Nearly-Optimal Compression Matrices for Signal Power Estimation

Daniel Romero and Roberto López-Valcarce

Department of Signal Theory and Communications, University of Vigo (Spain)
Email: {dromero,valcarce}@gts.uvigo.es

***Abstract*—We present designs for compression matrices minimizing the Cramér-Rao bound for estimating the power of a stationary Gaussian process, whose second-order statistics are known up to a scaling factor, in the presence of (possibly colored) Gaussian noise. For known noise power, optimum designs can be found assuming either low or high signal-to-noise ratio (SNR). In both cases the optimal schemes sample the frequency bins with highest SNR, suggesting near-optimality for all SNR values. In the case of unknown noise power, optimal patterns in both SNR regimes sample two subsets of frequency bins with lowest and highest SNR, which also suggests that they are nearly-optimal for all SNR values.**

***Index Terms*—Compressive covariance sensing, sampler design, power estimation, spectrum sensing.**

I. Introduction

This paper addresses the design of sampling patterns for estimating the power of wide-sense stationary (WSS) signals whose second-order statistics are known up to a scaling factor. A linearly compressed version of this signal is to be processed to obtain such estimate. The noise corrupting the signal may have a known or unknown power and a white or colored power spectrum.

An important application of this setting is the estimation of the power of a signal acquired by means of an *analog-to-information converter* (AIC). These devices compute linear projections of an analog signal onto a compressed discrete subspace [1], [2] and are expected to substitute analog-to-digital converters (ADCs) in cases where the Nyquist rate is too demanding. A field of application is spectrum sensing for dynamic spectrum access [3], [4], where prior information about the power spectrum of primary transmissions (spectral masks, carrier frequencies, bandwidths, etc.) is often available since their waveforms typically observe public standards [5]–[8]. An estimate of the power of a primary signal allows to declare the corresponding user as active or inactive. Another application is to monitor the correct usage of a secondary network, where users are subject to stringent transmission power limits.

Work supported by the Galician Government, the Spanish Government and the European Regional Development Fund (ERDF) under projects TACTICA, CONSOLIDER-INGENIO 2010 CSD2008-00010, TEC2010-21245-C02/TCM, TEC2013-47020-C2-1-R, GRC2013/009, FPU grant AP2010-0149, and the agreement for funding the Atlantic Research Center for Information and Communication Technologies (AtlantTIC).

The design of sub-Nyquist sampling schemes has been addressed in several contexts, such as sampling of multiband signals [9], [10], compressed sensing [11], MIMO radar [12], etc. Whereas these approaches are geared towards reconstructing the original signal from the compressed observations, it is often the case that only a few parameters, and not the signal itself, are of interest. Thus the problem becomes that of *parameter estimation* (rather than signal reconstruction) from compressed measurements [6], [13]. In the model considered here, signals and noise are assumed WSS Gaussian; thus, only second-order statistics (SOS) are relevant.

A general framework to design samplers preserving the relevant SOS is provided in [14], [15], where *universal samplers* are defined as those allowing estimation of the SOS of any WSS process after compression. Although, by definition, the design of universal samplers does not require knowledge of the SOS of the processes involved, it is expected that tailoring a sampler for a particular process, whenever possible, should be beneficial in terms of compression ratios and/or estimation performance with respect to the universal case. Our goal is to provide such designs by exploiting the statistical information available about the signal to be compressed.

Since we wish our designs to be general, i.e., independent of any particular estimation method, our goal is the minimization of the Cramér-Rao bound (CRB) [16] for the parameter of interest. Our problem is related to [17]–[19], which analyze the impact of compression on the CRB for a different model (Gaussian model with known covariance matrix and parametric dependence of the mean vector). In contrast, the Gaussian model we study has zero mean and parametric dependence of the covariance. In addition, our goal is to explicitly obtain an optimal design of the compression matrix, whereas [17]–[19] focus on the effect of random matrices satisfying a restricted isometry property (RIP), such as those typically used in reconstruction-oriented applications.

The rest of the paper is structured as follows: Sec. II formulates the problem and establishes some results used in the remaining sections. Sections III and IV consider, respectively, the design of sampling matrices in the cases where the noise power is known or unknown. Finally, the main conclusions are illustrated in Sec. V by means of a numerical example and summarized in Sec. VI.

II. Problem Formulation

Consider a model where the uncompressed observations are collected in the vector $\boldsymbol{x} \in \mathbb{C}^L$, which is given by

$$\boldsymbol{x} = \alpha \boldsymbol{s} + \sigma \boldsymbol{w}, \tag{1}$$

where $\alpha\boldsymbol{s}$ and $\sigma\boldsymbol{w}$ respectively denote the signal and noise terms. Both $\boldsymbol{s}$ and $\boldsymbol{w}$ contain samples of independent zero-mean unit-variance WSS random processes with known power spectral density (PSD) and circularly complex Gaussian distribution. Thus, $\boldsymbol{s} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma_s})$ and $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma_w})$ with $\boldsymbol{\Sigma_s}$ and $\boldsymbol{\Sigma_w}$ known Hermitian Toeplitz covariance matrices with ones on the diagonal. Consequently, $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma_x})$, where $\boldsymbol{\Sigma_x} = \alpha^2\boldsymbol{\Sigma_s} + \sigma^2\boldsymbol{\Sigma_w}$. If L is large, then it is possible to write $\boldsymbol{\Sigma_s} \approx \boldsymbol{W}\boldsymbol{\Lambda_s}\boldsymbol{W}^H$ and $\boldsymbol{\Sigma_w} \approx \boldsymbol{W}\boldsymbol{\Lambda_w}\boldsymbol{W}^H$, where $\boldsymbol{W}$ is the unitary inverse discrete Fourier transform (IDFT) matrix, and the elements on the main diagonal of the matrices

$$\boldsymbol{\Lambda_s} = \operatorname{diag}\{\lambda_{\boldsymbol{s},1}, \cdots, \lambda_{\boldsymbol{s},L}\} \tag{2}$$

$$\boldsymbol{\Lambda_w} = \operatorname{diag}\{\lambda_{\boldsymbol{w},1}, \cdots, \lambda_{\boldsymbol{w},L}\} \tag{3}$$

are samples of the PSD of the associated processes [20], [21]. In some of the results that follow, it is necessary to assume that these coefficients are strictly positive.

With the goal of estimating α^2, a total of $K \leq L$ linear measurements of $\boldsymbol{x}$ are collected in $\boldsymbol{y}$:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \quad \in \mathbb{C}^K. \tag{4}$$

$\boldsymbol{\Phi} \in \mathbb{C}^{K\times L}$ is referred to as the *sampling* or *compression matrix* or, for brevity, just *sampler*. Clearly $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}, \bar{\boldsymbol{\Sigma}})$, with $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Phi}\boldsymbol{\Sigma_x}\boldsymbol{\Phi}^H$. By defining $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}} = \boldsymbol{\Phi}\boldsymbol{\Sigma_s}\boldsymbol{\Phi}^H$ and $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \boldsymbol{\Phi}\boldsymbol{\Sigma_w}\boldsymbol{\Phi}^H$, we can write

$$\bar{\boldsymbol{\Sigma}} = \alpha^2\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}} + \sigma^2\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}. \tag{5}$$

Intuitively, the estimation performance is related to the signal-to-noise ratio (SNR), defined as α^2/σ^2. Since $\sigma^2 \geq 0$ and $\alpha^2 \geq 0$, the minimum variance of any unbiased estimator of α^2 is given by its *constrained* CRB [22]. For simplicity we assume $\sigma^2 > 0$ and $\alpha^2 > 0$, which results in the well-known unconstrained CRB. This bound can be written in terms of the Fisher information matrix (FIM) of the parameter vector $\boldsymbol{\theta} = [\,\alpha^2\,\sigma^2\,]^T$, which, for the problem at hand, is given by Bang's formula [16], [23]:

$$\boldsymbol{F} = \begin{bmatrix} \operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\right) & \operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right) \\ \operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\right) & \operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right) \end{bmatrix}. \tag{6}$$

The CRB for α^2 can be derived from (6) in different settings. The problem is to design $\boldsymbol{\Phi}$ in order to minimize this CRB. To this end, we will repeatedly make use of the following result:

Lemma 1. *Let $\boldsymbol{P} \in \mathbb{C}^{L\times L}$ be an orthogonal projection matrix and $\boldsymbol{A} \in \mathbb{C}^{L\times L}$ a Hermitian positive semi-definite matrix. Then* $\operatorname{Tr}(\boldsymbol{PA}) \leq \operatorname{Tr}(\boldsymbol{A})$.

Proof: Let $\boldsymbol{P} = \boldsymbol{UDU}^H$ be an eigendecomposition of $\boldsymbol{P}$. Then, $\operatorname{Tr}(\boldsymbol{PA}) = \operatorname{Tr}\left(\boldsymbol{DU}^H\boldsymbol{AU}\right) = \sum_i d_i b_i$, where d_i and b_i are, respectively, the i-th entry on the diagonal of $\boldsymbol{D}$ and $\boldsymbol{B} = \boldsymbol{U}^H\boldsymbol{AU}$. Since d_i is either 0 or 1 and $b_i \geq 0$, it is clear that $\operatorname{Tr}(\boldsymbol{PA}) \leq \sum_i b_i = \operatorname{Tr}\left(\boldsymbol{U}^H\boldsymbol{AU}\right) = \operatorname{Tr}(\boldsymbol{A})$. ∎

III. Samplers for Known Noise Power

If σ^2 is known, the CRB for α^2 is directly given by the reciprocal of the $(1,1)$ element of the FIM (6):

$$(\boldsymbol{F}_{1,1})^{-1} = \left[\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right)\right]^{-1}. \tag{7}$$

Thus, minimizing the CRB w.r.t. $\boldsymbol{\Phi}$ amounts to maximizing $\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right)$. However, since this quantity depends on the unknown parameter α^2, a sensible approach is to focus on the low and high SNR cases, where this dependence fades away. If we then find that the optimal designs are similar in both cases, we will conclude that the influence of α^2 on the design the optimal $\boldsymbol{\Phi}$ is relatively small and any of these solutions can be regarded as nearly optimal for any SNR range.

A. Low SNR regime

When $\sigma^2 \gg \alpha^2$, the covariance matrix of the observations satisfies $\bar{\boldsymbol{\Sigma}} \approx \sigma^2\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$, which results in the following problem:

$$\underset{\boldsymbol{\Phi}}{\text{maximize}} \quad \operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right). \tag{8}$$

After straightforward operations, we find that

$$\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right) = \operatorname{Tr}\left((\boldsymbol{P_w}\boldsymbol{\Lambda}_{\text{SNR}})^2\right), \tag{9}$$

where $\boldsymbol{\Lambda}_{\text{SNR}} = \boldsymbol{\Lambda}_{\boldsymbol{w}}^{-1}\boldsymbol{\Lambda_s}$ is the spectral SNR matrix, and

$$\boldsymbol{P_w} = \boldsymbol{\Lambda}_{\boldsymbol{w}}^{1/2}\boldsymbol{W}^H\boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{\Lambda_w}\boldsymbol{W}^H\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{\Lambda}_{\boldsymbol{w}}^{1/2}$$

is the orthogonal projector onto the columns of $\boldsymbol{\Lambda}_{\boldsymbol{w}}^{1/2}\boldsymbol{W}^H\boldsymbol{\Phi}^H$, which is a *noise-weighted* frequency version of the sampler. Observe that by applying the change of variable $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{\Lambda}_{\boldsymbol{w}}^{1/2}$ we obtain $\boldsymbol{P_w} = \tilde{\boldsymbol{\Phi}}^H(\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Phi}}^H)^{-1}\tilde{\boldsymbol{\Phi}}$, which means that we can freely choose $\boldsymbol{P_w}$ to be any projection matrix on a subspace of dimension K. For this reason, we first accomplish the maximization of (9) in terms of the projector $\boldsymbol{P_w} \in \mathbb{C}^{L\times L}$, and then obtain $\boldsymbol{\Phi}$ from the optimal $\boldsymbol{P_w}$.

Being an orthogonal projector, $\boldsymbol{P_w}$ is Hermitian, idempotent ($\boldsymbol{P}_{\boldsymbol{w}}^2 = \boldsymbol{P_w}$), and its only eigenvalues are 1 (with multiplicity K) and 0 (with multiplicity $L-K$). Its (i,i)-th element equals the squared norm of the i-th row/column:

$$||\boldsymbol{P_w}\boldsymbol{e}_i||_2^2 = \boldsymbol{e}_i^H\boldsymbol{P}_{\boldsymbol{w}}^H\boldsymbol{P_w}\boldsymbol{e}_i = \boldsymbol{e}_i^H\boldsymbol{P_w}\boldsymbol{e}_i, \tag{10}$$

where $\boldsymbol{e}_i \in \mathbb{C}^L$ has a one at the i-th position and zeros elsewhere. The fact that the eigenvalues have a magnitude between zero and one means that $0 \leq \boldsymbol{e}_i^H\boldsymbol{P_w}\boldsymbol{e}_i \leq 1$.

Applying Lemma 1 to (9), an upper bound is found as

$$\operatorname{Tr}\left((\boldsymbol{P_w}\boldsymbol{\Lambda}_{\text{SNR}})^2\right) \leq \operatorname{Tr}\left(\boldsymbol{P_w}\boldsymbol{\Lambda}_{\text{SNR}}^2\right), \tag{11}$$

which holds with equality if $\boldsymbol{P_w}$ is diagonal. Our approach is to maximize the upper bound $\operatorname{Tr}\left(\boldsymbol{P_w}\boldsymbol{\Lambda}_{\text{SNR}}^2\right)$ w.r.t. $\boldsymbol{P_w}$. Clearly, if the resulting maximizer happens to be diagonal, then it must maximize the original objective function $\operatorname{Tr}\left((\boldsymbol{P_w}\boldsymbol{\Lambda}_{\text{SNR}})^2\right)$ as well. As we will see, it turns out that this is the case.

Let us introduce

$$\begin{aligned}\boldsymbol{\lambda} &= \operatorname{diag}\{\boldsymbol{\Lambda}_{\text{SNR}}\} = [\ \lambda_1 \ \ \cdots \ \ \lambda_L\ ]^T,\\ \boldsymbol{\lambda}^2 &= \operatorname{diag}\{\boldsymbol{\Lambda}_{\text{SNR}}^2\} = [\ \lambda_1^2 \ \ \cdots \ \ \lambda_L^2\ ]^T,\\ \boldsymbol{p} &= \operatorname{diag}\{\boldsymbol{P}_{\boldsymbol{w}}\} = [\ p_1 \ \ \cdots \ \ p_L\ ]^T.\end{aligned} \tag{12}$$

The right-hand side of (11) can be rewritten as $\boldsymbol{p}^T\boldsymbol{\lambda}^2$ and, according to the properties of $\boldsymbol{P}_{\boldsymbol{w}}$, it has to be maximized subject to[1] $\mathbf{0} \preceq \boldsymbol{p} \preceq \mathbf{1}$ and $\mathbf{1}^T\boldsymbol{p} = K$. It follows that the optimum is such that $p_i = 1$ for the indices i corresponding to the K largest entries of $\boldsymbol{\lambda}^2$ and $p_i = 0$ otherwise.

Since $p_i = \boldsymbol{e}_i^H\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{e}_i$, it follows from (10) that $\boldsymbol{P}_{\boldsymbol{w}}$ is diagonal. Thus (11) holds with equality, and $\boldsymbol{P}_{\boldsymbol{w}}$ solves (8). In order to compute $\boldsymbol{\Phi}$ from this solution, it suffices to take any $L \times K$ matrix $\boldsymbol{R}$ with the same column space as $\boldsymbol{P}_{\boldsymbol{w}}$ and then compute $\boldsymbol{\Phi}$ from $\boldsymbol{R} = \boldsymbol{\Lambda}_{\boldsymbol{w}}^{1/2}\boldsymbol{W}^H\boldsymbol{\Phi}^H$, which results in $\boldsymbol{\Phi} = \boldsymbol{R}^H\boldsymbol{\Lambda}_{\boldsymbol{w}}^{-1/2}\boldsymbol{W}^H$. However, since $\boldsymbol{P}_{\boldsymbol{w}}$ is diagonal and $\boldsymbol{R}$ arbitrary, one can disregard $\boldsymbol{\Lambda}_{\boldsymbol{w}}$ and directly make $\boldsymbol{\Phi} = \boldsymbol{R}^H\boldsymbol{W}^H$. Thus, the rows of $\boldsymbol{\Phi}$ are invertible linear combinations of the rows of the DFT matrix indexed by the p_i's which are equal to one, i.e., the optimal samplers measure linear combinations of the K frequency bins with largest spectral SNR.

B. High SNR regime

When $\alpha^2 \gg \sigma^2$, we use the second-order approximation

$$\bar{\boldsymbol{\Sigma}}^{-1} \approx \frac{1}{\alpha^2}\left[\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1} - \frac{\sigma^2}{\alpha^2}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\right]. \tag{13}$$

From this approximation and after neglecting second-order terms, maximizing $\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right)$ w.r.t. $\boldsymbol{\Phi}$ is found to be equivalent to the following problem:

$$\underset{\boldsymbol{\Phi}}{\text{minimize}} \quad \operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right). \tag{14}$$

Now one has $\operatorname{Tr}\left(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right) = \operatorname{Tr}\left(\boldsymbol{P}_{\boldsymbol{s}}\boldsymbol{\Lambda}_{\text{SNR}}^{-1}\right)$, where

$$\boldsymbol{P}_{\boldsymbol{s}} = \boldsymbol{\Lambda}_{\boldsymbol{s}}^{1/2}\boldsymbol{W}^H\boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{\Lambda}_{\boldsymbol{s}}\boldsymbol{W}^H\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{W}\boldsymbol{\Lambda}_{\boldsymbol{s}}^{1/2} \tag{15}$$

is the orthogonal projector onto the columns of $\boldsymbol{\Lambda}_{\boldsymbol{s}}^{1/2}\boldsymbol{W}^H\boldsymbol{\Phi}^H$. Following analogous steps to those in Sec. III-A, the minimizer is found to be a diagonal matrix $\boldsymbol{P}_{\boldsymbol{s}}$ with ones in the entries corresponding to the K smallest values of $\boldsymbol{\Lambda}_{\text{SNR}}^{-1}$, which yields the same solution as in Sec. III-A, i.e., the optimal sampling matrix acquires linear combinations of the K frequency bins with largest spectral SNR.

The fact that the same compression matrix is optimal for both the high and low SNR regimes suggests that nearly optimal results can be expected for intermediate SNR values as well.

IV. Samplers for Unknown Noise Power

An unknown σ^2 becomes a *nuisance parameter* and affects the CRB for α^2 [16], given by the $(1,1)$ element of $\boldsymbol{F}^{-1}$. Again, we consider the low and high SNR regimes to sidestep the dependence of the CRB with the unknown parameters.

[1]We say that $[\,a_1 \cdots a_n\,]^T \preceq [\,b_1 \cdots b_n\,]^T$ if $a_i \leq b_i$ for $i = 1, \ldots, n$.

A. Low SNR Regime

Using $\bar{\boldsymbol{\Sigma}} \approx \sigma^2\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ as in Sec. III-A, the CRB for α^2 becomes

$$(\boldsymbol{F}^{-1})_{1,1} \approx \frac{K\sigma^4}{K\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}})^2\right) - \operatorname{Tr}^2\left(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}\right)}. \tag{16}$$

Minimizing (16) w.r.t. $\boldsymbol{\Phi}$ amounts to maximizing

$$K\operatorname{Tr}\left((\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}})^2\right) - \operatorname{Tr}^2\left(\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}}\right), \tag{17}$$

which, according to Lemma 1, can be upper bounded as

$$\begin{aligned}K\operatorname{Tr}\left((\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}})^2\right) - \operatorname{Tr}^2\left(\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}}\right) &\leq \\ K\operatorname{Tr}\left(\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}}^2\right) &- \operatorname{Tr}^2\left(\boldsymbol{P}_{\boldsymbol{w}}\boldsymbol{\Lambda}_{\text{SNR}}\right).\end{aligned} \tag{18}$$

Clearly, equality holds in (18) if $\boldsymbol{P}_{\boldsymbol{w}}$ is diagonal. We now attempt to maximize the upper bound in (18), which can be rewritten as

$$K\boldsymbol{p}^T\boldsymbol{\lambda}^2 - (\boldsymbol{p}^T\boldsymbol{\lambda})^2, \tag{19}$$

with $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}^2$ and $\boldsymbol{p}$ as defined in (12). The next result provides the maximizer of this expression for the case where the spectral SNR coefficients λ_j are sorted in increasing order, i.e., $0 \leq \lambda_1 < \cdots < \lambda_L$. In other case, one must first sort them and revert this operation over the resulting optimal $\boldsymbol{p}$.

Let us define the sets

$$\mathcal{I}_0[i] = \{i+1, \cdots, i+D-1\}, \tag{20}$$

$$\mathcal{I}_1[i] = \{1, \cdots, i-1\} \cup \{i+D+1, \cdots, L\}, \tag{21}$$

where $D = L - K$. As we will see shortly, the optimal solution is essentially given by a sampler which measures linear combinations of a total of K frequency bins; of these, i_0 of them correspond to the bins with the smallest spectral SNR, whereas the remaining $K - i_0$ bins correspond to those with the largest spectral SNR. These 'active' bins are indexed by the set $\mathcal{I}_1[i_0] \cup \{i_0\}$. The following result in Theorem 1 formalizes this operation and provides the corresponding value of i_0. To this end, let us define

$$\bar{\beta}_1[i] = \frac{1}{K}\sum_{j\in\mathcal{I}_1[i]}\lambda_j, \tag{22}$$

which is an approximate average of the spectral SNR over the bins with smallest and largest values, and the thresholds

$$\gamma_0[i] = m[i] - \frac{\lambda_{i+D}}{K}, \qquad \gamma_1[i] = m[i] - \frac{\lambda_i}{K}, \tag{23}$$

$$\gamma_2[i] = m[i+1] - \frac{\lambda_i}{K}, \tag{24}$$

where

$$m[i] = \frac{\lambda_i + \lambda_{i+D}}{2}. \tag{25}$$

Theorem 1. *Let* $\boldsymbol{\lambda} = [\,\lambda_1 \cdots \lambda_L\,]^T$, *with* $0 \leq \lambda_1 < \cdots < \lambda_L$ *and let* i_0 *be the smallest integer satisfying* $\bar{\beta}_1[i_0] \leq \gamma_2[i_0]$. *Then, if* $\bar{\beta}_1[i_0] \geq \gamma_1[i_0]$, *the maximizer of* (19) *subject to* $\mathbf{0}_L \preceq \boldsymbol{p} \preceq \mathbf{1}_L$ *and* $\mathbf{1}_L^T\boldsymbol{p} = K$ *is given by*

$$p_i^\star = \begin{cases} 0 & \text{if } i \in \mathcal{I}_0[i_0] \cup \{i_0 + D\} \\ 1 & \text{if } i \in \mathcal{I}_1[i_0] \cup \{i_0\}. \end{cases}$$

Otherwise, if $\bar{\beta}_1[i_0] < \gamma_1[i_0]$, the maximizer is given by

$$p_i^\star = \begin{cases} 0 & \textit{if } i \in \mathcal{I}_0[i_0] \\ 1 & \textit{if } i \in \mathcal{I}_1[i_0] \\ \delta & \textit{if } i = i_0 \\ 1-\delta & \textit{if } i = i_0 + D, \end{cases}$$

where $\delta = (\bar{\beta}_1[i_0] - \gamma_0[i_0])/(\gamma_1[i_0] - \gamma_0[i_0])$.

Sketch of Proof: The proof is based on deriving the Karush-Kuhn-Tucker (KKT) conditions [24], which are necessarily satisfied at the optimum since the objective and constraint functions are differentiable. Then, it is seen that the solutions are of two forms:

- Solutions where $p_i^\star$ is 1 for K values of i and 0 otherwise.
- Solutions where $p_i^\star$ is 1 for $K-1$ values of i, 0 for $L-K-1$ values and in the open interval $0 < p_i^\star < 1$ for two values of this index.

For each case, it can be shown that the indices with $p_i^\star = 0$ must be adjacent and, for the second case, that the indices satisfying $0 < p_i^\star < 1$ separate those equal to one from those equal to zero. With this structure, the problem becomes finding i_0, which is seen to satisfy $\gamma_0[i_0] < \bar{\beta}_1[i_0] \leq \gamma_2[i_0]$. The existence and uniqueness of this value can also be established. Finally, according to whether $\bar{\beta}_1[i_0] < \gamma_1[i_0]$ or $\bar{\beta}_1[i_0] \geq \gamma_1[i_0]$, one of the two cases above occur. ■

Whereas for σ^2 known the optimum samplers capture those K frequency bins with largest SNR, Theorem 1 says that for unknown σ^2 we must sample the bins with highest and lowest SNR. The index i_0 specifies the number of bins of each class to be sampled. Note that the requirement that all λ_i's form a strictly increasing sequence is not a real drawback, since one could introduce arbitrarily small perturbations in equal λ_i's and obtain solutions arbitrarily close to the optimal one.

Recall that the objective function in Theorem 1 is an upper bound (19) on the CRB. When $\bar{\beta}_1[i_0] \geq \gamma_1[i_0]$, the elements of $\boldsymbol{p}^\star$ equal either 0 or 1, resulting in a diagonal $\boldsymbol{P_w}$ which minimizes the CRB in (16), again because for diagonal $\boldsymbol{P_w}$ equality holds in (18). On the other hand, when $\bar{\beta}_1[i_0] < \gamma_1[i_0]$ the maximizer of (19) contains elements different from zero and one, which means that the projector $\boldsymbol{P_w}$ maximizing the right-hand side of (18) is not exactly diagonal. The only non-null off-diagonal elements in $\boldsymbol{P_w}$ correspond to the indices (i_0, i_0+D) and (i_0+D, i_0) and, according to (10), equal $\sqrt{\delta(1-\delta)}$. Consequently, (18) is not satisfied with equality and the resulting non-diagonal $\boldsymbol{P_w}$ is only an approximate minimizer of the CRB. Still, a small loss can be expected since this matrix is nearly diagonal. The compression matrix $\boldsymbol{\Phi}$ can be retrieved in both cases from $\boldsymbol{P_w}$ by following the procedure introduced in Sec. III-A.

B. High SNR Regime

Using the first-order approximation $\bar{\boldsymbol{\Sigma}} \approx \alpha^2 \bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}$ results in

$$(\boldsymbol{F}^{-1})_{1,1} \approx \frac{\alpha^4 \operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}})^2\right)}{K \operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}})^2\right) - \operatorname{Tr}^2\left(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right)}. \quad (26)$$

Minimizing (26) is tantamount to minimizing

$$\frac{\operatorname{Tr}^2\left(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right)}{\operatorname{Tr}\left((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{s}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{w}})^2\right)} = \frac{\operatorname{Tr}^2\left(\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-1}\right)}{\operatorname{Tr}\left((\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-1})^2\right)} \quad (27)$$

with $\boldsymbol{P_s}$ defined as in (15). According to Lemma 1, (27) can be lower bounded as

$$\frac{\operatorname{Tr}^2\left(\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-1}\right)}{\operatorname{Tr}\left((\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-1})^2\right)} \geq \frac{\operatorname{Tr}^2\left(\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-1}\right)}{\operatorname{Tr}\left(\boldsymbol{P_s}\boldsymbol{\Lambda}_{\text{SNR}}^{-2}\right)} = \frac{(\boldsymbol{q}^T\boldsymbol{\lambda})^2}{\boldsymbol{q}^T\boldsymbol{\lambda}^2}, \quad (28)$$

where $\boldsymbol{q} = \operatorname{diag}\{\boldsymbol{P_s}\}$ and, in order to apply the definitions from Sec. IV-A, we have redefined $\boldsymbol{\lambda} = \operatorname{diag}\{\boldsymbol{\Lambda}_{\text{SNR}}^{-1}\}$ and $\boldsymbol{\lambda}^2 = \operatorname{diag}\{\boldsymbol{\Lambda}_{\text{SNR}}^{-2}\}$. Again, equality holds in (28) if $\boldsymbol{P_s}$ is diagonal. The following result gives the minimizer of the lower bound in (28). It makes use of the definition

$$\bar{\beta}_2[i] = \frac{1}{K}\sum_{j\in\mathcal{I}_1[i]} \lambda_j^2,$$

which, as opposed to $\bar{\beta}_1[i]$ from Sec. IV-A, is an approximate average of the *squared* content of the bins indexed by $\mathcal{I}_1[i]$; and the thresholds

$$\zeta_1[i] = \frac{\bar{\beta}_2[i]}{m[i]} - \frac{\lambda_i}{K}\left[1 - \frac{\lambda_i}{m[i]}\right],$$

$$\zeta_2[i] = \frac{\bar{\beta}_2[i]}{m[i]} + \frac{\lambda_{i+D}}{K}\left[1 - \frac{\lambda_i}{m[i]}\right],$$

with $m[i]$ defined in (25). We have the following:

Theorem 2. *Let $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_L]^T$, with $0 \leq \lambda_1 < \cdots < \lambda_L$ and let i_0 be the largest integer satisfying $\bar{\beta}_1[i_0] < \zeta_2[i_0]$. Then, if $\bar{\beta}_1[i_0] \leq \zeta_1[i_0]$, the minimizer of the right-hand side of (28) subject to $\boldsymbol{0}_L \preceq \boldsymbol{q} \preceq \boldsymbol{1}_L$ and $\boldsymbol{1}_L^T\boldsymbol{q} = K$ is given by*

$$q_i^\star = \begin{cases} 0 & \textit{if } \; i \in \mathcal{I}_0[i_0] \cup \{i_0+D\} \\ 1 & \textit{if } \; i \in \mathcal{I}_1[i_0] \cup \{i_0\}. \end{cases}$$

Otherwise, if $\bar{\beta}_1[i_0] > \zeta_1[i_0]$, the minimizer is given by

$$q_i^\star = \begin{cases} 0 & \textit{if } \; i \in \mathcal{I}_0[i_0] \\ 1 & \textit{if } \; i \in \mathcal{I}_1[i_0] \\ \delta & \textit{if } \; i = i_0 \\ 1-\delta & \textit{if } \; i = i_0 + D, \end{cases}$$

where $\delta = (\zeta_2[i_0] - \bar{\beta}_1[i_0])/(\zeta_2[i_0] - \zeta_1[i_0])$.

The proof follows similar guidelines as for Theorem 1 and it is omitted due to lack of space. Observe that the structure of the solutions revealed by Theorem 2 in the high SNR regime is analogous to that of the solutions from Theorem 1 in the low SNR case. The differences lay in the resulting index i_0. The remarks at the end of Sec. IV-A are applicable also here.

V. Numerical Example

Consider a setting with white noise ($\boldsymbol{\Sigma_w} = \boldsymbol{I}_L$) of unknown power and an auto-regressive signal term obtained according to $s_n = 0.5 \cdot s_{n-1} + z_n$, where z_n is zero-mean white complex Gaussian whose power is set so that $\operatorname{E}\left[|s_n|^2\right] = 1$. The CRB

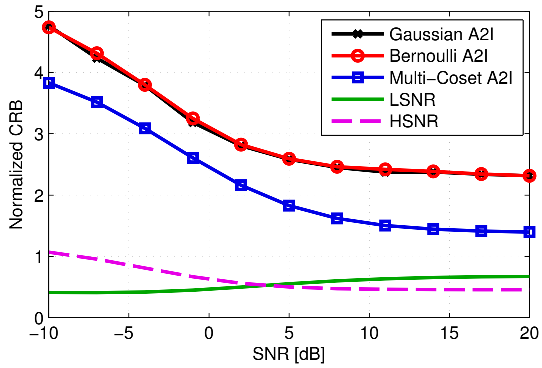

Fig. 1: The CRB associated with the proposed designs is much smaller than for conventional samplers.

is normalized by $\frac{L}{K}$ times the uncompressed CRB ($\boldsymbol{\Phi} = \boldsymbol{I}_L$). We take $L = 80$ and $K = 15$, yielding a compression rate of $L/K \approx 5.33$. Five different designs are considered: (i) $\boldsymbol{\Phi}$ has i.i.d. entries, drawn from a $\mathcal{CN}(0, 1)$ distribution; (ii) $\boldsymbol{\Phi}$ has i.i.d. entries taking the values ± 1 with equal probability; (iii) $\boldsymbol{\Phi}$ is a linear sparse ruler (LSR) sampler [14], [15]; (iv) the design form Sec. IV-A; and (v) the design from Sec. IV-B. For the random samplers, results are averaged over 500 realizations. As seen in Fig. 1, although the LSR sampler consistently outperforms the random samplers, the CRB for the proposed designs is considerably smaller than those of the other schemes, which do not exploit the a priori information, especially for low SNR. For SNR < 3.5 dB, the low-SNR design performs better than the high-SNR scheme. For larger SNRs, the situation is reversed. In any case, the low-SNR design is seen to be nearly optimal for the whole SNR range. The fact that the associated CRB increases with the SNR is a consequence of a sub-optimal choice of i_0 in high SNR.

VI. Conclusions

By exploiting spectral prior information, we have derived designs for compression matrices that approximately minimize the CRB for estimating the power of a signal in (possibly) colored noise. We considered the cases where the noise power is either known or unknown, and the cases of low and high SNR. The optimum sampling matrix computes linear combinations of the frequency bins with the highest SNR values when the noise power is known, and those with both the highest and lowest SNR values when the noise power is unknown. In the latter case, the fact that the CRB is similar for the low-SNR and high-SNR designs presented suggests that their performance is nearly optimal in the middle SNR regime. Practical implementation of these compression matrices may be feasible by making use of analog/mixed-signal based FFT processors [25]–[27] in the design of the AIC.

References

[1] 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 *IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 1959–1962.

[2] M. Mishali and Y. C. Eldar, "Sub-Nyquist sampling: bridging theory and practice," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 98–124, 2011.

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

[4] E. Axell, G. Leus, E.G. Larsson, and H.V. Poor, "Spectrum sensing for cognitive radio : State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[5] A.I. Perez-Neira, Miguel-Angel Lagunas, M.A. Rojas, and P. Stoica, "Correlation matching approach for spectrum sensing in open spectrum communications," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4823–4836, Dec. 2009.

[6] G. Vazquez-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 *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 2958–2961.

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

[8] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Signal Process.*, in press.

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

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

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

[12] Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing-based mimo radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5338–5352, 2011.

[13] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[14] D. Romero and G. Leus, "Compressive Covariance Sampling," in *Inf. Theory Appl. Workshop*, 2013.

[15] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *Submitted to IEEE Trans. on Inf. Theory. Available at http://arxiv.org/abs/1311.0737*.

[16] S. M. Kay, *Fundamentals of Statistical Signal Process., Vol. I: Estimation Theory*, Prentice-Hall, 1993.

[17] L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Sensitivity considerations in compressed sensing," in *Asilomar Conf. Signals, Syst., Comput.*, 2011, pp. 744–748.

[18] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive estimation in AWGN: General observations and a case study," in *Asilomar Conf. Signals, Syst., Comput.*, 2012, pp. 953–957.

[19] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramer-Rao bound for nonlinear parameter estimation after compressed sensing," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 6630–6634.

[20] S. M. Kay, *Fundamentals of Statistical Signal Process., Vol. II: Detection Theory*, Prentice-Hall, 1998.

[21] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[22] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[23] W. J. Bangs, "Array processing with generalized beamformers," *PhD Thesis, Yale University*, 1971.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[25] N. Sadeghi, V. Gaudet, and C. Schlegel, "Analog DFT processors for OFDM receivers: Circuit mismatch and system performance analysis," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 9, pp. 2123–2131, 2009.

[26] M. Lehne and S. Raman, "A discrete-time FFT processor for ultrawideband OFDM wireless transceivers: Architecture and behavioral modeling," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 11, pp. 3011–3022, 2010.

[27] S. P. Chepuri, R. de Francisco, and G. Leus, "Low-power architecture for wideband spectrum sensing," in *3rd Int. Workshop Cognitive Inform. Process.*, 2012.