Default Bayesian Estimation of the Fundamental Frequency

Jesper Kjær Nielsen, *Student Member, IEEE*, Mads Græsbøll Christensen, *Senior Member, IEEE*, and Søren Holdt Jensen, *Senior Member, IEEE*

***Abstract*—Joint fundamental frequency and model order estimation is an important problem in several applications. In this paper, a default estimation algorithm based on a minimum of prior information is presented. The algorithm is developed in a Bayesian framework, and it can be applied to both real- and complex-valued discrete-time signals which may have missing samples or may have been sampled at a non-uniform sampling frequency. The observation model and prior distributions corresponding to the prior information are derived in a consistent fashion using maximum entropy and invariance arguments. Moreover, several approximations of the posterior distributions on the fundamental frequency and the model order are derived, and one of the state-of-the-art joint fundamental frequency and model order estimators is demonstrated to be a special case of one of these approximations. The performance of the approximations are evaluated in a small-scale simulation study on both synthetic and real world signals. The simulations indicate that the proposed algorithm yields more accurate results than previous algorithms. The simulation code is available online.**

***Index Terms*—Fundamental frequency estimation, Bayesian model comparison, Zellner's g-prior.**

I. Introduction

An important and basic problem in time-series analysis is the estimation of the fundamental frequency and the number of harmonic components of a periodic signal. The problem is encountered in a wide range of science and engineering applications such as music processing [1], [2], speech processing [3], [4], sonar [5], and electrocardiography (ECG) [6]. In particular for musical applications, fundamental frequency estimation has been subject to extensive research for several decades [2]. This is primarily due to that a musical note is composed of the sum of a fundamental partial and a number of overtone partials. For harmonic instruments, these overtone partials are called harmonics since their frequencies $\{\omega_i\}_{i=2}^{l}$ are approximately related to the fundamental frequency ω of the fundamental partial by $\omega_i \approx i\omega$ for $i = 2, \cdots, l$ [1], [7].

Manuscript received September 29, 2011; revised March 04, 2012, July 04, 2012, and October 26, 2012; accepted November 01, 2012. Date of publication November 27, 2012; date of current version February 13, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Engin Erzin.

J.K. Nielsen and S.H. Jensen are with the Multimedia Information and Signal Processing Group, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: jkn@es.aau.dk; shj@es.aau.dk).

M.G. Christensen is with the Audio Analysis Lab, Department of Architecture, Design & Media Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: mgc@create.aau.dk).

Digital Object Identifier 10.1109/TASL.2012.2229979

Since the fundamental frequency is such an important physical attribute to musical applications, the more elegant term pitch is often used instead [2]. Therefore, the problem considered in this paper is often referred to as (single-)pitch estimation in the context of musical applications.

The problem of estimating the fundamental frequency is typically defined in the following way. A data set $\{x(t_n)\}_{n=0}^{N-1}$ originating from a discrete-time signal is observed and modelled as

$$x(t_n) = s(t_n) + e(t_n) \ , \quad n = 0, 1, \cdots, N-1 \qquad (1)$$

where $\{t_n\}_{n=0}^{N-1}$, $\{s(t_n)\}_{n=0}^{N-1}$, and $\{e(t_n)\}_{n=0}^{N-1}$ are the sampling times, the predictable part of the signal, and the non-predictable part of the signal, respectively. Usually, the sampling period T is assumed to be constant so that $t_n = nT$ for $t_0 = 0$. However, in order to allow for a non-uniform sampling scheme or missing samples, this assumption is not made here. The predictable part consists of l harmonic components and is at time t_n given by

$$s(t_n) = \begin{cases} \displaystyle\sum_{i=1}^{l} \alpha_i \exp(ji\omega t_n), & x(t_n) \in \mathbb{C} \\ \displaystyle\sum_{i=1}^{l} a_i \cos(i\omega t_n) + b_i \sin(i\omega t_n), & x(t_n) \in \mathbb{R} \end{cases} \qquad (2)$$

where $\mathbb{C}$ and $\mathbb{R}$ denote the set of complex and real numbers, respectively, and $j = \sqrt{-1}$ is the imaginary unit. For the i'th harmonic component, the complex amplitude α_i, the linear weights a_i and b_i, the amplitude A_i, and the phase ϕ_i are related by $\alpha_i = a_i + jb_i = A_i \exp(j\phi_i)$. Note that a real-valued signal of the form in (2) can be cast into the form of a complex-valued signal in (2) by computing its down-sampled analytic signal [8]. Provided that the frequencies of the first and last harmonics are not too close to zero and the Nyquist frequency (relative to N), respectively, the parameter estimates based on the down-sampled analytic signal are nearly identical to the corresponding parameter estimates based on the real-valued signal [2], [9]. In this paper, the focus is on the complex-valued signal model since it leads to simpler notation and faster algorithms [2], [10]. However, the results for the real-valued signal model are also given.

Numerous fundamental frequency estimation algorithms have been suggested in the literature. The simplest algorithms are the non-parametric methods based on, for example, the auto-correlation function [11], [12] or the cepstrum [13] (See [14], [15] for other non-parametric methods). The more advanced algorithms are based on a signal model of the observed signal and are therefore referred to as parametric methods. These are typically maximum likelihood-based (ML) methods [16], [17], subspace-based methods [10], [18], filtering methods [19], [20], or Bayesian methods [7], [21], [22]. We refer the interested reader to [2] for a review of many of the non-Bayesian methods. Only a few of the suggested methods assume that the number of harmonics is unknown. In order to perform model selection, these methods typically add an order dependent penalty term to the log-likelihood function [23]–[25], use the eigenvalues [26] or eigenvectors [27], [28], or compare the angle between subspaces [29]. A good overview over these and other methods can be found in [2]. In contrast to model comparison in which a probability for each model is computed, these methods are typically only designed for detecting the most likely model. On the other hand, model comparison enables us to account for model uncertainty in the estimation of unknown model parameters and the prediction of missing data points by using all models instead of just the most likely one. As demonstrated in, e.g., [30], model averaging increases the prediction performance.

In this paper, inference about the fundamental frequency and the number of harmonics are made in a Bayesian framework. The Bayesian framework is used for model comparison since it leads to consistent estimates under very mild conditions, naturally selects the simplest model which explains the data reasonably well (the principle of Occam's razor [31]), takes model uncertainty into account for estimation and prediction, and enables a more intuitive interpretation of the results [32], [33]. In a Bayesian framework, prior distributions on the unknown quantities must be elicited and their hyperparameters must be selected. In general, this is not a trivial problem since the prior information is usually not in the form of probability distributions, and prior information must therefore be turned into one or several probability distributions. For model comparison, this prior elicitation is very important since improper or vague priors may lead to indeterminate or bad answers [33]. Another difficulty of the Bayesian methods is that closed-form analytical solutions seldomly exist. Various numerical algorithms such as Markov chain Monte Carlo sampling [34] can overcome this limitation, but the computational load of running these algorithms is typically very high.

The primary aim of this paper is to develop a default estimation scheme for estimating the fundamental frequency and detecting the number of harmonics. Even though the number of harmonics might not be of interest by itself, it is still vital to detect it in order to avoid problems with for example pitch halving [10]. By the word *default*, we mean that an almost user-parameter free algorithm is developed which automatically follows from a minimum of prior information and a few minor approximations. The approximations are made so that closed-form expressions are obtained which have a computational load comparable to the methods suggested in [2]. Moreover, we show that a special case of the proposed approach is identical to the algorithm proposed in [2, Sec. 2.6]. Finally, we demonstrate through simulation examples that the proposed method is superior to the state-of-the-art ML-based and subspace-based methods. Note that we are here not concerned with the development of a full pitch detection and tracking system for speech or music applications such as YIN [11], RAPT [35], or NDF [36]. However, we believe that our estimator may be a useful component in such systems as well as in similar systems for other application domains.

The paper is organised as follows. We first formally define the inference problem in a Bayesian framework in Sec. II. In Sec. III, the prior information is turned into the observation model and prior distributions which are used to derive the posterior distributions on the fundamental frequency and the model order in Sec. IV. In Sec. V, various approximations of varying accuracy and computational load are developed, and in Sec. VI it is demonstrated that a state-of-the-art ML-based algorithm is a special case of one of these approximations. In Sec. VII, the approximations are evaluated on a synthetic signal, and the applicability of the algorithm is demonstrated for the spectral analysis of a speech signal.

II. PROBLEM FORMULATION AND BACKGROUND

The primary aim is to make inference about the fundamental frequency ω and the model order l given the prior information I and the N data points collected in the vector $\boldsymbol{x}$. That is, we wish to find the posterior densities[1]

$$p(\omega, l|\boldsymbol{x}, I) = p(\omega|\boldsymbol{x}, l, I)p(l|\boldsymbol{x}, I) \tag{3}$$

and some of their statistics such as the mode, the mean, and the variance. The model order l labels a unique model $\mathcal{M}_l$ with model parameters $\boldsymbol{\theta}_l \in \Theta_l$. For the problem at hand, ω is one of these parameters, and the remaining model parameters are nuisance parameters. The observation model $p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)$ describes the relationship between the data and the model. When viewed as a function of the model parameters, the observation model is referred to as the likelihood function, and it plays an important role in statistics where it is mainly used for parameter estimation. However, model comparison cannot only be based on comparing the highest likelihoods of the candidate models as more complex models can always fit the observed data better than simpler models. In a Bayesian framework, the model parameters and the model order are random variables with the prior pdf $p(\boldsymbol{\theta}_l|l, I)$ and pmf $p(l|I)$, respectively. The posterior pdf $p(\boldsymbol{\theta}_l|\boldsymbol{x}, l, I)$ and pmf $p(l|\boldsymbol{x}, I)$ are connected to these priors through Bayes' theorem

$$p(\boldsymbol{\theta}_l|\boldsymbol{x}, l, I) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)p(\boldsymbol{\theta}_l|l, I)}{p(\boldsymbol{x}|l, I)} \tag{4}$$

$$p(l|\boldsymbol{x}, I) = \frac{p(\boldsymbol{x}|l, I)p(l|I)}{p(\boldsymbol{x}|I)} \tag{5}$$

where

$$p(\boldsymbol{x}|l, I) = \int_{\Theta_l} p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)p(\boldsymbol{\theta}_l|l, I)d\boldsymbol{\theta}_l \tag{6}$$

is called the marginal likelihood or the evidence. For model comparison, the odds of two competing model orders k and i

[1] In (3) and the rest of the paper, the generic notation $p(\cdot)$ is used to denote both a probability density function (pdf) over a continuous parameter and a probability mass function (pmf) over a discrete parameter.

are often compared. In this connection, the posterior odds are often used, and they are given by

$$\frac{p(k|\boldsymbol{x}, I)}{p(i|\boldsymbol{x}, I)} = \text{BF}[k, i]\frac{p(k|I)}{p(i|I)} \tag{7}$$

where the Bayes' factor is $\text{BF}[k, i] = p(\boldsymbol{x}|k, I)/p(\boldsymbol{x}|i, I)$. Since the prior and posterior pdfs on the model order are discrete, it is easy to find the posterior odds and the posterior distribution once the Bayes' factors are known. Therefore, the main challenge in Bayesian model comparison is to compute the Bayes' factor for competing pairs of models. However, before Bayes' theorem can be used to make inference about the fundamental frequency and the model order in Sec. IV, the prior information I must first be turned into an observation model and prior distributions on the model parameters.

III. A Default Probability Model

As alluded to previously, we are here concerned with the development of an inference scheme which automatically follows from a minimum of prior information I. Thus, a fundamental problem is to specify a probability model which reflects I about which we assume the following.

Assumption 3.1: We are given N data points $\{x(t_n)\}_{n=0}^{N-1}$ from a zero-mean real- or complex-valued signal which has been sampled at the known time instances $\{t_n\}_{n=0}^{N-1}$. The signal is wide-sense stationary (WSS) and consists of a predictable part which is periodic, corrupted by additive noise, and bandlimited to the known angular frequency interval $[\omega_a, \omega_b]$.

For a given application, more prior information may be available which should be included in this assumption. For example in a pitch tracking system, the estimates of the last frame is known, and we might also know something about the correlation structure of the amplitudes of the harmonics and the noise based on, e.g., physical properties. However, we are here concerned with a default and application independent inference scheme so only the information I in Ass. 3.1 is assumed and used to elicit the probability model consisting of the observation model and the prior distributions on the model parameters. For notational convenience, the following vectors and matrix are defined

$$\boldsymbol{x} \triangleq \begin{bmatrix} x(t_0) & \cdots & x(t_{N-1}) \end{bmatrix}^T \tag{8}$$

$$\boldsymbol{e} \triangleq \begin{bmatrix} e(t_0) & \cdots & e(t_{N-1}) \end{bmatrix}^T \tag{9}$$

$$\boldsymbol{\alpha}_l \triangleq \begin{cases} \begin{bmatrix} \alpha_1 & \cdots & \alpha_l \end{bmatrix}^T , & \boldsymbol{x} \in \mathbb{C}^N \\ \begin{bmatrix} a_1 & \cdots & a_l & b_1 & \cdots & b_l \end{bmatrix}^T , & \boldsymbol{x} \in \mathbb{R}^N \end{cases} \tag{10}$$

$$\boldsymbol{z}_i \triangleq \begin{bmatrix} \exp(ji\omega t_0) & \cdots & \exp(ji\omega t_{N-1}) \end{bmatrix}^T \tag{11}$$

$$\boldsymbol{Z}_l \triangleq \begin{cases} \begin{bmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_l \end{bmatrix} , & \boldsymbol{x} \in \mathbb{C}^N \\ \begin{bmatrix} \text{Re}(\boldsymbol{z}_1) & \cdots & \text{Re}(\boldsymbol{z}_l) & \text{Im}(\boldsymbol{z}_1) & \cdots & \text{Im}(\boldsymbol{z}_l) \end{bmatrix} , & \boldsymbol{x} \in \mathbb{R}^N \end{cases} \tag{12}$$

where $(\cdot)^T$ denotes matrix transposition, and $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ take the real and imaginary part, respectively, of a complex number.

A. The observation model

In order to deduce the observation model, a model for the non-predictable part or the noise $\boldsymbol{e}$ given the prior information I must be selected in (1) which in vector notation is given by

$$\boldsymbol{x} = \boldsymbol{Z}_l\boldsymbol{\alpha}_l + \boldsymbol{e} \ . \tag{13}$$

Obviously, the distribution must integrate to one and have zero-mean, and the average power of the noise process must be finite since the signal has been sampled. Thus, the noise variance σ^2 is therefore finite, and the WSS property implies that σ^2 does not change with time. As advocated in [37], [38], the pdf which maximises the entropy under these constraints should be selected, and this pdf is the (complex) normal distribution with density

$$p(\boldsymbol{e}|\sigma^2, I) = \left[r\pi\sigma^2\right]^{-N/r} \exp\left(-\frac{\boldsymbol{e}^H\boldsymbol{e}}{r\sigma^2}\right) \tag{14}$$

$$= \begin{cases} \mathcal{CN}(\boldsymbol{e}; \boldsymbol{0}, \sigma^2\boldsymbol{I}_N) \ , & r = 1 \\ \mathcal{N}(\boldsymbol{e}; \boldsymbol{0}, \sigma^2\boldsymbol{I}_N) \ , & r = 2 \end{cases} \tag{15}$$

where $(\cdot)^H$ denotes conjugate matrix transposition, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and r is either 1 for $\boldsymbol{x} \in \mathbb{C}^N$ or 2 for $\boldsymbol{x} \in \mathbb{R}^N$. To simplify the notation, the non-standard notation $\mathcal{N}_r(\cdot)$ is used to refer to either the complex normal pdf $\mathcal{CN}(\cdot)$ for $r = 1$ or the real normal pdf $\mathcal{N}(\cdot)$ for $r = 2$. It is important to note that the noise variance σ^2 is a random variable. As opposed to the case where it is simply a fixed and unknown quantity, the noise distribution marginalised over this random noise variance is able to model noise with heavy tails and is robust towards outliers. In Sec. III-B1, the prior distribution on the noise variance is elicited. Note that (15) does not explicitly model any correlation structure in the noise. If prior information about such a structure is available, it should be included in the constraints to enable more accurate estimation results (see, e.g., [25], [39]). However, including these constraints lowers the entropy so if nothing is known about a correlation structure, (15) is the least informative distribution on the noise since it maximises the entropy [39], [40]. If the noise is known to be coloured, the proposed method is still useful if it is combined with a linear pre-filter which whitens the noise.

From (15), it follows that the observation model is

$$p(\boldsymbol{x}|\boldsymbol{\alpha}_l, \sigma^2, \omega, l, I) = \mathcal{N}_r(\boldsymbol{x}; \boldsymbol{Z}_l\boldsymbol{\alpha}_l, \sigma^2\boldsymbol{I}_N) \ . \tag{16}$$

In most of the literature on fundamental frequency estimation, the same observation model is used. However, the derivation presented here facilitates a different interpretation of this model. Namely, when nothing is known about the noise except that it is WSS and has a finite power, the white Gaussian noise assumption is the least informative or most conservative noise distribution.

B. The Prior Distributions

When the parametrisation is not given by the problem, the maximum entropy method cannot be used for the elicitation of a default prior distribution [41, Sec. 5.6.2]. For example, the noise variance σ^2 has so far been used in the parametrisation, but the standard deviation σ or the precision parameter $\lambda = \sigma^{-2}$ could have been used instead. Applying the maximum entropy principle to either of these three common representations leads to the unsatisfactory situation that the prior distribution is not invariant under the choice of parametrisations. In order to cope with the different representations, the invariances which the prior distribution must obey are often considered [37], [38]. That is, which transformations of the parameters do not change the prior knowledge? Another useful question to consider is which parameters are logically connected. That is, if the value of one parameter is known, would that change the state of knowledge about the other parameters? Although this is not a necessary question to consider, selecting a representation in which the parameters are not logically connected simplifies the prior elicitation [42, App. A]. In our representation, the parameters are the complex amplitudes $\boldsymbol{\alpha}_l$, the noise variance σ^2, the fundamental frequency ω, and the model order l. The fundamental frequency is clearly logically connected to the model order since it must be below ω_b/l. However, other dependencies between the parameters cannot be extracted from our prior information I, and the prior pdf is therefore factored as

$$p(\boldsymbol{\alpha}_l, \sigma^2, \omega, l|I) = p(\boldsymbol{\alpha}_l|I)p(\sigma^2|I)p(\omega|l,I)p(l|I) \ . \quad (17)$$

1) The Noise Variance: Since the choice of parametrisation is not obvious, the prior distribution on the noise variance is selected such that it does not depend on whether the noise variance, the precision parameter, or the standard deviation is used. For invariance under either of these representations, it is therefore required that

$$p(\sigma|I)d\sigma = p(\sigma^m|I)d\sigma^m \ , \qquad \forall\ m \neq 0 \quad (18)$$

which is satisfied for $p(\sigma|I) \propto \sigma^{-1}$. This improper prior pdf is very famous and known as the Jeffreys' prior [43]. It is improper since it does not integrate to one. In practice, however, the noise variance cannot go all the way to zero due to, for example, quantisation noise, and the noise variance is always upper bounded so a normalised prior pdf on the noise variance is

$$p(\sigma^2|I) = \begin{cases} \left[\ln(w/v)\sigma^2\right]^{-1} & v < \sigma^2 < w \\ 0 & \text{otherwise} \end{cases} . \quad (19)$$

The bounds on the noise variance have almost no influence on the inference so they are often selected as $v \to 0$ and $w \to \infty$ to simplify the analysis [42, App. A].

2) The Fundamental Frequency: For the elicitation of the prior distribution on the fundamental frequency, the arguments from Sec. III-B1 can be repeated. Whether the (angular) fundamental frequency ω, the ordinary fundamental frequency $f = \omega/(2\pi)$, or the fundamental period $\tau = f^{-1}$ is used, does not change the prior knowledge. From the prior information I, the signal is bandlimited to the interval $[\omega_a, \omega_b]$, and ω must therefore lie on the interval $\Omega_l = [\omega_a, \omega_b/l]$. Thus, the posterior pdf on the fundamental frequency is

$$p(\omega|l,I) = \begin{cases} (F_l\omega)^{-1} & \omega \in \Omega_l \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

where $F_l = \ln(\omega_b) - \ln(l\omega_a)$. This prior was also derived in [42, App. A] for a single sinusoid using a more ingenious argument.

3) The Complex Amplitudes: The sinusoidal model is typically parametrised by the Cartesian coordinates (a_i, b_i) or the polar coordinates (A_i, ϕ_i). Since neither of these representations change the state of knowledge, the prior pdf on the complex amplitudes is required to be invariant under the transformation between these two representations. That is,

$$p(a_i, b_i|I)da_i db_i = q(A_i, \phi_i|I)A_i dA_i d\phi_i \quad (21)$$

for $i = 1, \cdots, l$ where $p(a_i, b_i|I)$ and $q(A_i, \phi_i|I)$ are the pdfs on the Cartesian and polar coordinates, respectively. From the prior information I, the signal is assumed to be zero mean and WSS. In terms of the Cartesian coordinates, this implies that a_i and b_i are uncorrelated and both have zero mean and the same expected power $\sigma_\alpha^2/2$. For the polar coordinates, it implies that the phase is uniformly distributed on any continuous interval of length 2π and uncorrelated with the amplitude. Finally, since the phases $\{\phi_i\}_{i=1}^l$ are independent and uniformly distributed, the l harmonic components are uncorrelated [44, Ch. 4]. We note in passing that many of the same arguments are also used for the derivation of the covariance matrix model for a time series. The only distribution satisfying (21) and these properties is the normal distribution with pdf [38, Ch. 7]

$$p(a_i, b_i|\sigma_\alpha^2, I) = \mathcal{N}_2([a_i, b_i]^T; \mathbf{0}, (\sigma_\alpha^2/2)\boldsymbol{I}_2) \ . \quad (22)$$

Turning this bivariate real normal pdf into a univariate complex normal pdf on the complex amplitude α_i gives [45, Ch. 15]

$$p(\alpha_i|\sigma_\alpha^2, I) = \mathcal{N}_1(\alpha_i; 0, \sigma_\alpha^2) \ . \quad (23)$$

The joint pdf on $\boldsymbol{\alpha}_l$ is therefore

$$p(\boldsymbol{\alpha}_l|\sigma_\alpha^2, I) = \mathcal{N}_r(\boldsymbol{\alpha}_l; \mathbf{0}, (\sigma_\alpha^2/r)\boldsymbol{I}_{rl}) \ . \quad (24)$$

The derivation of the normal pdf given above is often called the Herschel-Maxwell derivation [38]. Since σ_α^2 is unknown, this hyperparameter is treated as a random variable. By using the same arguments as for the noise variance, the following hyperprior is obtained

$$p(\sigma_\alpha^2|I) = \begin{cases} \left[\ln(w/v)\sigma_\alpha^2\right]^{-1} & v < \sigma_\alpha^2 < w \\ 0 & \text{otherwise} \end{cases} . \quad (25)$$

4) The Model Order: Since the model order is a discrete parameter, the maximum entropy principle can be applied without worrying about the parametrisation. Under the constraint that the prior pmf of the model order must integrate to one, $p(l)$ is the uniform pmf on the set $l \in \{1, 2, \cdots, L\}$. As model orders larger than $\lfloor \omega_b/\omega_a \rfloor$ have zero support, L should not be chosen larger than this value. Note that the model order $l = 0$ is not in the support set since the prior information I states that a predictable part is present in the signal. However, later on, it is discussed how the proposed algorithm can cope with the detection of a predictable part.

C. The g-Prior

In the previous sections, the prior information I has been turned into a default probability model. Unfortunately, the prior probability model renders the inference problem analytically intractable. However, if a re-parametrisation and a few minor approximations are made, a prior on the same form as the Zellner's g-prior [46] is obtained, and this prior has some tractable analytical properties [33], [47]. For the re-parametrisation, the power of the i'th harmonic component is written as

$$\frac{\sigma_\alpha^2}{r} = \frac{rg\sigma^2}{N} \quad \iff \quad g = \frac{N\sigma_\alpha^2}{r^2\sigma^2} = \frac{N\eta}{rl} \tag{26}$$

where the signal-to-noise ratio (SNR) is defined as

$$\eta \triangleq \frac{E[|s(t_n)|^2]}{E[|e(t_n)|^2]} = \sum_{i=1}^{l} \frac{\sigma_\alpha^2}{r\sigma^2} = \frac{l\sigma_\alpha^2}{r\sigma^2} \ . \tag{27}$$

Thus, g may be interpreted as N/r times the average SNR. Note that although any prior dependency between the complex amplitudes and the noise variance was included in the factorisation in (17), the dependency automatically appears through g. As reviewed in [47], the hyperparameter g can be set to a fixed value or treated as a random variable. When g is a random variable, the prior pdf of g can be derived from (26), (25), and (19) to

$$p(g|I) = \begin{cases} \frac{\ln(w/v) - |\ln(r^2 g/N)|}{\ln^2(w/v)g} \ , & g \in \left[\frac{Nv}{r^2 w}, \frac{Nw}{r^2 v}\right] \\ 0 \ , & \text{otherwise} \end{cases} \tag{28}$$

which in the limit of $v \to 0$ and $w \to \infty$ reduces to the prior $p(g|I) \propto g^{-1}$ for $g > 0$.

To justify the approximations, which we introduce below, the following assumption is made.

Assumption 3.2: The number of data points N is large enough to justify that $(N/r)(\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1} \approx \boldsymbol{I}_{rl}$.

Ass. 3.2 is often used in connection with sinusoidal frequency estimation to lower the computational complexity of the inference algorithm significantly. It holds for a uniform sampling scheme and for sufficiently random non-uniform sampling schemes, and it stems from that sinusoids are asymptotically orthogonal for any set of distinct frequencies. That is,

$$\lim_{N\to\infty} \frac{r}{N} \boldsymbol{Z}_l^H \boldsymbol{Z}_l = \boldsymbol{I}_{rl} \ . \tag{29}$$

For a fixed N, the approximation gets progressively worse as the frequencies become smaller and closer [2]. Under Ass. 3.2 and the re-parametrisation in (26), the prior pdf on the complex amplitudes becomes

$$p(\boldsymbol{\alpha}_l|\sigma^2, \omega, g, I) = \mathcal{N}_r(\boldsymbol{\alpha}_l; \boldsymbol{0}, g\sigma^2(\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1}) \ . \tag{30}$$

Another consequence of Ass. 3.2 is that the likelihood function for the fundamental frequency is very sharply peaked around the ML estimate of ω. Therefore, the prior distribution on ω only has negligible effect on the posterior distribution [42, App. A], and it is therefore approximated by a uniform pdf on the interval Ω_l. That is,

$$p(\omega|l) = W_l^{-1} \mathbb{I}_{\Omega_l}(\omega) \tag{31}$$

where $W_l = \omega_b/l - \omega_a$ and $\mathbb{I}_{\Omega_l}(\omega)$ is the indicator function on the interval Ω_l.

As noted in Sec. III-B1, the bounds on the noise variance have almost no influence on the inference. They are therefore selected as $v \to 0$ and $w \to \infty$ so that the improper Jeffreys' prior $p(\sigma^2|I) \propto (\sigma^2)^{-1}$ is obtained for the noise variance. For Bayesian comparison of models with parameter spaces of different dimensions, proper prior distributions must be selected on the model parameters to make the Bayes' factor well-defined [33]. However, since the noise variance is a common parameter in all models, an improper prior may be used on it [48]. Since g is also a common parameter in all models, the prior $p(g|I) \propto g^{-1}$ may be used for $g > 0$. For example, this prior has been used in [49]. Although simple, this prior does not allow analytical marginalisation w.r.t. g in the inference step. However, the prior is a limiting case of the beta prime or inverted beta distribution with density

$$p(g|\epsilon, \delta, I) = \frac{(\delta - 1)\Gamma(\epsilon + \delta)}{\Gamma(\epsilon + 1)\Gamma(\delta)} g^\epsilon (1 + g)^{-\delta - \epsilon} \mathbb{I}_{\mathbb{R}+}(g) \tag{32}$$

which is proper for $\delta > 1$ and $\epsilon > -1$. Although this prior pdf enables analytical inference w.r.t. g, the special case for $\epsilon = 0$ is only used in the sequel to keep the results simpler. Moreover, this special case was also suggested in [47], and it involves only the single hyperparameter δ. Since it is proper, it can be used to detect if a predictable part is present. In the limit of $\delta \to 1$, the improper and user-parameter free prior $p(g|I) \propto (1+g)^{-1}$ is obtained, and it has been shown in [50] that the joint prior $p(g, \sigma^2|I) \propto [\sigma^2(1+g)]^{-1}$ is the Jeffreys' prior and the reference prior [51] for a linear regression model. As this improper prior is a special case of the proper prior on g, the algorithm is derived in the next section for the proper prior. This means that the developed algorithm is able to cope with the detection of predictable part.

IV. Bayesian Inference

Bayes' theorem is now used to compute the posterior distributions on the quantities of interest which are the fundamental frequency for every candidate model and the model order. In order to cope with the detection of a predictable part in the signal, the proper prior distribution on g in (32) with $\epsilon = 0$ is used. The joint posterior pdf on all model parameters and the number of harmonics is[2]

$$\begin{aligned} p(\boldsymbol{\alpha}_l, \sigma^2 \omega, g|\boldsymbol{x}, l) &\propto p(\boldsymbol{x}|\boldsymbol{\alpha}_l, \sigma^2, \omega, g, l) p(\boldsymbol{\alpha}_l|\sigma^2, \omega, g, l) \\ &\quad \times p(\sigma^2) p(\omega|l) p(g) \\ &\propto \mathcal{N}_r(\boldsymbol{\alpha}_l; c\hat{\boldsymbol{\alpha}}_l, \sigma^2 \boldsymbol{C}_l) \text{Inv-}\mathcal{G}(\sigma^2; N/r, N\hat{\sigma}_l^2/r) \\ &\quad \times \frac{\Gamma(N/r)\mathbb{I}_{\Omega_l}(\omega)\mathbb{I}_{\mathbb{R}+}(g)}{(\pi N \hat{\sigma}_l^2)^{N/r} W_l (1+g)^{l+\delta}} \end{aligned} \tag{33}$$

[2]To keep the notation uncluttered, the explicit dependence on the prior information I is omitted in the rest of the paper.

where Inv-$\mathcal{G}$ is the inverse gamma pdf. Moreover, we have defined

$$\hat{\boldsymbol{\alpha}}_l \triangleq (\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1} \boldsymbol{Z}_l^H \boldsymbol{x} \quad (34)$$

$$c \triangleq g(1+g)^{-1} \quad (35)$$

$$\boldsymbol{C}_l \triangleq c(\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1} \quad (36)$$

$$\hat{\sigma}_l^2 \triangleq \frac{\boldsymbol{x}^H(\boldsymbol{I}_N - c\boldsymbol{P}_l)\boldsymbol{x}}{N} \quad (37)$$

$$R_l^2(\omega) \triangleq \frac{\boldsymbol{x}^H \boldsymbol{P}_l \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} \ . \quad (38)$$

When the ML estimate of the fundamental frequency is used and $c = 1$, $\hat{\sigma}_l^2$ is the ML estimate of the noise variance. The matrix $\boldsymbol{P}_l$ is the orthogonal projection matrix onto the space spanned by the columns of $\boldsymbol{Z}_l$, and $R_l^2(\omega)$ resembles the coefficient of determination from linear regression analysis where it is used to measure the prediction performance. Integrating (33) over the noise variance and the complex amplitudes gives

$$p(\omega, g|\boldsymbol{x}, l) \propto \frac{m_0(\boldsymbol{x})(\delta - 1) f_l(\omega, g, \delta) \mathbb{I}_{\Omega_l}(\omega) \mathbb{I}_{\mathbb{R}^+}(g)}{W_l} \quad (39)$$

where

$$m_0(\boldsymbol{x}) \triangleq \Gamma(N/r)(\pi \boldsymbol{x}^H \boldsymbol{x})^{-N/r} \propto p(\boldsymbol{x}|l=0) \quad (40)$$

$$f_l(\omega, g, \delta) \triangleq (1+g)^{N/r-l-\delta}[1 + g(1 - R_l^2(\omega))]^{-N/r} \quad (41)$$

are the unnormalised marginal likelihood for the noise-only model and a very important function in the sequel, respectively. In the case where g is a known parameter, the marginal posterior pdf on ω under model order l is proportional to this function

$$p(\omega|\boldsymbol{x}, g, l) = \frac{p(\omega, g|\boldsymbol{x}, l)}{p(g)} \propto f_l(\omega, g, 0) \mathbb{I}_{\Omega_l}(\omega) \ . \quad (42)$$

When g is an unknown parameter and $l > 1 - \delta$, it can be integrated out of (39) so that the marginal posterior pdf on ω under model order l is obtained as

$$p(\omega|\boldsymbol{x}, l) = \int_0^\infty p(\omega, g|\boldsymbol{x}, l) dg \propto \int_0^\infty f_l(\omega, g, \delta) \mathbb{I}_{\Omega_l}(\omega) dg$$
$$\propto {}_2F_1(N/r, 1; l + \delta; R_l^2(\omega)) \mathbb{I}_{\Omega_l}(\omega) \quad (43)$$

where ${}_2F_1$ is the Gaussian hypergeometric function [52, p. 314]. The condition $l > 1 - \delta$ ensures that the integral in (43) converges, and it is satisfied for all l when the prior on g is proper, i.e., $\delta > 1$, and for any $l > 0$ even for the improper prior on g with $\delta \to 1$. The marginal pmf on the model order is given by

$$p(l|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|l)p(l)}{p(\boldsymbol{x})} = \frac{\text{BF}[l, 0]p(l)}{\sum_{i=0}^L \text{BF}[i, 0]p(i)} \quad (44)$$

where $p(\boldsymbol{x}|l, I)$ is the marginal likelihood and

$$\text{BF}[l, 0] = \frac{p(\boldsymbol{x}|l)}{p(\boldsymbol{x}|l=0)} = \frac{m_l(\boldsymbol{x})}{m_0(\boldsymbol{x})} \quad (45)$$

is the Bayes' factor. Here, the noise-only model is used as the base model so the prior distribution on g must be proper. When the noise-only model is not in the set of candidate models, the model with a single harmonic component is used as the base model. In this case, the prior on g can be improper, and the Bayes' factor is given by

$$\text{BF}[l, 1] = \lim_{\delta \to 1} \frac{\text{BF}[l, 0]}{\text{BF}[1, 0]} \ . \quad (46)$$

When g is a known parameter, the Bayes' factor has the following integral representation

$$\text{BF}[l, 0|g] = \frac{p(l=0)\int_{\Omega_l} p(\omega, g|\boldsymbol{x}, l) d\omega}{p(l=0|\boldsymbol{x})p(g)}$$
$$= \frac{1}{W_l} \int_{\Omega_l} f_l(\omega, g, 0) d\omega \ , \quad (47)$$

and when g is an unknown parameter, the Bayes' factor is

$$\text{BF}[l, 0] = \frac{\delta - 1}{W_l} \int_{\Omega_l} \int_0^\infty f_l(\omega, g, \delta) dg d\omega \quad (48)$$
$$= \frac{\delta - 1}{W_l(l + \delta - 1)} \int_{\Omega_l} {}_2F_1(N/r, 1; l + \delta; R_l^2(\omega)) d\omega \ . \quad (49)$$

Unfortunately, the modes and the moments of the posterior pdf on the fundamental frequency are not available in closed-form due to the non-linear way that ω parametrises the pdfs in (42) and (43). Moreover, the posterior model order probabilities are not available in closed-form since the integrals in (47) and (49) cannot be computed analytically. In Sec. V, various approximate ways of finding these modes, moments, and posterior probabilities are discussed.

A. Selecting a Value for g

In order to facilitate an easier evaluation of the posterior pdfs for the fundamental frequency and the model order, the parameter g is often assumed to be a deterministic parameter rather than a random variable when it is unknown. Thus, instead of marginalising over g, a value for g is selected or estimated. There exist several ways of selecting the value of g, and two popular choices are considered here. Selecting $g^{\text{BIC}} = N$ approximately corresponds to the Bayesian information criterion (BIC) [47]. Alternatively, an empirical Bayesian method can be used in which the unknown hyperparameter g is estimated from the data. The value of g can then be estimated as the maximum likelihood estimate of the joint pdf $p(\boldsymbol{x}, \boldsymbol{\alpha}_l, \sigma^2, \omega|g, l)$ integrated w.r.t. the unknown parameters. However, since the marginalisation over the fundamental frequency cannot be done in closed-form, the marginalisation is only carried out over the complex amplitudes and the noise variance, and the fundamental frequency is simply replaced with its MAP estimate $\hat{\omega}$ which is derived in the next section. That is,

$$g_l^{\text{EB}} = \arg\max_{g \in \mathbb{R}^+} p(\boldsymbol{x}, \hat{\omega}|g, l) = \arg\max_{g \in \mathbb{R}^+} f_l(\hat{\omega}, g, 0)$$
$$= \max\left(\frac{NR_l^2(\hat{\omega}) - rl}{(1 - R_l^2(\hat{\omega}))rl}, 0\right) \ . \quad (50)$$

There are several other ways of selecting the value of g, and the interested reader is referred to the excellent review in [47] and the references therein.

V. Approximations

In this section, several approximations of (42), (43), (47), and (49) are derived for various choices of g . The accuracy of these approximations is evaluated in a small-scale simulation study in Sec. VII.

A. Numerical Integration

Since the integrals in (47) and (49) are one dimensional integrals, they can easily be evaluated using numerical integration techniques. For example, the integrals in (47) and (49) can be approximately evaluated by computing

$$\mathrm{BF}[l,0|g] \approx \sum_{k=1}^{K} \frac{f_l(\omega_k, g, 0)}{K} \tag{51}$$

$$\mathrm{BF}[l,0] \approx \sum_{k=1}^{K} \frac{(\delta-1)\,{}_2F_1(N/r,1;l+\delta;R_l^2(\omega))}{K(l+\delta-1)}\ , \tag{52}$$

respectively, where $\{\omega_k\}_{k=1}^K$ are K equidistant candidate frequencies from the set Ω_l with $W_l/K = \omega_{k+1} - \omega_k$, $\omega_1 = \omega_a$, and $\omega_K = \omega_b/l - W_l/K$. However, the functions $f_l(\omega, g, 0)$ and ${}_2F_1(N/r, 1; l+\delta; R_l^2(\omega))$ are usually very sharply peaked around their modes so the pdfs have to be evaluated over a fine frequency grid to make the approximation accurate. Moreover, the computation of $f_l(\omega_k, g, 0)$ and, in particular, ${}_2F_1(N/r, 1; l+\delta; R_l^2(\omega_k))$ is quite costly since either $\boldsymbol{x}^H\boldsymbol{P}_l\boldsymbol{x}$ or ${}_2F_1$ has to be computed for all K candidate frequencies. Even under Ass. 3.2, the limit in (29) cannot be used to justify the approximation

$$\boldsymbol{x}^H\boldsymbol{P}_l\boldsymbol{x} \approx \frac{r}{N}\|\boldsymbol{Z}_l^H\boldsymbol{x}\|^2 \tag{53}$$

since the value of $f_l(\omega, g, \delta)$ is very sensitive to even small perturbations in $R_l^2(\omega)$ when it is close to one and the SNR is large. Thus, the numerical integration of (47) and (49) may entail a too high computational load, and some analytical approximations are therefore also considered since they can reduce this computational load significantly.

B. The Distribution on the Fundamental Frequency

Although closed-form expressions (up to a constant of proportionality) have been derived for the pdf of the fundamental frequency for a known and an unknown g in (42) and (43), respectively, neither the moments, the modes nor the normalisation constants can be found in closed-form. To find approximate expressions for these, we therefore assume the following.

Assumption 5.1: The posterior pdfs $p(\omega|\boldsymbol{x}, g, l)$ and $p(\omega|\boldsymbol{x}, l)$ of the fundamental frequency for a known and an unknown g, respectively, can be approximated by the pdf of a normal distribution.

As we demonstrate in Sec. VII, this assumption holds for moderate and high SNRs. Under adverse signal conditions such as a low SNR, the pdfs $p(\omega|\boldsymbol{x}, g, l)$ and $p(\omega|\boldsymbol{x}, l)$ have several significant peaks and Ass. 5.1 does therefore not hold. In this case, the distribution on the fundamental frequency may be approximated by a Gaussian mixture model instead [53, Ch. 12]. However, this is not explored any further in this paper. The normal approximation of $p(\omega|\boldsymbol{x}, g, l)$ is

$$p(\omega|\boldsymbol{x}, g, l) \approx \mathcal{N}_2(\omega; \hat{\omega}, s_l(\hat{\omega}|g)) \tag{54}$$

where $\hat{\omega}$ is the mode of $p(\omega|\boldsymbol{x}, g, l)$ corresponding to the MAP estimate of the fundamental frequency, and

$$s_l(\hat{\omega}|g) = -\left[\left.\frac{\partial^2 \ln p(\omega|\boldsymbol{x}, g, l)}{\partial\omega^2}\right|_{\omega=\hat{\omega}}\right]^{-1}\ . \tag{55}$$

The normal approximation

$$p(\omega|\boldsymbol{x}, l) \approx \mathcal{N}_2(\omega; \hat{\omega}, s_l(\hat{\omega})) \tag{56}$$

has the same mean, but its variance is

$$s_l(\hat{\omega}) = -\left[\left.\frac{\partial^2 \ln p(\omega|\boldsymbol{x}, l)}{\partial\omega^2}\right|_{\omega=\hat{\omega}}\right]^{-1}\ . \tag{57}$$

As stated above, the MAP estimate of the fundamental frequency under model order l does not depend on whether the value of g is known or not. It is given as the solution to

$$\begin{aligned}\hat{\omega} &= \arg\max_{\omega\in\Omega_l} p(\omega|\boldsymbol{x}, g, l) = \arg\max_{\omega\in\Omega_l} p(\omega|\boldsymbol{x}, l)\\ &= \arg\max_{\omega\in\Omega_l} R_l^2(\omega) = \arg\max_{\omega\in\Omega_l} \boldsymbol{x}^H\boldsymbol{P}_l\boldsymbol{x}\ ,\end{aligned} \tag{58}$$

and it is the same as the ML estimate [44, Ch. 4]. Unfortunately, it is costly from a computational point of view to find the ML estimate since the cost-function in (58) has a complicated multi-modal shape and is very sharply peaked around $\hat{\omega}$, especially for a high SNR. Typically, the ML estimate is found by first evaluating the cost-function on a fine grid and then performing a local optimisation around the maximum value of the cost-function on this grid. However, the computational complexity of this procedure may be too high since the projection matrix $\boldsymbol{P}_l$ must be evaluated for every candidate frequency. The computational cost can be significantly reduced by making the approximation in (53). This leads to the following approximate MAP-estimate

$$\hat{\omega} \approx \arg\max_{\omega\in\Omega_l} \boldsymbol{x}^H\boldsymbol{Z}_l\boldsymbol{Z}_l^H\boldsymbol{x} = \arg\max_{\omega\in\Omega_l} \|\boldsymbol{Z}_l^H\boldsymbol{x}\|_2^2 \tag{59}$$

which under a uniform sampling frequency can be computed efficiently using a single FFT [2]. To get the MAP estimate in (58), the approximate MAP estimate in (59) may be used as the starting point of a local optimisation using the exact cost-function in (58). The local optimisation can also be substituted for faster and approximate techniques based on, e.g., interpolation [54].

In order to find the variances $s_l(\hat{\omega}|g)$ and $s_l(\hat{\omega})$, the second order derivatives of $\ln p(\omega|\boldsymbol{x}, g, l)$ and $\ln p(\omega|\boldsymbol{x}, l)$ must be found and evaluated at the mode $\hat{\omega}$. The first order derivatives are given by

$$\frac{\partial \ln p(\omega|\boldsymbol{x}, g, l)}{\partial\omega} = \frac{c}{r\hat{\sigma}_l^2}\frac{\partial C_l(\omega)}{\partial\omega} \tag{60}$$

$$\begin{aligned}\frac{\partial \ln p(\omega|\boldsymbol{x}, l)}{\partial\omega} &= \frac{{}_2F_1(N/r+1, 2; l+\delta+1; R_l^2(\omega))}{r\hat{\sigma}_0^2(l+\delta)\,{}_2F_1(N/r, 1; l+\delta; R_l^2(\omega))}\\ &\quad\times\frac{\partial C_l(\omega)}{\partial\omega}\end{aligned} \tag{61}$$

where

$$C_l(\omega) \triangleq \boldsymbol{x}^H \boldsymbol{P}_l \boldsymbol{x} \ . \tag{62}$$

Note that for $l = 1$, $C_l(\omega)$ is the periodogram. Evaluated at the mode $\hat{\omega}$, the second-order derivatives are

$$\left.\frac{\partial^2 \ln p(\omega|\boldsymbol{x}, g, l)}{\partial \omega^2}\right|_{\omega=\hat{\omega}} = \frac{c}{r\hat{\sigma}_l^2} d_2 \tag{63}$$

$$\left.\frac{\partial^2 \ln p(\omega|\boldsymbol{x}, l)}{\partial \omega^2}\right|_{\omega=\hat{\omega}} = \frac{{}_2F_1(N/r+1, 2; l+\delta+1; R_l^2(\hat{\omega}))}{r\hat{\sigma}_0^2(l+\delta){}_2F_1(N/r, 1; l+\delta; R_l^2(\hat{\omega}))} \times d_2 \ . \tag{64}$$

where

$$d_2 \triangleq \left.\frac{\partial^2 C_l(\omega)}{\partial \omega^2}\right|_{\omega=\hat{\omega}} . \tag{65}$$

Both of these second order derivatives consist of the second-order derivative of $C_l(\omega)$. It is given by

$$\begin{aligned} d_2 = 2\text{Re}\Big[&\hat{\boldsymbol{e}}^H \boldsymbol{D}_2 \hat{\boldsymbol{\alpha}}_l - 2\hat{\boldsymbol{e}}^H \boldsymbol{D}_1 (\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1} \boldsymbol{Z}_l^H \boldsymbol{D}_1 \hat{\boldsymbol{\alpha}}_l\Big] \\ &+ 2\hat{\boldsymbol{e}}^H \boldsymbol{D}_1 (\boldsymbol{Z}_l^H \boldsymbol{Z}_l)^{-1} \boldsymbol{D}_1^H \hat{\mathrm{e}} - 2\hat{\boldsymbol{\alpha}}_l^H \boldsymbol{D}_1^H \boldsymbol{P}_l^\perp \boldsymbol{D}_1 \hat{\boldsymbol{\alpha}}_l \end{aligned} \tag{66}$$

where $\boldsymbol{P}_l^\perp = \boldsymbol{I}_N - \boldsymbol{P}_l$ and

$$\hat{\boldsymbol{e}} \triangleq \boldsymbol{x} - \boldsymbol{Z}_l \hat{\boldsymbol{\alpha}}_l \tag{67}$$

$$\boldsymbol{D}_1 \triangleq \left.\frac{\partial \boldsymbol{Z}_l}{\partial \omega}\right|_{\omega=\hat{\omega}} = j^r (\boldsymbol{1}_r^T \otimes \boldsymbol{t}\boldsymbol{l}^T) \odot \boldsymbol{Z}_l (\boldsymbol{J}_r \otimes \boldsymbol{I}_l) \tag{68}$$

$$\boldsymbol{1}_r \triangleq \begin{cases} 1 \ , & r = 1 \\ \begin{bmatrix} 1 & -1 \end{bmatrix}^T \ , & r = 2 \end{cases} \tag{69}$$

$$\boldsymbol{D}_2 \triangleq \left.\frac{\partial^2 \boldsymbol{Z}_l}{\partial \omega^2}\right|_{\omega=\hat{\omega}} = -(\boldsymbol{1}_r^T \otimes \boldsymbol{t}\boldsymbol{l}^T) \odot (\boldsymbol{1}_r^T \otimes \boldsymbol{t}\boldsymbol{l}^T) \odot \boldsymbol{Z}_l \tag{70}$$

$$\boldsymbol{t} \triangleq \begin{bmatrix} t_0 & t_1 & \cdots & t_{N-1} \end{bmatrix}^T \tag{71}$$

$$\boldsymbol{l} \triangleq \begin{bmatrix} 1 & 2 & \cdots & l \end{bmatrix}^T \ . \tag{72}$$

The operators $\otimes$ and $\odot$ are the Kronecker and Hadamard products, respectively, and $\boldsymbol{J}_r$ is the $r \times r$ exchange matrix. In order to decrease the computational cost of finding the variances $s_l(\hat{\omega}|g)$ and $s_l(\hat{\omega})$, a simpler, but only approximate, expression for the second-order derivative of $C_l(\omega)$ is also derived. Under Ass. 5.1 and at the mode $\hat{\omega}$, it follows that $\|\hat{\boldsymbol{\alpha}}_l\|^2 \gg \|\hat{\boldsymbol{e}}\|^2$. Thus, the second order derivative of $C_l(\omega)$ can be approximated by only the last term in (66). That is,

$$d_2 \approx -2\hat{\boldsymbol{\alpha}}_l^H \boldsymbol{D}_1^H \boldsymbol{P}_l^\perp \boldsymbol{D}_1 \hat{\boldsymbol{\alpha}}_l \ . \tag{73}$$

If the limit in (29) is used as an approximation, d_2 reduces to

$$\begin{aligned} d_2 \approx & -2\hat{\boldsymbol{\alpha}}_l^H \boldsymbol{D}_1^H \boldsymbol{D}_1 \hat{\boldsymbol{\alpha}}_l + \frac{2r}{N} \hat{\boldsymbol{\alpha}}_l^H \boldsymbol{D}_1^H \boldsymbol{Z}_l \boldsymbol{Z}_l^H \boldsymbol{D}_1 \hat{\boldsymbol{\alpha}}_l \\ \approx & -\frac{2}{r} \hat{\boldsymbol{\alpha}}_l^H \text{diag}(\boldsymbol{1}_r \otimes \boldsymbol{l})^2 \hat{\boldsymbol{\alpha}}_l \sum_{n=0}^{N-1} t_n^2 \\ & + \frac{2}{rN} \hat{\boldsymbol{\alpha}}_l^H \text{diag}(\boldsymbol{1}_r \otimes \boldsymbol{l})^2 \hat{\boldsymbol{\alpha}}_l \left[\sum_{n=0}^{N-1} t_n\right]^2 \\ = & \frac{2}{r} \sum_{i=1}^{l} \hat{A}_i^2 i^2 \left(\frac{1}{N} \left[\sum_{n=0}^{N-1} t_n\right]^2 - \sum_{n=0}^{N-1} t_n^2 \right) \end{aligned} \tag{74}$$

where $\text{diag}(\cdot)$ transforms a vector into a diagonal matrix. The second approximation follows from the limits

$$\lim_{N\to\infty} r \boldsymbol{D}_1^H \boldsymbol{Z}_l \left[\sum_{n=0}^{N-1} t_n\right]^{-1} = (-j)^r \text{diag}(\boldsymbol{1}_r \otimes \boldsymbol{l})(\boldsymbol{J}_r \otimes \boldsymbol{I}_l) \tag{75}$$

$$\lim_{N\to\infty} r \boldsymbol{D}_1^H \boldsymbol{D}_1 \left[\sum_{n=0}^{N-1} t_n^2\right]^{-1} = \text{diag}(\boldsymbol{1}_r \otimes \boldsymbol{l})^2 \ . \tag{76}$$

Under a uniform sampling frequency with no missing samples, $t_n = nT$ and the second-order derivative of $C_l(\omega)$ at $\hat{\omega}$ can be simplified even further since [45, p. 42]

$$\sum_{n=0}^{N-1} t_n = T \sum_{n=0}^{N-1} n = \frac{TN(N-1)}{2} \tag{77}$$

$$\sum_{n=0}^{N-1} t_n^2 = T^2 \sum_{n=0}^{N-1} n^2 = \frac{T^2 N(N-1)(2N-1)}{6} \ . \tag{78}$$

Inserting this into (74) leads to the approximation

$$d_2 \approx -\frac{T^2 N(N^2-1)}{6r} \sum_{i=1}^{l} \hat{A}_i^2 i^2 \ . \tag{79}$$

For a known g, this result has an interesting interpretation since the variance of the fundamental frequency under this approximation is

$$s_l(\hat{\omega}|g) \approx \frac{6r^2 \hat{\sigma}_l^2}{cT^2 N(N^2-1) \sum_{i=1}^{l} \hat{A}_i^2 i^2} \tag{80}$$

which for $c = 1$ is the same as the asymptotic Cramér-Rao lower bound of the fundamental frequency with the true values of the complex amplitudes and the noise variance replaced by their maximum likelihood estimates [10]. For a single real-valued sinusoidal signal, the approximate variance in (80) was also derived in [40] using a different approach.

In summary, an exact expression in (66) and an approximate expression in (74) have been derived for the second-order derivative of $C_l(\omega)$ at $\hat{\omega}$. These expressions are used for computing the variances $s_l(\hat{\omega}|g)$ in (55) and $s_l(\hat{\omega})$ in (57) of the normal approximation to the pdfs $p(\omega|\boldsymbol{x}, g, l)$ and $p(\omega|\boldsymbol{x}, l)$, respectively.

C. Model Comparison

By approximating $p(\omega|\boldsymbol{x}, g, l)$ and $p(\omega|\boldsymbol{x}, l)$ by the normal pdfs derived in the previous section, the integrals in (47) and (49) can be evaluated analytically. An approximation of this form is known as the Laplace approximation. Under the Laplace approximation, the Bayes' factors in (47) and (49) are

$$\text{BF}[l, 0|g] \approx W_l^{-1} f(\hat{\omega}, g, 0) \sqrt{2\pi s_l(\hat{\omega}|g)} \tag{81}$$

$$\text{BF}[l, 0] \approx \frac{(\delta-1){}_2F_1(N/r, 1; l+\delta; R_l^2(\hat{\omega}))\sqrt{2\pi s_l(\hat{\omega})}}{W_l(l+\delta-1)} \ . \tag{82}$$

D. The Gaussian Hypergeometric Function

Unfortunately, the Gaussian hypergeometric function is slow to evaluate so from a computational point of view it might not be advantageous to marginalise g analytically in (43) and (49). Moreover, the use of other priors over g than the hyper-g prior may prohibit analytical marginalisation. Using the Laplace approximation, an approximate way of marginalising (43) and (49) w.r.t. g is therefore derived. Since the marginal posterior pdf over g is not symmetric and in order to avoid edge effect near $g = 0$, the re-parametrisation $\tau = \ln g$ with the Jacobian $dg/d\tau = \exp(\tau)$ is first made [47]. This re-parametrisation suggest that the posterior distribution over g is approximately a log-normal distribution. With this re-parametrisation, the Laplace approximation of the integral in (48) is

$$\int_{\Omega_l}\int_0^\infty f(\omega, g, \delta)dgd\omega$$
$$= \int_{\Omega_l}\int_{-\infty}^\infty \exp(\tau)f(\omega, \exp(\tau), \delta)d\tau d\omega \quad (83)$$
$$= 2\pi\exp(\hat{\tau})f(\hat{\omega}, \exp(\hat{\tau}), \delta)\sqrt{s_l(\hat{\omega}|\exp(\hat{\tau}))\gamma_l(\hat{\tau}|\hat{\omega})} \quad (84)$$

where the mode $\hat{\tau}$ and the variance $\gamma_l(\hat{\tau}|\hat{\omega})$ are given by

$$\hat{\tau} = \ln\left[\frac{\sqrt{\beta_\tau^2 - 4\alpha_\tau} + \beta_\tau}{-2\alpha_\tau}\right] \quad (85)$$

$$\gamma_l(\hat{\tau}|\hat{\omega}) = \frac{r}{\hat{g}}\left[\frac{N(1-R_l^2(\hat{\omega}))}{[1+\hat{g}(1-R_l^2(\hat{\omega}))]^2} - \frac{(N - rl - r\delta)}{(1+\hat{g})^2}\right]^{-1} \quad (86)$$

where $\hat{g} \triangleq \exp(\hat{\tau})$ and

$$\alpha_\tau \triangleq (1 - R_l^2(\hat{\omega}))(1 - l - \delta) \quad (87)$$
$$\beta_\tau \triangleq (N/r - 1)R_l^2(\hat{\omega}) - l - \delta + 2 \ . \quad (88)$$

Thus, the Bayes' factor in (49) is approximately

$$\mathrm{BF}[l, 0] \approx \frac{2\pi(\delta-1)\hat{g}f(\hat{\omega}, \hat{g}, \delta)\sqrt{s_l(\hat{\omega}|\hat{g})\gamma_l(\hat{\tau}|\hat{\omega})}}{W_l} \ , \quad (89)$$

and the normal approximation of the pdf of the fundamental frequency in (43) is approximately

$$p(\omega|\boldsymbol{x}, l) \approx \mathcal{N}_2(\omega; \hat{\omega}, s_l(\hat{\omega}|\hat{g})) \ . \quad (90)$$

VI. Comparison to an ML Estimator

Before evaluating the proposed inference scheme, it is compared to the joint fundamental frequency and model order estimators [2, Sec. 2.6] which is based on the asymptotic MAP rule in [55], [56] and is similar to the rules in, e.g, [24], [25]. Although derived in a ML framework, the method can also be interpreted as an optimal filtering method [20]. Moreover, the same algorithm can be obtained as a special case of one of our approximations based on the BIC-like model selection rule. As stated earlier, the MAP estimate of the fundamental frequency coincides with the ML estimate of the fundamental frequency. Thus, the proposed point estimator of the fundamental frequency is the same as the suggested point estimate in [2]. However, as we treat the fundamental frequency as a random variable, we have also been able to calculate an approximate variance of the fundamental frequency. For model comparison, [2] does not explicitly work with a Bayes' factor. However, it is easy to rewrite their model order estimator as a Bayes' factor. In our notation, it is given by

$$\mathrm{BF}[l, 0] \approx \frac{(\hat{\sigma}_0^2)^N}{(\hat{\sigma}_l^2|_{c=1})^N\sqrt{N^3}N^l} \ . \quad (91)$$

where $\hat{\sigma}_0^2 = \boldsymbol{x}^H\boldsymbol{x}/N$. This Bayes' factor has been derived for complex-valued data using the asymptotic MAP rule proposed in [55], [56]. For a fixed g, a uniform sampling frequency, a complex-valued signal, $T = 1$, and the expression for the variance $s_l(\hat{\omega})$ in (80), our expression for the Bayes' factor may be written as

$$\mathrm{BF}[l, 0|g] \approx \frac{(\hat{\sigma}_0^2)^N\sqrt{2\pi}}{(1+g)^l W_l(\hat{\sigma}_l^2)^N}\sqrt{\frac{(1+g)6\hat{\sigma}_l^2}{gN(N^2-1)\sum_{i=1}^l \hat{A}_i^2 i^2}} \ . \quad (92)$$

For $g^{\mathrm{BIC}} = N$ and $N \gg 1$, $\mathrm{BF}[l, 0|g]$ is

$$\mathrm{BF}[l, 0|g] \approx \sqrt{\frac{12\pi}{W_l^2\sum_{i=1}^l \hat{A}_i^2 i^2}}\frac{(\hat{\sigma}_0^2)^N}{(\hat{\sigma}_l^2|_{c=1})^N\sqrt{N^3}N^l} \ . \quad (93)$$

Comparing this with (91), we see that the model order estimator in [2] implicitly assumes that $g^{\mathrm{BIC}} = N$, $N \gg 1$, and

$$\sqrt{\frac{12\pi}{W_l^2\sum_{i=1}^l \hat{A}_i^2 i^2}} \approx 1 \ . \quad (94)$$

VII. Simulations

In this section, the accuracy of the various approximations introduced in Sec. V is first evaluated on a synthetic signal. All possible combinations of the approximations are not evaluated, but only the most important ones. These are the various approximations of the posterior pdfs on the fundamental frequency and model order, respectively, for an unknown value of g. Second, the proposed inference scheme is evaluated on a female speech signal[3].

A. Synthetic Signal

To evaluate the accuracy of the various approximations introduced in Sec. V, Monte Carlo simulations was used for various SNRs. Every Monte Carlo realisation consisted of $N = 100$ data points and was sampled uniformly from a complex-valued, periodic, and synthetic signal. The SNR of the signal was varied in steps of 1 dB from -10 dB to 10 dB, and 500 realisations were generated for every SNR. The fundamental frequency ω was assumed to be smaller than $\omega_b = \pi(lT)^{-1}$ so that the frequency of the highest harmonic component was below the Nyquist frequency. For numerical reasons, ω was also assumed to be larger than $\omega_a = 2\pi(NT)^{-1}$.

An overview over the various approximations are given in Table I. In the rows marked with ■, an estimate of g is used whereas g is treated as a random variable in the rows marked

[3]The Matlab code used to generate the simulation results are available at http://kom.aau.dk/~jkn/publications/publications.php. Moreover, a real-time Python implementation of the algorithm is also available.

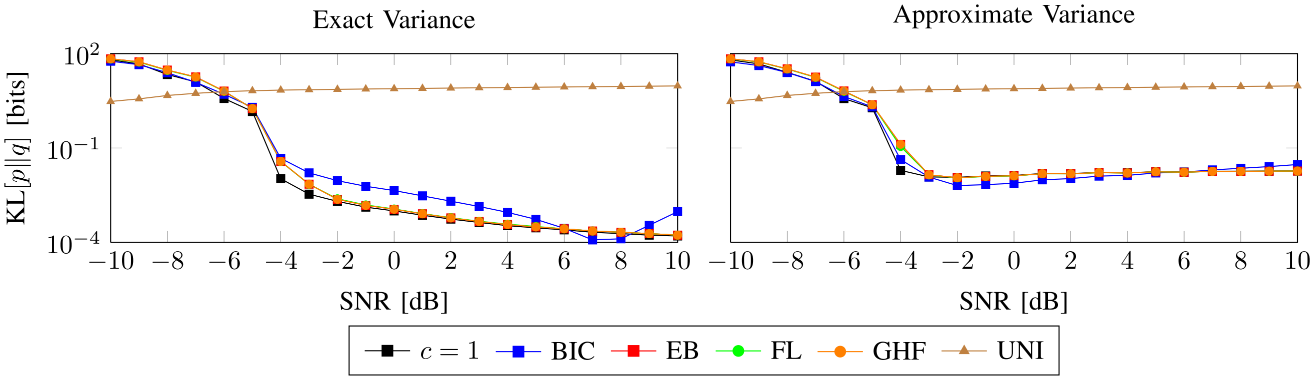

Fig. 1. The accuracy of the normal approximation of the fundamental frequency under the variance calculations in (66) and (74), respectively, for various SNRs and choices of g. Note that 'EB', 'FL', and 'GHF' are almost coinciding in the two plots.

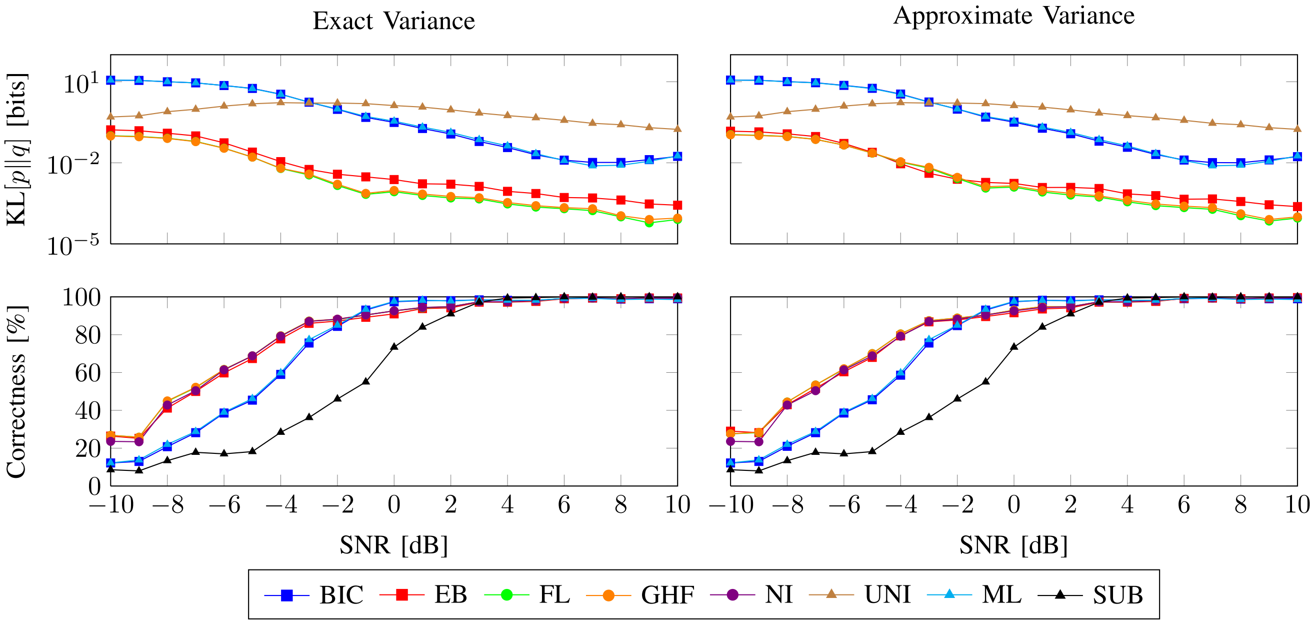

Fig. 2. The accuracy of the various approximation of the posterior pdf on the model order under the variance calculations in (66) and (74), respectively, for various SNRs and choices of g. Note that the curve labelled 'UNI' and the curves labelled 'ML' and 'SUB' are only in the plots in the top and bottom row, respectively.

with •. The rows marked with ▲ are used for reference and comparison to other algorithms. For the first five rows, either the exact or the approximate expressions can be used for the second order derivative of $C_l(\omega)$ given by (66) and (74), respectively.

1) The Distribution on the Fundamental Frequency: In order to measure the distance between $p(\omega|\boldsymbol{x}, l)$ and its normal approximation, the relative entropy or Kullback-Leibler (KL) divergence was used. It is given by [57]

$$\mathrm{KL}(p\|q) = \int_{\Omega_l} p(\omega|\boldsymbol{x}, l) \log_2 \left[\frac{p(\omega|\boldsymbol{x}, l)}{q(\omega|\boldsymbol{x}, l)}\right] d\omega \tag{95}$$

where $q(\omega|\boldsymbol{x}, l)$ is an approximation of $p(\omega|\boldsymbol{x}, l)$. The KL divergence is finite only if the support of $p(\omega|\boldsymbol{x}, l)$ is contained in Ω_l. Moreover, the KL divergence satisfies that $\mathrm{KL}(p\|q) \geq 0$ with equality if and only if $p(\omega|\boldsymbol{x}, l) = q(\omega|\boldsymbol{x}, l)$. For the true pdf, (49) was used, and the KL divergence was evaluated using numerical integration on a fine uniform grid consisting of 10,000 points. Fig. 1 shows the average KL divergence between $p(\omega|\boldsymbol{x}, l)$ and $q(\omega|\boldsymbol{x}, l)$ for a known model order of $l = 4$. For low SNRs, the KL divergence is large since $q(\omega|\boldsymbol{x}, l)$ consists of multiple significant peaks which cannot

ID	type	$p(\omega\|\boldsymbol{x}, l)$	BF$[l, 1]$	g
$c = 1$	■	(54)		∞
BIC	■	(54)	(81), (46)	N
EB	■	(54)	(81), (46)	(50)
FL	•	(54)	(89), (46)	(85)
GHF	•	(56)	(82), (46)	
NI	•		(52), (46)	
UNI	▲	W_l^{-1}	L^{-1}	
ML	▲		(91), (46)	
SUB	▲		[2, Sec. 4.6]	

TABLE I
OVERVIEW OVER THE VARIOUS APPROXIMATIONS.

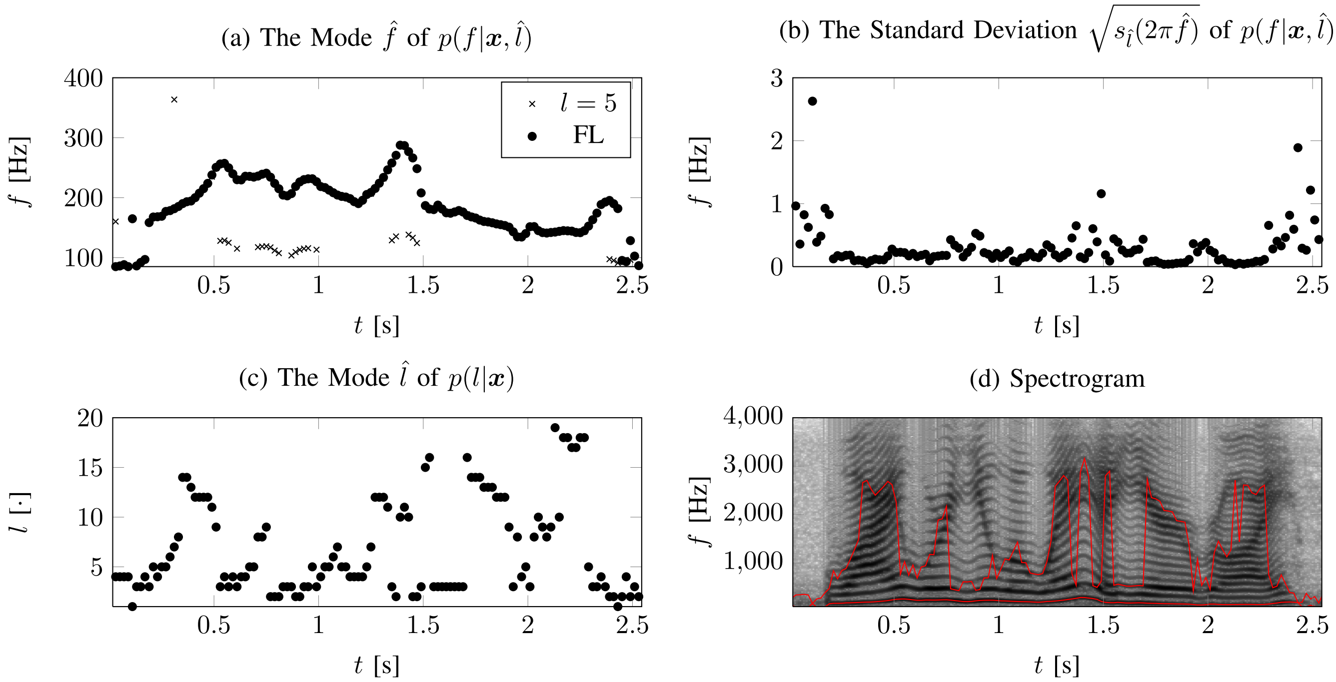

Fig. 3. The estimation of the fundamental frequency and the model order for a speech signal. Plot (a) and (b) show the estimated fundamental frequency $\hat{f}$ and its standard deviation, respectively, for the estimated model order $\hat{l}$ which is shown in plot (c). Plot (d) shows the spectrogram of the speech signal. The spectrogram has been overlaid with the estimated frequencies for the fundamental and largest harmonic components, respectively.

be approximated accurately by the normal approximation assumed in Ass. 5.1. As the SNR increases, however, only a single peak remains significant, and the KL divergence therefore decreases, meaning that Ass. 5.1 does approximately hold for an SNR larger than approximately -4 dB (in this simulation example). Below an SNR of approximately -4 dB, the KL divergence is insensitive to the choice of the variance for the normal approximation. However, above -4 dB, the choice matters. For the approximate variance, the KL divergence seems to exhibit a thresholding effect caused by the use of the approximations in (53), (75), and (76). This threshold will be lowered if N is increased.

2) Model Comparison: In order to evaluate the accuracy of the posterior pmf on the model order, we used the same procedure as in the previous section. Moreover, the model selection properties of the proposed inference scheme was also evaluated and compared to the ML-based algorithm in [2, Sec. 2.6] and the subspace-based algorithm in [2, Sec. 4.6]. The discrete version of the KL divergence is given by [57]

$$\mathrm{KL}(p\|q) = \sum_{l=1}^{L} p(l|\boldsymbol{x}) \ln \left[\frac{p(l|\boldsymbol{x})}{q(l|\boldsymbol{x})} \right] \ , \qquad (96)$$

and it is used to assess the accuracy of the posterior pmf $q(l|\boldsymbol{x})$ on the model order for the various approximations in Table I. For the true pmf $p(l|\boldsymbol{x})$, the 'NI' approximation based on the numerical integration on a very fine frequency grid was used. For the prior pmf $p(l)$ over the model order, a uniform prior was used so that the posterior pmf on the model order is proportional to the Bayes' factor. The same Monte Carlo simulation setup as above was used but with an unknown model order. Specifically, for each Monte Carlo run, the model order was generated from its prior with the minimum and maximum model order being 1 and $L = 10$, respectively. Since the all-noise model was not in the set of candidate models, the improper prior was used on g, and it is obtained by letting $\delta = 1$. The top row of Fig. 2 shows the results of measuring the average KL divergence between $p(l|\boldsymbol{x})$ and $q(l|\boldsymbol{x})$. For all SNRs, the full Laplace 'FL' and the 'GHF' approximations performs slightly better than the approximation based on the emperical bayes 'EB' estimate of g. All of these three approximations perform much better than the 'ML' and the 'BIC' approximations. As shown in Sec. VI, the 'ML' approximations is a special case of the 'BIC' approximation which explains why the 'ML' and the 'BIC' approximations seem to have the same accuracy. In each Monte Carlo run, the most probable model was selected and compared to the true model, and the bottom row of Fig. 2 shows the proportion of correctly selected model orders for the various SNRs. For SNRs below -2 dB, the 'FL', 'GHF', and 'NI' approximations were better than the other approximations. However, from -2 dB to approximately 3 dB, the 'ML' and 'BIC' approximations were slightly better at finding the true model order. For an SNR above 3 dB, all of the models performed equally well.

Thus, for model selection purposes, there is no best method for all SNRs. However, for problems such as model averaging and estimation in which all models are used, the approximations based on a random g seem to outperform the other approximations for all SNRs.

B. Speech Signal

In the last simulation, the applicability of the proposed algorithm was demonstrated to the problem of estimating the fundamental frequency and model order of a speech signal. The speech signal originates from a female voice uttering *"Why were you away a year, Roy?"* which has been sampled at a uniform sampling frequency of 8 kHz. Since the signal is real, the down-sampled analytic signal was first computed as described in the introduction. Subsequently, the signal was partitioned into consecutive frames of 20 ms corresponding to $N = 80$ samples. The minimum and maximum candidate model order were set to 1 and $L = 20$, respectively, and the bandwidth of the signal was set to the interval $[85\text{ Hz}, 4000\text{ Hz}]$ where the lower limit is the typical lower limit of human voiced speech [58, Ch. 6]. For the estimation of the fundamental frequency, the approximate MAP estimate was first estimated using (59). Second, a refined estimate was found using a Dichotomous search with the exact cost-function in (58). The posterior pmf for the model order was estimated using the 'FL' approximation (see Table I) with the approximate variance in (80). We have found that the above algorithm provides a good balance between computational load and estimation accuracy. The results of running the algorithm is shown in Fig. 3. Plot (a) and (b) show the MAP estimate and the standard deviation, respectively, of the fundamental frequency for the estimated model order which is shown in plot (c). In plot (a), the estimated fundamental frequency is also shown for a fixed model order of $l = 5$. We clearly see that the estimator based on a fixed model order suffers from pitch halving, and this illustrates why model order selection is important even if only the estimate of the fundamental frequency is interesting. In plot (d), the frequencies of the fundamental and largest harmonic components are shown on top of the spectrogram of the speech signal. We clearly see that the algorithm provided accurate estimates of the fundamental frequency and the model order even though the signal is not perfectly periodic.

VIII. CONCLUSION

In the first part of this paper, we have argued for and derived a default probability model for both a real- and complex-valued periodic signal in additive noise. Using Jaynes' principles of maximum entropy and transformation groups, the prior information in Ass. 3.1 was turned into an observation model and prior distributions on the model parameters. Subsequently, the prior distributions were turned into a more convenient prior of the same form as the g-prior using a few minor approximations on the signal-to-noise-ratio (SNR) and the number of observations. The g-prior is parametrised by the parameter g which is very important for performing model comparison. Several ways of estimating a value for it was given, and it was also treated as a random variable.

In the second part, the posterior distribution was derived for the fundamental frequency. Moreover, an integral representation of the posterior distributions on the model order was derived for both a known and an unknown value of g. Several approximations to these posterior distributions was also suggested, and it was shown that the state-of-the-art ML estimator is a special case of the approximation based on the Bayesian information criterion.

In the last part of this paper, the various approximations were compared in the simulation section on a synthetic signal. The simulations indicated that the value of g is not important for the posterior distribution on the fundamental frequency. For model comparison, however, the value of g was very important, and the most accurate approximations was obtained when g was treated as a random variable. The BIC approximation is worse than the other approximations. It was also demonstrated that one of the approximations was able to accurately estimate the fundamental frequency and model order of a voiced speech segment which was not perfectly periodic.

REFERENCES

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. Springer, Jun. 1998.

[2] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, B. H. Juang, Ed. Morgan & Claypool, 2009.

[3] H. Dudley, "The carrier nature of speech," *Bell System Technical Journal*, vol. 19, no. 4, pp. 495–515, Oct. 1940.

[4] R. J. Sluijter, "The development of speech coding and the first standard coder for public mobile telephony," Ph.D. dissertation, Techniche Universiteit Eindhoven, 2005.

[5] G. L. Ogden, L. M. Zurk, M. E. Jones, and M. E. Peterson, "Extraction of small boat harmonic signatures from passive sonar," *J. Acoust. Soc. Am.*, vol. 129, no. 6, pp. 3768–3776, Jun. 2011.

[6] V. K. Murthy, L. J. Haywood, J. Richardson, R. Kalaba, S. Salzberg, G. Harvey, and D. Vereeke, "Analysis of power spectral densities of electrocardiograms," *Mathematical Biosciences*, vol. 12, no. 1–2, pp. 41–51, Oct. 1971.

[7] M. Davy, S. J. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2498–2517, Apr. 2006.

[8] S. L. Marple, Jr., "Computing the discrete-time "analytic" signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[9] M. G. Christensen, "Accurate estimation of low fundamental frequencies from real-valued measurements," 2012, unpublished manuscript.

[10] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1635–1644, Jul. 2007.

[11] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002.

[12] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 1, pp. 24–33, Feb. 1977.

[13] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41, no. 2, pp. 293–309, Feb. 1967.

[14] W. Hess, *Pitch Determination of Speech Signals: Algorithms and Devices*. Springer, Apr. 1983.

[15] B. Gold and L. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *J. Acoust. Soc. Am.*, vol. 46, no. 2B, pp. 442–448, Aug. 1969.

[16] B. G. Quinn and P. J. Thomson, "Estimating the frequency of a periodic function," *Biometrika*, vol. 78, no. 1, pp. 65–74, Mar. 1991.

[17] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88, no. 4, pp. 972–983, Apr. 2008.

[18] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation using harmonic music," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2006, pp. 521–524.

[19] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1124–1138, Oct. 1986.

[20] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP J. on Advances in Signal Process.*, vol. 13, pp. 1–18, Jun. 2011.

[21] S. J. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 2002, pp. 1769–1772.
[22] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 679–694, Mar. 2006.
[23] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[24] B. G. Quinn, "Estimating the number of terms in a sinusoidal regression," *J. of Time Series Analysis*, vol. 10, no. 1, pp. 71–75, Jan. 1989.
[25] L. Kavalieris and E. J. Hannan, "Determining the number of terms in a trigonometric regression," *J. of Time Series Analysis*, vol. 15, no. 6, pp. 613–625, Nov. 1994.
[26] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
[27] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 450–458, Feb. 2006.
[28] J.-M. Papy, L. De Lathauwer, and S. Van Huffel, "A shift invariance-based order-selection technique for exponential data modelling," *IEEE Signal Process. Lett.*, vol. 14, no. 7, pp. 473–476, Jul. 2007.
[29] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. on Advances in Signal Process.*, vol. 2009, pp. 1–11, Nov. 2009.
[30] P. Stoica, Y. Selén, and J. Li, "Multi-model approach to model selection," *Digital Signal Process.*, vol. 14, no. 5, pp. 399–412, Sep. 2004.
[31] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, Jun. 2002.
[32] J. O. Berger and L. R. Pericchi, "The intrinsic Bayes factor for model selection and prediction," *J. Amer. Statistical Assoc.*, vol. 91, no. 433, pp. 109–122, Mar. 1996.
[33] ——, "Objective Bayesian methods for model selection: Introduction and comparison," *Institute of Mathematical Statistics Lecture Notes – Monograph Series*, vol. 38, pp. 135–207, 2001.
[34] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer-Verlag New York, Inc., Jul. 2004.
[35] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Elsevier, 1995, vol. 495, ch. 14, pp. 495–518.
[36] H. Kawahara, A. de Cheveigné, H. Banno, T. Takahashi, and T. Irino, "Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT," in *Proc. Interspeech*, 2005, pp. 537–540.
[37] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 3, pp. 227–241, 1968.
[38] ——, *Probability Theory : The Logic of Science*, G. L. Bretthorst, Ed. Cambridge University Press, Apr. 2003.
[39] G. L. Bretthorst, "The near-irrelevance of sampling frequency distributions," in *Max. Entropy and Bayesian Methods*, 1999, pp. 21–46.
[40] E. T. Jaynes, "Bayesian spectrum and chirp analysis," in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, C. R. Smith and G. J. Erickson, Eds. D. Reidel, Dordrecht-Holland, 1987, pp. 1–37.
[41] J. M. Bernardo and A. Smith, *Bayesian Theory*, 1st ed. John Wiley and Sons Ltd, 1994.
[42] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. Springer-Verlag, Berlin Heidelberg, 1988.
[43] H. Jeffreys, *Theory of Probability*. Oxford University Press, 1939.
[44] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Prentice Hall, May 2005.
[45] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall PTR, Mar. 1993.
[46] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*. Elsevier, 1986.
[47] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *J. Amer. Statistical Assoc.*, vol. 103, pp. 410–423, Mar. 2008.
[48] R. Strachan and H. K. v. Dijk, "Improper priors with well defined Bayes' factors," Department of Economics, University of Leicester, Discussion Papers in Economics 05/4, 2005.
[49] J. M. Marin and C. P. Robert, *Bayesian core: a practical approach to computational Bayesian statistics*, 1st ed. Springer, Feb. 2007.
[50] R. Guo and P. L. Speckman, "Bayes factor consistency in linear models," in *The 2009 International Workshop on Objective Bayes Methodology*, Jun. 2009.
[51] J. O. Berger, J. M. Bernardo, and D. Sun, "The formal definition of reference priors," *Ann. Stat.*, vol. 37, no. 2, pp. 905–938, Apr. 2009.
[52] I. S. Gradshteĭn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*. Academic Press, 2000.
[53] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Jul. 2003.
[54] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 141–148, Jan. 1998.
[55] P. M. Djuric, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1744–1751, Jul. 1996.
[56] ——, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.
[57] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, Jul. 2006.
[58] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. Singular, 1999.

Jesper Kjær Nielsen (S'12) was born in Struer, Denmark, in 1982. He received the B.Sc., M.Sc. (Cum Laude), and Ph.D. degrees in electrical engineering from Aalborg University, Aalborg, Denmark, in 2007, 2009, and 2012, respectively.

He is currently with the Department of Electronic Systems, Aalborg University, and Bang & Olufsen as an industrial postdoctoral researcher. He has been a Visiting Scholar at the Signal Processing and Communications Laboratory, University of Cambridge and at the Department of Computer Science, University of Illinois at Urbana-Champaign. His research interests include spectral estimation, (sinusoidal) parameter estimation as well as statistical and Bayesian methods for signal processing.

Mads Græsbøll Christensen (S'00–M'05–SM'11)) was born in Copenhagen, Denmark, in March 1977. He received the M.Sc. and Ph.D. degrees from Aalborg University, Denmark, in 2002 and 2005, respectively.

He was formerly with the Department of Electronic Systems, Aalborg University, and is currently an Associate Professor in the Department of Architecture, Design and Media Technology. He has been a Visiting Researcher at Philips Research Labs, Ecole Nationale Supérieure des Télécommunications (ENST), University of California, Santa Barbara (UCSB), and Columbia University. He has published more than 100 papers in peer-reviewed conference proceedings and journals and is coauthor (with A. Jakobsson) of the book *Multi-Pitch Estimation* (Morgan & Claypool Publishers, 2009). His research interests include digital signal processing theory and methods with application to speech and audio, in particular parametric analysis, modeling, enhancement, separation, and coding.

Dr. Christensen has received several awards and prestigious grants, including an ICASSP Student Paper Award, the Spar Nord Foundation's Research Prize for his Ph.D. thesis, a Danish Independent Research Council postdoc grant and Young Researcher's Award, and a Villum Foundation Young Investigator Programme grant. He is an Associate Editor for the IEEE Signal Processing Letters.

Søren Holdt Jensen (S'87–M'88–SM'00) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNI•C), Lyngby; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University.

He is Full Professor and is currently heading a research team working in the area of numerical algorithms, optimization, and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications.

Prof. Jensen was an Associate Editor for the IEEE Transactions on Signal Processing and Elsevier Signal Processing, and is currently Associate Editor for the IEEE Transactions on Audio, Speech and Language Processing and EURASIP Journal on Advances in Signal Processing. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and Founder and Chairman of the IEEE Denmark Section's Signal Processing Chapter. He is member of the Danish Academy of Technical Sciences and was in January 2011 appointed as member of the Danish Council for Independent Research—Technology and Production Sciences by the Danish Minister for Science, Technology and Innovation.