomposed as $k = m_k + b_k N$, where $m_k \in \mathcal{M}$ and $b_k \in \{0,\dots,B-1\}$. Denote as $m_p$, $m_q$, $b_p$ and $b_q$ the corresponding coefficients of the decomposition of $p$ and $q$. Therefore, the condition above becomes

$$\forall \delta \in \{N(B-1),\dots,NB-1\}, \; \exists m_p, m_q \in \mathcal{M} \qquad (86)$$

$$\text{and} \quad b_p, b_q \in \{0,\dots,B-1\} \text{ s.t. } m_q - m_p + (b_q-b_p)N = \delta.$$

Since $m_q - m_p \leq N-1$ and $\delta \geq N(B-1)$, it is clear that $b_q - b_p$ must equal $B-1$ in order for the condition $m_q - m_p + (b_q-b_p)N = \delta$ to hold. Then, after subtracting $N(B-1)$, the following equivalent expression arises:

$$\forall \delta \in \{0,\dots,N-1\}, \; \exists m_p, m_q \in \mathcal{M} \text{ s.t. } m_q - m_p = \delta.$$

Hence, $\mathcal{M}$ is a sparse ruler.

# APPENDIX E
## PROOF OF THEOREM 6

Assume that $L$ is odd. The proof for $L$ even follows similar lines. If $\Delta(\mathcal{K}) = \{0,\dots,L-1\}$, then the matrix from Theorem 1 is given by:

$$R = \begin{bmatrix} 1 & 0_{\bar{L}}^T & 0_{\bar{L}}^T \\ 0_{\bar{L}} & I_{\bar{L}} & -\jmath I_{\bar{L}} \\ 0_{\bar{L}} & K_{\bar{L}} & \jmath K_{\bar{L}} \\ 1 & 0_{\bar{L}}^T & 0_{\bar{L}}^T \\ 0_{\bar{L}} & I_{\bar{L}} & \jmath I_{\bar{L}} \\ 0_{\bar{L}} & K_{\bar{L}} & -\jmath K_{\bar{L}} \end{bmatrix} \qquad (87)$$

where $\bar{L} = \frac{L-1}{2}$, $0_{\bar{L}}$ is an $\bar{L} \times 1$ vector with all zeros and $K_{\bar{L}}$ is an $\bar{L} \times \bar{L}$ Hankel matrix with ones in the antidiagonal and zeros elsewhere, i.e., its $(m,n)$-th element equals 1 if $m+n = \bar{L}-1$ and 0 otherwise. All the columns are linearly independent so that $\text{rank } R = L$ and, according to Theorem 1, $\mathcal{K}$ is an $\mathcal{S}_C$-covariance sampler.

Now consider removing elements from $\Delta(\mathcal{K})$. It can readily be seen that the rank is not maximum iff there is some $\delta \in \{0,\dots,L\}$ such that $\delta \notin \Delta(\mathcal{K})$ and $L-\delta \notin \Delta(\mathcal{K})$. Equivalently, we can say that the rank is maximum if and only if $\Delta_L(\mathcal{K}) = \{0,\dots,L-1\}$.

# APPENDIX F
## PROOF OF THEOREM 7

Let us start by showing that if $\mathcal{M}$ is a circular sparse ruler, then $\mathcal{K}$ is a periodic circular sparse ruler or, in other words, if $\Delta_N(\mathcal{M}) = \{0,\dots,N-1\}$, then $\Delta_{NB}(\mathcal{K}) = \{0,\dots,NB-$

1}. Consider any $\delta \in \{0, \ldots, N-1\}$. Since $\delta \in \Delta_N(\mathcal{M})$, at least one of the following two conditions will hold:

C1: $\exists m_1, m_2 \in \mathcal{M}, \ m_2 \geq m_1$ such that $\qquad$ (88)
$$(m_2 - m_1)_N = m_2 - m_1 = \delta$$
C2: $\exists m_1, m_2 \in \mathcal{M}, \ m_2 < m_1$ such that $\qquad$ (89)
$$(m_2 - m_1)_N = N + m_2 - m_1 = \delta$$

We next show that, in both cases, all the elements of the form $\delta + bN$, with $b = 0, \ldots, B-1$, are in $\Delta_{NB}(\mathcal{K})$:

- C1: consider $k_2 = m_2 + bN$ and $k_1 = m_1$ for any $b = 0, \ldots, B-1$. Since $k_1, k_2 \in \mathcal{K}$, it follows that $(k_2 - k_1)_{NB} = m_2 + bN - m_1 = \delta + bN \in \Delta_{NB}(\mathcal{K})$.
- C2: first take $k_1 = m_1$ and $k_2 = m_2 + N + bN$ with $b = 0, \ldots, B-2$. Since $k_1, k_2 \in \mathcal{K}$, then $(k_2 - k_1)_{NB} = m_2 + N + bN - m_1 = \delta + bN \in \Delta_{NB}(\mathcal{K})$. It suffices only to show that $\delta + bN \in \Delta_{NB}(\mathcal{K})$ when $b = B-1$. To this end, consider $k_1 = m_1$ and $k_2 = m_2$, which results in $(k_2 - k_1)_{NB} = NB + m_2 - m_1 = N(B-1) + N + m_2 - m_1 = N(B-1) + \delta \in \Delta_{NB}(\mathcal{K})$.

To sum up, we have shown that $\delta + bN \in \Delta_{NB}(\mathcal{K})$ for any $\delta = 0, \ldots, N-1$ and $b = 0, \ldots, B-1$, which establishes that $\mathcal{K}$ is a circular sparse ruler.

To show the converse statement, assume that $\mathcal{K}$ is a circular sparse ruler, i.e., $\Delta_{NB}(\mathcal{K}) = \{0, \ldots, NB-1\}$. In particular, all modular distances of the form $\delta = \{0, \ldots, N-1\}$ are present in $\Delta_{NB}(\mathcal{K})$, which means that one or both of the following two conditions will be satisfied:

C1': $\exists k_1, k_2 \in \mathcal{K}, \ k_2 \geq k_1$ such that $\qquad$ (90)
$$(k_2 - k_1)_{NB} = k_2 - k_1 = \delta,$$
C2': $\exists k_1, k_2 \in \mathcal{K}, \ k_2 < k_1$ such that $\qquad$ (91)
$$(k_2 - k_1)_{NB} = NB + k_2 - k_1 = \delta.$$

It is therefore to be shown that $\delta \in \Delta_N(\mathcal{M})$ in both cases, where $\mathcal{M}$ is defined as $\mathcal{M} = \mathcal{K} \cap \{0, \ldots, N-1\}$.

- C1': clearly, we can assume without any loss of generality that $k_1 \in \mathcal{M}$. According to whether $k_2$ is also in $\mathcal{M}$ or not, we distinguish two scenarios:
  - $k_2 \in \mathcal{M}$: in this case, it is clear that $(k_2 - k_1)_N = k_2 - k_1 \in \Delta_N(\mathcal{M})$.
  - $k_2 \notin \mathcal{M}$: since $0 \leq \delta < N$, it follows that $k_2$ can be written as $k_2 = m + N$ for some $m \in \mathcal{M}$ with $m < k_1$. Therefore, $(m - k_1)_N = N + m - k_1 = k_2 - k_1 = \delta \in \Delta_N(\mathcal{M})$.
- C2': since $0 \leq \delta < N$, it can be seen that $N(B-1) < k_1 - k_2 \leq NB$, which in turn requires $k_2 \in \mathcal{M}$ and $k_1 = m + N(B-1)$ for some $m \in \mathcal{M}$ with $m > k_2$. Now consider the circular distance between $m$ and $k_2$:

$$(k_2 - m)_N = N + k_2 - m = N + k_2 - [k_1 - N(B-1)]$$
$$= k_2 - k_1 + NB = \delta \in \Delta_N(\mathcal{M}).$$

Therefore, we have shown that $\delta \in \Delta_N(\mathcal{M})$ for all $\delta = 0, \ldots, N-1$, which means that $\mathcal{M}$ is a circular sparse ruler.

## APPENDIX G
## PROOF OF THEOREM 8

If we form the matrix $\boldsymbol{R}$ in Theorem 1 using the matrices from (16), we conclude that $\{0, \ldots, d\} \subset \Delta(\mathcal{K})$ in order for $\mathcal{K}$ to define a covariance sampler. As we did to prove Theorem 4, we write the following necessary and sufficient condition:

$\forall \delta \in \{0, \ldots, d\}, \ \exists m_1, m_2 \in \mathcal{M}$ and $b_1, b_2 \in \{0, \ldots, B-1\}$
such that $m_2 - m_1 + (b_2 - b_1)N = \delta$. $\qquad$ (92)

We start by showing that if $\mathcal{M}$ is a circular sparse ruler, then (92) holds true, i.e., $\mathcal{K}$ is a covariance sampler. More specifically, we show that $\delta \in \Delta(\mathcal{K}) \ \forall \delta \in \{0, \ldots, N(B-1)\}$. Consider two different cases:

- Case $0 \leq \delta < N(B-1)$: It suffices to write $\delta$ as $\delta = m_\delta + b_\delta N$, with $m_\delta \in \{0, \ldots, N-1\}$ and $b_\delta \in \{0, \ldots, B-2\}$. Since $m_\delta \in \Delta_N(\mathcal{M})$, then $m_\delta$ can be represented either as $m_{\delta,2} - m_{\delta,1}$ or as $N + m_{\delta,2} - m_{\delta,1}$, where $m_{\delta,1}, m_{\delta,2} \in \mathcal{M}$. In the former case just make $m_2 = m_{\delta,2}, m_1 = m_{\delta,1}$, $b_2 = b_\delta$ and $b_1 = 0$. In the later case make $m_2 = m_{\delta,2}$, $m_1 = m_{\delta,1}, b_2 = b_\delta + 1$ and $b_1 = 0$.
- Case $\delta = N(B-1)$: this is trivial since $N(B-1) \in \Delta(\mathcal{K})$ for any non-empty choice of $\mathcal{M}$.

Now, in order to show the converse theorem, we prove that if $\{0, \ldots, N-1\} \subset \Delta(\mathcal{K})$, then $\{0, \ldots, N-1\} \subset \Delta_N(\mathcal{M})$. Let us consider some $\delta \in \{0, \ldots, N-1\}$. Since $\delta \in \Delta(\mathcal{K})$, it is clear that there exist some $m_1, m_2 \in \mathcal{M}$ and $b_1, b_2 \in \{0, \ldots, B-1\}$ such that $m_2 - m_1 + (b_2 - b_1)N = \delta$. In particular, $(b_2 - b_1)$ can be either 0 or 1. Therefore, for any $\delta \in \{0, \ldots, N-1\}$, there exists $m_1, m_2 \in \mathcal{M}$ such that either $m_2 - m_1 = \delta$ or $N + m_2 - m_1 = \delta$. Noting that this condition is equivalent to the condition $\{0, \ldots, N-1\} \subset \Delta_N(\mathcal{M})$ concludes the proof.

## REFERENCES

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

[2] 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.

[3] Y. P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 45, no. 3, pp. 340–351, Mar. 1998.

[4] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.

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

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

[7] 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. Info. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

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

[9] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[10] 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.

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

[12] 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.

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

[14] M. Wakin, S. Becker, E. Nakamura, M. Grant, E. Sovero, D. Ching, Y. Juhwan, J. Romberg, A. Emami-Neyestanak, and E. Candès, "A nonuniform sampler for wideband spectrally-sparse environments," *Emerging Select. Topics Circuits Syst., IEEE J. on*, vol. 2, no. 3, pp. 516–529, 2012.

[15] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *2007 IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 1959–1962.

[16] J. Yoo, S. Becker, M. Monge, M. Loh, E. Candes, and A. Emami-Neyestanak, "Design and implementation of a fully integrated compressed-sensing signal acquisition system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 5325–5328.

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

[18] D. Pearson, S. U. Pillai, and Y. Lee, "An algorithm for near-optimal placement of sensor elements," *IEEE Trans. Info. Theory*, vol. 36, no. 6, pp. 1280–1284, 1990.

[19] P. Wild, "Difference basis systems," *Discrete mathematics*, vol. 63, no. 1, pp. 81–90, 1987.

[20] S. U. Pillai and F. Haber, "Statistical analysis of a high resolution spatial spectrum estimator utilizing an augmented covariance matrix," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 35, no. 11, pp. 1517–1523, 1987.

[21] 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. Proc.*, vol. 46, no. 9, pp. 2458–2471, 1998.

[22] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. Part II: Partially augmentable arrays," *IEEE Trans. Sig. Proc.*, vol. 47, no. 6, pp. 1502–1521, 1999.

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

[24] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *IEEE/SP 15th Workshop on Statistical Signal Process.*, 2009, pp. 626–629.

[25] Y. Wang and G. Leus, "Space-time compressive sampling array," in *Proc. of IEEE Sensor Array Multichannel Sig. Proc. Workshop (SAM)*, 2010, pp. 33–36.

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

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

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

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

[30] G. Vázquez-Vilar, R. López-Valcarce, C. Mosquera, and N. González-Prelcic, "Wideband spectral estimation from compressed measurements exploiting spectral a priori information in cognitive radio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 2958–2961.

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

[32] D. Romero, S.-J. Kim, R. López-Valcarce, and G. B. Giannakis, "Spectrum cartography using quantized observations," *IEEE Int. Conf. Acoust., Speech, Signal Process. (Accepted)*, 2015.

[33] D. Romero, S.-J. Kim, and G. B. Giannakis, "Online spectrum cartography via quantized measurements," *Submitted to Conf. Info. Sci. Syst.*, 2015.

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

[35] D. D. Ariananda, D. Romero, and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals," in *Proc. IEEE Intl. Workshop Computational Advances Multi-Sensor Adaptive Proc. (CAMSAP)*, San Martin, France, Dec. 2013.

[36] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Digit. Signal Process. Workshop Signal Process. Educ. Workshop (DSP/SPE), 2011 IEEE*, Jan. 2011, pp. 289–294.

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

[38] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Intl. Workshop Sig. Proc. Advances Wireless Commun. (SPAWC)*, Jun. 2012, pp. 525–529.

[39] L. Rédei and A. Rényi, "On the representation of the numbers 1,2,...,n by means of differences (Russian)," *Matematicheskii sbornik*, vol. 66, no. 3, pp. 385–389, 1949.

[40] J. Leech, "On the representation of 1, 2,..., n by differences," *J. London Mathematical Society*, vol. 31, no. 2, pp. 160–169, 1956.

[41] B. Wichmann, "A note on restricted difference bases," *J. London Mathematical Society*, vol. s1-s38, no. 1, pp. 465–466, 1963.

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

[43] H. C. Pumphrey, "Design of sparse arrays in one, two, and three dimensions," *J. Acoust. Society America*, vol. 93, pp. 1620, 1993.

[44] P. Pal and P. P. Vaidyanathan, "Nested arrays in two dimensions. Part I: Geometrical considerations," *IEEE Trans. Sig. Proc.*, vol. 60, no. 9, pp. 4694–4705, 2012.

[45] M. E. Domínguez-Jiménez and N. González-Prelcic, "A class of circular sparse rulers for compressive power spectrum estimation," in *European Signal Process. Conf. (EUSIPCO)*, 2013.

[46] D. S. Bernstein, *Matrix mathematics. Theory, facts, and formulas. 2nd expanded ed*, Princeton University Press., 2009.

[47] D. Romero and G. Leus, "Compressive covariance sampling," in *Proc. Info. Theory Appl. Workshop*, San Diego, CA, Feb. 2013, pp. 1–8.

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

[49] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1996, vol. 3, pp. 1688–1691 vol. 3.

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

[51] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[52] R. M. Gray, *Toeplitz and circulant matrices: A review*, Now Pub, 2006.

[53] P. Billingsley, "Probability and measure," *Jown Wiley & Sons*, 1995.

[54] J. C. P. Miller, "Difference bases: Three problems in additive number theory," *Computers in Number Theory*, pp. 299–322, 1971.

[55] P. Xia, S. Zhou, and G. B. Giannakis, "Achieving the welch bound with difference sets," *IEEE Trans. Info. Theory*, vol. 51, no. 5, pp. 1900–1907, 2005.

[56] 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.

[57] M. E. Dominguez-Jimenez, N. Gonzalez-Prelcic, G. Vázquez-Vilar, and R. López-Valcarce, "Design of universal multicoset sampling patterns for compressed sensing of multiband sparse signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 3337–3340.

[58] E. Masry, "Poisson sampling and spectral estimation of continuous-time processes," *IEEE Trans. Info. Theory*, vol. 24, no. 2, pp. 173–183, Mar. 1978.