Bayesian Interpretation of Periodograms

Jean-François Giovannelli and Jérôme Idier

***Abstract*—The usual nonparametric approach to spectral analysis is revisited within the regularization framework. Both usual and windowed periodograms are obtained as the squared modulus of the minimizer of regularized least squares criteria. Then, particular attention is paid to their interpretation within the Bayesian statistical framework. Finally, the question of unsupervised hyperparameter and window selection is addressed. It is shown that maximum likelihood solution is both formally achievable and practically useful.**

***Index Terms*—Hyperparameters, penalized criterion, periodograms, quadratic regularization, spectral analysis, windowing, window selection, zero-padding.**

NOMENCLATURE

FT	Fourier transform.
IFT	Inverse Fourier transform.
CFT	Continuous frequency.
DF	Discrete frequency.
UP	Usual periodogram.
WP	Windowed periodogram.
L^2	$L^2_{\mathbb{C}}([0,1])$.
H^Q	$H^Q_{\mathbb{C}}([0,1])$.
ℓ^2	$\ell^2_{\mathbb{C}}(\mathbb{Z})$
$\mathcal{F}$	Discrete time FT ($\ell^2 \to L^2$).
$\mathcal{W}_N$	Truncated IFT ($L^2 \to \mathbb{C}^N$).
$\mathcal{W}_N^\dagger$	Adjoint operator of $\mathcal{W}_N$.
F_P	Square Fourier matrix ($\mathbb{C}^P \to \mathbb{C}^P$).
W_{NP}	Truncated IFT matrix ($\mathbb{C}^P \to \mathbb{C}^N, N \leq P$).
$W_{NP}^\dagger$	Hermitian matrix of W_{NP}.
$\mathbb{N}_N$	$\{0,1,\ldots,N-1\}$.

I. INTRODUCTION

SPECTRAL analysis is a fundamental problem in signal processing. Historical papers such as [1], tutorials such as [2] and books such as [3] and [4] are evidence of the basic role of spectral analysis, whether it is parametric or not.

The nonparametric approach has recently prompted renewed interest [5] (see also [6]) within the regularization framework, and the present contribution brings a new look at these methods. It provides statistical principles rather than empirical ones in order to derive periodogram estimators. From this standpoint, the major contribution of the paper is twofold. First, it proposes new coherent interpretations of existing periodograms and modern justification for windowing techniques. Second, it introduces a maximum likelihood method for automatic selection of the window shape.

Moreover, [5] suffers from a twofold limitation. On the one hand, the proposed model relies on the discrete frequency, whereas the frequency is a continuous variable. On the other hand, restriction to separable regularization functions does not allow spectral smoothness to be accounted for. The present contribution overcomes such limitations.

It takes advantage of a natural model in spectral analysis of complex discrete-time series: the sum of side-by-side pure frequencies. Two cases are investigated.

1) the continuous frequency (CF) case, which relies on an infinite number of pure frequencies $\nu \in [0,1[$ with amplitudes $a(\nu), a \in L^2$;
2) the discrete frequency (DF) one, which relies on a finite number, say P (usually large), of equally spaced pure frequencies $\nu = p/P$, with amplitudes a_p. Let us note that $\boldsymbol{a} = [a_0,\ldots,a_{P-1}] \in \mathbb{C}^P$, and $\boldsymbol{\nu} = [\nu_0,\ldots,\boldsymbol{\nu}_{P-1}] \in [0,1[^P$.

For N complex observed samples $\boldsymbol{y} = [y_0,\ldots,y_{N-1}] \in \mathbb{C}^N$, such models read

$$\begin{aligned} \text{CF: } y_n &= \int_0^1 a(\nu) e^{2i\pi\nu n} d\nu + b_n \\ \text{DF: } y_n &= P^{-1/2} \sum_{p=0}^{P-1} a_p e^{2i\pi pn/P} + b_n \end{aligned} \tag{1}$$

where $\boldsymbol{b} = [b_0,\ldots,b_{N-1}] \in \mathbb{C}^N$ accounts for model and observation uncertainties. Let us introduce $\mathcal{W}_N$ and W_{NP}:

$$\begin{aligned} &\text{CF: } \mathcal{W}_N : L^2 \longrightarrow \mathbb{C}^N \\ &\text{DF: } W_{NP} : \mathbb{C}^P \longrightarrow \mathbb{C}^N \end{aligned} \tag{2}$$

the CF and DF truncated IFT so that

$$\begin{aligned} &\text{CF: } \boldsymbol{y} = \mathcal{W}_N a + \boldsymbol{b} \\ &\text{DF: } \boldsymbol{y} = W_{NP}\boldsymbol{a} + \boldsymbol{b}. \end{aligned} \tag{3}$$

The current problem consists in estimating the amplitudes a and/or $\boldsymbol{a}$. Thanks to the linearity of these models w.r.t. the amplitudes, the problem clearly falls in the class of linear estimation problems [7]–[9]. However, in practice, estimation relies on a finite, maybe small, number of data N. As a consequence, in the CF case, a continuous frequency function a lying in L^2 must be selected from only N data. Such a problem is known to be ill-posed in the sense of Hadamard [8]. In the same way, under the DF formulation, since the amplitudes outnumber the available data, the problem is underdeterminate.

This kind of problem is nowadays well identified [8], [10] and can be fruitfully tackled by means of the regularization

Manuscript received October 24, 2000; revised March 7, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jian Li.

The authors are with the Laboratoire des Signaux et Systèmes SUPÉLEC, Gif-sur-Yvette, France (emaux: giova@lss.supelec.fr; idier@lss.supelec.fr).

Publisher Item Identifier S 1053-587X(01)05353-3.

approach. This approach rests on a compromise between fidelity to the data and fidelity to some prior information about the solution. As mentioned above, such an idea has already been introduced in several papers [5], [11]–[14]. In the autoregressive spectral estimation problem, [11] proposes to account for spectral smoothness as a function of autoregressive coefficients. Otherwise, high-resolution spectral estimation has been addressed within the regularization framework, founded on the Poisson-Gaussian model [14]. The present paper deepens Gaussian models and is organized as follows.

Section II focuses on the interpretation of usual periodograms (UPs), and Section III deals with the interpretation of windowed periodograms (WPs), both using penalized approaches with quadratic regularization. Results are exposed in four propositions, and the corresponding proofs are given in Appendix A. A Bayesian interpretation is presented in Section IV, whereas the problem of parameter estimation and window selection are addressed in Section V. Finally, conclusions and perspectives for future works are presented in Section VI.

II. Usual Periodogram

A. Continuous Frequency

The problem at stake consists of estimating $a \in L^2$ given data $\boldsymbol{y}$ such that (3). A first possible approach is founded on the least squares (LS) criterion

$$(\boldsymbol{y}-\mathcal{W}_N a)^\dagger(\boldsymbol{y}-\mathcal{W}_N a)=\sum_{n=0}^{N-1}\left|y_n-\int_0^1 a(\nu)e^{2i\pi\nu n}d\nu\right|^2$$

but since $\mathcal{W}_N$ is one-to-many and not many-to-one, there exists an infinity of solutions in L^2. Here, the preferred solution for raising the indetermination relies on regularized least squares (RLS). The simplest RLS criterion is founded on quadratic "separable regularization"

$$\mathcal{Q}_u(a)=(\boldsymbol{y}-\mathcal{W}_N a)^\dagger(\boldsymbol{y}-\mathcal{W}_N a)+\lambda\int_0^1|a(\nu)|^2d\nu \quad (4)$$

where "u" stands for usual. The regularization parameter $\lambda \geq 0$ balances the tradeoff between confidence in the data and confidence in the penalization term. For any $\lambda > 0$, the proposition below gives the minimizer $\hat{a}^\lambda$ of (4).

Proposition 1 (CF/UP): For any $\lambda > 0$, the unique minimizer of (4) reads

$$\hat{a}^\lambda(\nu)=(1+\lambda)^{-1}\sum_{n=0}^{N-1}y_n e^{-2i\pi\nu n}. \quad (5)$$

■

Proof: See Appendix A.

B. Discrete Frequency

This subsection investigates the DF counterpart of the previous result. In the DF approach, the LS criterion reads

$$(\boldsymbol{y}-W_{NP}\boldsymbol{a})^\dagger(\boldsymbol{y}-W_{NP}\boldsymbol{a}) \quad (6)$$

but since W_{NP} is one-to-many and not many-to-one, there also exists an infinity of solutions in $\mathbb{C}^P$. According to the quadratic "separable regularization," the corresponding RLS criterion is

$$Q_u(\boldsymbol{a})=(\boldsymbol{y}-W_{NP}\boldsymbol{a})^\dagger(\boldsymbol{y}-W_{NP}\boldsymbol{a})+\lambda\boldsymbol{a}^\dagger\boldsymbol{a} \quad (7)$$

with optimum given in the next proposition.

Proposition 2—(DF/UP): For any $\lambda > 0$, the unique minimizer of (7) reads

$$\hat{\boldsymbol{a}}^\lambda=(1+\lambda)^{-1}F_P\tilde{\boldsymbol{y}}_P \quad (8)$$

where $\tilde{\boldsymbol{y}}_P$ denotes the vector $\boldsymbol{y}$ zero-padded up to size P. ■

Proof: See Appendix A.

C. Usual Periodogram: Concluding Remarks

In the CF cases, the squared modulus of the penalized solutions $|\hat{a}^\lambda(\nu)|^2$ is proportional to the usual zero-padded periodogram. Moreover, $|\hat{a}^\lambda|^2$ is[1] a discretized version of $|\hat{a}^\lambda(\nu)|^2$ over the frequency grid $\boldsymbol{\nu}$. Therefore, within the proposed framework, *separable quadratic regularization* leads to the *usual zero-padding* technique associated with the practical computation of periodograms. Moreover, when λ tends to zero, the proportionality factor tends to one. It is noticeable that in this case, the criteria (4) and (7) degenerate, but their minimizer does not. They are the solution of the constraint problems

$$\text{CF: } \min_{a\in L^2}\int_0^1|a(\nu)|^2d\nu \text{ s.t. } \boldsymbol{y}=\mathcal{W}_N a$$
$$\text{DF: } \min_{\boldsymbol{a}\in\mathbb{C}^P}\boldsymbol{a}^\dagger\boldsymbol{a} \qquad \text{s.t. } \boldsymbol{y}=W_{NP}\boldsymbol{a}$$

i.e., solution of the noiseless problems adressed in [5] and [6].

III. Windowed Periodogram

The previous section investigates the relationships between the separable regularizers and the usual (nonwindowed) periodograms. The present section focuses on smoothing regularizers and windowed periodograms (see [15], which analyzes dozens of windows to compute smoothed periodograms).

A. Continuous Spectra

This subsection generalizes the usual norm in L^2 to the Sobolev [16] regularizer

$$\mathcal{R}_Q(a)=\int_0^1\sum_{q=0}^{Q}\alpha_q\left|\frac{d^q a}{d\nu^q}(\nu)\right|^2 d\nu$$

which can be interpreted as a measure of spectral smoothness. The α_q are positive real coefficients and can be generalized to positive real functions [8]. $\mathcal{R}_Q$ is defined onto the Sobolev space [16] $H^Q \subset L^2$. Note that $H^0 = L^2$ and that the usual norm invoked in Section II-A is the regularizer $\mathcal{R}_0$ with $\alpha_0 = 1$.

Remark 1: Strictly speaking, $\mathcal{R}_Q(a)$ is not a spectral smoothness measure since it is not a function of $|a(\nu)|$ but a function of $a(\nu)$, including phase. A true spectral smoothness measure does not depend on the phase of $a(\nu)$ and does not yield a quadratic criterion. The same remark holds for the definition of spectral smoothness proposed by Kitagawa and Gersh [11].

[1] If $u \in \mathbb{C}^P$, $|u|^2$ denotes the vector of the squared moduli of the component of u.

Accounting for spectral smoothness by means of $\mathcal{R}_Q(a)$ yields a new penalized criterion

$$\mathcal{Q}_s(a) = (\boldsymbol{y} - \mathcal{W}_N a)^\dagger (\boldsymbol{y} - \mathcal{W}_N a) + \lambda \mathcal{R}_Q(a) \tag{9}$$

where the index "s" stands for smoothness.

Proposition 3—(CF/WP): With the previous notations and definitions, the minimizer of (9) reads

$$\hat{a}^\omega(\nu) = \sum_{n=0}^{N-1} \omega_n y_n e^{-2i\pi\nu n} \tag{10}$$

i.e., a windowed FT. The window shape is

$$\omega_n = (1 + \lambda\varepsilon_n)^{-1} \tag{11}$$

$$\text{with } \varepsilon_p = \sum_{q=0}^{Q} \alpha_q (2\pi p)^{2q} \text{ for } p \in \mathbb{Z}. \tag{12}$$

Proof: See Appendix A. ∎

B. Discretized Spectra

This subsection is devoted to the generalization of criterion (7) to nonseparable penalization

$$Q_s(\boldsymbol{a}) = (\boldsymbol{y} - W_{NP}\boldsymbol{a})^\dagger (\boldsymbol{y} - W_{NP}\boldsymbol{a}) + \lambda \boldsymbol{a}^\dagger \Pi_a \boldsymbol{a}. \tag{13}$$

Given that the sought spectrum is circular periodic, the penalization term has to be designed under circularity constraint. As a consequence, Π_a is a circular matrix, and its eigenvalues, denoted $e_p, p \in \mathbb{N}_p$, can be calculated as the FT of the first row of Π_a. Moreover, without loss of generality, we assume that the diagonal elements of Π_a^{-1} are equal to one, and any scaling factor is integrated in the parameter λ.

Proposition 4—(DF/WP): The minimizer of (13) reads

$$\hat{\boldsymbol{a}}^w = F_P \breve{\boldsymbol{y}} \tag{14}$$

where the $\breve{\boldsymbol{y}}_p = w_p \tilde{y}_p$ for $p \in \mathbb{N}_P$ and

$$w_p = (1 + \lambda e_p)^{-1}.$$

Proof: See Appendix A ∎

C. Windowed Periodograms: Concluding Remarks

Hence, in the CF case, the squared modulus of the penalized solution $\hat{a}^\omega$ is the windowed periodogram associated with window ω_n. Moreover, the DF solution $\hat{a}^\omega$ is a discretized version of $\hat{a}^\omega$ as soon as the e_n are identified with the ε_n. As a conclusion, *quadratic smoothing regularizers* interpret *windowed periodograms*. Moreover, it is noteworthy that $\hat{a}^\omega(\nu)$ and $\hat{a}^\omega$ only depend on e_n and ε_n for $n \in \mathbb{N}_N$.

Remark 2—Empirical Power: One can easily show that

$$\text{CF: } \int_0^1 |a(\nu)|^2 d\nu = \sum_{n=0}^{N-1} \omega_n^2 |y_n|^2$$

$$\text{DF: } \boldsymbol{a}^\dagger \boldsymbol{a} = \sum_{n=0}^{N-1} w_n^2 |y_n|^2. \tag{15}$$

Hence, the empirical power of the estimated spectra is smaller than the empirical power of the observed data, and equality holds if and only if $\lambda = 0$.

Example 1—Zero-Order Penalization: The most simple example consists in retrieving the nonwindowed case of Section II-A and B. Let us apply the previous Propositions 3 and 4 with regularizers

$$\text{CF: } \int_0^1 |a(\nu)|^2 d\nu \quad \text{i.e., } Q = 0 \text{ and } \alpha_0 = 1$$

$$\text{DF: } \boldsymbol{a}^\dagger \boldsymbol{a} \quad \text{i.e., } \Pi_a = I_P. \tag{16}$$

Then, we have $\varepsilon_n = e_n = 1$; the criteria (9) and (13), respectively, become (4) and (7), and the solutions (10) and (14), respectively, become (5) and (8). As expected, the nonwindowed solutions are retrieved. A more interesting example is the one given below.

Example 2—First-Order Penalization: Let the penalization term be

$$\text{CF: } \int_0^1 |a'(\nu)|^2 d\nu$$

$$\text{DF: } \frac{1}{2} P^2 \sum_{k=0}^{P} |a_k - a_{k-1}|^2 \tag{17}$$

with $a_P = a_0$ for notational convenience of the circularity assumption. Application of Propositions 3 and 4, respectively, yields $\varepsilon_n = 4\pi^2 n^2$ (CF case) and $e_n = (1 - \cos 2\pi n/P)$ (DF case). The corresponding windows read

$$\text{CF: } \omega_n = (1 + 4\pi^2 n^2 \lambda)^{-1}$$

$$\text{DF: } w_n = (1 + \lambda - \lambda \cos 2\pi n/P)^{-1}. \tag{18}$$

In the following, we refer to them as the Cauchy and the inverse cosine windows. Moreover, for a finer discretization of the spectral domain, $\lim_{P\to\infty} e_n = \varepsilon_n$, and one can retrieve the Cauchy window as the limit of the inverse cosine window (see Figs. 1 and 2).

IV. Bayesian Interpretation

This section is devoted to Bayesian interpretations of the penalized solutions presented in Propositions 1, 2, 3, and 4. Moreover, since usual nonwindowed forms are particular cases of windowed forms, we focus on the latter.

Since the considered criteria are quadratic, their Bayesian interpretations rely on Gaussian laws. Therefore, the Bayesian interpretations only require the characterization of means and correlation structures for the stochastic models at work.

A. Discrete Frequency Approach

In the DF case, i.e., in the finite dimension vector space, the Bayesian interpretation of the criteria (7) and (13) as a *posterior* co-log-likelihood is a classical result [10]. Within this probabilistic framework, the likelihood of the parameters $\boldsymbol{a}$ attached to the data $\boldsymbol{y}$ is

$$f(\boldsymbol{y}|\boldsymbol{a}) = (\pi r_b)^{-N} \exp \frac{-1}{r_b} (\boldsymbol{y} - W_{NP}\boldsymbol{a})^\dagger (\boldsymbol{y} - W_{NP}\boldsymbol{a}).$$

From a statistical viewpoint, it essentially results from the linearity of the model (3) and from the hypothesis of a zero-mean, circular (in the statistical sense), stationary, white, and Gaussian noise vector $\boldsymbol{b}$, with variance r_b.

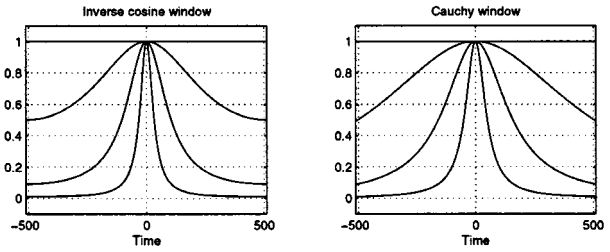

Fig. 1. Inverse cosine (lhs) and Cauchy windows (rhs) as a function of λ. In both cases, $\lambda = 0$ yields a constant shape. Furthermore, for any $\lambda\omega_0 = \omega_0 = 1$. Otherwise, as λ increases, the window shape decreases faster to zero, and the corresponding spectrum is smoothed.

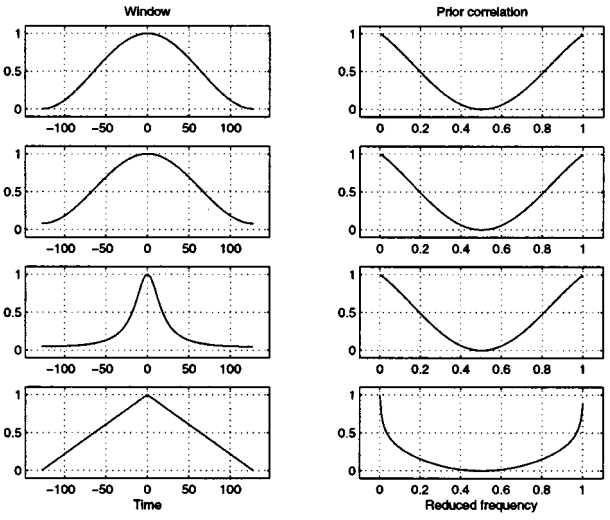

Fig. 2. Usual windows and the corresponding correlations. The lhs column shows the time window, and the rhs column shows the associated correlations. From top to bottom: the Hamming, the Hanning, the inverse cosine, and the triangular.

Moreover, in order to interpret the regularization term of (13), a zero-mean, circular, correlated Gaussian prior with covariance $R_a = r_a \Pi_a^{-1}$ is introduced.[2] Matrix Π_a^{-1} is the normalized covariance structure, i.e., all its diagonal elements are equal to 1, whereas r_a stands for the prior power. Therefore, the prior density reads

$$f(\boldsymbol{a}) = (\pi r_a)^{-N} \det \Pi_a \exp \frac{-1}{r_a} \boldsymbol{a}^\dagger \Pi_a \boldsymbol{a}.$$

The Bayes rule ensures the fusion of the likelihood and the prior into the *posterior* density

$$f(\boldsymbol{a}|\boldsymbol{y}) \propto \exp \frac{-1}{r_b} Q_s(\boldsymbol{a})$$

where Q_s is given by (13). The regularization parameter λ is clearly $\lambda = r_b/r_a$.

Thus, we have a Bayesian interpretation of the criterion (13) related to windowed periodograms. Interpretation of the criterion (7) related to usual ones results from a white prior: $\Pi_a = I_P$. Finally, interpretations of the RLS solutions (8) and (14) themselves result from the choice of the maximum *a posteriori* (MAP) as a punctual estimate. Moreover, thanks to the Gaussian character of *posterior* law, other basic Bayesian estimators such as *posterior* mean (PM) and marginal MAP (MMAP), are equal to the MAP solution itself.

[2]Rigorously speaking, this is possible only if Π_a is invertible.

B. Continuous Frequency Case

1) General Theory: In the CF case, the Bayesian interpretation is more subtle since it relies on continuous index stochastic processes. Indeed, no *posterior* likelihood for the parameter a is available. Therefore, there is no direct *posterior* interpretation of the criteria (4) and (9), nor is there MAP interpretation of the estimates (5) and (10). Roughly speaking, the *posterior* law vanishes everywhere. Nevertheless, there is a proper Bayesian interpretation of the estimates (5) and (10) as PM or MMAP, as shown below.

Let us introduce a zero-mean, circular (in the statistical sense) and Gaussian prior law [17] for a. This law is fully characterized by its correlation structure $\gamma_a(\nu), \nu \in [-1, 1]$, which is entirely described by its values for $\nu \in [0, 1]$ thanks to Hermitian symmetry. Furthermore, the usual circular-periodicity assumption for $a(\nu)$ results in another symmetry property: $\gamma_a(1/2 + \nu) = \gamma_a(1/2 - \nu)$ any $\nu \in [0, 1/2]$.

By assuming $\gamma_a \in L_2$, the latter can be expanded into a Fourier series

$$\gamma_a(\nu) = \sum_{p \in \mathbb{Z}} \mathring{\gamma}_a(p) e^{-2i\pi\nu p}, \quad \nu \in [0, 1]$$

with Fourier coefficients $\mathring{\gamma}_a \in \ell_2$ given by

$$\mathring{\gamma}_a(p) = \int_{[0,1]} \gamma_a(\nu) e^{-2i\pi\nu p}, \quad p \in \mathbb{Z}.$$

Let us note that $c_a(\nu) = \gamma_a(\nu)/r_a$ is the normalized correlation and that $\mathring{c}_a \in \ell_2$ is the corresponding Fourier sequence.

Proposition 5: With the previous notations and prior choice, the *posterior* mean of $a(\nu)$ is

$$E[a(\nu)|\boldsymbol{y}] = \hat{a}^\omega(\nu) = \sum_{n=0}^{N-1} \omega_n y_n e^{-2i\pi\nu n} \tag{19}$$

$$\text{with } \omega_n = [1 + \lambda \mathring{c}_a(n)^{-1}]^{-1}. \tag{20}$$

■

Proof: See Appendix A

Comparison of (19)–(20) and (10)–(11) immediately gives the Bayesian interpretation of windowed FT as PM[3]: $\mathring{c}_a(n) = \varepsilon_n^{-1}$, i.e., identification of the Fourier coefficients of the prior correlation $c_a(\nu)$ and the FT of the discrete correlation Π_a.

2) Example 3: The present subsection is devoted to a precise Bayesian interpretation of deterministic Examples 1 and 2. As we will see, there is a new obstacle in the Bayesian interpretation of these examples because the underlying correlations do not lie in L_2. In order to overcome this difficulty, we first interpret the penalization of both zero-order and first-order derivatives

$$\mathcal{R}_2(a) = \alpha_0 \int_0^1 |a(\nu)|^2 d\nu + \alpha_1 \int_0^1 |a'(\nu)|^2(\nu) d\nu. \tag{21}$$

The case of pure zero order and pure first order are obtained in Section IV-B.II.b and c as limit processes.

[3]Since $a(\nu) \mid y$ is a scalar Gaussian random variable, $E[a(\nu) \mid y]$ is also the MMAP.

As seen in Proposition 3, the associated coefficients are $\varepsilon_p = a_0 + 4\pi^2\alpha_1 p^2$, $p \in \mathbb{Z}$. According to Proposition 5, the Fourier series coefficients for $\gamma_a(\nu)$ are $\overset{\circ}{\gamma}_a(p) = \varepsilon_p^{-1}$. It is clear that $\overset{\circ}{\gamma} \in \ell_2$; hence, $\gamma_a \in L_2$ and

$$\gamma_a(\nu) = \sum_{p\in\mathbb{Z}} \frac{1}{\alpha_0 + 4\pi^2\alpha_1 p^2} e^{-2i\pi\nu p}, \quad \nu \in [0,1]. \tag{22}$$

It is shown in Appendix B that, with $\alpha = \sqrt{\alpha_0/\alpha_1}$ and $\alpha' = \sqrt{\alpha_0\alpha_1}$, $\gamma_a(\nu)$ reads

$$\gamma_a(\nu) = \frac{\cosh\alpha\,(|\nu| - 1/2)}{2\alpha' \sinh\alpha/2}, \quad \nu \in [-1,1] \tag{23}$$

and several analytic properties are straightforwardly deduced. In particular, γ_a has a continuous derivative over [-1,1]-{0}, and the slopes at $\nu = 0^-$ and $\nu = 0^+$ are, respectively, $1/\alpha_1$ and $-1/\alpha_1$. γ_a is minimum at $\nu = 1/2$ and maximum at $\nu = -1, \nu = 0$, and $\nu = 1$. Moreover, its integral from 0 to 1 remains constant and equals $1/\alpha_0$.

a) Markov Property: The present paragraph addresses the Markov property of the underlying prior process $a(\nu)$ [18], [19]. This process cannot be seen as a Markov chain since it is circular-periodic: "Future" frequency and "past" frequency cannot be independent. However, we show the Markov property for the conditional process $\bar{a}(\nu) = [a(\nu)|a(1)]_{\nu\in[0,1]}$. It is shown in Appendix B that its correlation structure reads

$$\gamma_{\bar{a}}(\nu,\nu') = \gamma_a(\nu-\nu') - \frac{\gamma_a(\nu)\gamma_a(\nu')}{\gamma_a(0)} \tag{24}$$

$$= \frac{\sinh\alpha\nu' \sinh\alpha(1-\nu)}{\alpha' \sinh\alpha} \tag{25}$$

for any $\nu, \nu' \in [0,1], \nu \geq \nu'$. According to the sufficient factorization of the correlation function proposed in [[20], p. 64], it turns out that $\bar{a}(\nu)$ is a Markov chain.

b) Limit Case as $\alpha_1 \to 0$: As α_1 tends to zero, it is easy to show that for each $\nu \in]0,1[$, the correlation $\gamma_a(\nu)$ tends to zero, i.e., there is no more correlation between $a(\nu_1)$ and $a(\nu_2)$ as soon as $\nu_1 \neq \nu_2$ and $(\nu_1,\nu_2) \neq (0,1)$. Moreover, $\gamma_a(0)$ and $\gamma_a(1)$ tend to infinity, whereas the integral of γ_a over [0,1] remains $1/\alpha_0$. Roughly speaking, the limit correlation is a Dirac distribution at $\nu = 0$ and $\nu = 1$ with weight $1/2\alpha_0$ i.e., the limit process is a circular white Gaussian noise with "pseudo-power" $1/\alpha_0$.

c) Limit Case as $\alpha_0 \to 0$: This case is more complex than the previous one since $\forall\nu \in [0,1], \gamma_a(\nu)$ tends to infinity as α_0 tends to zero. Therefore, we propose a characterization of the limit process *via* its increments. Let $\nu_1, \nu_2, \nu_1', \nu_2' \in [0,1], \nu_1 < \nu_2 < \nu_1' < \nu_2'$. Let us also note the frequency increments $\tau_\nu = \nu_2 - \nu_1$ and $\tau_\nu' = \nu_2' - \nu_1'$, and the vector of the increments themselves $i = [a(\nu_2) - a(\nu_1), a(\nu_4) - a(\nu_3)] \in \mathbb{C}^2$. This vector is clearly Gaussian and zero mean. Furthermore, it is shown in Appendix B that its covariance matrix reads

$$R_i = \frac{1}{2\alpha_1} \begin{bmatrix} \tau_\nu(1-\tau_\nu) & 2\tau_\nu\tau_\nu' \\ 2\tau_\nu\tau_\nu' & \tau_\nu'(1-\tau_\nu') \end{bmatrix}. \tag{26}$$

It turns out that the process $\tilde{a}(\nu) = a(\nu) - a(0)$ is a Brownian bridge [21, p. 36].

V. Hyperparameter and Window Slection

The problem of hyperparameter estimation within the regularization framework is a delicate one. It has been extensively studied, and numerous techniques have been proposed and compared [22]–[27]. The maximum likelihood (ML) approach is often chosen associated with the Bayesian interpretation. In the following subsections, we address regularization parameter estimation and automatic window selection using ML estimation.

A. Hyperparameters Estimation

In our context, the ML technique consists of integrating the amplitudes out of the problem and maximizing the resulting marginal likelihood w.r.t. the hyperparameters. Thanks to the linear and Gaussian assumptions, the marginal law for the data, namely, the likelihood function, is also Gaussian

$$f(\boldsymbol{y}; r_a, r_b) \propto (\det R_y)^{-1} \exp -\boldsymbol{y}^\dagger R_y^{-1} \boldsymbol{y}. \tag{27}$$

Moreover, the covariance structure R_y can be easily derived, as shown in the two following sections.

1) Discrete Frequency Marginal Covariance: In the present case, since all random quantities are in a finite dimensional linear space, the covariance is clearly

$$R_y = r_a(W_{NP}\Pi_a^{-1}W_{NP}^\dagger + \lambda I_N) = r_a\Sigma_y.$$

Accounting for the circular structure of the matrix Π_a, we have $\Pi_a = F_P\Lambda_\Pi F_P^\dagger$, where Λ_Π is the diagonal matrix of eigenvalues: $e_p, p \in \mathbb{N}_P$. Given the property (33) in Appendix B, Σ_y is shown to be diagonal

$$\Sigma_y = \text{diag}[\lambda + e_n^{-1}], \quad n \in \mathbb{N}_N. \tag{28}$$

2) Continuous Frequency Marginal Covariance: In the present case, the marginal covariance matrix R_y has already been derived in (32) in Appendix A. Hence, R_y and Σ_y are diagonal:

$$\Sigma_y = \frac{1}{r_a} R_y = \text{diag}[\lambda + \varepsilon_n^{-1}], \quad n \in \mathbb{N}_N. \tag{29}$$

Remark 3: In both cases, Σ_y only depends on e_n/ε_n for $n \in \mathbb{N}_N$. Consequently, the likelihood function and the ML parameter only depend on the N first coefficients.

3) Maximization: The opposite of the logarithm of the likelihood, namely, the co-log-likelihood (CLL)

$$\text{CLL}(r_a, \lambda) = N\log r_a + \log\det\Sigma_y + \frac{1}{r_a}\boldsymbol{y}^\dagger\Sigma_y^{-1}\boldsymbol{y} \tag{30}$$

must be minimized w.r.t. r_a and λ. Partial minimization is tractable w.r.t. r_a and yields $\hat{r}_a = \boldsymbol{y}^\dagger\sum_y^{-1}\boldsymbol{y}/N$. Substitution of $\hat{r}_a$ in (30) gives

$$\text{CLL}(\lambda) = \log\det\Sigma_y + N\log\boldsymbol{y}^\dagger\Sigma_y^{-1}\boldsymbol{y}. \tag{31}$$

Furthermore, since Σ_y is a diagonal matrix

$$\text{CLL}(\lambda) = \sum_{n=1}^N \log(\lambda + e_n^{-1}) + N\log\sum_{n=1}^N \frac{|y_n|^2}{\lambda + e_n^{-1}}$$

$$= \log\left\{\prod_{n=1}^N (\lambda + e_n^{-1}) \left[\sum_{n=1}^N \frac{|y_n|^2}{\lambda + e_n^{-1}}\right]^N\right\}$$

in the DF case. Substitution of e_n by ε_n yields the CF case. In both cases, CLL (λ) is the logarithm of the ratio of two degree $N-1$ polynomials of the variable λ with a strictly positive denominator. Minimization w.r.t. λ is not explicit, but it can be numerically performed.

4) Simulation Results: ML hyperparameter selection is illustrated for the problem of Section IV-B2. Computations have been performed on the basis of of 512 sample signals simulated by filtering standard Gaussian noises with the filter of impulse response $h = [1, -2, 3, -2, 1]$. Let us note that a^* as the true spectrum.

CLL has been computed on a (α_0, α_1)-grid of 100×100 logarithmically spaced values from 10^{-10} to 10^{10}. The first observation is that CLL is fairly regular and usually shows a unique minimum located between 10^{-1} and 10^{1} for α_0 and between 10^{-2} and 1 for α_1. However, a few "degenerated" cases have been observed for which $\hat{\alpha}_0^{ML}$ or $\hat{\alpha}_1^{ML}$ seem to be null or infinite. Let us note $\hat{\alpha}_0^{ML}$, $\hat{\alpha}_1^{ML}$ as the CLL minimizer[4] and $\hat{a}_{RLS}^{ML}$ as the corresponding RLS periodogram.

Since a^* is known in the proposed simulation study, various spectral distances [30] can be computed as functions of α_0 and α_1. L_1 distance, L_2 distance, the Itakura–Saito divergence (ISD) as well as the Itakura–Saito symmetric distance (SIS) have been considered. Each one provides an optimal couple $(\hat{\alpha}_0^{L_1}, \hat{\alpha}_1^{L_2}), (\hat{\alpha}_0^{L_2}, \hat{\alpha}_1^{L_2}), (\hat{\alpha}_0^{ISD}, \hat{\alpha}_1^{ISD})$, and $(\hat{\alpha}_0^{SIS}, \hat{\alpha}_1^{SIS})$, respectively. The corresponding spectra are, respectively, denoted $\hat{a}_{RLS}^{L_1}, \hat{a}_{RLS}^{ISD}$, and $\hat{a}_{RLS}^{SIS}$.

According to our experiments, as shown in Fig. 3, $\hat{a}_{RLS}^{L_2}, \hat{a}_{RLS}^{ISD}$, and the a^* can be graded by smoothness and estimation accuracy. From the smoothest to the roughest, the following gradation has always been observed: $\hat{a}_{RLS}^{L_2}, a^*$ and $\hat{a}_{RLS}^{ISD}$. Furthermore, $\hat{a}_{RLS}^{L_2}$ is systematically oversmoothed, whereas $\hat{a}_{RLS}^{ISD}$ is systematically undersmoothed. Moreover, the first one qualitatively approximates more precisely α^* in linear scale, whereas the second one reproduces more accurately α^* in a logarithmic scale and especially the two notches. This is due to the presence of the spectra ratio in the Itakura-Saito distance that emphasizes the small values of the spectra.

Finally, from our experience and as shown in Fig. 3, the maximum likelihood solution $\hat{a}_{RLS}^{ML}$ establishes a relevant compromise between $\hat{a}_{RLS}^{L_2}$ and $\hat{a}_{RLS}^{ISD}$ since it is smooth enough, whereas the two notches remain accurately described.

Quantitative comparisons have been conducted between the two practicable methods (when a^* is not known): the usual periodogram and the proposed method, i.e., the RLS solution with automatic ML hyperparameters. The obtained results are reported in Table I. They clearly show an improvement of about 40–50% for all the considered distances.

B. Window Selection

It has been shown that the ML technique allows the estimation of the regularization parameter. The problem of window selection is now addressed. Let us consider a set of K windows, i.e., K matrices Π_a^k for $k \in \mathbb{N}_K$. Index k becomes a new hyperparameter as well as λ and can be jointly estimated. The likelihood function (31) is now

$$\mathrm{CLL}(\lambda, k) = \log\det(\Sigma_y^k) + \log N \boldsymbol{y}^\dagger (\Sigma_y^k)^{-1} \boldsymbol{y}.$$

Maximization w.r.t. hyperparameters can be achieved in the same way as above for each value of $k \in \mathbb{N}_K$. The maximum maximorum can then be easily selected.

Numerous simulations have been performed. They are not reported here since they show similar results as the previous ones. However, it has been observed that the triangular window is the most often selected among Cauchy, inverse cosine, Hanning, Hamming, and triangle.

[4]Efficient algorithms are available in order to maximize the likelihood, such as gradient-based [28] or EM type [29]. They have not been implemented here as far as a mere feasibility study is concerned.

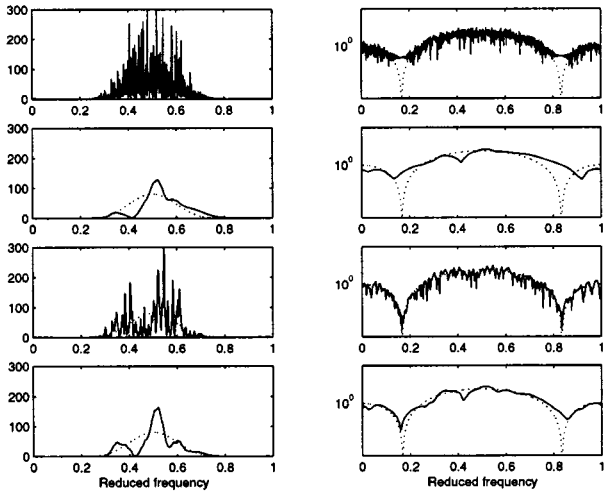

Fig. 3. Qualitative comparison. True spectra (dotted lines) and estimated ones (solid lines). The lhs column gives linear plots and the rhs column gives logarithmic plots. From top to bottom: Usual periodograms, $\hat{a}_{RLS}^{L_2}, \hat{a}_{RLS}^{ISD}$, and $\hat{a}_{RLS}^{ML}$.

TABLE I
QUANTITATIVE COMPARISON. THE FIRST LINE REFERS TO THE USUAL PERIODOGRAM, WHEREAS THE SECOND ONE REFERS TO THE RLS SOLUTION WITH ML HYPERPARAMETERS. THE THIRD LINE GIVES THE QUANTITATIVE IMPROVEMENT

	L_1	L_2	AIS	SIS
UP	0.766	1.14	751	750
RLS + ML	0.471	0.567	420	422
Gain	38.5%	50.3%	44.1%	43.8%

VI. CONCLUSION

In this paper, the usual nonparametric approach to spectral analysis has been revisited within the regularization framework. We have shown that usual and windowed periodograms could be obtained *via* the minimizer of regularized least squares criteria. In turn, penalized quadratic criteria are interpreted within the Bayesian framework so that periodograms are interpreted *via* Bayesian estimators. The corresponding prior is a zero-mean Gaussian process, fully specified by its correlation function. Particular attention is paid to the connection between correlation structure and window shape. With regard to *quadratic* regularization, the present study significantly deepens a recent contribution by Sacchi *et al.* [5], given that the latter addresses neither windowed periodograms, nor the continuous frequencial

setting. Extension to the *nonquadratic*[31] and two-dimensional (time–frequency) case would be of particular interest, and we are presently working on this issue.

Whereas the first part of our contribution provides interpretations of pre-existing tools for spectral analysis, new estimation schemes are derived in the second part: unsupervised hyperparameter and window selection. It is shown that maximum likelihood solutions are both formally achievable and practically useful.

APPENDIX A
PROOF OF PROPOSITIONS

A. Proof of Proposition 1

Several proofs are available, and the proposed one relies on variational principles [32]. Application of these principles to quadratic regularization of linear problem yields the functional (8)

$$-2\mathcal{W}_N^\dagger(\boldsymbol{y}-\mathcal{W}_N a)+2\lambda I_{L^2}a=0$$

where I_{L^2} stands for the identity application from L^2 onto itself, and $\mathcal{W}_N^\dagger$ stands for the adjoint application of $\mathcal{W}_N$ (see Appendix B). After elementary algebra, we find

$$(\mathcal{W}_N^\dagger\mathcal{W}_N+\lambda I_{L^2})a=\mathcal{W}_N^\dagger\boldsymbol{y}.$$

As shown in Appendix B, $\mathcal{W}_N\mathcal{W}_N^\dagger=I_N$; then, taking the FT and, next, the IFT gives

$$\hat{a}^\lambda(\nu)=(1+\lambda)^{-1}\mathcal{W}_N^\dagger\boldsymbol{y}=(1+\lambda)^{-1}\sum_{n=0}^{N-1}y_n e^{-2i\pi\nu n}.$$

B. Proof of Proposition 2

The minimizer of the RLS criterion (7) obviously is

$$\hat{\boldsymbol{a}}^\lambda=\left(W_{NP}^\dagger W_{NP}+\lambda I_P\right)^{-1}W_{NP}^\dagger\boldsymbol{y}.$$

Refer to Appendix B for a detailed calculus required to analyze the normal matrix $(W_{NP}^\dagger W_{NP}+\lambda I_P)$. $W_{NP}^\dagger W_{NP}$ and I_P are circulant matrices, and this property also holds for their sum, which hence is diagonal in the Fourier basis. Elementary algebra leads to

$$\begin{aligned}\hat{\boldsymbol{a}}^\lambda=&F_P\begin{bmatrix}(1+\lambda)^{-1}I_N & O_{N,P-N}\\ O_{P-N,N} & \lambda^{-1}I_{P-N}\end{bmatrix}\begin{bmatrix}I_N\\ O_{P-N,N}\end{bmatrix}\boldsymbol{y}\\ =&(1+\lambda)^{-1}F_P\tilde{\boldsymbol{y}}_P.\end{aligned}$$

C. Proof of Proposition 3

The proof is founded on a time domain version of the criterion (9), resulting from application of the Plancherel–Parseval theorem to the successive derivatives of a

$$\int_0^1\left|\frac{d^q a}{d\nu^q}(\nu)\right|^2 d\nu=\sum_{n\in\mathbb{Z}}(2\pi n)^{2q}|z_n|^2$$

where $z_n=\int_0^1 a(\nu)e^{2i\pi\nu n}d\nu$. Summation w.r.t. q and inversion of summation w.r.t. q and w.r.t. n gives

$$R_Q(a)=\sum_{n\in\mathbb{Z}}e_n|z_n|^2$$

where the weighting coefficients e_p fulfill (12). Hence, the time domain counterpart of criterion (4) reads

$$\mathcal{Q}_s(a)=(\boldsymbol{y}-\boldsymbol{z})^\dagger(\boldsymbol{y}-\boldsymbol{z})+\lambda\sum_{n\in\mathbb{Z}}e_n|z_n|^2.$$

Thanks to separability, the solution is easily derived: $\hat{z}_n^\omega=(1+\lambda e_n)^{-1}y_n$ if $n\in\mathbb{N}_N$ and $\hat{z}_n^\omega=0$ elsewhere. a^ω is the Fourier transform of the sequence $\{\hat{z}_n^\omega\}_{n\in\mathbb{Z}}$

$$\hat{a}^\omega(\nu)=\sum_{n=0}^{N-1}\hat{z}_n^\omega e^{-2i\pi\nu n}.$$

D. Proof of Proposition 4

Elementary linear algebra provides the minimizer of (13)

$$\hat{\boldsymbol{a}}^\omega=\left(W_{NP}^\dagger W_{NP}+\lambda\Pi_a\right)^{-1}W_{NP}^\dagger\boldsymbol{y}.$$

Accounting for its circular structure, the Fourier basis diagonalizes Π_a

$$\Pi_a=F_P\Lambda_\Pi F_P^\dagger$$

where Λ_Π is the diagonal matrix of the eigenvalues $e_0,\ldots,e_{P-1}$ of Π_a. Hence

$$\hat{\boldsymbol{a}}^\omega=F_P(I_P+\lambda\Lambda_\Pi)\tilde{\boldsymbol{y}}_P$$

and we easily find

$$\hat{\boldsymbol{a}}^\omega=F_P\breve{\boldsymbol{y}}$$

with $\breve{y}_p=\omega_p\tilde{y}_p$ for $p\in\mathbb{N}_P$, i.e., the data vector windowed by

$$\omega_n=(1+\lambda e_n)^{-1}.$$

E. Proof of Proposition 5

Let $\nu_0\in[0,1]$ and $a_0=a(\nu_0)$. Thanks to the linearity of the model (3) and thanks to the Gaussian assumption for a and $\boldsymbol{b}$, the joint law of $(a_0,\boldsymbol{y})$ is also Gaussian. Hence, the random variable $(a_0|\boldsymbol{y})$ is clearly Gaussian, and it is well known that its mean reads

$$\mathrm{E}[a_0|\boldsymbol{y}]=R_{a_0y}R_y^{-1}\boldsymbol{y}$$

where $R_{a_0y}=\mathrm{E}a_0y^\dagger$, and $R_{\boldsymbol{y}}=\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^\dagger]$. Elementary algebra and independence of a and $\boldsymbol{b}$ yield

$$\begin{aligned}R_{a_0y_n}=&\int_0^1\mathrm{E}[a(\nu_0)a(\nu)^*]e^{-2i\pi\nu n}d\nu+\mathrm{E}[a(\nu_0)b_n]\\ =&\mathring{\gamma}_a(\nu)e^{-2i\pi\nu_0 n}.\end{aligned}$$

Moreover, under the previously mentioned assumptions, the generic entry R_{mn} for R_y is

$$\begin{aligned}R_{mn}=\mathrm{E}[y_m y_n^*]=&\int\int_0^1\mathrm{E}[a(\nu)a(\nu')^*]\\ &\times\exp[2i\pi(\nu n-\nu' m)]d\nu'd\nu+r_b\delta_{n-m}\\ =&(\mathring{\gamma}_a(\nu)+r_b)\delta_{n-m}\end{aligned}\tag{32}$$

where δ_n stands for the Kronecker sequence. Therefore, R_y is a diagonal matrix with elements $\overset{\circ}{\gamma}_a(\nu) + r_b$. Hence

$$\hat{a}_0 = \sum_{n=0}^{N-1} \left[1 + \lambda\, \overset{\circ}{c}_a(n)^{-1}\right]^{-1} y_n e^{-2i\pi\nu_0 n}$$

with $\lambda = r_b/r_a$.

APPENDIX B
TECHNICAL RESULTS

This appendix collects several useful properties of Fourier operators. In particular, special attention is paid to W_{NP} and $\mathcal{W}_N$. Some of the stated properties are classical. We have reported them in order to make our notations and normalization conventions explicit. The other properties are less usual, but all of them have straightforward proofs.

A. Discrete Case

Structure of F_P: In the case of $N = P$, the matrix W_{NP} identifies with the square matrix $F_P^\dagger$, where F_P performs the discrete FT for vectors of size P. We have the well-known orthogonality relations $F_P^\dagger F_P = F_P F_P^\dagger = I_P$ and $F_P^t = F_P$.

Structure of W_{NP}: The matrix W_{NP} evaluates the FT on a discrete grid of P points for sequences of N points ($P \geq N$). Straightforward expansion of the product provides

$$W_{NP} F_P = [\, I_N \quad O_{N,P-N}\,]. \tag{33}$$

As a consequence, we obtain

$$W_{NP}^\dagger \boldsymbol{y} = F_P \begin{bmatrix} I_N \\ O_{P-N,N} \end{bmatrix} \boldsymbol{y} = F_P \tilde{\boldsymbol{y}}_P \tag{34}$$

where $\tilde{\boldsymbol{y}}_P$ is the zero-padded version of $\boldsymbol{y}$ up to length P.

Structure of $W_{NP}^\dagger W_{NP}$: The matrix $W_{NP} W_{NP}^\dagger$ has a very simple structure since, for $P \geq N$: $W_{NP} W_{NP}^\dagger = I_N$. Otherwise, $W_{NP}^\dagger W_{NP}$ is a non-negative, Hermitian, $P \times P$ circulant matrix. Circularity results from digonalization in the Fourier basis F_P

$$W_{NP}^\dagger W_{NP} = F_P \Lambda F_P^\dagger$$

and from (33)

$$\Lambda = \begin{bmatrix} I_N & O_{N,P-N} \\ O_{P-N,N} & O_{P-N,P-N} \end{bmatrix}.$$

As a consequence, $W_{NP}^\dagger W_{NP}$ has only two eigenvalues (1 and 0) of respective order N and $P - N$. Such a structure is useful in the proof of Propositions 2 and 4 in Appendix A.

B. Continuous Case

1) The $\mathcal{W}_N$ Operator: The linear application $\mathcal{W}_N : a \in L^2 \longrightarrow \boldsymbol{z} \in \mathbb{C}^N$ is defined by $z_n = \int_0^1 a(\nu) e^{2i\pi\nu n} d\nu$ for $n \in \mathbb{N}_N$. The adjoint operator $\mathcal{W}_N^\dagger : \boldsymbol{z} \in \mathbb{C}^N \longrightarrow a = \mathcal{W}_N^\dagger \boldsymbol{z}$ is the linear operator such that

$$\forall\, a \in L^2, \forall\, \boldsymbol{z} \in \mathbb{C}^N \quad \langle \mathcal{W}_N a, \boldsymbol{z} \rangle_{\mathbb{C}^N} = \langle a, \mathcal{W}_N^\dagger \boldsymbol{z} \rangle_{L^2}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{C}^N}$ and $\langle \cdot, \cdot \rangle_{L^2}$ stand for the standard inner product in $\mathbb{C}^N$ and L^2, respectively. It is given by

$$a(\nu) = \mathcal{W}^\dagger \boldsymbol{z} = \sum_{n=0}^{N-1} z_n e^{-2i\pi\nu n}.$$

This can be justified as follows: By inverting the order of the finite sum $\sum_0^{N-1}$ and the definite integral $\int_0^1$, we get

$$\langle W_N a, \boldsymbol{z} \rangle_{\mathbb{C}^N} = \int_0^1 a(\nu) \sum_{n=0}^{N-1} z_n^* e^{2i\pi\nu n} = \langle a, \mathcal{W}_N^\dagger \boldsymbol{z} \rangle_{L^2}.$$

Finally, elementary algebra shows that the composed application $\mathcal{W}_N \mathcal{W}_N^\dagger$ is the identity application from $\mathbb{C}^N$ onto itself.

2) Technical Results for the Example in Section IV-B2:

a) Fourier Series (22): The proof of (22) consists of three steps. The first one relies on the Fourier relationship between Cauchy and Laplace functions

$$\frac{2\beta}{\beta^2 + 4\pi^2 t^2} = \int_{\mathbb{R}} e^{-\beta|f|} e^{-2j\pi t f} df, \quad t \in \mathbb{R}.$$

The second step is founded on discrete time $t = n \in Z$ and expansion in a series of integrals

$$\begin{aligned} \frac{2\beta}{\beta^2 + 4\pi^2 n^2} &= \int_{\mathbb{R}} e^{-\beta|f|} e^{-2j\pi n f} df \\ &= \sum_{p \in \mathbb{Z}} \int_0^1 e^{-\beta|\nu - p|} e^{-2j\pi n \nu} d\nu \\ &= \int_0^1 \sum_{p \in \mathbb{Z}} e^{-\beta|\nu - p|} e^{-2j\pi n \nu} d\nu \end{aligned}$$

since the invoked series are convergent. The last step is a simple geometric series calculus

$$\sum_{p \in \mathbb{Z}} e^{-\beta|\nu - p|} = \frac{\cosh \beta\,(\nu - 1/2)}{\sinh \beta/2}, \quad \nu \in [0, 1]$$

which is easily obtained by rewriting the series as the sum of a series for $p \in \mathbb{Z}_-$ (i.e., $p \leq \nu$) and a series for $p \in \mathbb{Z}_+^*$ (i.e., $p \geq \nu$).

b) Conditional Process: Let us note $\nu, \nu' \in [0, 1], \nu > \nu'$. The partitioned vector $\bar{\boldsymbol{a}} = [a(\nu), a(\nu'), a(1)]^t = [\tilde{\boldsymbol{a}} | a_1]^t$ is clearly a zero-mean Gaussian vector with covariance

$$R_{\bar{\boldsymbol{a}}} = \begin{bmatrix} \gamma_a(0) & \gamma_a(\nu - \nu') & \gamma_a(\nu) \\ \gamma_a(\nu - \nu') & \gamma_a(0) & \gamma_a(\nu') \\ \gamma_a(\nu) & \gamma_a(\nu') & \gamma_a(0) \end{bmatrix}.$$

According to the conditional covariance matrix formula $R_{\tilde{\boldsymbol{a}}|a_1} = R_{\tilde{\boldsymbol{a}}} - R_{\tilde{\boldsymbol{a}}a_1}^t R_{a_1}^{-1} R_{\tilde{\boldsymbol{a}}a_1}$, we immediately get (24). Accounting for the explicit expression for $\gamma_a(\nu)$ given by (23), simple expansion of hyperbolic functions yields (25).

c) Law of Increments: We have $\nu_1, \nu_2, \nu_1', \nu_2' \in [0, 1]$, $\nu_1 < \nu_2 < \nu_1' < \nu_2'$. Let us introduce the collection of the four values $\underline{\boldsymbol{a}} = [a(\nu_1), a(\nu_2), a(\nu_1'), a(\nu_2')]$, which is clearly a zero-mean and Gaussian vector with covariance $R_{\underline{\boldsymbol{a}}}$. The increment vector $\boldsymbol{i} = [a(\nu_2) - a(\nu_1), a(\nu_2') - a(\nu_1')] \in \mathbb{C}^2$ is a linear transform of the vector $\underline{\boldsymbol{a}}$: $\boldsymbol{i} = H\underline{\boldsymbol{a}}$ with increment covariance $R_{\boldsymbol{i}}$

$$H = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad R_{\boldsymbol{i}} = H R_{\underline{\boldsymbol{a}}} H^t = \begin{bmatrix} r_i & \rho \\ \rho & r_i' \end{bmatrix}$$

with $r_i = 2(\gamma_a(0) - \gamma_a(\nu_2 - \nu_1))$, $r_i' = 2(\gamma_a(0) - \gamma_a(\nu_2' - \nu_1'))$, and $\rho = \gamma_a(\nu_2 - \nu_2') + \gamma_a(\nu_1 - \nu_1') - \gamma_a(\nu_1 - \nu_2') - \gamma_a(\nu_2 - \nu_1')$. Finally, Taylor development at $\alpha_0 = 0$ yields $r_i = (\nu_2 - \nu_1)(1 - (\nu_2 - \nu_1))/2\alpha_1$, $r_i' = (\nu_2' - \nu_1')(1 - (\nu_2' - \nu_1'))/2\alpha_1$, and $\rho = (\nu_2 - \nu_1)(\nu_2' - \nu_1')/\alpha_1$ and proves(26).

ACKNOWLEDGMENT

The first author is particularly thankful to Alain, Naomi, Philippe, and Denise for committed support and coaching.

REFERENCES

[1] E. R. Robinson, "A historical perspective of spectrum estimation," *Proc. IEEE*, vol. 9, pp. 885–907, Sept. 1982.
[2] S. M. Kay and S. L. Marple, "Spectrum analysis—a modern perpective," *Proc. IEEE*, vol. 69, pp. 1380–1419, Nov. 1981.
[3] S. L. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[4] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[5] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transforms," *IEEE Trans. Signal Processing*, vol. 46, pp. 31–38, January 1998.
[6] M. D. Sacchi and T. J. Ulrych, "Estimation of the discrete Fourier transform, a linear inversion approach," *Geophysics*, vol. 61, no. 4, pp. 1128–1136, 1996.
[7] H. W. Sorenson, *Parameter Estimation*. New York: Marcel Dekker, 1980, vol. 9.
[8] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.
[9] M. Z. Nashed and G. Wahba, "Generalized inverses in reproducing kernel spaces: An approach toregularization of linear operators equations," *SIAM J. Math. Anal.*, vol. 5, pp. 974–987, 1974.
[10] G. Demoment, "Image reconstruction and restoration: Overview of common estimationstructure and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.
[11] G. Kitagawa and W. Gersch, "A smoothness priors long AR model method for spectral estimation," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 57–65, Jan. 1985.
[12] G. Wahba, "Automatic smoothing of the log periodogram," *J. Amer. Statist. Assoc., Theory Methods Section*, vol. 75, no. 369, pp. 122–132, Mar. 1980.
[13] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, J. Berger, S. Fienberg, J. Gani, K. Krickeberg, and B. Singer, Eds. New York: Springer-Verlag, 1988, vol. 48.
[14] F. Dublanchet, J. Idier, and P. Duvaut, "Direction-of-arrival and frequency estimation using Poisson-Gaussian modeling," in *Proc. IEEE ICASSP*, Munich, Germany, 1997, pp. 3501–3504.
[15] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, Jan. 1978.
[16] A. Bertin, *Espaces de Hilbert*. Paris, France: l'ENSTA, 1993.
[17] H. Cramér and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. New York: Wiley, 1967.
[18] P. Brémaud, "Markov Chains. Gibbs fields, Monte Carlo Simulation, and Queues," in *Texts in Applied Mathematics 31*. New York: Spinger, 1999.
[19] J. M. F. Moura and G. Sauraj, "Gauss-Markov random fields (GMrf) with continuous indices," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1560–1573, Sept. 1997.
[20] E. Wong, "Stochastic processes in information and dynamical systems," in *Series in Systems Science*. New York: McGraw-Hil, 1971.
[21] R. N. Bhattacharya and E. C. Waymire, *Stochastic Processes with Applications*. New York: Wiley, 1990.
[22] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, May 1979.
[23] D. M. Titterington, "Common structure of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, no. 2, pp. 141–170, 1985.
[24] P. Hall and D. M. Titterington, "Common structure of techniques for choosing smoothing parameter in regression problems," *J. R. Statist. Soc. B*, vol. 49, no. 2, pp. 184–198, 1987.
[25] A. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 326–339, Apr. 1991.
[26] N. Fortier, G. Demoment, and Y. Goussard, "Comparison of GCV and ML methods of determining parameters in image restoration by regularization," *J. Vis. Commun. Image Repres.*, vol. 4, pp. 157–170, 1993.
[27] J.-F. Giovannelli, G. Demoment, and A. Herment, "A Bayesian method for long AR spectral estimation: a comparative study," *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.*, vol. 43, pp. 220–223, Mar. 1996.
[28] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
[29] R. Shumway and D. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, pp. 253–264, 1982.
[30] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Process.*, vol. 18, no. 4, pp. 349–369, Dec. 1989.
[31] P. Ciuciu, J. Idier, and J.-F. Giovannelli, "Markovian high resolution spectral analysis," in *Proc. IEEE ICASSP*, Phoenix, AZ, 1999, pp. 1601–1604.
[32] D. G. Luenberger, *Optimization by Vector Space Methods*, 1 ed. New York: John Wiley, 1969.

Jean-François Giovannelli was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications, Paris, France, in 1990. He received the Doctoral degree in physics from Université de Paris-Sud, Orsay, France, in 1995.

He is presently an assistant professor with the Département de Physique, the Laboratoire des Signaux et Systémes, Université de Paris-Sud. He is interested in regularization methods for inverse problems in signal and image processing, mainly in spectral characterization. His application fields essentially concern medical imaging.

Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from the École Supérieure d'Électricité, Gif-sur-Yvette, France, in 1988 and the Ph.D. degree in physics from the Université de Paris-Sud, Orsay, France, in 1991.

Since 1991, he has been with the Laboratoire des Signaux et Systèmes, Centre National de la Recherche Scientifique, Gif-sur-Yvette. His major scientific interest is in probabilistic approaches to inverse problems for signal and image processing.