Unsupervised Frequency Tracking Beyond the Nyquist Frequency Using Markov Chains

Jean-François Giovannelli, Jérôme Idier, Rédha Boubertakh, and Alain Herment

Abstract—This paper deals with the estimation of a sequence of frequencies from a corresponding sequence of signals. This problem arises in fields such as Doppler imaging, where its specificity is twofold. First, only short noisy data records are available (typically four sample long), and experimental constraints may cause spectral aliasing so that measurements provide unreliable, ambiguous information. Second, the frequency sequence is smooth. Here, this information is accounted for by a Markov model, and application of the Bayes rule yields the *a posteriori* density. The maximum *a posteriori* is computed by a combination of Viterbi and descent procedures. One of the major features of the method is that it is entirely unsupervised. Adjusting the hyperparameters that balance data-based and *prior*-based information is done automatically by maximum likelihood (ML) using an expectation-maximization (EM)-based gradient algorithm. We compared the proposed estimate to a reference one and found that it performed better: Variance was greatly reduced, and tracking was correct, even beyond the Nyquist frequency.

Index Terms—Aliasing inversion, Bayesian statistic, EM algorithm, forward-backward procedure, frequency tracking, hyperparameter estimation, maximum a posteriori, maximum likelihood, meteorological Doppler radar, regularization, ultrasonic Doppler velocimetry, Viterbi algorithm.

I. INTRODUCTION

FREQUENCY tracking (or mean frequency tracking) is currently of interest [1]–[6], especially in fields such as the ultrasonic characterization of biological tissues, synthetic aperture radar, and speech processing. Our main interest is its use in Doppler imaging (radars [7], ultrasound blood flow mapping [8]–[10]). There are two main features in this area.

1) One is that only short noisy data records are available (typically four sample long), and they are in a vectorial form. Moreover, the constraints on the sampling frequency may cause spectral aliasing so that measurements provide small amounts of ambiguous information.
2) The second is that there is information on the smoothness of the sought frequency sequence. This *a priori* information is the foundation of the proposed construction. It allows robust tracking, even beyond the Nyquist limit.

The most popular methods used for spectral characterization rely on periodogram and empirical correlations. The mean frequency is usually estimated by computing the mean frequency of the periodogram [8] over the standardized frequency range $\nu \in (-0.5, +0.5]$. Another popular estimate is proportional to the phase of the first empirical correlation lag [11], [12]. It is also provided by a first-order autoregression in a least squares framework [13], but better accuracy is obtained by using all the available estimated correlation lags in a Taylor series expansion of the correlation function [12], [14]. The resulting estimate is also the mean frequency of the periodogram. However, the estimated parameters vary greatly, particularly when short data records are used. Moreover, the estimated frequency approaches zero when the true frequency becomes near the Nyquist frequency $\nu \simeq \pm 0.5$ (due to the periodogram 1-periodicity) [8]. To reduce this bias, [15] uses the maximum of the periodogram instead of its mean (and yields a maximum likelihood (ML) estimate; see Section III-A and [16, p. 410]), and [8] iteratively shifts the frequency of the data. This results in greater variance so that no frequency tracking remains possible beyond $\nu = \pm 0.5$.

Thus, all the current methods have two drawbacks. First, the tracking problem is tackled by a (necessary suboptimal) two-step procedure:

1) Estimate frequencies in the aliased band $(-0.5, +0.5]$.
2) Detect and inverse aliasing.

Second, they are clearly based on empirical second-order statistics that perform poorly with short data records independently processed. Unfortunately, the inverse aliasing in step 2 often fails due to the great variations in the estimated aliased frequencies of step 1. This is usually compensated for by post-smoothing the aliased frequency sequence. This provides spatial continuity but affects the aliased frequency discontinuities, therefore limiting the capacity to detect aliasing. The proposed method copes with the great variation and aliasing in a single step; it models the whole data set (by noisy cisoids) and the smoothness of the frequency sequence (by a Markov random walk) in the regularization/Bayesian framework. It then becomes possible to smooth frequency sequence and invert aliasing at the same time, avoiding the pitfalls of chaining these operations.

We have found several papers [3], [17], [18] that adopt such a framework, and this study provides four additional features.

1) First, it deals with vectorial data records as they occur in Doppler imaging (see Section II).
2) Second, it enables tracking beyond the Nyquist frequency, whereas others have not investigated this problem.

3) Third, exact frequency likelihood functions are computed, whereas [17] uses a detection step, and [3] uses an approximation.

4) Last, the tracking method is entirely unsupervised with a maximum likelihood hyperparameter estimation. This is not a straightforward task in the context of frequency tracking since the nonlinear character of the data as functions of frequencies prevents the explicit handling of the likelihood function of the hyperparameter. We have developed an EM-like gradient procedure, inspired by [19]–[21]. It can be derived only after discretizing the frequencies on a finite grid.

The paper is organized as follows. The notation, signal model, and assumptions are defined in Section II. Section III contains the proposed regularized method, and Section IV gives a discrete approximation. Section V is devoted to the estimation of hyperparameters. The performance of the proposed method is demonstrated by the computer simulations in Section VI, whereas Section VII gives our conclusion and describes possible extensions.

II. STATEMENT, NOTATIONS AND ASSUMPTIONS

In Doppler imaging, the signals to be analyzed occur as a set of complex signals $\mathcal{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$ juxtaposed spatially in T range bins [22], [23]. The data record $\boldsymbol{y}_t = [y_t(1), \ldots, y_t(N)]^t$ ("t" denotes the matrix transpose) is extracted from a cisoid in additive complex noise. The amplitude and the frequency of the cisoid are $a_t \in \mathbb{C}$ and $\nu_t \in \mathbb{R}$:

$$\boldsymbol{y}_t = a_t \boldsymbol{z}(\nu_t) + \boldsymbol{b}_t = a_t [1, \ldots, e^{j2\pi\nu_t(N-1)}]^t + \boldsymbol{b}_t. \quad (1)$$

The vectors $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_T]^t$ and $\boldsymbol{a} = [a_1, \ldots, a_T]^t$ collect the frequencies and corresponding amplitudes. Finally, the true parameters are denoted with a star. This paper builds a robust estimate $\widehat{\boldsymbol{\nu}}$ for $\boldsymbol{\nu}^\star$ on the basis of data set $\mathcal{Y}$ (see Fig. 1 for a simulated example).

Remark 1: Model (1) is frequently used for spectral problems; it has three main features. First, while it is linear w.r.t. a_t, it is not so w.r.t. ν_t; the problem to be solved is nonlinear. Second, $\boldsymbol{z}(\nu_t)$ is a 1-periodic function w.r.t. ν_t, and this causes the difficulties of aliasing, frequency ambiguity, likelihood periodicity, etc. Last, this periodicity is also the keystone of the paper; aliasing is inverted, using a coherent statistical approach that takes periodicity into consideration.

The following definition of periodicity is used throughout the paper.

Definition 1: Let $A \subset \mathbb{R}^T$ and $\varphi : A \to \mathbb{R}$. Let us note $\mathbf{1} = [1, \ldots, 1]^t \in \mathbb{R}^T$. φ is said to be

• separately-1-periodic (S1P) if $\forall \boldsymbol{u} \in A$, $\forall \boldsymbol{k} \in \mathbb{Z}^T$ (such that $\boldsymbol{u} + \boldsymbol{k} \in A$): $\varphi(\boldsymbol{u}) = \varphi(\boldsymbol{u} + \boldsymbol{k})$;
• globally-1-periodic (G1P) if $\forall \boldsymbol{u} \in \mathbb{R}^T$, $\forall k_0 \in \mathbb{Z}$ (such that $\boldsymbol{u} + k_0 \mathbf{1} \in A$): $\varphi(\boldsymbol{u}) = \varphi(\boldsymbol{u} + k_0 \mathbf{1})$.

The proposed estimation method deals with periodicity and aliasing inversion thanks to the following assumptions. They are stated for the sake of simplicity and calculation tractability as well as coherence with the applications under the scope of this paper.
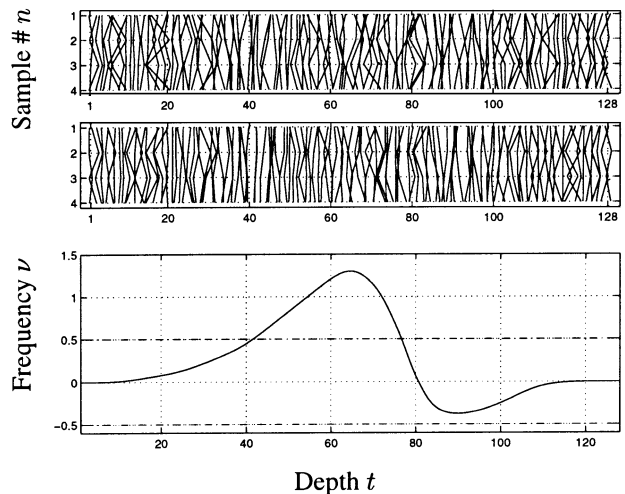

Fig. 1.   Simulated observations over $T = 128$ range bins with $N = 4$ samples per bin. From top to bottom: real parts, imaginary parts of the data $\boldsymbol{y}_t$, and the true frequency sequence $\nu_t^\star$.

• Parameter dependence.
– H_1: $\boldsymbol{a}$, $\boldsymbol{\nu}$ and the $\boldsymbol{b}_t$ are independent.
• Law for measurement and modeling noise $\boldsymbol{b}_t$.
– H_2^a: Each $\boldsymbol{b}_t$ is $\mathcal{N}(r_b I_N)$.
– H_2^b: The sequence of $\boldsymbol{b}_t$ is itself white.
• Law for parameters $\boldsymbol{a}$ and $\boldsymbol{\nu}$.
– H_3^a: $\boldsymbol{a}$ is $\mathcal{N}(r_a I_T)$, i.e., white.
– H_3^b: $\boldsymbol{\nu}$ is, on the contrary, correlated: $\mathcal{N}(R_\nu)$

where $\mathcal{N}(R)$ stands for a complex zero-mean Gaussian vector with covariance R, and $I_P, P \in \mathbb{N}^*$ denotes the $P \times P$ identity matrix.

The first assumption H_1 is quite natural since no information is available about the relative fluctuations of noise and objects. The assumptions H_2^a, and H_2^b are also natural since no correlation structure is expected in noise. Similarly, we have no information about the variation of the amplitude sequence; therefore, an independent law is used. A Gaussian law is preferred (H_3^a) to make the calculations tractable. Contrarily, the smoothness of the frequency sequence is modeled as a positive correlation. A Markovian structure (specified below) is a simple, useful way to account for it. Several choices are available, but the Gaussian one is also stated for the sake of simplicity (H_3^b).

III. PROPOSED METHOD

A. Likelihood

Assumption H_2^a yields a parametric structure for each likelihood function $f(\boldsymbol{y}_t \,|\, \nu_t, a_t)$:

$$f(\boldsymbol{y}_t \,|\, \nu_t, a_t) = (\pi r_b)^{-N} \exp\left[-\frac{1}{r_b} CLL(\nu_t, a_t)\right]$$

involving the opposite of the logarithm of the likelihood function (up to constant terms) i.e., the Co-Log-Likelihood (CLL):

$$\mathrm{CLL}(\nu_t, a_t) = [\boldsymbol{y}_t - a_t \boldsymbol{z}(\nu_t)]^\dagger [\boldsymbol{y}_t - a_t \boldsymbol{z}(\nu_t)].$$

From a deterministic standpoint, $\mathrm{CLL}(\nu_t, a_t)$ is clearly the least squares (LS) estimation criterion.

Considering the whole frequency vector $\boldsymbol{\nu}$ and the whole data set $\mathcal{Y}$, assumption H_2^b yields

$$f(\mathcal{Y} \,|\, \boldsymbol{\nu}, \boldsymbol{a}) = (\pi r_b)^{-NT} \exp\left[-\frac{1}{r_b} CLL(\boldsymbol{\nu}, \boldsymbol{a})\right] \qquad (2)$$

where the global CLL is a global LS criterion

$$CLL(\boldsymbol{\nu}, \boldsymbol{a}) = \sum_{t=1}^{T} CLL(\nu_t, a_t).$$

Remark 2: According to Definition 1, the likelihood function $CLL(\cdot, \boldsymbol{a})$ is S1P for all $\boldsymbol{a} \in \mathbb{C}^T$. Therefore, two configurations $\boldsymbol{\nu}$ and $\boldsymbol{\nu} + \boldsymbol{k}$ ($\boldsymbol{k} \in \mathbb{Z}^T$) for the frequency sequence are equilikelihood. As a consequence, an ML approach suffers from T independent frequency ambiguities.

B. Amplitude Law and Marginalization

The parameters of interest are the frequencies, whereas the amplitudes are nuisance parameters. These are integrated out of the problem in the usual Bayesian approach.

Given separability assumption H_1, one has $f(\boldsymbol{\nu}, \boldsymbol{a}) = f(\boldsymbol{\nu})f(\boldsymbol{a})$, and the marginal law can easily be deduced:

$$f(\mathcal{Y}, \boldsymbol{\nu}) = f(\boldsymbol{\nu}) \int_{\boldsymbol{a}} f(\mathcal{Y} \,|\, \boldsymbol{a}, \boldsymbol{\nu}) f(\boldsymbol{a}) d\boldsymbol{a} = f(\boldsymbol{\nu}) f(\mathcal{Y} \,|\, \boldsymbol{\nu}).$$

The joint law for the amplitudes is separable according to assumption H_3^a. Since likelihood (2) is also separable, marginalization can be performed independently.

$$f(\mathcal{Y} \,|\, \boldsymbol{\nu}) = \prod_{t=1}^{T} \int_{a_t} f(\boldsymbol{y}_t \,|\, \nu_t, a_t) f(a_t) da_t. = \prod_{t=1}^{T} f(\boldsymbol{y}_t \,|\, \nu_t). \tag{3}$$

The Gaussian amplitude assumption H_3^a results in analytic derivations and yield the marginal likelihood for the data $\boldsymbol{y}_t$, given ν_t, which is zero mean Gaussian vector. Its covariance R_t is given in Appendix A-B as well as its determinant (23) and its inverse (24). $f(\boldsymbol{y}_t \,|\, \nu_t)$ then reads

$$f(\boldsymbol{y}_t \,|\, \nu_t) = \beta \exp\left[-\gamma_t\right] \exp\left[\alpha P_t(\nu_t)\right] \qquad (4)$$

with $\alpha = N r_a / (r_b(N r_a + r_b))$, $\beta = \pi^{-N} r_b^{1-N} / (N r_a + r_b)$, $\gamma_t = \boldsymbol{y}_t^{\dagger} \boldsymbol{y}_t / r_b$, and P_t is the periodogram of vector $\boldsymbol{y}_t$

$$P_t(\nu_t) = \frac{1}{N} \left| \sum_{n=1}^{N} y_t(n) e^{-2j\pi\nu_t n} \right|^2.$$

The joint law for the whole data set given the frequency sequence is obtained by the product (3)

$$f(\mathcal{Y} \,|\, \boldsymbol{\nu}) = \beta^T \exp\left[-\gamma\right] \exp\left[-\alpha CLML(\boldsymbol{\nu})\right] \qquad (5)$$

where γ is the sum of the γ_t for $t \in \mathbb{N}_T^* = \{1, \ldots, T\}$, and where CLML is the co-log-marginal-likelihood

$$CLML(\boldsymbol{\nu}) = -\sum_{t=1}^{T} P_t(\nu_t) \qquad (6)$$

which is the opposite of the sum of the periodograms of data $\boldsymbol{y}_t$ at frequency ν_t in gate t.

Remark 3: This remark is the marginal counterpart of Remark 2. As well as $CLL(\cdot, \boldsymbol{a})$, $CLML(\cdot)$ is S1P. There are still many ambiguities as in the nonmarginal case. This was expected since no information about the frequency sequence has been accounted for in $CLML(\boldsymbol{\nu})$ w.r.t. $CLL(\boldsymbol{\nu}, \boldsymbol{a})$. In contrast, periodicity will be eliminated in the next subsection by accounting for the frequency sequence smoothness.

C. Prior Law for Frequency Sequence

Unlike amplitudes, the frequency sequence is smooth. A Markovian structure accurately accounts for this information, and there are many algorithms suited to computing this structure. The choice of the family law is not crucial for using these algorithms, but we have used the Gaussian family

$$f(\nu_{t+1} \,|\, \nu_t) = (2\pi r_\nu)^{-1/2} \exp\left[-\frac{1}{2r_\nu}(\nu_{t+1} - \nu_t)^2\right].$$

The complete law for the chain also involves the initial state. It is assumed to be uniformly distributed over a symmetric set $\mathbb{K}$ defined by $K \in \mathbb{N}^*$: $\mathbb{K} = [-K/2; +K/2]$. Therefore, $f(\nu_1) = (1/K)\mathbb{1}_K^0(\nu_1)$, where $\mathbb{1}_K^0$ is 1 in $\mathbb{K}$ and 0 outside.

The recursive conditioning rule immediately yields

$$f(\boldsymbol{\nu}) = (2\pi r_\nu)^{-(T-1)/2} \exp\left[-\frac{1}{2r_\nu} CLP(\boldsymbol{\nu})\right] \qquad (7)$$

where $CLP(\boldsymbol{\nu})$ is the co-log-prior

$$CLP(\boldsymbol{\nu}) = \widetilde{K} \mathbb{1}_K^\infty(\nu_1) + \sum_{t=1}^{T-1} (\nu_{t+1} - \nu_t)^2. \qquad (8)$$

$\widetilde{K} = 2r_\nu \log K$ and $\mathbb{1}_K^\infty$ is 1 in $\mathbb{K}$ and $+\infty$ outside. In the deterministic framework, $CLP(\boldsymbol{\nu})$ is a quadratic norm for the first-order differences, namely, a regularization term [24]–[26].

D. Posterior Law

Fusion of *prior*-based and data-based information is achieved by the Bayes rule, which provides the *a posteriori* density for $\boldsymbol{\nu}$

$$f(\boldsymbol{\nu} \,|\, \mathcal{Y}) = \frac{f(\mathcal{Y} \,|\, \boldsymbol{\nu}) f(\boldsymbol{\nu})}{f(\mathcal{Y})}.$$

The marginal law $f(\mathcal{Y})$ for the whole data set $\mathcal{Y}$ is not analytically tractable, essentially due to the nonlinearity of the periodogram w.r.t. ν_t and the correlated structure of $\boldsymbol{\nu}$. Fortunately, this p.d.f. does not depend on $\boldsymbol{\nu}$; therefore, the *a posteriori* density remains explicit up to a positive constant. *Prior* structure of (7) and (8) and likelihood structure of (5) and (6) immediately yield the *posterior* law

$$f(\boldsymbol{\nu} \,|\, \mathcal{Y}) \propto \exp\left[-\alpha CLPL(\boldsymbol{\nu})\right] \qquad (9)$$

where the co-log-posterior-likelihood function (CLPL) reads

$$CLPL(\boldsymbol{\nu}) = -\sum_{t=1}^{T} P_t(\nu_t) + \lambda \sum_{t=1}^{T-1} (\nu_{t+1} - \nu_t)^2 + \mathbb{1}_K^\infty(\nu_1) \tag{10}$$

where $\lambda = 1/2\alpha r_\nu$, up to irrelevant constants. In the deterministic framework, CLPL is a regularized least squares (RLS) criterion. It has three terms: one measures fidelity to the data, the

second measures fidelity to the *prior* smoothness, and the third enforces the first frequency $\nu_1 \in \mathbb{K}$. The regularization parameter λ (depending on hyperparameters $r = [r_a, r_b, r_\nu]$) balances the compromise between *prior*-based and data-based information.

E. Point Estimate

As a point estimate, a popular choice is the maximum *a posteriori* (MAP) i.e., the maximizer of the *posterior* law of (9) or the minimizer of the RLS criterion (10):

$$\widehat{\boldsymbol{\nu}}^{\mathrm{MAP}} = \arg\max_{\boldsymbol{\nu} \in \mathbb{R}} f(\boldsymbol{\nu} \mid \mathcal{Y}) = \arg\min_{\boldsymbol{\nu} \in \mathbb{R}} \mathrm{CLPL}(\boldsymbol{\nu}). \quad (11)$$

Remark 4: This remark is the posterior counterpart of Remarks 2 and 3. Whereas CLL and CLML are S1P, CLPL is not; regularization breaks periodicities, favors solutions according to prior probabilities, and enables some ambiguities to be removed. Nevertheless, a global indetermination remains: CLPL is a G1P function. This is essentially due to the facts that i) the marginal likelihood CLML is a S1P function, and ii) the regularization term CLP is a G1P function (since it only involves frequency differences). As a consequence, two frequency profiles, which are different from a constant integer level, remain equi-likelihood. Finally, the latter indeterminacy can be removed by choosing an appropriate K: $K = 1$ enforces the first frequency ν_1 to remain in $(-0.5, +0.5]$, and the corresponding CLPL is no longer G1P.

Proposition 1: With the previous notations and definitions, the MAP estimate is such that

$$|\widehat{\nu}_{t+1}^{\mathrm{MAP}} - \widehat{\nu}_t^{\mathrm{MAP}}| \leqslant \frac{1}{2} \quad \text{for } t \in \mathbb{N}_{T-1}^*. \quad (12)$$

Proof: See Appendix B. ∎

F. Optimization Stage

The proposed approach allows ambiguous periodicity to be removed at the expense of accepting local minima in the built energy (10). A gradient procedure [27] can achieve *local* minimization of (10) and CLPL gradient involves the periodograms derivatives

$$P_t'(\nu_t) = 2j\pi \sum_{n=1-N}^{N-1} n\widehat{c}_t(n) e^{2j\pi\nu_t n}$$

when rewriting $P_t(\nu_t)$ as a function of empirical correlation lags $\widehat{c}_t(n)$ of the signal $\boldsymbol{y}_t$. It is also possible to calculate the second-order derivative

$$P_t''(\nu_t) = -4\pi^2 \sum_{n=1-N}^{N-1} n^2 \widehat{c}_t(n) e^{2j\pi\nu_t n}$$

and to implement second-order descent algorithms.

There are several ways of coping with *global* optimization, e.g., graduated nonconvexity [28], [29] and stochastic algorithms such as simulated annealing [30], [31]. We have used a dynamic programming procedure for computational simplicity. It is based on a discrete approximation of the *prior* law for the frequencies. This approximation allows global optimization (on an arbitrary fine discrete frequency grid) and provides a convenient framework for estimating hyperparameters.

IV. DISCRETE STATE MARKOV CHAIN

This section is devoted to a discrete approximation for

1) maximizing *posterior* law for the frequency sequence $\boldsymbol{\nu}$;
2) building an ML procedure for estimating hyperparameters.

We have therefore introduced an equally spaced discretization of the frequency range $[\nu_m; \nu_M]$ in P states $\nu^1, \ldots, \nu^P$ ($\nu_M = -\nu_m = 2.5$ and $P = 128$ in our simulations).

A. Probabilities

Discretization and normalization of the *a priori* law (7) yields the state transition probabilities:

$$\begin{aligned}
\mathbb{P}_t(p, q) &= \Pr\left[\nu_{t+1} = \nu^p \mid \nu_t = \nu^q\right] \\
&= \frac{\exp\left(\frac{-(\nu^p - \nu^q)^2}{2r_\nu}\right)}{\sum_{p=1}^{P} \exp\left(\frac{-(\nu^p - \nu^q)^2}{2r_\nu}\right)}.
\end{aligned} \quad (13)$$

Note that $\mathbb{P}_t$ does not depend on t, i.e., the proposed chain is homogeneous $\mathbb{P}_t = \mathbb{P}$. The full state model also includes the initial probabilities $_{\mathbf{P}}(p)$ chosen constant over $(-0.5, +0.5]$ (see Remark 4).

The marginal (w.r.t. amplitudes) likelihood function for the observation sequence given by (4) yields the observation probability distribution $\mathbb{O}_t(p) = f(\boldsymbol{y}_t \mid \nu_t = \nu^p)$.

B. Available Algorithms

The Markov chain is now convenient for using algorithms given in [32] and [33]: the Viterbi and the Forward-Backward algorithms. They enable us to compute

- the MAP;
- the hyperparameters likelihood as well as its gradient.

1) Viterbi Algorithm: The Viterbi algorithm, which is shown in Appendix C-A, has been implemented to cope with global optimization (on a discrete grid) and performs a step-by-step optimization of the *posterior* law. The required observation probabilities are also readily precomputable by the FFT.

2) Forward–Backward Algorithm: We have used a normalized version of the procedure, as recommended in [34] and [35], to avoid computational problems. It is founded on forward and backward probabilities

$$\mathcal{F}_t(p) = \frac{\Pr\left[\mathcal{Y}_1^t, \nu_t = \nu^p\right]}{\Pr\left[\mathcal{Y}_1^t\right]}$$

$$\text{and} \quad \mathcal{B}_t(p) = \frac{\Pr\left[\mathcal{Y}_{t+1}^T \mid \nu_t = \nu^p\right]}{\Pr\left[\mathcal{Y}_{t+1}^T \mid \mathcal{Y}_1^t\right]}$$

where $\mathcal{Y}_t^{t'} = [\boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t'}]$ denotes the partial observation matrix from time t to t'.

The (count-up) Forward algorithm, which is given in Appendix C-B, computes non-normalized probabilities $\overline{\mathcal{F}}_t(p)$, normalization coefficients $\mathcal{N}_t$, and the $\mathcal{F}_t(p)$ themselves. As a result, the observation likelihood can be deduced

$$\Pr[\mathcal{Y}] = \prod_{t=1}^{P} \mathcal{N}_t. \quad (14)$$

It is useful for estimating ML hyperparameters in Section V. The (count-down) Backward step, which is described in Appendix C-C, yields marginal *a posteriori* probabilities (see [32, p. 10])

$$p_t(p) = \Pr[\nu_t = \nu^p \,|\, \mathcal{Y}] = \mathcal{F}_t(p)\mathcal{B}_t(p) \tag{15}$$

and double marginal *a posteriori* probabilities (see [32, p. 11])

$$\begin{aligned} p_t(p,q) &= \Pr[\nu_{t-1} = \nu^q, \nu_t = \nu^p \,|\, \mathcal{Y}] \\ &= \mathcal{N}_t\, \mathcal{F}_{t-1}(p)\, \mathcal{B}_t(q)\, \mathbb{P}(p,q)\, \mathbb{O}_t(q) \end{aligned} \tag{16}$$

which are both needed to calculate the likelihood gradient.

V. ESTIMATING HYPERPARAMETERS

The MAP estimate of (11) depends on a unique regularization parameter λ function of three hyperparameters $\boldsymbol{r} = [r_a, r_b, r_\nu]$. This section is devoted to their estimation using the available data set $\mathcal{Y}$.

Estimating hyperparameters within the regularization framework is generally a delicate problem. It has been extensively studied, several techniques have been proposed and compared [36]–[41] and the preferred strategy is founded on ML.

The ML estimation consists of i) expressing the hyperparameter likelihood (HL) as $HL_{\mathcal{Y}}(\boldsymbol{r}) = f(\mathcal{Y})$ and ii) maximizing the resulting function. Although we have chosen a simple Gaussian law, $\boldsymbol{\nu}$ cannot be marginalized in closed form because $\boldsymbol{\nu}$ enters $f(\mathcal{Y}|\boldsymbol{\nu})$ in a complex manner. Fortunately, the discrete state approximation of Section IV provides a satisfactory solution to this problem. It also allows us to devise several kinds of algorithms for local maximization of the likelihood. One such scheme is the acknowledged expectation-maximization (EM) algorithm, although its application reveals uneasy in the present context of a parametric model of hidden Markov chain ([19] provides a meaningful discussion of such situations; see also [20] and [21]). Section V-B is devoted to the EM framework, within which a gradient procedure is proposed. Section V-A deals with the computation of the likelihood and proposes a simple coordinatewise descent procedure.

A. Hyperparameter Likelihood

The hyperparameter likelihood $HL_{\mathcal{Y}}$ can be deduced from the joint law for $(\boldsymbol{\nu}, \mathcal{Y})$ by frequency marginalization:

$$HL_{\mathcal{Y}}(\boldsymbol{r}) = \sum_{p_1,\ldots,p_T=1}^{P} \Pr[\mathcal{Y}, \nu_1 = \nu^{p_1}, \ldots, \nu_T = \nu^{p_T}]$$

but the indices run over P^T states; therefore, the above summation is not directly tractable. However, the Forward procedure efficiently achieves a recursive marginalization; it yields $HL_{\mathcal{Y}}(\boldsymbol{r})$ according to (14) and requires about TP^2 calculations.

Let us introduce the co-log-HL (CLHL) to be minimized w.r.t. hyperparameters vector $\boldsymbol{r}$:

$$\widehat{\boldsymbol{r}}^{\mathrm{ML}} = \arg\min_{\boldsymbol{r}} \mathrm{CLHL}_{\mathcal{Y}}(\boldsymbol{r}).$$

One possible optimization scheme is a coordinatewise descent algorithm with a golden section line search [27], but a more efficient scheme may be a gradient algorithm [27].

B. Likelihood Gradient

The EM algorithm relies on an auxiliary function, which is usually denoted Q [42], [43] built on two hyperparameter vectors $\boldsymbol{r}$ and $\boldsymbol{r}'$ by completing the observed data set $\mathcal{Y}$ with parameters to be marginalized $\boldsymbol{\nu}$:

$$\begin{aligned} Q(\boldsymbol{r}, \boldsymbol{r}') &= E_{\boldsymbol{\nu}}\Big[\log(\Pr[\boldsymbol{\nu}, \mathcal{Y}; \boldsymbol{r}']) \,\Big|\, \mathcal{Y}; \boldsymbol{r} \Big] \\ &= \sum_{\boldsymbol{\nu}} \log \Pr[\boldsymbol{\nu}, \mathcal{Y}; \boldsymbol{r}'] \Pr[(\boldsymbol{\nu} \,|\, \mathcal{Y}; \boldsymbol{r})]. \end{aligned}$$

With the proposed notations, usual hidden Markov chains calculations yield

$$\begin{aligned} Q(\boldsymbol{r}, \boldsymbol{r}') = &\sum_{t=2}^{T} \sum_{p,q=1}^{P} p_t(p,q) \log \mathbb{P}'(p,q) \\ &+ \sum_{p=1}^{P} \mathbb{P}(p) \log \mathbb{P}'(p) + \sum_{t=1}^{T} \sum_{p=1}^{P} p_t(p) \log \mathbb{O}'_t(p) \end{aligned} \tag{17}$$

where we have the following.

- $(\mathbb{P}', \mathbb{P}', \mathbb{O}')$ and $(\mathbb{P}, \mathbb{P}, \mathbb{O})$ are parameters of the model under hyperparameters $\boldsymbol{r}'$ and $\boldsymbol{r}$, respectively.
- $p_t(p)$ and $p_t(p,q)$ denote the *a posteriori* marginal laws defined by (15) and (16), under hyperparameters $\boldsymbol{r}$.

The kth iteration of the EM scheme maximizes $Q(\boldsymbol{r}^{(k-1)}, \boldsymbol{r}')$ as a function of $\boldsymbol{r}'$ to yield $\boldsymbol{r}^{(k)}$ as the maximizer. Unfortunately, it seems impossible to derive an explicit expression for such a maximizer. However, an alternate route can be followed, given the key property

$$\left.\frac{\partial Q(\boldsymbol{r}, \boldsymbol{r}')}{\partial \boldsymbol{r}'}\right|_{\boldsymbol{r}'=\boldsymbol{r}} = -\frac{\partial \mathrm{CLHL}_{\mathcal{Y}}(\boldsymbol{r})}{\partial \boldsymbol{r}}.$$

As suggested by [19], this property enables us to calculate the gradient of $\mathrm{CLHL}_{\mathcal{Y}}(\boldsymbol{r})$ as the derivative of (17):

$$\frac{\partial Q}{\partial r'_a} = \sum_{t=1}^{T} \sum_{p=1}^{P} p_t(p) \frac{\partial \log \mathbb{O}'_t(p)}{\partial r'_a} \tag{18}$$

$$\frac{\partial Q}{\partial r'_b} = \sum_{t=1}^{T} \sum_{p=1}^{P} p_t(p) \frac{\partial \log \mathbb{O}'_t(p)}{\partial r'_b} \tag{19}$$

$$\frac{\partial Q}{\partial r'_\nu} = \sum_{t=2}^{T} \sum_{p,q=1}^{P} p_t(p,q) \frac{\partial \log \mathbb{P}'(p,q)}{\partial r'_\nu}. \tag{20}$$

The encountered derivatives $\partial \log \mathbb{O}'(p)/\partial r'_a$, $\partial \log \mathbb{O}'(p)/\partial r'_b$ and $\partial \log \mathbb{P}'(p,q)/\partial r'_\nu$, respectively, read

$$\frac{-N}{Nr'_a + r'_b} + \frac{N}{(Nr'_a + r'_b)^2} P_t(\nu^p)$$

$$\frac{1-N}{r'_b} - \frac{1}{Nr'_a + r'_b} + \frac{\boldsymbol{y}_t^\dagger \boldsymbol{y}_t}{r'^2_b} - \frac{Nr'_a(Nr'_a + 2r'_b)}{r'^2_b(Nr'_a + r'_b)^2} P_t(\nu^p)$$

$$\frac{1}{2r'^2_\nu}\left((\nu^q - \nu^p)^2 - \sum_{r=1}^{P}(\nu^r - \nu^p)^2 \mathbb{P}'(p,r)\right)$$

by derivation of (4) and (13). Finally, the likelihood gradient is readily calculated, and a gradient procedure can be applied.
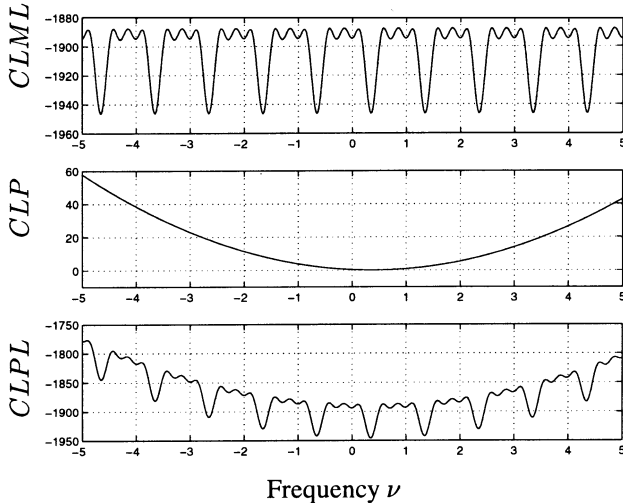
Fig. 2. Typical form of criteria. From top to bottom: CLML($\boldsymbol{\nu}$) (periodic), CLP($\boldsymbol{\nu}$) (quadratic), and CLPL($\boldsymbol{\nu}$) as a function of ν_t ($t = 50$). Regularization breaks periodicity.

VI. SIMULATION RESULTS AND COMPARISONS

The previous sections introduced a regularized method for frequency tracking and estimating hyperparameters. This section demonstrates the practical effectiveness of the proposed approach by processing[1] simulated signals shown in Fig. 1.

A. Hyperparameter Estimation

The hyperparameter likelihood function CLHL was first computed on a fine discrete grid of $25 \times 25 \times 25$ values, resulting in the level sets shown in Figs. 2 and 3. The function is fairly regular and has a single minimum.

The hyperparameters are tuned using two classes of descent algorithms:

- a coordinate-wise descent algorithm;
- a gradient descent algorithm.

The latter employs several descent directions: usual gradient, bisector correction, Vignes correction, and Polak-Ribière pseudo-conjugate direction. Two line search methods have also been implemented: usual dichotomy and quadratic interpolation. The starting point remains the empirical hyperparameter vector described in Appendix D.

All the strategies provide the correct minimizer, and they are compared in Table I and Fig. 3. The usual gradient generated zig-zagging trajectories and was slower than the other strategies. The three corrected direction strategies were 25 to 40% faster than the uncorrected ones with the Polak–Ribière pseudo-conjugate direction having a slight advantage. In contrast, interpolation did not result in any improvement within the corrected direction class.

The coordinate-wise descent algorithm performed well since it does not require any gradient calculation. Gradient calculus needs a lot more computation than the likelihood itself, due to summations in (18)–(20). Likelihood calculus took 0.05 s, whereas gradient calculus required 0.2 s., i.e., about four times more.

We have therefore adopted the two fastest methods: coordinate-wise and Polak-Ribière pseudo-conjugate gradient, which took less than 3.5 s. Fig. 3 also illustrates the convergence.

B. Frequency Tracking

The optimization procedure used to compute the MAP (given ML hyperparameters) consisted of applying the Viterbi algorithm (described in Section IV-B1). The solution was used as the starting point for the gradient or the Hessian procedure (described in Section III-F). The Viterbi algorithm explored the whole set of possible frequencies (on a discrete grid) and found the correct interval for each frequency, whereas the gradient or Hessian procedure locally refined the optimum. Table II shows the computation times. We adopted the Hessian procedure since it performed almost ten times faster.

Fig. 4 illustrates typical results. The ML strategy

- lacked robustness for two reasons: Estimation was performed independently at each depth, and N was small;
- could not be corrected by an unwrap-like post-processing since the ML solution was too rough (as already mentioned).

For the regularized solution (also given in Fig. 4), a simple qualitative comparison with the reference led to three conclusions.

- The estimated frequency sequence conformed much better to the true one. The frequency sequence was more regular since smoothness was introduced as a *prior* feature.
- The estimated frequency sequence remained close to the true one even beyond the usual Nyquist frequency. This was essentially due to the coherent accounting for the whole set of data and smoothness of the frequency sequence.
- The proposed strategy for estimating hyperparameters is adequate. A variation of 0.1 of the hyperparameters resulted in an almost imperceptible variation in the estimated frequency sequence. This is especially important for qualifying the robustness of the proposed method; the choice of $\boldsymbol{r}$ offers relatively broad leeway and can be reliably made.

VII. CONCLUSION AND PERSPECTIVES

This paper examines the problem of frequency tracking beyond the Nyquist frequency as it occurs in Doppler imaging when only short noisy data records are available. A solution is proposed in the Bayesian framework based on hidden Gauss–Markov models accounting for *prior* smoothness of the frequency sequence. We have developed a computationally efficient combination of dynamic programming and a Hessian procedure to calculate the maximum *a posteriori*. The method is entirely unsupervised and uses an ML procedure based on an original EM-based gradient procedure. The estimation of the ML hyperparameter is both formally achievable and practically useful.

This new Bayesian method allows tracking beyond the usual Nyquist frequency due to a coherent statistical framework that includes the whole set of data plus smoothness *prior*. To our
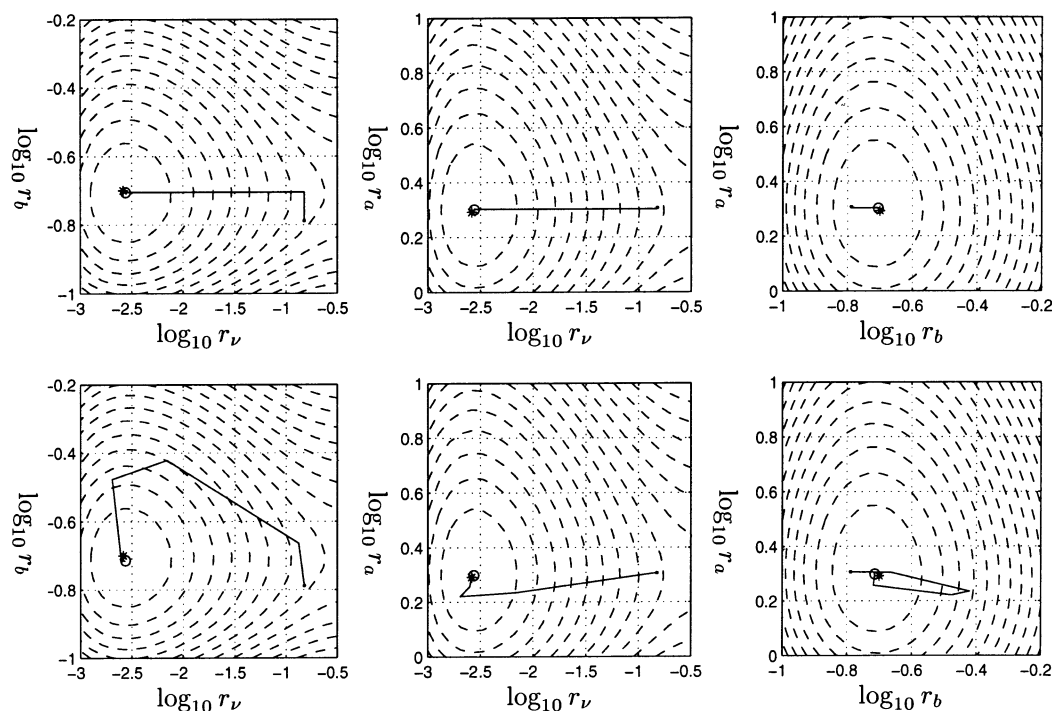
---

[1]Algorithms have been implemented using the computing environment *Matlab* on a Pentium III PC with a 450-MHz CPU and 128 MB of RAM.

Fig. 3. Hyperparameter likelihood: typical behavior. Level sets of CLHL are plotted as dashed lines (—). The minima are located by a star (∗), starting points (empirical estimates) by a dot, (.) and final estimate by a circle (o). The first row gives coordinate-wise algorithm, and the second row gives a gradient algorithm. First column: CLHL$(\widehat{r}_a^{\mathrm{ML}}, r_b, r_\nu)$; second column: CLHL$(r_a, \widehat{r}_b^{\mathrm{ML}}, r_\nu)$; third column: CLHL$(r_a, r_b, \widehat{r}_\nu^{\mathrm{ML}})$. Each figure is $\log_{10}$ scaled.

TABLE I

DESCENT ALGORITHM COMPARISON. THE FIRST COLUMN GIVES THE METHOD AT WORK: (1) USUAL GRADIENT, (2) VIGNES CORRECTION, (3) BISECTOR CORRECTION, AND (4) POLAK-RIBIÉRE PSEUDO-CONJUGATE DIRECTION. (A) NO INTERPOLATION AND (B) QUADRATIC INTERPOLATION. (5) COORDINATE-WISE DESCENT METHOD. FOLLOWING COLUMNS SHOW THE REACHED MINIMUM AND THE MINIMIZER. SIXTH COLUMN GIVES THE NUMBER OF GRADIENTS AND FUNCTION CALCULUS, WHEREAS THE LAST GIVES COMPUTATION TIMES IN SECONDS (s)

Method	Reached minimum	$\log_{10} \widehat{r}_a^{\mathrm{ML}}$	$\log_{10} \widehat{r}_b^{\mathrm{ML}}$	$\log_{10} \widehat{r}_\nu^{\mathrm{ML}}$	Grad./Fun.	Time (s)
(1a)	$4.513\ 10^2$	0.297	-0.685	-2.424	17/59	5.55
(1b)	$4.495\ 10^2$	0.297	-0.679	-2.519	13/87	5.92
(2a)	$4.494\ 10^2$	0.292	-0.678	-2.537	9 /49	3.77
(2b)	$4.494\ 10^2$	0.299	-0.681	-2.554	13/92	6.14
(3a)	$4.498\ 10^2$	0.297	-0.695	-2.589	9 /53	4.07
(3b)	$4.494\ 10^2$	0.298	-0.679	-2.547	13/92	6.21
(4a)	$4.497\ 10^2$	0.283	-0.674	-2.507	7 /40	3.12
(4b)	$4.500\ 10^2$	0.297	-0.685	-2.618	9 /75	4.84
(5)	$4.495\ 10^2$	0.300	-0.671	-2.559	0 /81	3.41

TABLE II

COMPUTATION TIMES COMPARISON FOR FREQUENCY ESTIMATE

Method	Time (s)
MAP Viterbi	0.13
MAP Gradient	4.82
MAP Hessian	0.51

knowledge, this capability is an original contribution to the field of frequency tracking.

Future work may include the extension to Gaussian DSP [9], to multiple frequencies tracking [3], [17], and to the two-dimensional (2-D) problem. The latter and its connection to 2-D phase unwrapping [44]–[46] is presently being investigated.

APPENDIX A
AMPLITUDE MARGINALIZATION

A. *Preliminary Results*

This Section includes two useful results: For $\boldsymbol{u} \in \mathbb{C}^N$

$$\det\left[I_N + \boldsymbol{u}\boldsymbol{u}^\dagger\right] = 1 + \boldsymbol{u}^\dagger\boldsymbol{u} \qquad (21)$$

$$\left(I_N + \boldsymbol{u}^\dagger\boldsymbol{u}\right)^{-1} = I_N - \frac{\boldsymbol{u}\boldsymbol{u}^\dagger}{1 + \boldsymbol{u}^\dagger\boldsymbol{u}} \qquad (22)$$

where I_N stands for the $N \times N$ identity matrix.

B. *Law for $(\boldsymbol{y}_t | \nu_t)$*

Linearity of model (1) w.r.t. amplitudes and assumptions for a_t and $\boldsymbol{b}_t$ allow easy marginalization of $(\boldsymbol{y}_t, a_t | \nu_t)$: $\boldsymbol{y}_t | \nu_t$
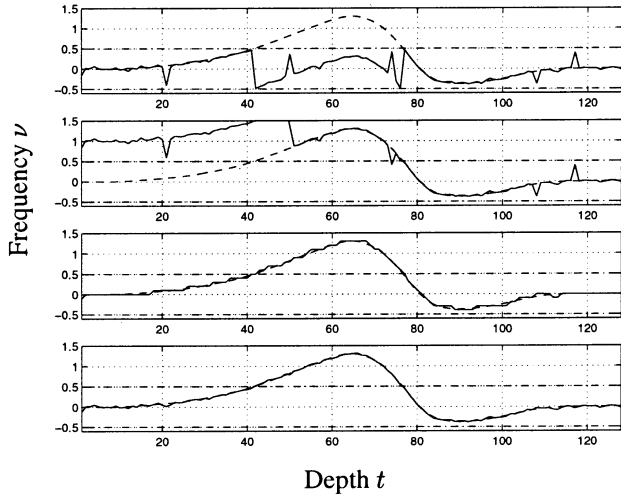
Fig. 4. Comparison of frequency profile estimates. From top to bottom: ML estimate (i.e., periodogram maximizer), unwrapped ML estimate, Viterbi-MAP estimate, and Hessian-MAP estimate.

is clearly a zero-mean and Gaussian vector with covariance $R_t = r_a \boldsymbol{z}(\nu_t)\boldsymbol{z}(\nu_t)^\dagger + r_b I_N$. From (21) and (22), its determinant and inverse reads

$$R_t^{-1} = \frac{1}{r_b}I_N - \frac{\alpha}{N}\boldsymbol{z}(\nu_t)\boldsymbol{z}(\nu_t)^\dagger \qquad (23)$$

$$\det R_t = r_b^{N-1}(r_b + N r_a). \qquad (24)$$

APPENDIX B
PROOF OF PROPOSITION 1

A. Preliminary Result

The proposed proof is based on the decimal part function D : $\mathbb{R} \longrightarrow [-0.5; +0.5[$ defined by

$$\begin{cases} D(x) = x, & \text{if } x \in [-0.5; +0.5) \\ D \text{ is } 1 & \text{-periodic} \end{cases} \qquad (25)$$

and the following straightforward properties:

$$D(x + k) = D(x), k \in \mathbb{Z} \qquad (26)$$

$$|D(x)| \leqslant |x| \qquad (27)$$

$$|D(x)| \leqslant \frac{1}{2} \qquad (28)$$

$$y = D(x) \Rightarrow \exists k \in \mathbb{Z} \text{ such that } y = x + k. \qquad (29)$$

B. Proof of Proposition

Let us define a frequency sequence $\boldsymbol{\nu}$ (with CLPL$(\boldsymbol{\nu}) < \infty$), which does not verify (12) of Proposition 1, i.e.,

$$\exists t_0 \in \mathbb{N}_{T-1}^* \text{ with } |\nu_{t_0+1} - \nu_{t_0}| > \frac{1}{2}. \qquad (30)$$

Let us recursively build a new frequency sequence $\widetilde{\boldsymbol{\nu}}$:

$$\widetilde{\nu}_1 = \nu_1 \qquad (31)$$

$$\widetilde{\nu}_{t+1} = \widetilde{\nu}_t + D(\nu_{t+1} - \widetilde{\nu}_t) \text{ for } t = 1,\ldots,T-1 \qquad (32)$$

and prove that (12) of Proposition 1 holds for $\widetilde{\boldsymbol{\nu}}$ and that the criterion CLPL reduces from $\boldsymbol{\nu}$ to $\widetilde{\boldsymbol{\nu}}$:

$$|\widetilde{\nu}_{t+1} - \widetilde{\nu}_t| \leqslant \frac{1}{2} \text{ for } t \in \mathbb{N}_{T-1}^* \qquad (33)$$

$$\text{CLPL}(\widetilde{\boldsymbol{\nu}}) < \text{CLPL}(\boldsymbol{\nu}). \qquad (34)$$

- Relation (33) is straightforward; by (32), one can see

$$\widetilde{\nu}_{t+1} - \widetilde{\nu}_t = D(\nu_{t+1} - \widetilde{\nu}_t) \text{ for } t \in \mathbb{N}_{T-1}^*$$

and hence, by Property (28)

$$|\widetilde{\nu}_{t+1} - \widetilde{\nu}_t| \leqslant \frac{1}{2} \text{ for } t \in \mathbb{N}_{T-1}^*.$$

- Proof of (34) takes three steps, corresponding to each term of CLPL (10). By (31) and (32) and Property (29), one can see

$$\exists k_t \in \mathbb{Z} \text{ such that } \widetilde{\nu}_t = \nu_t + k_t \text{ for } t \in \mathbb{N}_T^* \qquad (35)$$

(with $k_1 = 0$); therefore

$$P_t(\nu_t) = P_t(\widetilde{\nu}_t) \text{ for } t \in \mathbb{N}_T^*. \qquad (36)$$

By (32) and (35) and invoking Property (26), we have

$$\widetilde{\nu}_{t+1} - \widetilde{\nu}_t = D(\nu_{t+1} - \widetilde{\nu}_t) = D(\nu_{t+1} - \nu_t)$$

hence, accounting for Property (27)

$$|\widetilde{\nu}_{t+1} - \widetilde{\nu}_t| \leqslant |\nu_{t+1} - \nu_t|. \qquad (37)$$

Moreover, for $t = t_0$, we clearly have

$$|\widetilde{\nu}_{t_0+1} - \widetilde{\nu}_{t_0}| < |\nu_{t_0+1} - \nu_{t_0}| \qquad (38)$$

thanks to hypothesis (30). Finally, we have

$$\mathbb{1}_K^\infty(\nu_1) = \mathbb{1}_K^\infty(\widetilde{\nu}_1). \qquad (39)$$

Collecting (36)–(39) proves (34).

APPENDIX C
HMC ALGORITHMS

A. Viterbi Algorithm

- Precomputations

$$\mathcal{D}(p,q) = \lambda(\nu^p - \nu^q)^2 \qquad (p,q \in \mathbb{N}_P^*)$$

$$\mathcal{L}(p,t) = -P_t(\nu^p) \qquad (p \in \mathbb{N}_P^*, t \in \mathbb{N}_T^*).$$

- Initialization $(t = 1)$

$$\mathcal{C}_1(p) = \mathcal{L}(p,1)\,\mathbb{1}_1^\infty(\nu^p) \qquad (p \in \mathbb{N}_P^*).$$

- Iterations $(t = 2,\ldots,T)$

$$\widetilde{\mathcal{C}}_t(p,q) = \mathcal{C}_{t-1}(q) + \mathcal{D}(p,q) + \mathcal{L}(p,t) \qquad (p,q \in \mathbb{N}_P^*)$$

$$\mathcal{C}_t(p) = \min_q \widetilde{\mathcal{C}}_t(p,q) \qquad (p \in \mathbb{N}_P^*)$$

$$\mathcal{P}_t(p) = \arg\min_q \widetilde{\mathcal{C}}_t(p,q) \qquad (p \in \mathbb{N}_P^*).$$

- Termination $(t = T)$

$$\widehat{p}_T = \arg\min_p \mathcal{C}_T(p).$$

- Back tracking $(t = T - 1, \ldots, 1)$

$$\widehat{p}_t = \mathcal{P}_t(\widehat{p}_{t+1}).$$

B. Forward Algorithm

- Initialization $(t = 1)$

$$\overline{\mathcal{F}}_1(p) = \mathbb{O}_1(p)\,_{\mathbf{P}}(p) \qquad (p \in \mathbb{N}_P^*)$$

$$\mathcal{N}_1 = \sum_{q=1}^{P} \overline{\mathcal{F}}_1(q)$$

$$\mathcal{F}_1(p) = \frac{\overline{\mathcal{F}}_1(p)}{\mathcal{N}_1} \qquad (p \in \mathbb{N}_P^*).$$

- Iterations $(t = 2, \ldots, T)$

$$\overline{\mathcal{F}}_t(p) = \mathbb{O}_t(p) \sum_{q=1}^{P} \mathcal{F}_{t-1}(p)\mathbb{P}(q, p) \qquad (p \in \mathbb{N}_P^*)$$

$$\mathcal{N}_t = \sum_{q=1}^{P} \overline{\mathcal{F}}_t(q)$$

$$\mathcal{F}_t(p) = \frac{\overline{\mathcal{F}}_t(p)}{\mathcal{N}_t} \qquad (p \in \mathbb{N}_P^*).$$

C. The Backward Algorithm

- Initialization $(t = T)$

$$\overline{\mathcal{B}}_T(p) = 1 \qquad (p \in \mathbb{N}_P^*)$$
$$\mathcal{B}_T(p) = 1 \qquad (p \in \mathbb{N}_P^*).$$

- Iterations $(t = T - 1, \ldots, 1)$

$$\overline{\mathcal{B}}_t(p) = \sum_{q=1}^{P} \mathbb{O}_{t+1}(q)\mathcal{B}_{t+1}(p)\mathbb{P}(p, q) \qquad (p \in \mathbb{N}_P^*)$$

$$\mathcal{B}_t(p) = \frac{\overline{\mathcal{B}}_t(p)}{\mathcal{N}_{t+1}} \qquad (p \in \mathbb{N}_P^*).$$

APPENDIX D
EMPIRICAL ESTIMATION OF HYPERPARAMETERS

This section is devoted to the *empirical* estimation of hyperparameters used as a starting point in the maximization procedures of Section VI-A. These estimates are based on the correlation $r(n)$ of $\boldsymbol{y}_t | \nu_t$ and easily shown to verify $r(0) = r_a + r_b$, and $|r(1)| = r_a$, for all $t \in \mathbb{N}_T^*$. Empirical estimates $\widehat{r}(0)$ and $\widehat{r}(1)$ are computed from the whole data set $\mathcal{Y}$ and remain robust since T is large (even if N is small). Finally, one can compute $\widehat{r}_a = |\widehat{r}(1)|$, and $\widehat{r}_b = \widehat{r}(0) - |\widehat{r}(1)|$.

For r_ν, the estimation is based on the ML estimate of the frequency sequence in each range bin $t \in \mathbb{N}_T^*$. The proposed empirical estimate of r_ν is naturally the empirical variance of the differences between the ML frequencies. This procedure yields an overestimated value for r_ν. This result is expected since the sequence of ML frequencies varies greatly and has discontinuities, as mentioned above. Nevertheless, this estimate is a suitable starting point for the maximization procedures of Section VI-A.

REFERENCES

[1] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal – Part 1: Fundamentals," *Proc. IEEE*, vol. 80, pp. 519–538, Apr. 1992.
[2] ——, "Estimating and interpreting the instantaneous frequency of a signal – Part 2: Algorithms and applications," *Proc. IEEE*, vol. 80, pp. 539–568, Apr. 1992.
[3] R. F. Barret and D. A. Holdsworth, "Frequency tracking using hidden Markov models with amplitude and phase information," *IEEE Trans. Signal Processing*, vol. 41, pp. 2965–2975, Oct. 1993.
[4] P. Tichavský and A. Nehorai, "Comparative study of four adaptive frequency trackers," *IEEE Trans. Signal Processing*, vol. 45, pp. 1473–1484, June 1997.
[5] P. J. Kootsookos and J. M. Spanjaard, "An extended Kalman filter for demodulation of polynomial phase signals," *IEEE Signal Processing Lett.*, vol. 5, pp. 69–70, Mar. 1998.
[6] H. C. So, "Adaptive algorithm for discret estimation of sinusoidal frequency," *Electron. Lett.*, vol. 36, no. 8, pp. 759–760, Apr. 2000.
[7] J. M. B. Dias and J. M. N. Leitão, "Nonparametric estimation of mean Doppler and spectral width," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 271–282, Jan. 2000.
[8] A. Herment, G. Demoment, P. Dumée, J.-P. Guglielmi, and A. Delouche, "A new adaptive mean frequency estimator: Application to constant variance color flow mapping," *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.*, vol. 40, pp. 796–804, 1993.
[9] J.-F. Giovannelli, J. Idier, B. Querleux, A. Herment, and G. Demoment, "Maximum likelihood and maximum a posteriori estimation of Gaussian spectra. Application to attenuation measurement and color Doppler velocimetry," in *Proc. Int. Ultrason. Symp.*, vol. 3, Cannes, France, Nov. 1994, pp. 1721–1724.
[10] D. Hann and C. Greated, "The measurement of sound fileds using laser Doppler anemometry," *Acustica*, vol. 85, pp. 401–411, 1999.
[11] C. Kasai, K. Namekawa, A. Koyano, and R. Omoto, "Real-time two-dimensional blood flow imaging using an autocorrelation technique," *IEEE Trans. Sonics Ultrason.*, vol. SU-32, pp. 458–464, May 1985.
[12] R. F. Woodman, "Spectral moment estimation in MST radars," *Radio Sci.*, vol. 20, no. 6, pp. 1185–1195, Nov. 1985.
[13] T. Loupas and W. N. McDicken, "Low-order complex AR models for mean and maximum frequency estimation in the context of Doppler color flow mapping," *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.*, vol. 37, pp. 590–601, Nov. 1990.
[14] B. A. J. Angelsen and K. Kristoffersen, "Discrete time estimation of the mean Doppler frequency in ultrasonic blood velocity measurement," *IEEE Trans. Biomed. Eng.*, vol. BME-30, pp. 207–214, 1983.
[15] F.-K Li, D. N. Held, H. C. Curlander, and C. Wu, "Doppler parameter estimation for spaceborne synthetic-aperture radars," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-23, pp. 47–56, Jan. 1985.
[16] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[17] R. L. Streit and R. F. Barret, "Frequency line tracking using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 38, pp. 586–598, Apr. 1990.
[18] E. S. Chornoboy, "Optimal mean velocity estimation for Doppler weather radars," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 575–586, May 1993.
[19] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic function of a Markov process to automatic speech processing," *Bell Syst. Tech. J.*, vol. 62, no. 4, pp. 1035–1074, Apr. 1982.
[20] K. Lange, "A gradient algorithm locally equivalent to the EM algorithm," *J. R. Statist. Soc. B*, vol. 57, no. 2, pp. 425–437, 1995.
[21] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[22] H. E. Talhami and R. I. Kitney, "Maximum likelihood frequency tracking of the audio pulsed Doppler ultrasound signal using a Kalman filter," *Ultrasound Med. Biol.*, vol. 14, no. 7, pp. 599–609, 1988.
[23] D. K. Barton and S. Leonov, *Radar Technology Encyclopedia*. Norwell, MA: Artech House, 1997.

[24] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.

[25] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.

[26] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Commun.*, vol. C-26, pp. 219–229, Mar. 1977.

[27] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[28] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[29] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Processing*, vol. 7, pp. 571–585, Apr. 1998.

[30] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[31] C. Robert, *Méthodes de Monte-Carlo par Chaînes de Markov*. Paris, France: Economica, 1996.

[32] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE Acoust., Speech, Signal Processing Mag.*, pp. 4–16, 1986.

[33] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[34] P. A. Devijver and M. Dekessel, "Champs aléatoires de Pickard et modélization d'images digitales," *Traitement du Signal*, vol. 5, no. 5, pp. 131–150, 1988.

[35] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recognit. Lett.*, vol. 3, pp. 369–373, Dec. 1985.

[36] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometr.*, vol. 21, no. 2, pp. 215–223, May 1979.

[37] D. M. Titterington, "Common structure of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, no. 2, pp. 141–170, 1985.

[38] 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.

[39] 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.

[40] N. Fortier, G. Demoment, and Y. Goussard, "Comparison of GCV and ML methods of determining parameters in image restoration by regularization," *J. Visual Commun. Image Repres.*, vol. 4, pp. 157–170, 1993.

[41] 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. 220–233, Mar. 1996.

[42] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.

[43] L. A. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 729–734, Sept. 1982.

[44] D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping*. New York: Wiley Interscience, 1998.

[45] M. Servin, J. L. Marroquin, D. Malacara, and F. J. Cueva, "Phase unwrapping with a regularized phase-tracking system," *Appl. Opt.*, vol. 37, no. 10, pp. 1917–1923, Apr. 1998.

[46] G. Nico, G. Palubinskas, and M. Datcu, "Bayesian approaches to phase unwrapping: Theoretical study," *IEEE Trans. Signal Processing*, vol. 48, pp. 2545–2556, Sept. 2000.

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 and received the Doctorat degree in physics from the Laboratoire des Signaux et Systèmes, Université de Paris-Sud, Orsay, France, in 1995.

He is presently Assistant Professor with the Département de Physique, Université de Paris-Sud. He is interested in regularization methods for inverse problems in signal and image processing, mainly in spectral characterization. Application fields essentially concern radar and medical imaging.

Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from the École Supérieure d'Électricité, Paris, 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 Centre National de la Recherche Scientifique, assigned to the Laboratoire des Signaux et Systèmes, Université de Paris-Sud. His major scientific interests are in probabilistic approaches to inverse problems for signal and image processing.

Rédha Boubertakh was born in Algiers, Algeria, in 1975. He received the diploma degree in electrical engineering from the École Nationale Polytechnique d'Alger, Algiers, in 1996. He is currently pursuing the Ph.D. degree at the INSERM Unit 494, Hôpital Pitié-Salpétrière, Paris, France.

He is interested in signal and image processing, mainly in the fields of magnetic resonance imaging.

Alain Herment was born in Paris, France, in 1948. He graduated from ISEP Engineering School, Paris, in 1971. He received the Doctorat d'État degree in physics from ISEP Engineering School in 1984.

Initially, he worked as an engineer at the Centre National de la Recherche Scientifique. In 1977, was a researcher at the Institut National pour la Santé et la Recherche Médicale (INSERM), Paris. He is currently in charge of the department of cardiovascular imaging at the INSERM Unit 66, Hôpital Pitié, Paris. He is interested in signal and image processing for extracting morphological and functional information from images sequences, mainly in the fields of ultrasound investigations, X-ray CT, and digital angiography.