

# Asymptotic Analysis of Synchronous Signal Processing

Marc Vilà-Insa , and Jaume Riba , *Senior Member, IEEE*

**Abstract**—This paper extends various theoretical results from stationary data processing to cyclostationary (CS) processes under a unified framework. We first derive their asymptotic eigenbasis, which provides a link between their Fourier and Karhunen-Loève (KL) expansions, through a unitary transformation dictated by the cyclic spectrum. By exploiting this connection and the optimalities offered by the KL representation, we study the asymptotic performance of smoothing, filtering and prediction of CS processes, without the need for deriving explicit implementations. We obtain minimum mean squared error expressions that depend on the cyclic spectrum and include classical limits based on the power spectral density as particular cases. We conclude this work by applying the results to a practical scenario, in order to quantify the achievable gains of synchronous signal processing.

**Index Terms**—Cyclostationary processes, cyclic Wiener filter, Karhunen-Loève expansion, asymptotic bounds.

## I. INTRODUCTION

**M**ANY random phenomena in nature and engineering exhibit statistical properties that vary periodically with time. Notable examples are encountered in Earth sciences (e.g. hydrology, oceanography and climatology [1]) as well as in human activities (e.g. econometrics, electronic design and mechanical engineering [1]). Of particular interest are the signals used in digital communications and radar systems, where periodicities occur due to the use of constant symbol rates and carrier frequencies [2, Ch. 12].

These cyclic behaviors are typically modeled statistically as *cyclostationary* (CS) or *periodically correlated* stochastic processes [3]. The research on signal processing methods that exploit their unique characteristics is broad, rich and mature [4]. Numerous applications have emerged in areas in which these phenomena typically arise, which include signal extraction/separation from spatial or temporal mixtures [2, Ch. 14], modeling [5, Ch. 7], blind system equalization and identification [6], spectrum sensing [7] and time-delay estimation [3]. They are usually based on exploiting the diversity offered by such structured statistical change, and their performance is generally affected by the accuracy in prior knowledge of the inherent periods of variation.

This work was (partially) funded by project RODIN (PID2019-105717RB-C22) by MCIU/AEI/10.13039/501100011033, project MAYTE (PID2022-136512OB-C21) by MCIU/AEI/10.13039/501100011033 and ERDF “A way of making Europe”, grant 2021 SGR 01033 and grant 2022 FI SDUR 00164 by Departament de Recerca i Universitats de la Generalitat de Catalunya.

The authors are with the Signal Processing and Communications Group (SPCOM), Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain (e-mail: {marc.vila.insa, jaume.riba}@upc.edu).

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

Concerning linear filtering, it is well known that the *cyclic Wiener filter* (CWF) [8] is the optimal periodically time-variant linear estimator for CS signals. It can be understood as a combination of several linear shift-invariant filters operating on replicas of the signal of interest under different frequency shifts (FRESH) [9, Ch. 10], thus providing a variety of improved designs [7]. Compared to simpler methods that assume stationarity, detectors and estimators derived from CS modeling perform better at discriminating signal from interference, by being aware of the intrinsic period of the signals involved, either the one of interest [10], [11], or the one considered interference [12], [13].

The derivation of asymptotic performance limits of linear time-invariant (LTI) filters with unconstrained length has been a fundamental tool within the classical theory of stationary signals. For a given problem, they provide the ultimate achievable performance in advance and inform about how far a constrained implementation is from the optimum, thus shedding light on the achieved complexity/performance trade-off. Not less important is their role in unveiling key physical aspects of the signal that are the most relevant for the pursued processing performance.

In this sense, several bounds are well-known for stationary processes, on account of their asymptotic uncorrelatedness through Fourier analysis. This property, which is in fact a corollary of the more general *Szegö limit theorems*, comes from the asymptotic diagonalization of Toeplitz matrices through the unitary Fourier basis [14]. Examples of these limits are expressions of minimum mean squared error (MMSE) through integral operators involving spectral coherences (i.e. Wiener smoothing in the frequency domain [15, Sec. 12.3]). The *Kolmogorov-Szegö formula* [16, Ch. 6] on the MMSE for asymptotic one-step linear prediction is another remarkable limit expression, and involves integral operators over the log-spectrum. When causality is a design restriction, such as in applications requiring small processing latency, causal Wiener filtering becomes relevant. The performance limit in this case involves, as in linear prediction, integral operators on the log-spectrum [17]. Since these causal formulas resemble Shannon capacity expressions, information theoretic interpretations have been explored in the literature [18]. A fundamental result, known as the *I-MMSE formula* (or *Guo-Shamai-Verdú (GSV) theorem*), was obtained in [19] and establishes a direct path between the MMSE of causal and non-causal processing for a very general kind of filtering.

Given the prevalence of CS signals in both theoretical [20] and applied [21] research fields, one would expect to find results parallel to the ones mentioned for stationary data. However, this is mostly not the case, possibly due to the lack of more

general asymptotic properties of their autocorrelation structures. Therefore, the main purpose of this paper is going in the direction of filling this gap and providing new insights within the context of filtering CS signals. The goal is to understand the ultimate performance gain achievable from synchronous signal processing, under the assumption of attainable synchronization with the periodic phenomenon that is inherent to the desired noisy data.

Classical treatments of CS signals represent them through their *Cramér-Loève (CL)* spectral expansion [9, Ch. 10], which provides a direct physical interpretation of the problem at hand. This approach, however, leads to processes with correlated spectral increments for non-stationary data. In contrast, we follow a fully uncorrelated approach based on the asymptotic eigendecomposition of periodic Toeplitz matrices [22]. To this end, we leverage known results for stationary data and extend them without breaking the inherent structure of the signals, such that the resulting formulas remain interpretable. In particular, we avoid the classical polyphase [3] and Gladyshev decompositions [23, Sec. 7.2] of CS signals onto multivariate low-rate stationary data, which are useful for implementation purposes but blur the connection between performance bounds and physical parameters.

Moreover, we capitalize on the general GSV theorem to explore the causal CWF, instead of studying the causal constraint specific to the CS problem. By following this alternative route, we obtain mathematical expressions of performance bounds that depend on the cyclic spectrum explicitly, thus becoming insightful generalizations that contain classical limits based on the spectrum as particular cases. The tools developed in the paper allow to obtain theoretical results for any particular application: they provide knowledge from the problem without the need for deriving explicit implementations. The exposition is probabilistic and focused on discrete-time signals and systems encountered in signal processing and communications, while maintaining a continuous spectrum interpretation. As a result, it is possible to plot performance limits against the relevant signal parameters in a very direct manner, as integral expressions.

The main contributions of this paper are listed below:

- We revisit and refine results obtained in [22]. Using an incremental formulation [9, Ch. 8], we obtain the asymptotic *Karhunen-Loève (KL)* basis of CS processes. This allows an interpretation of such KL expansion as a decorrelation of their CL representation, bridging them through a unitary transformation.
- We extend classical results of stationary filtering to CS processes. In particular, we provide a new sense of optimality for the CWF by proving its equivalence to the Wiener filter in the KL domain, which exploits the energy compaction properties of such representation.
- We develop a unified theoretical analysis of smoothing, filtering and prediction of CS processes thanks to the discrete-time approach and the use of [19, Th. 8] to connect the three regimes. We quantify the MMSE and maximum achievable gain with synchronous processing in each case.
- We relate the obtained MMSE expressions to the *spectral coherence* and *coherence matrix* [24] in an explicit manner.

The text is organized as follows. Section II establishes the basic notions and concepts used in the rest of the derivations. Section III develops the theoretical results regarding the KL expansion of CS processes, as well as its connection to the CL expansion and other properties of interest. The filtering problem is explored throughout the rest of the paper: Section IV presents the model, while Sections V to VII study various asymptotic performance metrics for smoothing, filtering and prediction, respectively. Finally, some numerical illustrations of the theoretical contents are displayed in Section VIII.

*Notation:* Vectors and matrices are denoted by boldface lowercase and uppercase letters. An element  $(r, c)$  from a matrix  $\mathbf{A}$  is indicated by  $[\mathbf{A}]_{r,c}$ . The conjugate, transpose and conjugate transpose operators are  $\cdot^*$ ,  $\cdot^T$  and  $\cdot^H$ , respectively. The inverse of  $\mathbf{A}$  is  $\mathbf{A}^{-1}$ . The trace of  $\mathbf{A}$  is  $\text{Tr}\{\mathbf{A}\}$ , while its determinant is  $|\mathbf{A}|$ . The identity matrix of size  $N$  is  $\mathbf{I}_N$ , and an all-zeros vector of size  $M$  is  $\mathbf{0}_M$ . A product of matrices that depend on the same parameter  $\mathbf{A}(s)\mathbf{B}(s)\cdots\mathbf{C}(s)$  is shortened by  $(\mathbf{A}\mathbf{B}\cdots\mathbf{C})(s)$ . Operator  $\text{Diag}(a, b, \dots, c)$  constructs a diagonal matrix from the set  $\{a, b, \dots, c\}$ , while  $\text{diag}(\mathbf{A})$  performs the opposite action. The imaginary unit is  $j$ . The expectation operator is  $\mathbb{E}[\cdot]$ . Dirac delta is  $\delta(s)$  and Kronecker delta is  $\delta_{a,b}$ . The Dirac comb with separation  $T$  is  $\text{III}_T(s)$ . If a set  $a$  majorizes another set  $b$ , it is indicated as  $a \succ b$ . The element-wise product between matrices is  $\circ$ .

## II. PRELIMINARY DEFINITIONS AND BACKGROUND

Let  $x(n) \in \mathbb{C}$  be a discrete-time random process with time index  $n \in \mathbb{Z}$ . For a clearer exposition, all the considered processes will be assumed zero-mean and *proper* (i.e.  $\mathbb{E}[x(n)x(n+m)] = 0, \forall m \in \mathbb{Z}$ ) [9, Sec. 1.6.2]. The autocorrelation function of  $x(n)$  is defined as

$$R_x(n, m) \triangleq \mathbb{E}[x(n+m)x^*(n)], \quad (1)$$

where index  $m$  is referred to as *lag*. Various classes of random processes can be identified based on properties of their autocorrelation. In particular, CS processes display a periodic structure with period  $P$  within their autocorrelation function:

$$R_x(n, m) \equiv R_x(n+lP, m), \quad \forall l \in \mathbb{Z}. \quad (2)$$

When  $P = 1$ , a CS process becomes *wide-sense stationary (WSS)*, whose autocorrelation function is independent from the time index:  $R_x(n, m) \equiv R_x(m)$  (i.e. it displays *shift-invariant second-order statistics* [9, Ch. 8]).

The following definition will prove useful in the sequel. The *cyclic spectrum* is the two-dimensional Fourier transform of the autocorrelation function, both in the time and lag domains. Particularized for CS processes, it is expressed as

$$S_X^{(\frac{k}{P})}(f) = \frac{1}{P} \sum_{n=0}^{P-1} \sum_{m \in \mathbb{Z}} R_x(n, m) e^{-j2\pi(\frac{nk}{P} + mf)}, \quad (3)$$

for all  $k \in \{0, \dots, P-1\}$ . This is known as the *cyclic Wiener-Khinchin relationship* [25, Eq. (39a)].

While these expressions provide a statistical representation of the properties of  $x(n)$ , in many signal processing applications it is convenient to have a spectral representation of the process

itself. In the following sections, the CL and KL expansions are presented.

#### A. CL spectral representation

A complex second-order random process [23, Ch. 1]  $x(n)$  can be expressed as

$$x(n) = \int_0^1 e^{j2\pi n f} d\nu_x(f), \quad (4)$$

where  $d\nu_x(f)$  are the increments of a process in the spectral domain. This Riemann-Stieltjes integral representation is known as the CL expansion, and processes that can be expressed in such a way are called *harmonizable* [23, Ch. 5].

The CL expansion is of great conceptual interest since it decomposes  $x(n)$  onto the Fourier basis (complex exponentials), thus preserving the frequency interpretation of the transformed domain. If the derivative process  $X(f) \triangleq \frac{d\nu_x(f)}{df}$  existed, it would be the discrete-time Fourier transform (DTFT) of  $x(n)$  [26, Eq. (4.13)]. However,  $x(n)$  is not square-integrable in general and its DTFT might be ill-defined. Thus we resort to the increment definition<sup>1</sup>:

$$d\nu_x(f) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x(n) e^{-j2\pi n f}. \quad (5)$$

In a similar manner to (1), a correlation function between frequently displaced spectral increments can be defined [9, Sec. 9.2]:

$$S_X(f, \alpha) df d\alpha \triangleq E[d\nu_x(f) d\nu_x^*(f - \alpha)], \quad (6)$$

known as *spectral correlation*. When  $\alpha = 0$ , it is convenient to define the *power spectral density (PSD)* [9, Sec. 9.2.2]:

$$S_X(f) df \triangleq E[|d\nu_x(f)|^2]. \quad (7)$$

Many stochastic processes can be described from properties of their CL spectral representation. For instance, WSS processes display the unique characteristic of having orthogonal increments [9, Sec. 8.1], which translates into

$$S_X^{(WSS)}(f, \alpha) df d\alpha = S_X(f) \delta(\alpha) df d\alpha. \quad (8)$$

On the contrary, CS processes present the following spectral correlation [9, Sec. 10.1.1]:

$$S_X^{(CS)}(f, \alpha) df d\alpha = \sum_{k=0}^{P-1} S_X^{(\frac{k}{P})}(f) \delta(\alpha - \frac{k}{P}) df d\alpha, \quad (9)$$

in terms of the cyclic PSD from (3).

Representing a non-WSS process over the Fourier domain comes at the cost of losing the orthogonality property (8). For this reason, it might be useful to study an alternative spectral expansion that preserves it.

### III. KL REPRESENTATION OF CS PROCESSES

Given a second-order random process  $x(n)$ , its KL expansion [9, Sec. 9.1] is defined as

$$x(n) = \int_0^1 \phi(n, \lambda) d\xi_x(\lambda), \quad (10)$$

such that  $d\xi_x(\lambda)$  are increments of a random process in the KL spectral domain. The set  $\{\phi(n, \lambda)\}$ , for all  $n \in \mathbb{Z}$  and  $\lambda \in [0, 1)$ , contains eigenfunctions that satisfy the *orthonormality condition* [27, Def. 2.17] and *completeness relation* [27, Th. 2.13], which are

$$\begin{cases} \int_0^1 \phi(n, \lambda) \phi^*(n', \lambda) d\lambda = \delta_{n-n'} \\ \sum_{n \in \mathbb{Z}} \phi(n, \lambda) \phi^*(n, \lambda') = \delta(\lambda - \lambda') \end{cases}, \quad (11)$$

respectively. They are obtained by solving the following eigenequation [28, Ch. 6]:

$$\sum_{l \in \mathbb{Z}} R_x(l, k-l) \phi(l, \lambda) = \mathcal{S}_X(\lambda) \phi(k, \lambda). \quad (12)$$

The term  $\mathcal{S}_X(\lambda)$  is the *KL spectrum* of  $x(n)$ . By combining (11) and (12), it is straightforward to obtain explicit conversion formulas:

$$\begin{aligned} R_x(l, k-l) &= \int_0^1 \phi(k, \lambda) \mathcal{S}_X(\lambda) \phi^*(l, \lambda) d\lambda \\ \mathcal{S}_X(\lambda) \delta(\lambda - \lambda') &= \sum_{k, l \in \mathbb{Z}} \phi(l, \lambda) R_x(l, k-l) \phi^*(k, \lambda'). \end{aligned} \quad (13)$$

Complementary to (10), the KL transform<sup>2</sup> of  $x(n)$  is defined by projecting the random process onto its KL basis  $\{\phi(n, \lambda)\}$ :

$$d\xi_x(\lambda) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x(n) \phi^*(n, \lambda) \triangleq \text{KL}\{x(n)\}(\lambda). \quad (14)$$

The defining characteristic of the KL expansion is the orthogonality between spectral increments, *i.e.*

$$\begin{aligned} \mathcal{S}_X(\lambda, \beta) d\lambda d\beta &\triangleq E[d\xi_x(\lambda) d\xi_x^*(\lambda - \beta)] \\ &= \mathcal{S}_X(\lambda) \delta(\beta) d\lambda d\beta, \end{aligned} \quad (15)$$

which can be derived from (14) and (13). This property generalizes (8) and will prove to be fundamental in subsequent derivations of simple MMSE expressions. The trade-off for spectral uncorrelatedness is the fact that each class of random processes has a different eigenbasis, compared to the simplicity of the CL representation. WSS processes are special in this sense because their KL and CL decompositions coincide and are both asymptotically given by the orthonormal Fourier basis.

The KL representation of CS processes of period  $P$  was obtained in [22]. To define it, the full KL domain  $[0, 1)$  is segmented into  $P$  non-overlapping domains and  $\lambda$  is replaced by the pair  $(p, \sigma)$ . The continuous variable  $\sigma$  is defined in a spectral sub-band  $[0, 1/P)$  and  $p \in \{0, 1, \dots, P-1\}$  is an index that selects one of these subintervals, *i.e.*  $\lambda \triangleq \sigma + p/P$ .

<sup>2</sup>Expression (10) is sometimes referred to as *backward KL transform* or *synthesis*. Similarly, *forward KL transform* or *analysis* are used for (14).

<sup>1</sup>We assume  $N \in 2\mathbb{N}$  without loss of generality.

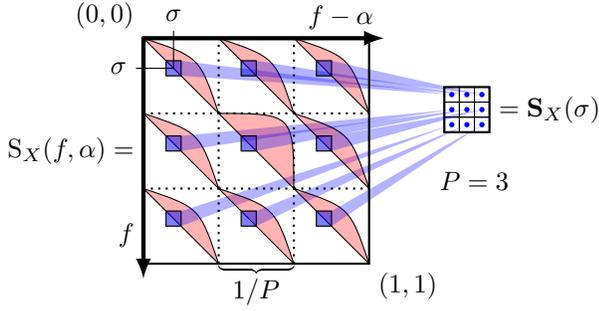


Figure 1. Graphical representation of the construction of  $\mathbf{S}_X(\sigma)$  from the spectral correlation  $S_X(f, \alpha)$  of a CS process, which is only defined across  $\delta$ -ridges spaced by  $1/P$  both horizontally and vertically [9, Sec. 10.1.1].

*Theorem 1:* (Riba & Vilà-Insa [22, Th. 1]) The KL eigenbasis of a CS process of period  $P$  takes the form:

$$\phi_{\text{CS}}^{(p)}(n, \sigma) = \sum_{q=0}^{P-1} b_X^{(p)}(q, \sigma) e^{j2\pi n(\sigma + \frac{q}{P})}. \quad (16)$$

Weighting coefficients  $b_X^{(p)}(q, \sigma)$  are grouped in vectors

$$\mathbf{b}_X^{(p)}(\sigma) \triangleq [b_X^{(p)}(0, \sigma), \dots, b_X^{(p)}(P-1, \sigma)]^T, \quad (17)$$

and obtained by solving the following eigenproblem:

$$\mathbf{S}_X(\sigma) \mathbf{b}_X^{(p)}(\sigma) = \mathcal{S}_X^{(p)}(\sigma) \mathbf{b}_X^{(p)}(\sigma). \quad (18)$$

Terms  $\mathcal{S}_X^{(p)}(\sigma) \triangleq \mathcal{S}_X(\sigma + p/P)$  are the eigenvalues of the Hermitian non-negative definite cyclic PSD matrix of the process,  $\mathbf{S}_X(\sigma) \in \mathbb{C}^{P \times P}$ , given by

$$[\mathbf{S}_X(\sigma)]_{r,c} \triangleq \mathcal{S}_X^{(\frac{r-c}{P})}(\sigma + \frac{r}{P}). \quad (19)$$

*Proof:* See Appendix A.

The cyclic PSD matrix can be equivalently defined as

$$\begin{aligned} [\mathbf{S}_X(\sigma)]_{r,c} d\sigma &= \mathbb{E}[d\nu_x^{(r)}(\sigma) d\nu_x^{(c)*}(\sigma)] \\ &= S_X(\sigma + \frac{r}{P}, \frac{r-c}{P}) d\sigma d\sigma, \end{aligned} \quad (20)$$

where  $d\nu_x^{(p)}(\sigma) \triangleq d\nu_x(\sigma + p/P)$ . This shows how  $\mathbf{S}_X(\sigma)$  can be constructed from  $P^2$  equally spaced elements of the two-dimensional spectral correlation  $S_X(f, \alpha)$ , as illustrated in Fig. 1. The next result relates the CL and KL expansions of a CS process.

*Corollary 1:* (Connection between CL and KL expansions of CS processes) The KL spectral representation of a CS random process  $x(n)$  is related to its CL spectral representation through the following equality:

$$\tilde{\mathbf{x}}(\sigma) = \mathbf{B}_X^H(\sigma) \check{\mathbf{x}}(\sigma), \quad (21)$$

where

$$\begin{aligned} \tilde{\mathbf{x}}(\sigma) &\triangleq [d\xi_x^{(0)}(\sigma), \dots, d\xi_x^{(P-1)}(\sigma)]^T \\ \check{\mathbf{x}}(\sigma) &\triangleq [d\nu_x^{(0)}(\sigma), \dots, d\nu_x^{(P-1)}(\sigma)]^T \end{aligned} \quad (22)$$

are constructed from  $P$  samples of the KL and CL spectral processes, respectively, and  $d\xi_x^{(p)}(\sigma) \triangleq d\xi_x(\sigma + p/P)$ . Matrix  $\mathbf{B}_X(\sigma)$  contains the eigenbasis of  $\mathbf{S}_X(\sigma)$ :

$$\mathbf{B}_X(\sigma) \triangleq [\mathbf{b}_X^{(0)}(\sigma), \dots, \mathbf{b}_X^{(P-1)}(\sigma)]. \quad (23)$$

*Proof:* See Appendix B. ■

This result implies the KL transform of a CS process can be thought of as a CL transform followed by a decorrelation through unitary matrix  $\mathbf{B}_X(\sigma)$ . A straightforward consequence of this is summarized in the following corollary.

*Corollary 2:* (Connection between cyclic PSD and KL-PSD matrices of CS processes) The KL-PSD matrix

$$\begin{aligned} \mathcal{S}_X(\sigma) d\sigma &\triangleq \text{Diag}(\mathcal{S}_X^{(0)}(\sigma), \dots, \mathcal{S}_X^{(P-1)}(\sigma)) d\sigma \\ &= \mathbb{E}[\tilde{\mathbf{x}}(\sigma) \tilde{\mathbf{x}}^H(\sigma)] \end{aligned} \quad (24)$$

contains the  $P$  eigenvalues of the cyclic PSD matrix  $\mathbf{S}_X(\sigma)$ .

*Proof:* Using relationships (20) and (21), we have:

$$\begin{aligned} \mathcal{S}_X(\sigma) d\sigma &= \mathbf{B}_X^H(\sigma) \mathbb{E}[\check{\mathbf{x}}(\sigma) \check{\mathbf{x}}^H(\sigma)] \mathbf{B}_X(\sigma) \\ &= (\mathbf{B}_X^H \mathbf{S}_X \mathbf{B}_X)(\sigma) d\sigma. \end{aligned} \quad (25)$$

### A. Temporal dependence of KL transforms

While the reference time  $n_0$  has been assumed 0 and omitted from the previous analysis, its impact on KL representations should be addressed. In general, an arbitrary time shift will affect the KL expansion of a random process.

The two classes of stochastic processes considered here are special in this sense. The KL basis of a WSS process remains unchanged by a temporal shift, whereas that of a CS process depends on the reference time [22]:

$$\phi_{\text{CS}}^{(p)}(n, \sigma, n_0) = \sum_{q=0}^{P-1} b_X^{(p)}(q, \sigma) e^{j2\pi n_0 \frac{q}{P}} e^{j2\pi n(\sigma + \frac{q}{P})}. \quad (26)$$

Regarding their spectral representation, a translation in time becomes a phase shift in the KL domain for both of them:

$$\begin{aligned} d\xi_x^{(p)}(\sigma, n_0) &\triangleq \text{KL}\{x(n + n_0)\}(p, \sigma) \\ &= \text{KL}\{x(n)\}(p, \sigma) e^{j2\pi n_0 \sigma}. \end{aligned} \quad (27)$$

This property is derived for CS processes in Appendix C<sup>3</sup>.

On the contrary, the KL spectrum of both classes is invariant to temporal translations, since the effect of the phase in (27) disappears. This phenomenon is easily seen by using definition (15):

$$\begin{aligned} \mathcal{S}_X^{(p)}(\sigma, n_0) d\sigma &= \mathbb{E}[|d\xi_x^{(p)}(\sigma, n_0)|^2] = \mathbb{E}[|d\xi_x^{(p)}(\sigma) e^{j2\pi n_0 \sigma}|^2] \\ &= \mathbb{E}[|d\xi_x^{(p)}(\sigma)|^2] \equiv \mathcal{S}_X^{(p)}(\sigma) d\sigma. \end{aligned} \quad (28)$$

Since we are interested in obtaining asymptotic theoretical bounds, this property simplifies the subsequent derivations.

### B. Parseval's Theorem in the KL domain

The average power of a random process  $x(n)$  is defined as

$$P_x \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \mathbb{E}[|x(n)|^2]. \quad (29)$$

<sup>3</sup>Refer to [26, Eq. (4.33)] for the WSS case.

Since the KL transform is unitary, it preserves power in the spectral domain. A result similar to *Parseval's theorem* for the frequency domain can thus be stated [26, Eq. (2.80)]:

$$P_x = \int_0^1 \mathbb{E}[|d\xi_x(\lambda)|^2] = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) d\sigma. \quad (30)$$

Using the segmented KL spectrum allows to more compactly express the previous integral, by transforming the summation into a trace and applying Corollary 2:

$$P_x = \int_0^{\frac{1}{P}} \text{Tr}\{\mathcal{S}_{\mathcal{X}}(\sigma)\} d\sigma = \int_0^{\frac{1}{P}} \text{Tr}\{\mathbf{S}_X(\sigma)\} d\sigma. \quad (31)$$

This way, we have obtained the power of a CS process in terms of its cyclic PSD matrix.

### C. Energy compaction and minimum representation entropy

While most unitary transforms exhibit some sort of energy concentration [26, Sec. 9.2.3], the KL transform does it optimally. Let  $\{\theta(n, \lambda)\}$  form an arbitrary unitary basis for a process  $x(n)$ :

$$x(n) = \int_0^1 \theta(n, \lambda) d\eta_x(\lambda), \quad (32)$$

where  $d\eta_x(\lambda)$  is its spectral representation, *i.e.*

$$d\eta_x(\lambda) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x(n) \theta^*(n, \lambda). \quad (33)$$

Since we are dealing with a unitary transform, the power of  $x(n)$  can be obtained in the spectral domain due to Parseval's theorem:

$$P_x = \int_0^1 \mathbb{E}[|d\eta_x(\lambda)|^2] \triangleq \int_0^1 \mathcal{S}_{\mathcal{X}, \theta}(\lambda) d\lambda. \quad (34)$$

We define the power contained in a portion  $[0, \rho)$  of the spectrum:

$$P_x(\rho, \{\theta(n, \lambda)\}) \triangleq \int_0^{\rho} \mathcal{S}_{\mathcal{X}, \theta}(\lambda) d\lambda, \quad 0 < \rho < 1. \quad (35)$$

Of all possible bases that fulfill (11), we aim for the one that maximizes it for all  $\rho$ ; *i.e.*

$$\begin{aligned} \{\phi(n, \lambda)\} &= \arg \max_{\{\theta(n, \lambda)\}} P_x(\rho, \{\theta(n, \lambda)\}) \\ \text{s.t.} \quad &\sum_{n \in \mathbb{Z}} \theta(n, \lambda) \theta^*(n, \lambda') = \delta(\lambda - \lambda'). \end{aligned} \quad (36)$$

The solution to this problem, which can be obtained through Lagrange multipliers [26, Th. 9.1], is the set of eigenfunctions  $\{\theta(n, \lambda)\}$  that satisfies

$$\sum_{l \in \mathbb{Z}} R_x(l, k-l) \theta(l, \lambda) = \chi(\lambda) \theta(k, \lambda). \quad (37)$$

By definition (see (12)), this is the KL basis  $\{\phi(n, \lambda)\}$ , and  $\chi(\lambda)$  must be the KL-PSD  $\mathcal{S}_{\mathcal{X}}(\lambda)$ .

This optimality in energy compaction can be quantified with a metric known as *representation entropy* [29] or *spectral entropy* [30], which indicates the compression of information achieved by a spectral representation. Let  $p_{x, \theta}(\lambda) \triangleq$

$\mathcal{S}_{\mathcal{X}, \theta}(\lambda)/P_x$  be the normalized spectrum of  $x(n)$  for a given orthonormal basis  $\{\theta(n, \lambda)\}$ . Since  $p_{x, \theta}(\lambda)$  is nonnegative and its integral is 1, it can be interpreted as a probability density function. We may then define the differential entropy associated with its corresponding representation as

$$h(\{\theta(n, \lambda)\}) \triangleq - \int_0^1 p_{x, \theta}(\lambda) \cdot \ln(p_{x, \theta}(\lambda)) d\lambda. \quad (38)$$

*Proposition 1: (Representation entropy)* The KL expansion has the minimum representation entropy:

$$h(\{\phi(n, \lambda)\}) \leq h(\{\theta(n, \lambda)\}), \quad \forall \{\theta(n, \lambda)\}. \quad (39)$$

*Proof:* Let  $f: [0, 1] \mapsto \mathbb{R}_{\geq 0}$  be the continuous concave function  $f(a) \triangleq -a \cdot \ln(a)$ . Knowing that

$$P_x(\rho, \{\phi(n, \lambda)\}) \geq P_x(\rho, \{\theta(n, \lambda)\}), \quad \forall \{\theta(n, \lambda)\}, \quad (40)$$

using [31, Def. 14.H.1] it can be stated that  $\mathcal{S}_{\mathcal{X}, \theta} \prec \mathcal{S}_{\mathcal{X}, \phi}$ . By [31, Prop. 14.H.1.a], we can say that

$$\int_0^1 f(p_{x, \phi}(\lambda)) d\lambda \leq \int_0^1 f(p_{x, \theta}(\lambda)) d\lambda, \quad \forall \{\theta(n, \lambda)\}, \quad (41)$$

which is equivalent to (39).  $\blacksquare$

Representation entropy is a measure of spectral redundancy, such as the one emerging from the spectrally correlated components of CS processes. It is of interest to determine this form of frequency diversity [9, Sec. 9.2.3] since it is precisely the signal characteristic that synchronous processing exploits.

## IV. CYCLIC WIENER FILTERING: PROBLEM STATEMENT

The previous formulation of CS processes can be applied to the theory of Wiener filtering in order to obtain asymptotic bounds on the achievable performance. We will focus on the following framework. We deal with an observation  $x(n)$  of a reference signal  $d(n)$  distorted by additive noise  $z(n)$ :

$$x(n) = d(n) + z(n). \quad (42)$$

They are assumed uncorrelated (*i.e.*  $\mathbb{E}[d(n)z^*(n')] = 0, \forall n, n'$ ). Either  $d(n)$  or  $z(n)$  will be WSS with known PSD, while the other one will be CS. This combination of random processes presents many desirable properties. By passing  $x(n)$  through a stable LTI filter, its WSS component can be whitened [15, Sec. 11.3.3], while the CS properties of the other one are preserved, as proven in [23, Sec. 6.8.1]. Therefore, and without loss of generality, the WSS process will be assumed white from this point onwards.

White WSS processes are of great interest because they admit any orthonormal KL eigenbasis. This allows the following simplification:

$$\begin{aligned} d\xi_x^{(p)}(\sigma) &= \text{KL}\{x(n)\}(p, \sigma) = \text{KL}\{d(n) + z(n)\}(p, \sigma) \\ &= \text{KL}\{d(n)\}(p, \sigma) + \text{KL}\{z(n)\}(p, \sigma) \\ &= d\xi_d^{(p)}(\sigma) + d\xi_z^{(p)}(\sigma). \end{aligned} \quad (43)$$

This simple model will provide various insights on the filtering problem and can be adapted to represent different cases of interest:

- **CS signal  $d(n)$  and white WSS noise  $z(n)$ :** this is a typical scenario in digital wireless communications, which usually involve the reception of CS signals hindered by additive noise [8].
- **Wideband signal  $d(n)$  and narrowband CS interference  $z(n)$ :** in this scenario, which is often encountered in spread-spectrum systems, the high degree of predictability of  $z(n)$  is used against the unpredictability of the desired term  $d(n)$  to improve signal separation. The CS nature of  $z(n)$  can be exploited beyond the spectral domain to further reduce the distortion of the desired component after a cancellation filter [32].

We will only develop results for the first case for brevity. Extending the following analysis to the second one is straightforward and yields complementary expressions.

Our objective is to design a filter  $w(\cdot)$ , such that it minimizes the average filtering error power, or mean squared error (MSE), for a time interval  $N_1 \leq n \leq N_2$ :

$$\begin{aligned} \text{MSE}(N_1, N_2) &\triangleq \frac{1}{N_2 - N_1} \sum_{n=N_1}^{N_2} \text{E}[|d(n) - \hat{d}(n)|^2] \\ &= \frac{1}{N_2 - N_1} \sum_n \text{MSE}(n). \end{aligned} \quad (44)$$

The filtered signal is  $\hat{d}(n) \triangleq w(x(n))$  and  $e(n) \triangleq d(n) - \hat{d}(n)$  is the error process. Three different problems will be studied in this setting [28, Sec. 1.2]: *smoothing*, *filtering* and *prediction*.

## V. ASYMPTOTIC SMOOTHING PERFORMANCE

When future samples of process  $x(n)$  are available, the previous problem is referred to as *non-causal* or *smoothing* [28, Sec. 9.1]. This case is common in applications without latency restrictions. Since we are interested in the asymptotically achievable performance, the limit

$$\text{MSE}_{\text{nc}} \triangleq \lim_{N \rightarrow \infty} \text{MSE}\left(-\frac{N}{2}, \frac{N}{2} - 1\right) \quad (45)$$

is considered. Using Parseval's theorem, the error power can be expressed as

$$\text{MSE}_{\text{nc}} = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \text{E}[|d\xi_e^{(p)}(\sigma)|^2], \quad (46)$$

where  $d\xi_e^{(p)}(\sigma)$  is the KL transform of the error process  $e(n)$ :

$$d\xi_e^{(p)}(\sigma) \triangleq \text{KL}\{d(n) - \hat{d}(n)\}(p, \sigma). \quad (47)$$

For a more convenient analysis, and without loss of generality<sup>4</sup>, we set the filter in the KL domain to be a linear operator applied onto the observed signal  $x(n)$ :

$$d\xi_{\hat{d}}^{(p)}(\sigma) \triangleq \mathcal{W}_p^*(\sigma) \cdot d\xi_x^{(p)}(\sigma), \quad (48)$$

where  $d\xi_{\hat{d}}^{(p)}(\sigma)$  is the KL representation of the filtered signal  $\hat{d}(n)$ . The previous expression can be seen as a natural extension of the transfer function in the frequency domain, commonly employed to deal with WSS processes through LTI

systems: it relates every input to its corresponding output across the full spectrum.

Using (48), we can express the spectral representation of  $e(n)$  in terms of the ones of  $d(n)$  and  $\hat{d}(n)$ :

$$d\xi_e^{(p)}(\sigma) = d\xi_d^{(p)}(\sigma) - d\xi_{\hat{d}}^{(p)}(\sigma). \quad (49)$$

This decoupled structure is achieved by both signals sharing the same KL basis, due to the noise component being white. We can harness this simplicity to design filter  $\mathcal{W}_p^*(\sigma)$ , such that it minimizes the average error power at every point  $(p, \sigma)$  of the spectrum:

$$\text{MSE}_p(\sigma) \triangleq \text{E}[|d\xi_e^{(p)}(\sigma)|^2]. \quad (50)$$

To obtain it, the derivative of each MSE term must be nullified:

$$\frac{d \text{MSE}_p(\sigma)}{d\mathcal{W}_p^*(\sigma)} = \mathcal{W}_p(\sigma) \cdot \text{E}[|d\xi_x^{(p)}(\sigma)|^2] - \text{E}[d\xi_d^{(p)*}(\sigma)d\xi_x^{(p)}(\sigma)] \equiv 0 \quad (51)$$

$$\mathcal{W}_p(\sigma) = \frac{\text{E}[d\xi_d^{(p)*}(\sigma)d\xi_x^{(p)}(\sigma)]}{\text{E}[|d\xi_x^{(p)}(\sigma)|^2]} \triangleq \frac{\mathcal{S}_{\mathcal{D}\mathcal{X}}^{(p)*}(\sigma)d\sigma}{\mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)d\sigma}. \quad (52)$$

This solution is the KL extension of the classical non-causal Wiener filter in the frequency domain [28, Eq. (8.144)].

### A. Non-causal MMSE and coherence

Plugging (52) back into (46) produces the MMSE<sup>5</sup>:

$$\text{MMSE}_{\text{nc}} = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \mathcal{S}_{\mathcal{D}}^{(p)}(\sigma)(1 - |\gamma_p(\sigma)|^2)d\sigma, \quad (53)$$

where

$$|\gamma_p(\sigma)|^2 \triangleq \frac{|\mathcal{S}_{\mathcal{D}\mathcal{X}}^{(p)}(\sigma)|^2}{\mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)\mathcal{S}_{\mathcal{D}}^{(p)}(\sigma)} \in [0, 1] \quad (54)$$

is the KL version of the *squared spectral coherence* in the frequency domain [24, Eq. (3.6)]. This is a measure of linear dependency between observation  $d\xi_x^{(p)}(\sigma)$  and reference  $d\xi_d^{(p)}(\sigma)$ . Its value is directly related to the MMSE: as clearly stated in (53), the closer  $|\gamma_p(\sigma)|^2$  is to 1 for all  $(p, \sigma)$ , the lower the achievable MMSE will be.

Representing (53) in matrix form will provide valuable insights on the filtering problem. Let

$$\mathcal{S}_{\mathcal{D}\mathcal{X}}(\sigma) \triangleq \text{Diag}(\mathcal{S}_{\mathcal{D}\mathcal{X}}^{(0)}(\sigma), \dots, \mathcal{S}_{\mathcal{D}\mathcal{X}}^{(P-1)}(\sigma)) \quad (55)$$

be the KL correlation matrix between  $d(n)$  and  $x(n)$ . Then, the summation in (53) can be expressed as a trace:

$$\begin{aligned} \text{MMSE}_{\text{nc}} &= \\ &\int_0^{\frac{1}{P}} \text{Tr}\{\mathcal{S}_{\mathcal{D}}(\sigma)(\mathbf{I}_P - (\mathcal{S}_{\mathcal{D}}^{-1}\mathcal{S}_{\mathcal{D}\mathcal{X}}\mathcal{S}_{\mathcal{X}}^{-1}\mathcal{S}_{\mathcal{D}\mathcal{X}}^H)(\sigma))\}d\sigma. \end{aligned} \quad (56)$$

Using the circularity property of the trace operator together with decomposition

$$\mathcal{S}_{\mathcal{D}\mathcal{X}}(\sigma) \triangleq (\mathbf{B}_{\mathcal{D}}\mathcal{S}_{\mathcal{D}\mathcal{X}}\mathbf{B}_{\mathcal{X}}^H)(\sigma), \quad (57)$$

<sup>4</sup>This is ensured by the uncorrelatedness between spectral components provided by the KL representation.

<sup>5</sup>Notice how (53) resembles [33, Eq. (13.115)].

which is a straightforward consequence of Corollary 2, we can express the non-causal MMSE in terms of cyclic PSD matrices<sup>6</sup>:

$$\text{MMSE}_{\text{nc}} = \int_0^{\frac{1}{P}} \text{Tr}\{\mathbf{S}_D(\sigma)(\mathbf{I}_P - \mathbf{C}_{DX}(\sigma)\mathbf{C}_{DX}^H(\sigma))\}d\sigma, \quad (58)$$

where

$$\mathbf{C}_{DX}(\sigma) \triangleq (\mathbf{S}_D^{-1/2}\mathbf{S}_{DX}\mathbf{S}_X^{-1/2})(\sigma), \quad (59)$$

is the *coherence matrix* [24, Def. 3.3]. It generalizes the notion of spectral coherence presented in (54): the more  $\mathbf{C}_{DX}(\sigma)$  resembles a unitary matrix for all  $\sigma$ , the lower the MMSE will be. This mathematical structure is very recurrent in the literature and is strongly related to canonical correlations and principal angles [24, Ch. 3].

On a final note, we can particularize the non-causal MMSE (53) for our additive model, in which the white noise has KL-PSD  $\mathcal{S}_Z^{(p)}(\sigma) = P_z$ , for all  $(p, \sigma)$ . The filter expression (52) simplifies to

$$\mathcal{W}_p(\sigma) = \frac{\mathcal{S}_D^{(p)}(\sigma)}{\mathcal{S}_D^{(p)}(\sigma) + P_z}, \quad (60)$$

with which we obtain

$$\begin{aligned} \text{MMSE}_{\text{nc}} &= \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \frac{\mathcal{S}_D^{(p)}(\sigma)P_z}{\mathcal{S}_D^{(p)}(\sigma) + P_z} d\sigma \\ &= P_z \int_0^{\frac{1}{P}} \text{Tr}\{\mathbf{S}_D(\sigma)(\mathbf{S}_D(\sigma) + P_z\mathbf{I}_P)^{-1}\}d\sigma. \end{aligned} \quad (61)$$

### B. Equivalence between KL Wiener filter and CWF

It is known that the CWF, which is the optimum filter for CS signals, is periodically time variant [8]. It is usually implemented through a FRESH structure, which has the following formulation in the frequency domain [9, Sec. 10.3]:

$$\begin{aligned} d\nu_d^{(p)}(\sigma) &\triangleq \sum_{q=0}^{P-1} \mathbf{W}_q^*(\sigma + \frac{p}{P})d\nu_x^{(q)}(\sigma + \frac{p}{P}) \\ &\triangleq \check{\mathbf{w}}_p^H(\sigma)\mathbf{P}_\Pi^p\check{\mathbf{x}}(\sigma), \end{aligned} \quad (62)$$

where  $\mathbf{P}_\Pi$  is a *cyclic permutation matrix* [34, Eq. (4.1.1)]:

$$\mathbf{P}_\Pi \triangleq \begin{bmatrix} \mathbf{0}_{(P-1)} & \mathbf{I}_{(P-1)} \\ 1 & \mathbf{0}_{(P-1)}^T \end{bmatrix}. \quad (63)$$

Applying Parseval's theorem onto (45), we may express the (non-causal) error power in the CL domain:

$$\text{MSE}_{\text{nc}} = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \text{E}[|d\nu_d^{(p)}(\sigma) - \check{\mathbf{w}}_p^H(\sigma)\mathbf{P}_\Pi^p\check{\mathbf{x}}(\sigma)|^2]d\sigma. \quad (64)$$

To obtain the optimal FRESH filter, we must nullify the derivative of the previous expression with respect to  $\check{\mathbf{w}}_p^*(\sigma)$ :

$$\begin{aligned} \nabla_{\check{\mathbf{w}}_p^*(\sigma)} \text{MSE}_{\text{nc}}^{(p)}(\sigma) &= \mathbf{P}_\Pi^p(\mathbf{S}_X(\sigma)\mathbf{P}_\Pi^{-p}\check{\mathbf{w}}_p(\sigma) - \mathbf{s}_{XD}^{(p)}(\sigma)) \equiv 0 \\ \check{\mathbf{w}}_p(\sigma) &= \mathbf{P}_\Pi^p\mathbf{S}_X^{-1}(\sigma)\mathbf{s}_{XD}^{(p)}(\sigma), \end{aligned} \quad (65)$$

<sup>6</sup>The matrix inside the trace in (58) is the *error covariance matrix* in [24, Sec. 3.6].

where

$$\mathbf{s}_{XD}^{(p)}(\sigma)d\sigma \triangleq \text{E}[\check{\mathbf{x}}(\sigma)d\nu_d^{(p)*}(\sigma)] \quad (66)$$

is the  $p$ th column of  $\mathbf{S}_{DX}^H(\sigma)d\sigma$ . The following result relates the KL Wiener filter derived in (52) with the CWF obtained herein.

*Theorem 2:* (CWF in the KL domain) The CWF is equivalent to the KL Wiener filter.

*Proof:* The proof is based on comparing the output signals obtained from the cyclic and KL Wiener filters. Regarding the former, we simply plug (65) into (62):

$$d\nu_d^{(p)}(\sigma) = (\mathbf{s}_{XD}^{(p)H}\mathbf{S}_X^{-1}\check{\mathbf{x}})(\sigma). \quad (67)$$

As for the latter, we apply (52) to (48):

$$d\xi_d^{(p)}(\sigma) = \frac{\mathcal{S}_{DX}^{(p)}(\sigma)}{\mathcal{S}_X^{(p)}(\sigma)}d\xi_x^{(p)}(\sigma). \quad (68)$$

To compare them, we must express both filtered signals in the same domain. We achieve this by stacking  $P$  samples of  $d\nu_d^{(p)}(\sigma)$  as

$$\begin{aligned} \check{\mathbf{d}}(\sigma) &\triangleq [d\nu_d^{(0)}(\sigma), \dots, d\nu_d^{(P-1)}(\sigma)]^T \\ &= (\mathbf{S}_{DX}\mathbf{S}_X^{-1}\check{\mathbf{x}})(\sigma), \end{aligned} \quad (69)$$

and using Corollaries 1 and 2:

$$\begin{aligned} \mathbf{B}_D^H(\sigma)\check{\mathbf{d}}(\sigma) &= (\mathbf{B}_D^H\mathbf{S}_{DX}\mathbf{B}_X\mathbf{S}_X^{-1}\mathbf{B}_X^H\check{\mathbf{x}})(\sigma) \\ &= (\mathbf{S}_{DX}\mathbf{S}_X^{-1}\check{\mathbf{x}})(\sigma). \end{aligned} \quad (70)$$

Since the  $p$ th term of the transformed vector corresponds to (68), we can assert that both expansions correspond to the same process, making the cyclic and linear KL Wiener filters equivalent. This completes the proof. ■

This result bears remarkable implications in the context of this paper. By studying a linear filtering problem in the KL domain, we have accessed asymptotic performance results of periodically time variant filtering (namely (53)) without having to explicitly derive its expressions nor deal with the FRESH structure.

### C. Second order characterization of the error process

The error process resulting after filtering (42) with the CWF has the following spectral representation:

$$\begin{aligned} \check{\mathbf{e}}(\sigma) &\triangleq [d\nu_e^{(0)}(\sigma), \dots, d\nu_e^{(P-1)}(\sigma)]^T \\ &= \check{\mathbf{d}}(\sigma) - \mathbf{S}_D(\sigma)(\mathbf{S}_D(\sigma) + P_z\mathbf{I}_P)^{-1}(\check{\mathbf{d}}(\sigma) + \check{\mathbf{z}}(\sigma)) \\ &= (\mathbf{S}_D(\sigma) + P_z\mathbf{I}_P)^{-1}(P_z\check{\mathbf{d}}(\sigma) - \mathbf{S}_D(\sigma)\check{\mathbf{z}}(\sigma)). \end{aligned} \quad (71)$$

Notice we have used the same notation as in Section III. Its cyclic spectrum matrix is

$$\mathbf{S}_E(\sigma)d\sigma = \text{E}[\check{\mathbf{e}}(\sigma)\check{\mathbf{e}}^H(\sigma)] = (P_z^{-1}\mathbf{I}_P + \mathbf{S}_D^{-1}(\sigma))^{-1}d\sigma. \quad (72)$$

Since this matrix is not diagonal for  $P_z > 0$ , it implies the error process of the CWF is CS.

### D. Synchronous gain

To complement this analysis, it is of interest to define the *synchronous gain*: the improvement in MSE that can be harnessed by taking cyclostationarity into account. Returning to the additive model (42) with CS reference, recall the MMSE expression in the Fourier domain (64). If  $d(n)$  is erroneously treated as WSS, the classical non-causal Wiener filter [28, Eq. (8.149)] can be employed, which yields

$$\text{MMSE}_{\text{nc,WSS}} = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \frac{S_D(\sigma + \frac{p}{P}) P_z}{S_D(\sigma + \frac{p}{P}) + P_z} d\sigma, \quad (73)$$

as stated in [19, Eq. (106)]. Notice how (73), which is a function of the diagonal elements of  $\mathbf{S}_D(\sigma)$ , is related to (61), which depends on its eigenvalues instead. With the two expressions, the synchronous gain is defined as

$$\zeta_{\text{nc}} \triangleq \frac{\text{MMSE}_{\text{nc}}}{\text{MMSE}_{\text{nc,WSS}}}. \quad (74)$$

One might expect this value to be less than 1, implying a reduction in MSE achieved by a more nuanced processing. This intuition can be formally proved using the theory of majorization [31], as in Proposition 1. We know that the eigenvalues of a Hermitian matrix majorize its diagonal elements [31, Th. 9.B.1.], *i.e.*  $\text{diag}(\mathbf{S}_D(\sigma)) \prec \text{diag}(\mathbf{S}_D(\sigma))$ . Therefore, by [31, Prop. 4.B.1.], we can assert that  $\text{MMSE}_{\text{nc}} \leq \text{MMSE}_{\text{nc,WSS}}$ . This ensures that there is a MSE improvement in performing a synchronous processing, compared to the conventional Wiener filter treatment.

## VI. ASYMPTOTIC FILTERING PERFORMANCE

Non-causal filters are often referred to as *unrealizable* [35, Sec. 4.1] because it may be unrealistic to assume the availability of future samples from a process. For this reason, the results obtained in the previous section are to be understood as bounds on how accurate synchronous filtering can get, rather than an achievable performance in practice.

We define the signal-to-noise ratio (SNR) as the ratio between the average power of  $d(n)$  and  $z(n)$ . If we assume  $d(n)$  has unit power, it reduces to  $\text{SNR} \triangleq 1/P_z$ . We may express the MMSE (61) obtained from the non-causal CWF in terms of this SNR:

$$\text{MMSE}_{\text{nc}}(\text{SNR}) = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \frac{S_D^{(p)}(\sigma)}{\text{SNR} \cdot S_D^{(p)}(\sigma) + 1} d\sigma. \quad (75)$$

Let  $\{x(n)\}_{n_a}^{n_b}$  be a sequence of consecutive samples from  $x(n)$ , for  $n_a \leq n \leq n_b$ . The infinite past *causal MMSE* at any  $n_b$  given  $\{x(n)\}_{n_a}^{n_b}$  is known to be [36, Sec. 4.2.1]

$$\text{mmse}_c(n_b, \text{SNR}) \triangleq \text{E}[\lvert d(n_b) - \text{E}[d(n_b) \mid \{x(n)\}_{-\infty}^{n_b}; \text{SNR}] \rvert^2], \quad (76)$$

and its time average is

$$\text{MMSE}_c(\text{SNR}) \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=n_b-N+1}^{n_b} \text{mmse}_c(n, \text{SNR}). \quad (77)$$

Its derivation can be circumvented by employing the following fundamental result in nonlinear filtering.

*Theorem 3:* (Guo-Shamai-Verdú [19, Th. 8]) The average (per unit time) MMSE for filtering can be obtained by averaging the smoothing MMSE over all possible values of SNR:

$$\text{MMSE}_c(\text{SNR}) = \text{E}[\text{MMSE}_{\text{nc}}(\Gamma)], \quad (78)$$

where  $\Gamma$  is distributed uniformly in the interval  $[0, \text{SNR}]$ .

In Section V-C, we have proved the error process after smoothing signal (42) with the CWF is CS. Hence, (75) is valid for the (asymptotic) non-causal MMSE averaged over a single period  $P$ . Since Theorem 3 links it with its causal counterpart, this implies  $\text{MMSE}_c(\text{SNR})$  averaged over a single period is equivalent to (77), which is averaged over the full infinite past.

To obtain (78) from (75), we perform the following expectation:

$$\begin{aligned} \text{MMSE}_c(\text{SNR}) &= \frac{1}{\text{SNR}} \int_0^{\text{SNR}} \text{MMSE}_{\text{nc}}(\Gamma) d\Gamma \\ &= \frac{1}{\text{SNR}} \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \int_0^{\text{SNR}} \frac{S_D^{(p)}(\sigma)}{\Gamma \cdot S_D^{(p)}(\sigma) + 1} d\Gamma d\sigma \\ &= \frac{1}{\text{SNR}} \int_0^{\frac{1}{P}} \sum_{p=0}^{P-1} \ln(\text{SNR} \cdot S_D^{(p)}(\sigma) + 1) d\sigma. \end{aligned} \quad (79)$$

Transforming the sum of logarithms into a log-determinant, and making use of Corollary 2, we can express this result in terms of the cyclic PSD and the coherence matrix (59):

$$\begin{aligned} \text{MMSE}_c(\text{SNR}) &= \frac{1}{\text{SNR}} \int_0^{\frac{1}{P}} \ln|\text{SNR} \cdot \mathbf{S}_D(\sigma) + \mathbf{I}_P| d\sigma \\ &= \frac{1}{\text{SNR}} \int_0^{\frac{1}{P}} \ln|\text{SNR} \cdot \mathbf{S}_D(\sigma) + \mathbf{I}_P| d\sigma \\ &= \frac{-1}{\text{SNR}} \int_0^{\frac{1}{P}} \ln|\mathbf{I}_P - \mathbf{C}_{DX}(\sigma) \mathbf{C}_{DX}^H(\sigma)| d\sigma. \end{aligned} \quad (80)$$

This formula is related to various problems that require the detection of correlation and cyclostationarity [24, Ch. 8]. It also provides a link between MMSE, the coherence matrix and mutual information, since it involves a generalization of the *information-theoretic coherence* from [24, Sec. 11.4].

### A. Synchronous gain

The causal MMSE assuming  $d(n)$  is WSS is known to be [28, Eq. (8.177)]

$$\text{MMSE}_{c,\text{WSS}}(\text{SNR}) = \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \frac{\ln(\text{SNR} \cdot S_D(\sigma + \frac{p}{P}) + 1)}{\text{SNR}} d\sigma. \quad (81)$$

Using the same rationale based on majorization theory as in Section V-D, we once again state that the MSE can be reduced by synchronous processing, *i.e.*

$$\zeta_c \triangleq \frac{\text{MMSE}_c}{\text{MMSE}_{c,\text{WSS}}} \leq 1. \quad (82)$$

### B. High SNR regime

Suppose that  $\mathcal{S}_{\mathcal{D}}(\lambda)$  in (79) is concentrated inside a band  $[0, B)$ , for  $0 < B < 1$ , such that it is null outside of it:

$$\text{MMSE}_c(\text{SNR}) = \int_0^B \frac{\ln(\mathcal{S}_{\mathcal{D}}(\lambda) + \frac{1}{\text{SNR}})}{\text{SNR}} d\lambda + B \cdot \frac{\ln \text{SNR}}{\text{SNR}}. \quad (83)$$

In the limit of  $\text{SNR} \rightarrow \infty$ , the causal MMSE decays as<sup>7</sup>

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \text{MMSE}_c(\text{SNR}) &= \lim_{\text{SNR} \rightarrow \infty} \int_0^B \frac{\ln \mathcal{S}_{\mathcal{D}}(\lambda)}{\text{SNR}} d\lambda \\ &+ \lim_{\text{SNR} \rightarrow \infty} B \cdot \frac{\ln \text{SNR}}{\text{SNR}} = \lim_{\text{SNR} \rightarrow \infty} B \cdot \frac{\ln \text{SNR}}{\text{SNR}}. \end{aligned} \quad (84)$$

This implies that, the more concentrated the KL spectrum of  $d(n)$  is (*i.e.* the smaller  $B$  is), the lower  $\text{MMSE}_c$  will be at high SNR. The same idea applies to its non-causal counterpart; it can be clearly seen by observing the limit of (75) at high SNR under mild conditions:

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \text{MMSE}_{nc}(\text{SNR}) &= \\ \lim_{\text{SNR} \rightarrow \infty} \int_0^B \frac{\mathcal{S}_{\mathcal{D}}(\lambda)}{\text{SNR} \cdot \mathcal{S}_{\mathcal{D}}(\lambda) + 1} d\lambda &= \lim_{\text{SNR} \rightarrow \infty} \frac{B}{\text{SNR}}. \end{aligned} \quad (85)$$

From (84) and (85) we conclude that the causal MMSE asymptotically approaches the rate of decay of the non-causal MMSE, which is hyperbolic.

## VII. ASYMPTOTIC PREDICTION PERFORMANCE

The final problem that will be analyzed is the one-step linear prediction of  $x(n)$  based on  $N$  past samples:

$$\hat{x}_N(n) = \mathbf{w}_N^H(n) \mathbf{x}_N(n-1), \quad (86)$$

where

$$\mathbf{x}_N(n-1) \triangleq [x(n-N), \dots, x(n-2), x(n-1)]^T. \quad (87)$$

The MMSE in this case is given by

$$\text{MMSE}_p^{(N)}(n) \triangleq \min_{\mathbf{w}_N(n)} \text{E}[|x(n) - \hat{x}_N(n)|^2]. \quad (88)$$

This error is achieved by the cyclic Wiener predictor, which is periodic and implements a synchronous processing. It is known to be [16, Eq. (2.10)]

$$\begin{aligned} \text{MMSE}_p^{(N)}(n) &= \mathbf{R}_x(n, 0) - \mathbf{r}_x^{(N)H}(n) (\mathbf{R}_x^{(N)}(n-1))^{-1} \mathbf{r}_x^{(N)}(n), \end{aligned} \quad (89)$$

where the autocorrelation matrix and vector are respectively defined as

$$\begin{aligned} \mathbf{R}_x^{(N)}(n-1) &\triangleq \text{E}[\mathbf{x}_N(n-1) \mathbf{x}_N^H(n-1)] \\ \mathbf{r}_x^{(N)}(n) &\triangleq \text{E}[\mathbf{x}_N(n-1) x^*(n)]. \end{aligned} \quad (90)$$

Notice that the correlation matrix of size  $N+1$  at instant  $n$  can be expressed blockwise as

$$\mathbf{R}_x^{(N+1)}(n) = \begin{bmatrix} \mathbf{R}_x^{(N)}(n-1) & \mathbf{r}_x^{(N)}(n) \\ \mathbf{r}_x^{(N)H}(n) & \mathbf{R}_x(n, 0) \end{bmatrix}. \quad (91)$$

<sup>7</sup>We have assumed that  $\ln \mathcal{S}_{\mathcal{D}}(\lambda)$  is Riemann-integrable in  $[0, B)$  [37, Def. 9.5.1], which is closely related to the *regularity condition* in [38, Th. 4.3].

Using the determinant formula for block matrices, we can express the MMSE as follows [16, Eq. (2.41)]:

$$\text{MMSE}_p^{(N)}(n) = \frac{|\mathbf{R}_x^{(N+1)}(n)|}{|\mathbf{R}_x^{(N)}(n-1)|}. \quad (92)$$

We are interested in the behavior of this error as past samples grow without bound, *i.e.*

$$\text{MMSE}_p \triangleq \lim_{N \rightarrow \infty} \text{MMSE}_p^{(N)}(n). \quad (93)$$

*Theorem 4:* (Riba & Vilà-Insa [22, Th. 2]) A lower bound on the MMSE of one-step prediction of a cyclostationary signal is given by

$$\text{MMSE}_p = \exp \int_0^{\frac{1}{P}} \ln |\mathbf{S}_X(\sigma)| d\sigma. \quad (94)$$

*Proof:* See Appendix D. ■

A metric that is closely related to the MMSE in prediction is the *spectral flatness* of WSS processes [16, Def. 6.1], which we extend to CS ones by defining the *KL spectral flatness*:

$$\omega_x^2 \triangleq \frac{\text{MMSE}_p}{P_x} \in [0, 1]. \quad (95)$$

This ratio measures the shape (*i.e.* flatness) of the KL spectrum of  $x(n)$  and is directly linked to its representation entropy and predictability through synchronous processing [16, Sec. 6.6.1].

### A. Synchronous gain

The minimum MMSE achievable from classical Wiener prediction is given by the *Kolmogorov-Szegö theorem* [16, Eq. (6.23)]:

$$\text{MMSE}_{p, \text{WSS}} = \exp \int_0^{\frac{1}{P}} \ln |\bar{\mathbf{S}}_X(\sigma) \bar{\mathbf{S}}_X^H(\sigma)| d\sigma. \quad (96)$$

where  $\bar{\mathbf{S}}_X(\sigma) \triangleq (\mathbf{I}_P \circ \mathbf{S}_X(\sigma))^{1/2}$ . In this case, the synchronous gain

$$\zeta_p \triangleq \frac{\text{MMSE}_p}{\text{MMSE}_{p, \text{WSS}}} \leq 1 \quad (97)$$

can be expressed compactly as a function of the *spectral coherence matrix* [24, Sec. 8.2]:

$$\bar{\mathbf{C}}_X(\sigma) \triangleq (\bar{\mathbf{S}}_X^{-1} \mathbf{S}_X \bar{\mathbf{S}}_X^{-H})(\sigma). \quad (98)$$

This fact was stated in [22]:

$$\begin{aligned} \zeta_p &= \exp \int_0^{\frac{1}{P}} \ln |\mathbf{S}_X(\sigma)| - \ln |\bar{\mathbf{S}}_X(\sigma) \bar{\mathbf{S}}_X^H(\sigma)| d\sigma \\ &= \exp \int_0^{\frac{1}{P}} \ln |\bar{\mathbf{C}}_X(\sigma)| d\sigma. \end{aligned} \quad (99)$$

### B. High SNR regime

If  $x(n)$  is obtained from the additive model (42), we may relate the prediction and filtering MMSE as follows:

$$\begin{aligned} \text{MMSE}_p(\text{SNR}) &= \exp \int_0^{\frac{1}{P}} \ln |\mathbf{S}_{DD}(\sigma) + \frac{1}{\text{SNR}} \mathbf{I}_P| d\sigma \\ &= \frac{\exp \int_0^{\frac{1}{P}} \ln |\text{SNR} \cdot \mathbf{S}_{DD}(\sigma) + \mathbf{I}_P| d\sigma}{\text{SNR}} = \frac{e^{\text{MMSE}_c \cdot \text{SNR}}}{\text{SNR}}. \end{aligned} \quad (100)$$

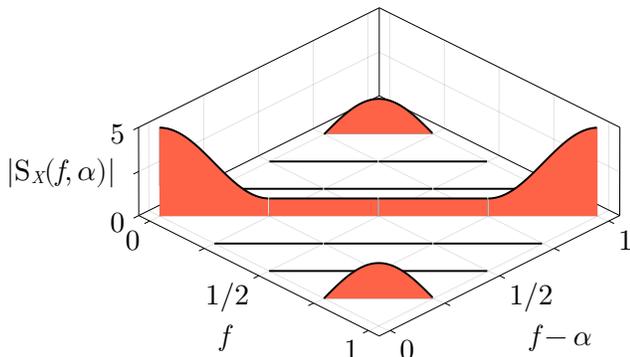


Figure 2. Spectral correlation of  $x(n)$  from model (104), for  $P = 4$ ,  $\Delta = 0$  and  $P_z = 1$ .

With it, we can easily replicate the high SNR analysis from Section VI-B for prediction MMSE:

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \text{MMSE}_p(\text{SNR}) &= \lim_{\text{SNR} \rightarrow \infty} \frac{\exp(B \cdot \frac{\ln \text{SNR}}{\text{SNR}} \cdot \text{SNR})}{\text{SNR}} \\ &= \lim_{\text{SNR} \rightarrow \infty} \frac{\text{SNR}^B}{\text{SNR}} = \lim_{\text{SNR} \rightarrow \infty} \frac{1}{\text{SNR}^{(1-B)}}. \end{aligned} \quad (101)$$

Some remarks about this result should be made. The decay of (101) is slower than hyperbolic and is controlled by the occupied spectral band  $B$ . When  $B \rightarrow 0$ , *i.e.* very low  $\omega_x^2$ , it approaches the rate of decay of  $\text{MMSE}_{\text{nc}}$  at high SNR, as seen in (85). When  $B = 1$ , *i.e.*  $d(n)$  occupies the full KL spectrum (high flatness), the desired signal becomes unpredictable and  $\text{MMSE}_p$  does not vanish at  $\text{SNR} \rightarrow \infty$ .

### VIII. NUMERICAL ILLUSTRATION

In this final section, we are going to apply the previous results to a linearly modulated digital communication system, in order to validate them numerically. Consider a pulse-amplitude modulation signal of cycle period  $P$ ,  $d(n)$ , affected by an independent additive noise component  $z(n)$ , which is WSS and white and has autocorrelation  $R_z(m) = P_z \delta_m$ . The transmitted complex baseband signal is constructed as follows:

$$d(n) = \sum_{k \in \mathbb{Z}} a(k)b(n - kP). \quad (102)$$

The symbols  $a(k)$  are uncorrelated and have power  $P$  (*i.e.*  $R_a(m) = P \delta_m$ ). The shaping pulse  $b(n)$  is a 100% excess bandwidth square-root raised cosine [39, Ex. 5-22], with DTFT

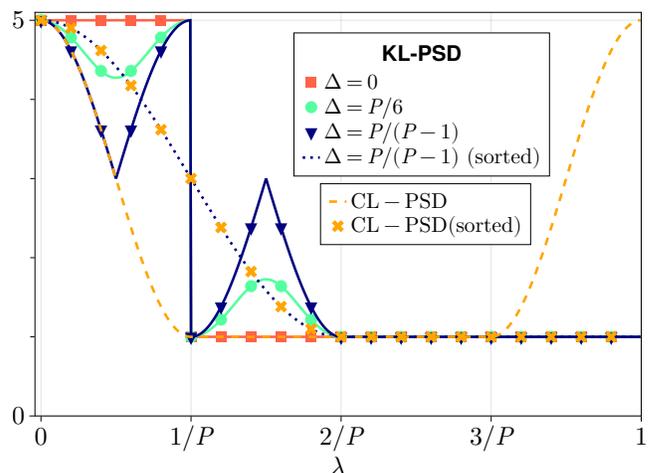
$$B(f) = \sqrt{P} \cos\left(\frac{\pi}{2} P f\right), \quad f \in \left[-\frac{1}{P}, \frac{1}{P}\right]. \quad (103)$$

The signal model of interest is as follows:

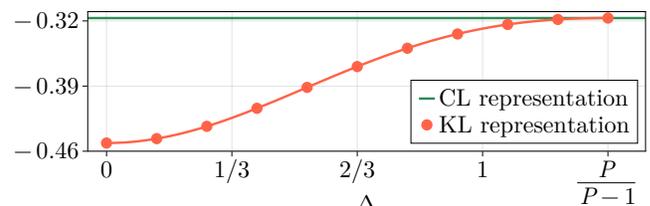
$$x(n) = d(n - \varepsilon) + z(n), \quad (104)$$

where  $\varepsilon \sim \mathcal{U}[0, \Delta)$  is a random delay. The spectral correlation of  $x(n)$  is derived in Appendix E and represented graphically in Fig. 2. It is only defined on  $\delta$ -ridges parallel to the stationary manifold at  $\alpha = k/P$  for  $k = 0, \dots, P-1$  [9, Sec. 10.1.1]. From (122), its cyclic spectrum is

$$S_X^{(\frac{k}{P})}(\sigma) = P_z \delta_k + B(\sigma)B^*\left(\sigma - \frac{k}{P}\right) \frac{1 - e^{-j2\pi\Delta\frac{k}{P}}}{j2\pi\Delta\frac{k}{P}}, \quad (105)$$



(a) PSD and KL-PSD of  $x(n)$  for various values of  $\Delta$ .



(b) Representation entropy of  $x(n)$  in bits.

Figure 3. Spectral density and representation entropy of  $x(n)$  for  $P = 4$  and  $P_z = 1$ .

with which we construct the cyclic PSD matrix  $\mathbf{S}_X(\sigma)$  for all  $\sigma \in [0, 1/P)$ , as explained in Section III (see Fig. 1). By obtaining its eigenvalues  $\mathcal{S}_X(\sigma)$  for each  $\sigma$ , the KL-PSD can be derived, since  $[\mathcal{S}_X(\sigma)]_{p,p} = \mathcal{S}_X(\sigma + p/P)$ .

The maximum delay  $\Delta$  allows to control the statistical properties of  $x(n)$ . Notice that its impact cannot be observed on the PSD; indeed, if we take (105) and set  $k = 0$ , we obtain  $S_X(\sigma) = P_z + |B(\sigma)|^2$ , which does not depend on  $\Delta$ . On the contrary, the KL-PSD does behave differently for different values of  $\Delta$ , as shown in Fig. 3a. For  $\Delta = 0$ , (*i.e.*  $d(n)$  is fully CS), the signal component of  $\mathcal{S}_X(\lambda)$  is contained in a spectral band of width  $1/P$ , while the rest of the spectrum is only occupied by noise. As  $\Delta$  increases (*e.g.*  $\Delta = P/6$ ), some amount of the signal KL-PSD leaks to the band  $[1/P, 2/P)$ . When  $\Delta = P/(P-1)$ , the leakage is maximum and  $d(n)$  becomes WSS, in which case the KL-PSD coincides with the CL-PSD. To see it, we sort them both in decreasing order so we can compare the two spectra regardless of the domain they belong to.

A different way to analyze the effect of  $\Delta$  onto the statistical behavior of  $x(n)$  is through the representation entropy defined in Section III-C (see Fig. 3b). As expected, the lowest entropy is achieved for  $\Delta = 0$ , since the KL spectrum of  $x(n)$  is the most compact (*i.e.* CS). Its value increases with  $\Delta$  until it reaches a maximum for  $\Delta = P/(P-1)$ , which coincides with the entropy of the CL representation of  $x(n)$ , since it is WSS.

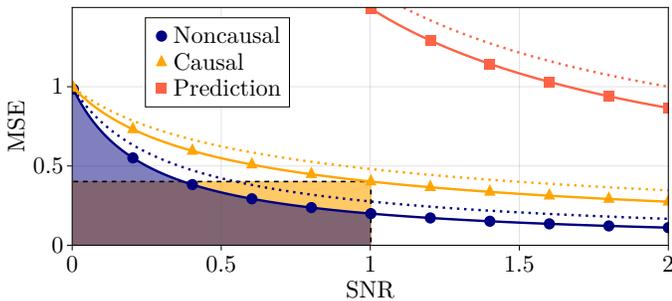


Figure 4. MMSE in synchronous (solid lines) and asynchronous (dotted lines) processing.  $\Delta = 0$  and  $P = 4$ .

### A. Signal processing performance

We are now interested in numerically assessing the performance of smoothing, filtering and prediction, given model (104). Fig. 4 displays the MMSE of both synchronous and asynchronous processing in the three modalities for different SNR values, making very clear the gains achievable by exploiting cyclostationarity. The relationship between non-causal and causal filtering from [19, Th. 8], explored in Section VI, is illustrated as well. The area under  $\text{MMSE}_{\text{nc}}$  in the interval  $\text{SNR} \in [0, 1]$  is the same as the one in the colored box, which coincides with  $\text{MMSE}_{\text{c}}$  at  $\text{SNR} = 1$ , as expected from (79).

In Fig. 5, there is a plot of MMSE times SNR against SNR for synchronous processing in the three modalities. Regarding smoothing, its corresponding curve is a horizontal straight line, since its MMSE decays hyperbolically. We can observe the high SNR approximation derived in Section VI-B is remarkably accurate. Its ordinate is dictated by the bandwidth parameter in (85) which, in the model considered, depends on the signal period (*i.e.*  $B = 1/P$ ). Relating to causal filtering, its curve increases at a slower rate the higher the SNR is. This implies its MMSE presents asymptotically hyperbolic decay, as predicted in (84). Its approximation improves as the SNR grows. Finally, the prediction curves increase linearly with SNR, with a slope equal to  $1/P$ : this is the performance loss with respect to hyperbolic decay, as expected from (101). Notice that the approximation curve displays the same slope as the real one.

To assess the gain achievable by performing a synchronous treatment of the signal at high SNR we refer to Fig. 6. The horizontal axis is a scan across all the possible values of  $\Delta \in [0, P/(P-1)]$ ; it represents the synchronization reliability at the receiver (*i.e.* 0 is perfect synchronization while 1 is the maximum uncertainty). In both smoothing and filtering, the achievable gain is not affected by the signal period  $P$  when perfect synchronization is available. The gain degradation as the uncertainty increases, however, is less sensitive to  $\Delta$  in filtering. The prediction case bears some comment. Under perfect synchronization, the achievable gains are the highest among the three processing modes and are remarkably affected by  $P$ : the higher the period, the lower the gain. This phenomenon can be explained in terms of spectral flatness. For  $P = 2$ , the KL-PSD of  $d(n)$  is concentrated in a band of width  $1/2$ . On the contrary, its CL-PSD occupies the full spectrum, *i.e.* it is significantly flatter. Therefore, implementing a perfectly synchronized processing yields a notable performance gain. As

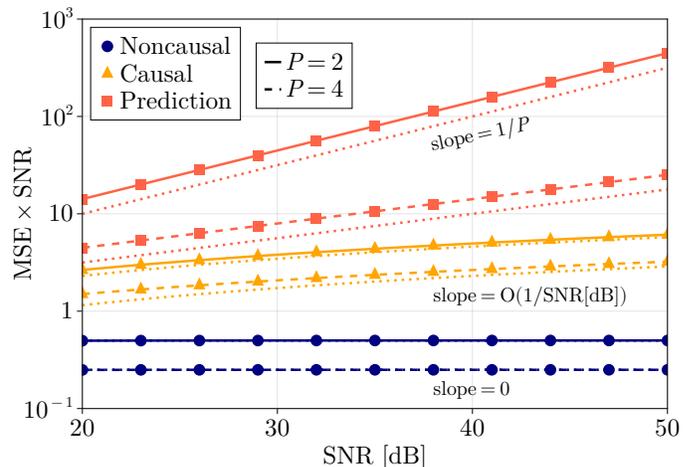


Figure 5.  $\text{MMSE} \times \text{SNR}$  vs. SNR for synchronous processing. The dotted lines are the corresponding approximations in the high SNR regime.

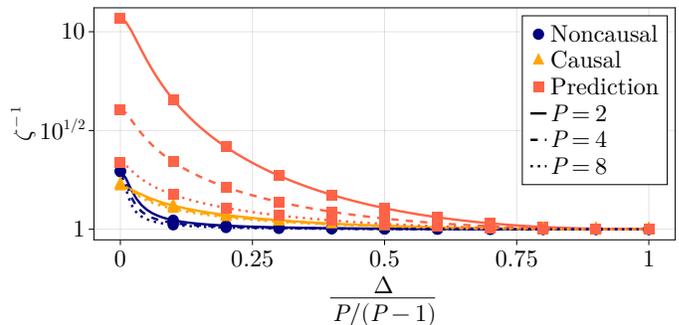


Figure 6. Processing gain in terms of synchronization reliability at  $\text{SNR} = 30\text{dB}$ .

$P$  increases, however, the CL-PSD concentrates in a smaller band and the synchronous gain shrinks.

## IX. CONCLUSIONS

This work has explored the problem of synchronous processing of CS signals. By using their KL representation, we have leveraged its desirable properties to conduct various theoretical analyses. In particular, its uncorrelated eigendecomposition and time-shift invariant spectral density have proven to be very valuable in the extraction of asymptotic performance bounds for filtering applications.

This spectral treatment has enabled to address smoothing, filtering and prediction under the same unified framework and study cyclic Wiener processing without having to delve into specific architectures. The unitary transformation that connects the CL and KL expansions has played a central role in obtaining compact MMSE expressions, as integrals of traces and determinants of the cyclic spectrum matrix. With them, we have also quantified the achievable synchronous gain, which is of remarkable interest in prediction applications. Indeed, the optimal energy compaction of the KL spectrum is intimately linked to improved predictability.

The toolset developed in the present work can be applied in various new directions. Interesting extensions can be found in *almost CS* and more general classes of stochastic processes [40],

[41], which relax different aspects of the periodic structure assumed herein, such as the integer period. From a theoretical point of view, the presence of coherence statistics in MMSE expressions should be investigated further. This would provide deep insights in the relation between synchronous signal processing and other problems in information theory.

## APPENDIX

### A. Proof of Theorem 1

The following proof is an alternative version of the one presented in [22]. It is based on checking that (12) holds for CS processes with the proposed basis:

$$\mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)\phi_{\text{CS}}^{(p)}(k, \sigma)d\sigma = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=-\frac{N}{2}}^{\frac{N}{2}-1} R_x(l, k-l)\phi_{\text{CS}}^{(p)}(l, \sigma). \quad (106)$$

Applying definition (1) and the CL expansion of  $x(k)$ , we have

$$\begin{aligned} \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)\phi_{\text{CS}}^{(p)}(k, \sigma)d\sigma & \quad (107) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_l \mathbb{E}[x(k)x^*(l)] \sum_{c=0}^{P-1} b_X^{(p)}(c, \sigma) e^{j2\pi l(\sigma + \frac{c}{P})} \\ &= \sum_c b_X^{(p)}(c, \sigma) \mathbb{E} \left[ \left( \sum_{r=0}^{P-1} \int_0^{\frac{1}{P}} e^{j2\pi(\sigma' + \frac{r}{P})k} d\nu_x^{(r)}(\sigma') \right) \right. \\ & \quad \times \left. \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_l x(l) e^{-j2\pi l(\sigma + \frac{c}{P})} \right)^* \right] \\ &= \sum_{r,c} b_X^{(p)}(c, \sigma) \int_0^{\frac{1}{P}} e^{j2\pi k(\sigma' + \frac{r}{P})} \mathbb{E}[d\nu_x^{(r)}(\sigma') d\nu_x^{(c)*}(\sigma)], \end{aligned}$$

where  $d\nu_x^{(p)}(\sigma)$  is defined in (20). Using the spectral correlation of a CS process (9) yields:

$$\begin{aligned} \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)\phi_{\text{CS}}^{(p)}(k, \sigma)d\sigma &= \sum_{r,c} b_X^{(p)}(c, \sigma)d\sigma \\ & \quad \times \int_0^{\frac{1}{P}} e^{j2\pi k(\sigma' + \frac{r}{P})} S_X^{(\frac{r-c}{P})}(\sigma' + \frac{r}{P}) \delta(\sigma' - \sigma) d\sigma' \quad (108) \\ &= \sum_{r,c} e^{j2\pi k(\sigma + \frac{r}{P})} b_X^{(p)}(c, \sigma) S_X^{(\frac{r-c}{P})}(\sigma + \frac{r}{P}) d\sigma. \end{aligned}$$

Afterwards, we substitute the proposed basis on the LHS of the equality and cancel the common exponential terms that depend on  $\sigma$ :

$$\begin{aligned} \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) \sum_r b_X^{(p)}(r, \sigma) e^{j2\pi k \frac{r}{P}} \\ &= \sum_{r,c} e^{j2\pi \frac{r}{P} k} b_X^{(p)}(c, \sigma) S_X^{(\frac{r-c}{P})}(\sigma + \frac{r}{P}). \quad (109) \end{aligned}$$

Finally, each term of the sum with respect to  $r$  is equated and stacked in vector form:

$$\begin{aligned} \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) \mathbf{b}_X^{(p)}(r, \sigma) &= \sum_c b_X^{(p)}(c, \sigma) S_X^{(\frac{r-c}{P})}(\sigma + \frac{r}{P}) \\ \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) \mathbf{b}_X^{(p)}(\sigma) &= \mathbf{S}_X(\sigma) \mathbf{b}_X^{(p)}(\sigma). \quad (110) \end{aligned}$$

In order to fulfill this eigenequation,  $\mathbf{b}_X^{(p)}(\sigma)$  must be an eigenvector of  $\mathbf{S}_X(\sigma)$  and  $\mathcal{S}_{\mathcal{X}}^{(p)}(\sigma)$  its corresponding eigenvalue. This completes the proof. ■

### B. Proof of Corollary 1

The proof is obtained by particularizing (14) for the CS basis (16) and applying the definition of the CL transform (5):

$$\begin{aligned} d\xi_x^{(p)}(\sigma) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x(n) \sum_{q=0}^{P-1} b_X^{(p)*}(q, \sigma) e^{-j2\pi n(\sigma + \frac{q}{P})} \\ &= \sum_q b_X^{(p)*}(q, \sigma) \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_n x(n) e^{-j2\pi n(\sigma + \frac{q}{P})} \right) \\ &= \sum_q b_X^{(p)*}(q, \sigma) d\nu_x^{(q)}(\sigma) = \mathbf{b}_X^{(p)\text{H}}(\sigma) \check{\mathbf{x}}(\sigma). \quad (111) \end{aligned}$$

■

### C. Time-shift effect on the KL transform of CS processes

Changing the reference time of a CS process adds a phase shift to its KL representation:

$$\begin{aligned} d\xi_{x, \text{CS}}^{(p)}(\sigma, n_0) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x(n+n_0) \sum_{q=0}^{P-1} b_X^{(p)*}(q, \sigma) \\ & \quad \times e^{-j2\pi n_0 \frac{q}{P}} e^{-j2\pi n(\sigma + \frac{q}{P})} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n' \in \mathbb{Z}} x(n') \sum_q b_X^{(p)*}(q, \sigma) e^{-j2\pi n'(\sigma + \frac{q}{P})} e^{j2\pi n_0 \sigma} \\ &= d\xi_{x, \text{CS}}^{(p)}(\sigma) e^{j2\pi n_0 \sigma}. \quad (112) \end{aligned}$$

### D. Proof of Theorem 4

This proof is based on the one provided in [22], adapted to the present notation and setting in this work. We define value

$$J(n, N) \triangleq \frac{1}{N} \ln |\mathbf{R}_x^{(N)}(n)| \quad (113)$$

and represent (92) in terms of it:

$$\text{MMSE}_p^{(N)}(n) = \frac{\exp((N+1) \cdot J(n, N+1))}{\exp(N \cdot J(n-1, N))}. \quad (114)$$

The determinant of  $\mathbf{R}_x^{(N)}(n)$  can be expressed as the product of its eigenvalues, which are obtained as

$$\mathcal{S}_{\mathcal{X}}^{(k)}(n) \triangleq \boldsymbol{\varphi}_k^{\text{H}}(n) \mathbf{R}_x^{(N)}(n) \boldsymbol{\varphi}_k(n), \quad (115)$$

where  $\boldsymbol{\varphi}_k(n) \triangleq [\varphi_{k,1}(n), \dots, \varphi_{k,N}(n)]^{\text{T}}$  are its unitary eigenvectors. Developing the previous quadratic form, we have

$$\mathcal{S}_{\mathcal{X}}^{(k)}(n) = \sum_{l,m=0}^{N-1} \varphi_{k,l}^*(n) R_x(n-N+m, l-m) \varphi_{k,m}(n). \quad (116)$$

Notice the similarities between this expression and the second identity from (13). Indeed, with appropriate index shifts in the autocorrelation function and the eigenvectors, in the limit

$$\lim_{N \rightarrow \infty} \mathcal{S}_{\mathcal{X}}^{(k)}(n) = \lim_{N \rightarrow \infty} \mathcal{S}_{\mathcal{X}}(\frac{k}{N}), \quad (117)$$

where the dependence on  $n$  has been dropped, since the spectrum of CS processes is shift-invariant, as proved in Section III-A.

Returning to (113), we have

$$\begin{aligned} \lim_{N \rightarrow \infty} J(n, N) &= \lim_{N \rightarrow \infty} \frac{1}{N} \ln \prod_{k=0}^{N-1} \mathcal{S}_{\mathcal{X}}^{(k)}(n) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_k \ln \mathcal{S}_{\mathcal{X}}\left(\frac{k}{N}\right) = \int_0^1 \ln \mathcal{S}_{\mathcal{X}}(\lambda) d\lambda \triangleq J. \end{aligned} \quad (118)$$

We have converted a Riemann sum into an integral, by assuming this limit exists and setting  $k/N \rightarrow \lambda \in [0, 1)$  and  $1/N \rightarrow d\lambda$ . The same analysis applies to  $J(n-1, N)$  and  $J(n, N+1)$ . Using these results in (114), we finally have

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{MMSE}_p^{(N)}(n) &= \lim_{N \rightarrow \infty} \frac{\exp((N+1) \cdot J)}{\exp(N \cdot J)} = e^J \quad (119) \\ &= \exp \sum_{p=0}^{P-1} \int_0^{\frac{1}{P}} \ln \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) d\sigma = \exp \int_0^{\frac{1}{P}} \ln \prod_p \mathcal{S}_{\mathcal{X}}^{(p)}(\sigma) d\sigma \\ &= \exp \int_0^{\frac{1}{P}} \ln |\mathcal{S}_{\mathcal{X}}(\sigma)| d\sigma = \exp \int_0^{\frac{1}{P}} \ln |\mathbf{S}_X(\sigma)| d\sigma. \quad \blacksquare \end{aligned}$$

### E. Spectral correlation of $x(n)$ in Section VIII

The CL expansion of  $d(n)$  is obtained as follows:

$$\begin{aligned} d\nu_d(f) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n, k=-\frac{N}{2}}^{\frac{N}{2}-1} a(k) b(n-kP) e^{-j2\pi f n} \\ &= \left( \sum_{n'} b(n') e^{-j2\pi f n'} \right) \left( \sum_k a(k) e^{-j2\pi f k P} \right) df \quad (120) \\ &= B(f) df \cdot \sum_k a(k) e^{-j2\pi f k P}, \end{aligned}$$

where we have used the change of variable  $n' \triangleq n - kP$ . We can now compute the spectral correlation of  $x(n)$ :

$$\begin{aligned} S_X(f, \alpha) &= S_Z(f, \alpha) + B(f) B^*(f - \alpha) \mathbb{E}[e^{-j2\pi \alpha \varepsilon}] \\ &\quad \times \sum_{k, k' \in \mathbb{Z}} \mathbb{E}[a(k) a^*(k')] e^{j2\pi (f(k' - k) - \alpha k') P}, \end{aligned} \quad (121)$$

due to  $d(n)$  and  $z(n)$  being independent. Since the symbols are uncorrelated and the noise is WSS and white, we have

$$\begin{aligned} S_X(f, \alpha) &= P_z \delta(\alpha) + \frac{B(f) B^*(f - \alpha)}{\Delta} \int_0^\Delta e^{-j2\pi \alpha \varepsilon} d\varepsilon \\ &\quad \times \sum_{k, k'} P \delta_{k-k'} e^{j2\pi (f(k' - k) - \alpha k') P} \quad (122) \\ &= P_z \delta(\alpha) + B(f) B^*(f - \alpha) \frac{1 - e^{-j2\pi \alpha \Delta}}{j2\pi \alpha \Delta} P \sum_k e^{j2\pi \alpha k P} \\ &= P_z \delta(\alpha) + B(f) B^*(f - \alpha) \frac{1 - e^{-j2\pi \alpha \Delta}}{j2\pi \alpha \Delta} \text{III}_{\frac{1}{P}}(\alpha). \end{aligned}$$

### REFERENCES

- [1] A. Napolitano, "Cyclostationarity: New trends and applications," *Signal Processing*, vol. 120, pp. 385–408, Mar. 2016.
- [2] W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1986.
- [3] G. B. Giannakis, "Cyclostationary Signal Analysis," in *Digital Signal Processing fundamentals*, 2nd ed., V. K. Madisetti, Ed. Boca Raton, FL, USA: CRC Press, 2010.
- [4] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, no. 4, pp. 639–697, Apr. 2006.
- [5] A. Napolitano, *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Hoboken, NJ, USA: Wiley, 2012.
- [6] O. A. Yeste Ojeda and J. Grajal, "The Relationship Between the Cyclic Wiener Filter and Fractionally Spaced Equalizers," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4333–4341, Aug. 2019.
- [7] R. Chopra, D. Ghosh, and D. K. Mehra, "FRESH Filter-Based Spectrum Sensing in the Presence of Cyclic Frequency Offset," *IEEE Wireless Communications Letters*, vol. 5, no. 2, pp. 124–127, Apr. 2016.
- [8] W. Gardner, "Cyclic Wiener filtering: theory and method," *IEEE Transactions on Communications*, vol. 41, no. 1, pp. 151–163, 1993.
- [9] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, U.K.: Cambridge University Press, Feb. 2010.
- [10] J. Riba, J. Font-Segura, J. Villares, and G. Vazquez, "Frequency-Domain GLR Detection of a Second-Order Cyclostationary Signal Over Fading Channels," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1899–1912, Apr. 2014.
- [11] D. Ramírez, P. J. Schreier, J. Vía, I. Santamaría, and L. L. Scharf, "Detection of Multivariate Cyclostationarity," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5395–5408, Oct. 2015.
- [12] Y.-R. Chien, J.-L. Lin, and H.-W. Tsao, "Cyclostationary Impulsive Noise Mitigation in the Presence of Cyclic Frequency Offset for Narrowband Powerline Communication Systems," *Electronics*, vol. 9, no. 6, p. 988, Jun. 2020.
- [13] M. Elgenedy, M. Sayed, N. Al-Dhahir, and R. C. Chabaan, "Cyclostationary Noise Mitigation for SIMO Powerline Communications," *IEEE Access*, vol. 6, pp. 5460–5484, 2018.
- [14] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Boston, MA, USA: Now, 2005.
- [15] A. V. Oppenheim and G. C. Verghese, *Signals, Systems and Inference*. Pearson Education, Limited, 2015.
- [16] P. P. Vaidyanathan, *Theory of Linear Prediction*. San Rafael, CA, USA: Morgan & Claypool, 2007.
- [17] B. Anderson and S. Chirarattananon, "Smoothing as an improvement on filtering: a universal bound," *Electronics Letters*, vol. 7, no. 18, p. 524, 1971.
- [18] T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 368–371, Jul. 1971.
- [19] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [20] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "The Distortion Rate Function of Cyclostationary Gaussian Processes," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3810–3824, May 2018.
- [21] A. Pries, D. Ramirez, and P. J. Schreier, "LMPIT-Inspired Tests for Detecting a Cyclostationary Signal in Noise With Spatio-Temporal Structure," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6321–6334, Sep. 2018.
- [22] J. Riba and M. Vila, "On Infinite Past Predictability of Cyclostationary Signals," *IEEE Signal Processing Letters*, vol. 29, pp. 647–651, 2022.
- [23] A. M. Harry L. Hurd, *Periodically Correlated Random Sequences: Spectral Theory and Practice*. Hoboken, NJ, USA: Wiley, Sep. 2007.
- [24] D. Ramírez, I. Santamaría, and L. Scharf, *Coherence*. Springer International Publishing, 2022.
- [25] W. A. Gardner, *Cyclostationarity in communications and signal processing*. Piscataway, NJ, USA: IEEE Press, 1994.
- [26] R. Wang, *Introduction to Orthogonal Transforms*. Cambridge, U.K.: Cambridge University Press, Mar. 2012.
- [27] P. S. Rodney A. Kennedy, *Hilbert Space Methods in Signal Processing*. Cambridge, U.K.: Cambridge University Press, Sep. 2013.
- [28] K. L. B. Harry L. van Trees, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. Hoboken, NJ, USA: Wiley, Apr. 2013.
- [29] P. Mitra, C. Murthy, and S. Pal, "Unsupervised feature selection using feature similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 301–312, 2002.
- [30] W. Yang, J. Gibson, and T. He, "Coefficient rate and lossy source coding," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 381–386, Jan. 2005.
- [31] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. Springer New York, 2011.
- [32] N. J. Bershad, E. Eweda, and J. C. M. Bermudez, "Stochastic Analysis of an Adaptive Line Enhancer/Canceler With a Cyclostationary Input,"

- IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 104–119, Jan. 2016.
- [33] W. A. Gardner, *Introduction to Random Processes: With Applications to Signals and Systems*. New York, NY, USA: Macmillan, 1986.
  - [34] D. S. Bernstein, *Scalar, Vector, and Matrix Mathematics: Theory, Facts, and Formulas*, 3rd ed. Princeton, NJ, USA: Princeton University Press, 2018.
  - [35] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing*. London, England: Pearson Education, Limited, 2013.
  - [36] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer US, 2008.
  - [37] A. D. R. Choudary and C. P. Niculescu, *Real Analysis on Intervals*, 1st ed. New Delhi, India: Springer India, 2014.
  - [38] J. L. Doob, *Stochastic processes*. Hoboken, NJ, USA: Wiley, 1990.
  - [39] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. New York, NY, USA: Springer US, 2004.
  - [40] A. Napolitano, “Generalizations of Cyclostationarity: A New Paradigm for Signal Processing for Mobile Communications, Radar, and Sonar,” *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 53–63, Nov. 2013.
  - [41] —, “Cyclostationarity: Limits and generalizations,” *Signal Processing*, vol. 120, pp. 323–347, Mar. 2016.