AALBORG UNIVERSITY

Accurate Estimation of Low Fundamental Frequencies from Real-Valued Measurements

Christensen, Mads Græsbøll

Publication date:
2013

Document Version
Early version, also known as pre-print

Link to publication from Aalborg University

Accurate Estimation of Low Fundamental Frequencies from Real-Valued Measurements

Mads Græsbøll Christensen, *Senior Member, IEEE*

Abstract— In this paper, the difficult problem of estimating low fundamental frequencies from real-valued measurements is addressed. The methods commonly employed do not take the phenomena encountered in this scenario into account and thus fail to deliver accurate estimates. The reason for this is that they employ asymptotic approximations that are violated when the harmonics are not well-separated in frequency, something that happens when the observed signal is real-valued and the fundamental frequency is low. To mitigate this, we analyze the problem and present some exact fundamental frequency estimators that are aimed at solving this problem. These estimators are based on the principles of nonlinear least-squares, harmonic fitting, optimal filtering, subspace orthogonality, and shift-invariance, and they all reduce to already published methods for a high number of observations. In experiments, the methods are compared and the increased accuracy obtained by avoiding asymptotic approximations is demonstrated.

I. INTRODUCTION

Signals that are periodic can be decomposed into a sum of sinusoids having frequencies that are integer multiples of a fundamental frequency, much like the well-known Fourier series, except that real-life signals are noisy and are not observed over an integer number of periods. The problem of finding this fundamental frequency is referred to as fundamental frequency estimation or sometimes as pitch estimation, with the latter term referring to the perceptual attribute that is associated with sound waves exhibiting periodicity. Many signals that can be encountered by the signal processing practitioner are periodic or approximately so. This is, for example, the case in speech processing, where voiced speech exhibits such characteristics, and in music processing for tones produced by musical instruments. Also in the analysis of some bird calls and various other biological signals, like vital signs [1], such signals can be encountered. Moreover, they occur in radar applications for rotating targets [2] and in passive detection, localization, and identification of boats and helicopters [3]. It is then also not surprising that a host of methods have been proposed over the years including methods based on the principles of maximum likelihood, least-squares (LS), and weighted least-squares (WLS) [4]–[8], auto-/cross-correlation and related methods [9]–[13], linear prediction [14], filtering [2], [15]–[17], and subspace methods [18], [19]. We note in passing that several of the cited methods can be interpreted in more than one way and may therefore be

considered as belonging to several of the categories above. For an introduction to the fundamental frequency estimation problem and an overview of fundamental frequency estimators, we refer the interested reader to [20].

We are here concerned with a specific problem in determining the fundamental frequency under certain circumstances. When the fundamental frequency of a periodic signal is low as compared to the number of samples, the harmonics of the signal are closely spaced in its spectrum, as the distance between harmonics is given by the fundamental frequency. A similar effect comes into play when the observed signal is real (when we say that some quantity is real, we mean that it is real-valued, i.e., its imaginary part is zero). In this case, harmonics occur in the spectrum not only at positive integer multiples of a fundamental, but also for negative, as complex conjugate pairs of complex sinusoids combine to yield real signals. Again, the distance between the individual complex sinusoids is given by the fundamental frequency. The problem here is that when harmonics are close in frequency, and they are far from being orthogonal, they will interact. As such, this is not really a problem, but most of the parametric methods in the literature ignore this. The reason for this is simple: by ignoring the interaction, one obtains simpler estimators that can be implemented efficiently using, for example, the fast Fourier transform (FFT) or polynomial rooting methods. An example of this is the so-called harmonic summation method [4], in which an approximate maximum likelihood estimate of the fundamental frequency is obtained by summing the power spectral density sampled at candidate fundamental frequencies and picking the one that yields the highest power. This method is accurate when the number of samples approaches infinity, but it fails to take the interaction into account for finite length signals. From the above discussion, it should also be clear that when the fundamental frequency is high relative to the number of available samples, there is essentially no error in using a complex model for a real-valued signal.

Interestingly, the problem of taking the nature of real signals into account has been addressed in the frequency estimation literature, i.e., for the case where sinusoids are not constrained to being integer multiples of a fundamental frequency. Some examples of adaptations of well-known estimators to this problem are for maximum likelihood methods [21], [22], subspace methods [23], [24], Capon's method [25], and the linear prediction [26] method.

It is possible to bound the performance of estimators by computing the Cramér-Rao lower bound (CRLB), which is a lower bound on the variance of an unbiased estimator. This has also been done for the problem of estimating the fundamental frequency [2], [18]. These show that the expected

performance (of an optimal estimator) does not depend on the fundamental frequency. At first glance, this seems to contradict the premise of this paper. However, upon closer inspection, it turns out that these bounds were derived based on asymptotic approximations relying on the number of samples approaching infinity or being *sufficiently* large. In former case, the support of spectrum of the sinusoids reduces to a single point, and, hence, the interaction between sinusoids will be zero as long as the fundamental frequency is different from zero, a trivial case that is of no interest anyway.

In this paper, we aim to analyze and solve this problem in a systematic manner. We define the problem of interest with complex and real signal models and analyze it using using what we refer to as the exact CRLB. Then, a number of solutions to the problem are presented, some of which are new, some of which are known, namely a nonlinear least-squares method, an optimal filtering method, a subspace method based on angles between subspaces, and, finally, a method based on a WLS fitting of unconstrained frequencies (called harmonic fitting). The presented methods have in common that they avoid the use of asymptotic approximations, whenever possible, and they take the real-valued nature of the observed signal into account. The nonlinear least-squares method is well-documented in the literature having been applied to many problems, including also frequency estimation and fundamental frequency estimation [5], [6], [8]. The optimal filtering method, which is based on constrained optimization, was originally proposed in [8], but only for complex signals. Here, the underlying constraints are modified to fit real signals. The method based on angles between subspaces is an exact version of the MUSIC-based methods of [8], [18] both of which employ an approximate measure of subspace orthogonality as introduced in [27]. The connection between the exact and approximate measures of the angles between subspaces was first analyzed in [28], but was only used for deriving an approximate, normalized measure for order estimation and, hence, not for fundamental frequency estimation. The harmonic fitting method was originally proposed in [6], but employed a weighting of the individual harmonics derived based on asymptotic properties, while we here avoid using these. In simulations, the effectiveness of these methods is then investigated and their performance compared to the exact CRLB, and the problem is analyzed via comparisons of the asymptotic and exact CRLBs.

The remainder of the present paper is organized as follows: In Section II, we introduce the problem and the signal models and proceed to derive the corresponding CRLB. In the section that follows, namely Section III, we present some methods for solving the problem. We then present the experimental results in Section IV, after which we conclude on our work in Section V.

II. PRELIMINARIES

A. Model and Problem Definition

We will now proceed to define the problem of interest and the associated signal model. The observed real signal $x(n)$ is composed of a set of L sinusoids having frequencies that are integer multiples of a fundamental frequency ω_0, real amplitude $A_l > 0$, and phases $\phi_l \in [0, 2\pi)$. Aside from the sinusoids, we assume that an additive noise source $e(n)$ is present. This noise source represents all stochastic signal components, even those that are integral parts of natural signals that may be of interest to us in other cases. It is here assumed to be white Gaussian distributed having variance σ^2 and zero mean, although this is strictly speaking not necessary for all the presented methods. Mathematically, the observed signal can be expressed for $n = 0, \ldots, N-1$ as

$$x(n) = \sum_{l=1}^{L} A_l \cos\left(\omega_0 l n + \phi_l\right) + e(n). \qquad (1)$$

The problem is then to estimate ω_0 from $x(n)$. For a given L, the fundamental frequency can be in the range $\omega_0 \in (0, \frac{\pi}{L})$. Regarding the remaining unknown parameters, some comments are in order. The model order, L, (also referred to as the number of harmonics) can be found a variety of ways and it is possible to solve jointly for the fundamental frequency and the model order, something that has been done for all the methodologies employed here (see [20]), and the extension of these principles to the estimator presented herein is fairly straightforward for which reason we defer from any further discussion of this problem. Once the fundamental frequency and the model order L has been found, the corresponding phases and amplitudes can be found using one of the many existing amplitude estimators [20], [29]. Compared to the problem of estimating the fundamental frequency, this is fairly easy, as these parameters are linear. We note that for $L = 1$, the model above reduces to a single real sinusoid and the associated estimation problem to the usual frequency estimation problem.

Regarding the realism of the model (1), there are several issues that may be a concern. First, the amplitudes, phases and frequencies are assumed to be constant for the duration of the N samples. Since natural sources most often are time-varying, N should be chosen sufficiently low so that the model is a good approximation of the observed signal. Second, the frequencies of the harmonics are assumed to be integer multiples of the fundamental frequency. This should be considered an approximation too, as natural signals may exhibit deviations from this for variety of reasons. We note in passing that a number of modified signal models that take this into account exist [20], [30]. Since these are widely application and signal specific and we wish to retain the generality of the presented material, we will not go further into details on this matter. Third, the noise was assumed to be Gaussian and white. Regarding the Gaussian assumption, this appears to be the norm in the literature, and, in our experience, it does not appear to be a major shortcoming of existing methods used in speech and audio processing. It should also be noted that even though several of the estimators herein are derived based on this assumption, the estimators may still be accurate, at least asymptotically so, even if the assumption does not hold [31]. Moreover, the white Gaussian distribution can be shown to be the one that maximizes the entropy of the noise [32], i.e., it is a worst case scenario. For colored noise, one can apply pre-

whitening [5], [33], i.e., a filtering, to render the noise white, or, at least, more close to being white than it was prior to the pre-whitening. Fourth, the noise was assumed to have zero mean and no DC offset (0 frequency component) is included in the deterministic part of (1). This is mostly done for simplicity. The presence of such a component can, though, be addressed in several ways: a) the presented estimators can be extended by including the zero frequency component having an unknown amplitude [31]; b) the mean can be estimated (and removed) a priori as it is typically caused by calibration errors in microphones and constant outside $n = 0, \ldots, N-1$; c) the signal of interest can be preprocessed using a simple DC blocking filter.

This signal model in (1) can also be expressed using complex sinusoids as

$$x(n) = \sum_{l=-L}^{L} a_l e^{j\omega_0 ln} + e(n), \quad (2)$$

with $a_l = a_{-l}^*$ and $a_0 = 0$. In this notation, the phase and amplitude have been combined into a complex amplitude as $a_l = \frac{A_l}{2} e^{j\phi_l}$ and $(\cdot)^*$ denotes the complex-conjugate. It should be stressed that no additional assumptions have been used in going from (1) to (2), which means that (2) is exact. The error in applying a complex model arises when modifying (2) into $x(n) \approx \sum_{l=1}^{L} a_l e^{j\omega_0 ln} + e(n)$, i.e., when assuming that only half the complex sinusoids are there. This essentially ignores the interaction between the complex sinusoids having frequencies $\{\omega_0 l\}_{l=1}^{L}$ and $\{-\omega_0 l\}_{l=1}^{L}$. Another frequently used approach is to convert (1) into a complex model via the Hilbert transform, which can be used to compute the so-called discrete-time analytic signal. However, the error committed in this process is essentially the same (aside from the suboptimality of the finite-length Hilbert transform), and they are both accurate under the same conditions, namely that ω_0 is not close to 0 relative to N.

B. Cramér-Rao Lower Bound and Further Definitions

An estimator is said to be unbiased if the expected value of its estimate $\hat{\theta}_i$ of the ith parameter θ_i of the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$ is identical to the true parameter for all possible values of the true parameter, i.e., $\mathrm{E}\left\{\hat{\theta}_i\right\} = \theta_i \ \forall \theta_i$. The difference, i.e., $\theta_i - \mathrm{E}\left\{\hat{\theta}_i\right\}$, is referred to as the bias. The CRLB is a lower bound on the variance of an unbiased estimate of a parameter, say θ_i, and it is given by $\mathrm{var}(\hat{\theta}_i) \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{ii}$. Here, the notation $[\mathbf{I}(\boldsymbol{\theta})]_{il}$ means the ilth entry of the matrix $\mathbf{I}(\boldsymbol{\theta})$ and $\mathrm{var}(\cdot)$ denotes the variance. Furthermore, $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix defined as

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}\left\{\frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right\}, \quad (3)$$

where $p(\mathbf{x}; \boldsymbol{\theta})$ is the likelihood function of the observed signal parametrized by the parameters $\boldsymbol{\theta}$. For the case of Gaussian signals with $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{Q})$ where $\mathbf{Q}$ is the noise covariance matrix (which is not parametrized by any of the parameters in

$\boldsymbol{\theta}$) and $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean, the likelihood function is given by

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{\det\left(2\pi\mathbf{Q}\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{Q}^{-1}(\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))}. \quad (4)$$

For this case, Slepian-Bang's formula [34] can be used for determining a more specific expression for the Fisher information matrix. More specifically, it is given by

$$[\mathbf{I}(\boldsymbol{\theta})]_{nm} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial \theta_n} \mathbf{Q}^{-1} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_m}. \quad (5)$$

For the problem and signal model considered here, the involved quantities are given by:

$$\mathbf{x} \triangleq \left[\begin{array}{ccc} x(0) & \cdots & x(N-1) \end{array}\right]^T$$
$$\mathbf{Q} \triangleq \sigma^2 \mathbf{I}$$
$$\boldsymbol{\theta} \triangleq \left[\begin{array}{ccccc} \omega_0 & A_1 & \phi_1 & \cdots & A_L & \phi_L \end{array}\right]^T$$
$$\boldsymbol{\mu}(\boldsymbol{\theta}) \triangleq \mathbf{Z}\mathbf{a}$$
$$\mathbf{Z} \triangleq \left[\begin{array}{cccc} \mathbf{z}(\omega_0) & \mathbf{z}^*(\omega_0) & \cdots & \mathbf{z}(\omega_0 L) & \mathbf{z}^*(\omega_0 L) \end{array}\right],$$
$$\mathbf{a} \triangleq \frac{1}{2}\left[\begin{array}{ccccc} A_1 e^{j\phi_1} & A_1 e^{-j\phi_1} & \cdots & A_L e^{j\phi_L} & A_L e^{-j\phi_L} \end{array}\right]^T$$
$$\mathbf{z}(\omega_0 l) \triangleq \left[\begin{array}{cccc} 1 & e^{j\omega_0 l 1} & \cdots & e^{j\omega_0 l (N-1)} \end{array}\right]^T.$$

Note that we will make extensive use of these definitions later. In relation to the problem at hand, some observations about the nature of the matrix $\mathbf{Z}$ can be made: Firstly, for $\omega_0 \neq 0$ and $\omega_0 \in (0, \frac{\pi}{L})$, $\mathbf{Z}$ has full rank. However, for $\omega_0 = 0$, it will be rank deficient and as $\omega_0 \to 0$, the condition number of $\mathbf{Z}$ will tend to infinity and the involved estimation problem is basically ill-posed.

With the above in place, we now have to determine the following derivatives:

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \omega_0} = \frac{\partial \mathbf{Z}}{\partial \omega_0}\mathbf{a}, \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial A_l} = \mathbf{Z}\frac{\partial \mathbf{a}}{\partial A_l}, \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \phi_l} = \mathbf{Z}\frac{\partial \mathbf{a}}{\partial \phi_l}, \quad (6)$$

which, in turn, require that the following be computed:

$$\frac{\partial \mathbf{Z}}{\partial \omega_0} = \left[\begin{array}{ccccc} \frac{\partial \mathbf{z}(\omega_0)}{\partial \omega_0} & \frac{\partial \mathbf{z}^*(\omega_0)}{\partial \omega_0} & \cdots & \frac{\partial \mathbf{z}(\omega_0 L)}{\partial \omega_0} & \frac{\partial \mathbf{z}^*(\omega_0 L)}{\partial \omega_0} \end{array}\right]$$
$$\frac{\partial \mathbf{z}(\omega_0 l)}{\partial \omega_0} = \left[\begin{array}{cccc} 0 & jle^{j\omega_0 l} & \cdots & j(N-1)le^{j\omega_0 l(N-1)} \end{array}\right]^T \quad (7)$$
$$\frac{\partial \mathbf{a}}{\partial A_l} = \frac{1}{2}\left[\begin{array}{cccccccc} 0 & \cdots & 0 & e^{j\phi_l} & e^{-j\phi_l} & 0 & \cdots & 0 \end{array}\right]^T$$
$$\frac{\partial \mathbf{a}}{\partial \phi_l} = \frac{1}{2}\left[\begin{array}{cccccccc} 0 & \cdots & 0 & jA_l e^{j\phi_l} & -jA_l e^{-j\phi_l} & 0 & \cdots & 0 \end{array}\right]^T.$$

For simplicity, we introduce the following definitions:

$$\frac{\partial \mathbf{Z}}{\partial \omega_0}\mathbf{a} \triangleq \boldsymbol{\alpha}_0$$
$$\mathbf{Z}\frac{\partial \mathbf{a}}{\partial A_l} = \mathrm{Re}\left\{e^{j\phi_l}\mathbf{z}(\omega_0 l)\right\} \triangleq \boldsymbol{\beta}_l \quad (8)$$
$$\mathbf{Z}\frac{\partial \mathbf{a}}{\partial \phi_l} = -A_l \mathrm{Im}\left\{e^{j\phi_l}\mathbf{z}(\omega_0 l)\right\} \triangleq \boldsymbol{\gamma}_l.$$

Here, $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary values, respectively. Note that all the quantities above are real. The

entries in the Fisher information matrix can now be expressed in terms of inner products between these quantities as:

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\alpha}_0^T\boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_0^T\boldsymbol{\beta}_1 & \boldsymbol{\alpha}_0^T\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\alpha}_0^T\boldsymbol{\beta}_L & \boldsymbol{\alpha}_0^T\boldsymbol{\gamma}_L \\ \boldsymbol{\beta}_1^T\boldsymbol{\alpha}_0 & \boldsymbol{\beta}_1^T\boldsymbol{\beta}_1 & \boldsymbol{\beta}_1^T\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\beta}_1^T\boldsymbol{\beta}_L & \boldsymbol{\beta}_1^T\boldsymbol{\gamma}_L \\ \boldsymbol{\gamma}_1^T\boldsymbol{\alpha}_0 & \boldsymbol{\gamma}_1^T\boldsymbol{\beta}_1 & \boldsymbol{\gamma}_1^T\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_1^T\boldsymbol{\beta}_L & \boldsymbol{\gamma}_1^T\boldsymbol{\gamma}_L \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\beta}_L^T\boldsymbol{\alpha}_0 & \boldsymbol{\beta}_L^T\boldsymbol{\beta}_1 & \boldsymbol{\beta}_L^T\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\beta}_L^T\boldsymbol{\beta}_L & \boldsymbol{\beta}_L^T\boldsymbol{\gamma}_L \\ \boldsymbol{\gamma}_L^T\boldsymbol{\alpha}_0 & \boldsymbol{\gamma}_L^T\boldsymbol{\beta}_1 & \boldsymbol{\gamma}_L^T\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_L^T\boldsymbol{\beta}_L & \boldsymbol{\gamma}_L^T\boldsymbol{\gamma}_L \end{bmatrix} \tag{9}$$

The CRLB can now be determined from this by computing the inverse of this matrix and inspecting its diagonal elements. The simple closed form expressions for CRLBs obtained in [2], [18] can be found using the asymptotic orthogonality of complex sinusoids in computing the inner products above. However, we here do not employ this technique as we wish to take into account that the sinusoids are not orthogonal for low fundamental frequencies, and we therefore refer to this CRLB as the exact CRLB. For reference, the asymptotic CRLB for the problem at hand is given by

$$\text{var}(\hat{\omega}_0) \geq \frac{24\sigma^2}{N^3 \sum_{l=1}^L A_l^2 l^2}. \tag{10}$$

The lower bound can be seen to be determined by the signal-to-noise ratio (SNR) defined (in dB) as

$$SNR = 20\log_{10}\frac{\sum_{l=1}^L A_l^2 l^2}{\sigma^2} \, [\text{dB}] \,. \tag{11}$$

An interesting observation can be made from (9): it can be seen that the noise variance is simply a constant factor, and the effect of noise is, hence, unrelated to the problem of low fundamental frequencies. In this connection, it should be noted that this is also the case when the noise variance is uknown [35].

III. METHODS

A. Nonlinear Least-Squares

We will now present a number of estimators for solving the problem of interest. The first such method is the nonlinear least-squares (NLS) method, which is based on the principle of maximum likelihood estimation. It is an adaptation of a type of estimator that has appeared in many forms and contexts throughout the years to the problem at hand [4], [5], [8]. The maximum likelihood estimator for the parameters $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta}). \tag{12}$$

Under the assumption that $\mathbf{x}$ is Gaussian distributed and the noise is white, i.e., $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \sigma^2\mathbf{I})$, the likelihood function is given by (4). By inserting (4) into (12), taking the logarithm and dropping all constant terms, we obtain:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})\|^2, \tag{13}$$

where $\|\cdot\|^2$ denotes the vector 2-norm. This shows the well-known result that when the noise is white and Gaussian distributed, the LS method is the maximum likelihood estimator. As before, the mean is determined by the harmonic signal

model, i.e., $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{a}$ and the unknown parameters are in this case the fundamental frequency ω_0 that completely characterizes $\mathbf{Z}$ and the vector $\mathbf{a}$ containing the complex amplitudes. This results in the following problem:

$$(\hat{\omega}_0, \hat{\mathbf{a}}) = \arg\min_{\omega_0, \mathbf{a}} \|\mathbf{x} - \mathbf{Z}\mathbf{a}\|^2. \tag{14}$$

Since we are not really interested in the complex amplitudes, we will substitute these by their maximum likelihood estimate (for a given ω_0), which is $\hat{\mathbf{a}} = \left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{x}$, with $(\cdot)^H$ denoting the Hermitian-transposition. The resulting estimator depends only on ω_0:

$$\hat{\omega}_0 = \arg\max_{\omega_0} \mathbf{x}^T\boldsymbol{\Pi}_Z\mathbf{x}. \tag{15}$$

with $\boldsymbol{\Pi}_Z$ being the orthogonal projection matrix for the space spanned by the columns of $\mathbf{Z}$, i.e., $\boldsymbol{\Pi}_Z = \mathbf{Z}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H$. This is the estimator that we will here refer to as the NLS estimator. For each fundamental frequency candidate it involves operations of complexity $\mathcal{O}(L^2N) + \mathcal{O}(L^3) + \mathcal{O}(LN^2) + \mathcal{O}(N^2)$. The estimator does not, however, require any initialization[1], unlike the methods to follow. It should be noted that in assessing the complexity of the various methods, we treat the involved variables, here N and L, as independent variables, although they may not be. The matrix $\mathbf{Z}$ has full rank as long as $\omega_0 \neq 0$ and that $N \geq L$ for the inverse $\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}$ to exist. However, for very small ω_0, numerical effects may render the estimates useless.

The harmonic summation method [4] follows from this by using *the fact* that the columns of $\mathbf{Z}$ are orthogonal asymptotically in N [20]. Although this leads to a computationally efficient implementation based on the fast Fourier transform, this ultimately also leads to the failure of this method for low ω_0 and N.

B. Harmonic Fitting

The idea behind the following method is quite intuitive and appealing due to its simplicity. It is based on the principle of [36] used in [6]. Many different and good methods exist for finding frequencies of sinusoids in an unconstrained manner, meaning that they find frequencies that are not constrained to being integer multiples of a fundamental frequency. The question is then how to find an estimate of the fundamental frequency from these frequencies.

Suppose we find a set of parameter estimates $\hat{\boldsymbol{\eta}}$ from $\mathbf{x}$, and assuming that a maximum likelihood estimator with sufficiently large N is used (and that some regularity conditions are satisfied), the estimates $\hat{\boldsymbol{\eta}}$ are distributed as (see, e.g., [34])

$$\hat{\boldsymbol{\eta}} \sim \mathcal{N}(\boldsymbol{\eta}, \mathbf{I}^{-1}(\boldsymbol{\eta})) \tag{16}$$

where $\mathbf{I}(\boldsymbol{\eta})$ is the Fisher information matrix for the likelihood function for $\boldsymbol{\eta}$ (here, $\boldsymbol{\eta}$ are the true values). Now, suppose that we are not interested in these parameters, but rather in a

---

[1]In the context of complexity analysis, by initialization we mean that quantities that have to be computed before numerical optimization can be performed to obtain the parameters of interest, i.e. the computation of quantities other than the signal of interest.

different set $\boldsymbol{\theta}$ and that we can find a linear transformation $\mathbf{S}$ that relates these two. Mathematically, this can be stated as

$$\boldsymbol{\eta} = \mathbf{S}\boldsymbol{\theta}. \tag{17}$$

In the following, we assume that $\mathbf{S}$ is real, has full rank and is tall and that both parameter sets are real. Since $\hat{\boldsymbol{\eta}}$ are estimates of $\boldsymbol{\eta}$ and are distributed according to (16), the difference $\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta}$ is distributed as $\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\eta}))$. We can now use this to pose a probability density function of $\hat{\boldsymbol{\eta}}$ as

$$p(\hat{\boldsymbol{\eta}}; \boldsymbol{\theta}) = \frac{1}{\det\left(2\pi\mathbf{I}^{-1}(\boldsymbol{\eta})\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta})^T \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta})}, \tag{18}$$

which can be seen to be parametrized by the unknown parameters $\boldsymbol{\theta}$ and is, hence, a likelihood function. Proceeding now as in Subsection III-A, we can state the maximum likelihood estimate of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln p(\hat{\boldsymbol{\eta}}; \boldsymbol{\theta}) \tag{19}$$

$$= \arg\min_{\boldsymbol{\theta}} (\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta})^T \mathbf{I}(\boldsymbol{\eta}) (\hat{\boldsymbol{\eta}} - \mathbf{S}\boldsymbol{\theta}), \tag{20}$$

which can be seen to be a WLS estimator. Since the signal model is linear, the problem has a closed form solution, which is given by $\hat{\boldsymbol{\theta}} = \left(\mathbf{S}^T \mathbf{I}(\boldsymbol{\eta})\mathbf{S}\right)^{-1} \mathbf{S}^T \mathbf{I}(\boldsymbol{\eta})\hat{\boldsymbol{\eta}}$. At this point, some remarks are in order. Firstly, the estimator takes on the form of the solution to a linear LS problem regardless of the original distribution of $\mathbf{x}$. Secondly, the estimates $\hat{\boldsymbol{\eta}}$ need not follow the exact distribution in (16) for (20) to hold; the estimate covariance can be off by a multiplicative factor without affecting the form of the estimator. The principle used in arriving at (20) is known as the extended invariance principle (EXIP), or just the invariance principle, depending on the exact problem [36] (see also [31], [34], [37]). The principle has been applied to fundamental frequency estimation for a complex model and using asymptotic approximations of $\mathbf{I}(\boldsymbol{\eta})$ in [6]. Here, we will use it for a real model and without making use of the aforementioned asymptotic approximation.

The problem now remains to cast the problem of interest in this framework and determine $\boldsymbol{\eta}$, $\mathbf{S}$, and $\mathbf{I}(\boldsymbol{\eta})$. Firstly, for the case of a sinusoidal model with no harmonic constraint, we obtain a set of frequencies $\{\Omega_l \in (0, \pi)\}_{l=1}^{L}$. Moreover, we assume that the frequencies are organized as $\Omega_1 < \ldots < \Omega_L$. Next, we define a parameter set containing the corresponding parameters as $\boldsymbol{\eta} \triangleq \left[\begin{array}{cccccc} \Omega_1 & C_1 & \Phi_1 & \cdots & \Omega_L & C_L & \Phi_L \end{array}\right]^T$, where $\{C_l\}$ are the corresponding amplitudes and $\{\Phi_l\}$ the phases. It should be noted that there is no reason to include both positive and negative frequencies as these will be identical (as will the corresponding amplitudes and phases) for estimators tailored to real measurements. The transformation $\mathbf{S} \in \mathbb{R}^{3L \times 2L+1}$

relating these to $\boldsymbol{\theta}$ can easily be confirmed to be given by

$$\mathbf{S} = \begin{bmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ 0 & & 1 & & & \\ 2 & & & & & \\ 0 & & & 1 & & \mathbf{0} \\ 0 & & & & 1 & \\ \vdots & & & & & \vdots \\ L & & \mathbf{0} & & & \\ 0 & & & & 1 & \\ 0 & & & & & 1 \end{bmatrix}. \tag{21}$$

We can now express $\boldsymbol{\eta}$ in terms of $\boldsymbol{\theta}$ as in (17). The estimator in (20) requires that the true parameters are known to find $\mathbf{I}(\boldsymbol{\eta})$. Instead, we can use an approximation based on the parameter estimates $\hat{\boldsymbol{\eta}}$ (see [37], [38]), i.e., $\mathbf{I}(\boldsymbol{\eta}) \approx \mathbf{I}(\hat{\boldsymbol{\eta}})$, which, for Gaussian signals, is given by

$$[\mathbf{I}(\hat{\boldsymbol{\eta}})]_{nm} = \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\eta})}{\partial \eta_n} \mathbf{Q}^{-1} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\eta})}{\partial \eta_m}\bigg|_{\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}}, \tag{22}$$

with $\mathbf{Q}$ being the covariance matrix of the observation noise and $\boldsymbol{\mu}(\boldsymbol{\eta})$ being the mean of the same signal parametrized in terms of $\boldsymbol{\eta}$. The approximation above is essentially valid due to the maximum likelihood estimates $\hat{\boldsymbol{\eta}}$ being consistent estimates of $\boldsymbol{\eta}$ and $\mathbf{I}(\boldsymbol{\eta})$ being a continuous function. For the particular parametrization used here, i.e., the unconstrained model, $\mathbf{I}(\hat{\boldsymbol{\eta}})$ can be shown to be the following (see, e.g., [34]):

$$\mathbf{I}(\hat{\boldsymbol{\eta}}) = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\Xi}_{11} & \ldots & \boldsymbol{\Xi}_{1L} \\ \vdots & & \vdots \\ \boldsymbol{\Xi}_{L1} & \ldots & \boldsymbol{\Xi}_{LL} \end{bmatrix}, \tag{23}$$

where the individual blocks are given by

$$\boldsymbol{\Xi}_{kl} = \begin{bmatrix} \boldsymbol{\delta}_k^T \boldsymbol{\delta}_l & \boldsymbol{\delta}_k^T \boldsymbol{\epsilon}_l & \boldsymbol{\delta}_k^T \boldsymbol{\zeta}_l \\ \boldsymbol{\epsilon}_k^T \boldsymbol{\delta}_l & \boldsymbol{\epsilon}_k^T \boldsymbol{\epsilon}_l & \boldsymbol{\epsilon}_k^T \boldsymbol{\zeta}_l \\ \boldsymbol{\zeta}_k^T \boldsymbol{\delta}_l & \boldsymbol{\zeta}_k^T \boldsymbol{\epsilon}_l & \boldsymbol{\zeta}_k^T \boldsymbol{\zeta}_l \end{bmatrix}. \tag{24}$$

The entries in these blocks involve a number of quantities defined as

$$\boldsymbol{\delta}_l \triangleq \hat{C}_l \operatorname{Re}\left\{ e^{j\hat{\Phi}_l} \left[\begin{array}{cccc} 0 & je^{j\hat{\Omega}_l} & \cdots & j(N-1)e^{j\hat{\Omega}_l(N-1)} \end{array}\right]^T \right\}$$

$$\boldsymbol{\epsilon}_l \triangleq \operatorname{Re}\left\{ e^{j\hat{\Phi}_l} \mathbf{z}(\hat{\Omega}_l) \right\} \tag{25}$$

$$\boldsymbol{\zeta}_l \triangleq -\hat{C}_l \operatorname{Im}\left\{ e^{j\hat{\Phi}_l} \mathbf{z}(\hat{\Omega}_l) \right\}.$$

We note that the usually used expression for the CRLB for the unconstrained model is obtained, as before, by applying asymptotic approximations. More specifically, this leads to a block-diagonal structure in (23) as the off-diagonal blocks are approximately equal to zero, i.e., $\boldsymbol{\Xi}_{kl} = \mathbf{0}$, for $l \neq k$. Moreover, the individual blocks on the diagonal exhibit a block-diagonal structure themselves, hence leading to simple closed-form expressions.

Returning to the task at hand, we, finally, arrive at the estimator:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{S}^T \mathbf{I}(\hat{\boldsymbol{\eta}})\mathbf{S}\right)^{-1} \mathbf{S}^T \mathbf{I}(\hat{\boldsymbol{\eta}})\hat{\boldsymbol{\eta}}. \tag{26}$$

The processing steps of the estimator can be summarized as follows: First, estimate the parameters in $\hat{\boldsymbol{\eta}}$ and, second,

compute $\mathbf{I}(\hat{\boldsymbol{\eta}})$ from these parameters. Third, compute the parameter of $\hat{\boldsymbol{\theta}}$ from the aforementioned quantities along with $\mathbf{S}$, which is not signal-dependent. The fundamental frequency can now simply be extracted from the first element of $\hat{\boldsymbol{\theta}}$. Obviously, this process can be simplified somewhat if only the fundamental frequency is desired by determining only the first row of the matrix $\left(\mathbf{S}^T\mathbf{I}(\hat{\boldsymbol{\eta}})\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{I}(\hat{\boldsymbol{\eta}})$. As was demonstrated in [6], this methodology proved quite successful even with a number of asymptotic approximation, and we thus also expect it to perform well for our problem. Given the initial estimates $\hat{\boldsymbol{\eta}}$, the estimator has complexity $\mathcal{O}(L^3)$, but unlike the NLS method, it is in closed-form.

C. Optimal Filtering

The next solution to the problem under consideration is based on optimal filtering, which was first used for fundamental frequency estimation in [8] (see also [16]). Before providing more details on this, we introduce some notation and definitions. First, we define the output signal $\hat{x}(n)$ of the length M filter having real coefficients $h(n)$ as

$$\hat{x}(n) = \sum_{m=0}^{M-1} h(m)x(n-m) \triangleq \mathbf{h}^T\mathbf{x}(n), \qquad (27)$$

with $\mathbf{h}$ being a vector containing the filter coefficients of the filter defined as $\mathbf{h} = [\, h(0) \cdots h(M-1) \,]^T$ and $\mathbf{x}(n) = [\, x(n)\, x(n-1) \cdots x(n-M+1) \,]^T$. For our signal model, the output signal $\hat{x}(n)$ can be thought of as an estimate of the periodic parts of the signal. The output power of the filter can be expressed in terms of the covariance matrix $\mathbf{R}$ as $\mathrm{E}\left\{|\hat{x}(n)|^2\right\} = \mathbf{h}^T\mathbf{R}\mathbf{h}$. The question is now how to design the filter such that $\hat{x}(n)$ actually resembles a periodic signal. Such a filter should have a frequency response that allows the periodic components to pass undistorted while suppressing everything else. This means that the frequency response should be one for all the harmonic frequencies, and, since we are here concerned with real signals, this should be the case also for the negative frequencies. One can think of filters having these properties as a kind of comb filter. Mathematically, we can state this as the following optimization problem:

$$\min_{\mathbf{h}} \mathbf{h}^T\mathbf{R}\mathbf{h} \quad \text{s.t.} \quad \mathbf{Z}^H\mathbf{h} = \mathbf{1} \qquad (28)$$

with $\mathbf{1} = [\, 1 \cdots 1 \,]^T \in \mathbb{R}^{2L}$. We here remind the reader that $\mathbf{Z} \in \mathbb{C}^{M \times 2L}$ contains all the sinusoids of the real signal model, so the constraints state that the frequency response of the filter must be one for both positive and negative frequencies.

To solve the optimization problem, we introduce the Lagrange multipliers $\boldsymbol{\lambda} = [\, \lambda_1 \cdots \lambda_{2L} \,]^T$, and the Lagrangian dual function associated with the problem, which can be written as $\mathcal{L}(\mathbf{h}, \boldsymbol{\lambda}) = \mathbf{h}^T\mathbf{R}\mathbf{h} - \boldsymbol{\lambda}^T\left(\mathbf{Z}^H\mathbf{h} - \mathbf{1}\right)$. Taking the derivative with respect to the filter coefficients and the Lagrange multipliers and setting the result equal to zero and solving for the unknowns, leads to the optimal filter $\mathbf{h} = \mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1}$. The output power of this filter can then be expressed compactly as $\mathbf{h}^T\mathbf{R}\mathbf{h} = \mathbf{1}^H\left(\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1}$. Since the optimal filter depends on the observed signal via $\mathbf{R}$, the resulting filter can be thought of as an adaptive comb filter.

The filter can be used for determining the fundamental frequency in the following way: for a candidate fundamental frequency, the filter passes the candidate harmonics while it suppresses everything else. Therefore, the fundamental frequency can be identified as the value for which the output power of the filter is the highest. In math, this can be stated as

$$\hat{\omega}_0 = \arg\max_{\omega_0} \mathbf{1}^H\left(\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1}. \qquad (29)$$

For complex signals, this type of solution was demonstrated to have excellent performance under very adverse conditions in [8], effectively decoupling the multi-pitch estimation problem into a set of single-pitch problems. The estimator in (29) requires initialization of complexity $\mathcal{O}(M^3)$ for computing $\mathbf{R}^{-1}$ while for each fundamental frequency candidate, it requires computations of complexity $\mathcal{O}(L^3) + \mathcal{O}(ML^2) + \mathcal{O}(M^2L)$.

The method requires that the covariance matrix is replaced by an estimate. We use here the usual estimator, the sample covariance matrix, i.e.,

$$\mathbf{R} \approx \frac{1}{N-M+1} \sum_{n=M-1}^{N-1} \mathbf{x}(n)\mathbf{x}^T(n). \qquad (30)$$

Since the method also requires that this matrix is invertible, it follows that the filter length must be chosen such that $M < \frac{N}{2} + 1$, although it is well-documented in the literature that M in practice should not be chosen too close to this bound. Moreover, we also require that $M \geq L$ for the matrix inverse in (29) to exist. Combined, this allows us to bound M as $2L \leq M \leq \frac{N}{2} + 1$. It should also be noted that M should be chosen proportionally to N for the estimator to be consistent. This is also the case for the other methods presented later.

D. Angles between Subspace

The next method is a subspace method reminiscent of MUSIC [27], a method that has previously been applied to the fundamental frequency estimation problem in [8], [18]. It builds on more recent ideas presented in [20], [28]. In MUSIC, an estimate of a basis for the noise subspace is obtained via the eigenvalue decomposition of the sample covariance matrix. This is then used for estimation purposes by choosing the candidate model that is the closest to being orthogonal to that space. This is also the idea we here pursue, although the present methods differs in a fundamental way, namely in terms of how the angles between the subspaces are measured. Let $\mathbf{x}(n) = [\, x(n)\, x(n+1) \cdots x(n+M-1) \,]^T$. We can then express this vector as

$$\mathbf{x}(n) = \mathbf{Z}\mathbf{a} + \mathbf{e}(n), \qquad (31)$$

with $\mathbf{Z} \in \mathbb{C}^{M \times 2L}$ being defined as in (6) except that the columns have length M and $\mathbf{e}(n) = [\, e(n)\, e(n+1) \cdots e(n+M-1) \,]^T$. The covariance matrix[2] of this vector is given by

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\} = \mathbf{Z}\mathbf{P}\mathbf{Z}^H + \sigma^2\mathbf{I} \qquad (32)$$

---

[2]The reader should be aware that our definitions of $\mathbf{x}(n)$ and $\mathbf{R}$ here differ from those in Section III-C.

where $\mathrm{E}\left\{\mathbf{a}\mathbf{a}^H\right\} = \mathbf{P}$, which is given by

$$\mathbf{P} = \mathrm{E}\left\{\begin{bmatrix} a_1 a_1^* & a_1^* a_1^* & \dots & a_1 a_L^* & a_1^* a_L^* \\ a_1 a_1 & a_1^* a_1 & \dots & a_1 a_L & a_1^* a_L \\ \vdots & \vdots & & \vdots & \vdots \\ a_L a_1^* & a_L^* a_1^* & \dots & a_L a_L^* & a_L^* a_L^* \\ a_L a_1 & a_L^* a_1 & \dots & a_L a_L & a_L^* a_L \end{bmatrix}\right\}. \quad (33)$$

This matrix can be seen to involve block-matrices of the following form:

$$\mathbf{P}_{kl} = \mathrm{E}\left\{\begin{bmatrix} a_k a_l^* & a_k^* a_l^* \\ a_k a_l & a_k^* a_l \end{bmatrix}\right\}. \quad (34)$$

Next, we will analyze the behavior of this matrix assuming that the phases ϕ_l are uniformly distributed and independent over l. This means that $\mathrm{E}\left\{\frac{A_k}{2}e^{j\phi_k}\right\} = 0$ and that $\mathrm{E}\left\{\frac{A_k}{2}e^{j\phi_k}\frac{A_l}{2}e^{-j\phi_l}\right\} = \frac{A_k}{2}\mathrm{E}\left\{e^{j\phi_k}\right\}\frac{A_l}{2}\mathrm{E}\left\{e^{-j\phi_l}\right\} = 0$ for $k \neq l$. Hence, we obtain that for $k \neq l$, the matrix $\mathbf{P}_{kl}$ is simply $\mathbf{P}_{kl} = \mathbf{0}$. For $k = l$, we obtain

$$\mathbf{P}_{ll} = \begin{bmatrix} \frac{A_l^2}{4} & 0 \\ 0 & \frac{A_l^2}{4} \end{bmatrix}, \quad (35)$$

as $\mathrm{E}\left\{\frac{A_l}{2}e^{j\phi_l}\frac{A_l}{2}e^{-j\phi_l}\right\} = \frac{A_l^2}{4}$ and $\mathrm{E}\left\{\frac{A_l}{2}e^{j\phi_l}\frac{A_l}{2}e^{j\phi_l}\right\} = \frac{A_l^2}{4}\mathrm{E}\left\{e^{2j\phi_l}\right\} = 0$. Therefore, the amplitude covariance matrix $\mathbf{P}$ takes on the form $\mathbf{P} = \frac{1}{4}\mathrm{diag}\left(\begin{bmatrix} A_1^2 & A_1^2 & \cdots & A_L^2 & A_L^2 \end{bmatrix}\right)$, which means that the diagonal structure obtained for complex signals is retained for real signals, and the so-called covariance matrix model, therefore, still holds. We note that the assumptions that lead to this model are sufficient but not necessary conditions.

The eigenvalue decomposition (EVD) of the covariance matrix is $\mathbf{R} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{U}^H$, where $\boldsymbol{\Gamma}$ is a diagonal matrix containing the positive eigenvalues, γ_k, ordered as $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_M$. Moreover, it can easily be seen that $\gamma_{2L+1} = \dots = \gamma_M = \sigma^2$. The covariance matrix is positive definite and symmetric by construction. Therefore, $\mathbf{U}$ contains the M orthonormal vectors, which are eigenvectors of $\mathbf{R}$. We will denote these as $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_M \end{bmatrix}$. Let $\mathbf{S}$ be formed from a subset of the columns of this matrix as

$$\mathbf{S} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_{2L} \end{bmatrix}. \quad (36)$$

We denote the subspace spanned by the columns of $\mathbf{S}$ as $\mathcal{S} = \mathcal{R}(\mathbf{S})$ and refer to it as the signal subspace. Similarly, let $\mathbf{G}$ be formed from the remaining eigenvectors as

$$\mathbf{G} = \begin{bmatrix} \mathbf{u}_{2L+1} & \cdots & \mathbf{u}_M \end{bmatrix}. \quad (37)$$

We refer to the space $\mathcal{G} = \mathcal{R}(\mathbf{G})$ as the noise subspace. Using these definitions, we now obtain $\mathbf{U}\left(\boldsymbol{\Gamma} - \sigma^2\mathbf{I}\right)\mathbf{U}^H = \mathbf{Z}\mathbf{P}\mathbf{Z}^H$ as the identity matrix is diagonalized by an arbitrary orthonormal basis. Introducing $\boldsymbol{\Psi} = \mathrm{diag}(\begin{bmatrix} \gamma_1 - \sigma^2 & \cdots & \gamma_{2L} - \sigma^2 \end{bmatrix})$, this leads to the following partitioning of the EVD:

$$\mathbf{R} = \begin{bmatrix} \mathbf{S} & \mathbf{G} \end{bmatrix}\left(\begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \sigma^2\mathbf{I}\right)\begin{bmatrix} \mathbf{S}^H \\ \mathbf{G}^H \end{bmatrix}, \quad (38)$$

which shows that we may write $\mathbf{S}\boldsymbol{\Psi}\mathbf{S}^H = \mathbf{Z}\mathbf{P}\mathbf{Z}^H$. As the columns of $\mathbf{S}$ and $\mathbf{G}$ are orthogonal and $\mathcal{R}(\mathbf{Z}) = \mathcal{R}(\mathbf{S})$, it

follows that $\mathbf{Z}^H\mathbf{G} = \mathbf{0}$, which is the subspace orthogonality principle used in the MUSIC algorithm [27], [39].

In practice, the estimated noise subspace eigenvectors will not be perfect due to the observation noise and finite observation lengths. The above relation is, therefore, only approximate and a measure must be introduced to determine how close a candidate model $\mathbf{Z}$ is to being orthogonal to $\mathbf{G}$. Traditionally, this has been done using the Frobenius norm. However, this only measures the sum of cosine to the non-trivial angles squared between the two spaces for orthogonal vectors in both $\mathbf{Z}$ and $\mathbf{G}$, and, since we are here concerned with low frequencies, the asymptotic orthogonality of the column of $\mathbf{Z}$ is not accurate. We therefore measure the orthogonality as follows. The principal angles $\{\xi_k\}$ between the two subspaces $\mathcal{Z}$ and $\mathcal{G}$ are defined recursively for $k = 1, \dots, K$ as [40]

$$\cos(\xi_k) = \max_{\mathbf{u}\in\mathcal{Z}}\max_{\mathbf{v}\in\mathcal{G}}\frac{\mathbf{u}^H\mathbf{v}}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2} \triangleq \mathbf{u}_k^H\mathbf{v}_k, \quad (39)$$

where K is the minimal dimension of the two subspaces, i.e., $K = \min\{2L, M - 2L\}$ and $\mathbf{u}^H\mathbf{u}_i = 0$ and $\mathbf{v}^H\mathbf{v}_i = 0$ for $i = 1, \dots, k - 1$. The angles are bounded and ordered as $0 \leq \xi_1 \leq \dots \leq \xi_K \leq \frac{\pi}{2}$. Given the orthogonal projection matrices for $\mathcal{Z}$ and $\mathcal{G}$, denoted $\boldsymbol{\Pi}_Z$ and $\boldsymbol{\Pi}_G$, respectively, the expression in (39) can be written as

$$\cos(\xi_k) = \max_{\mathbf{y}}\max_{\mathbf{z}}\frac{\mathbf{y}^H\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G\mathbf{z}}{\|\mathbf{y}\|_2\|\mathbf{z}\|_2} \quad (40)$$

$$= \mathbf{y}_k^H\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G\mathbf{z}_k = \kappa_k. \quad (41)$$

As can be seen, $\{\kappa_k\}$ are the ordered singular values of the matrix product $\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G$, and the two sets of vectors $\{\mathbf{y}\}$ and $\{\mathbf{z}\}$ are the left and right singular vectors of the matrix product, respectively. The singular values are related to the Frobenius norm of $\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G$ and hence its trace, denoted with $\mathrm{Tr}\{\cdot\}$, as $\|\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G\|_F^2 = \sum_{k=1}^{K}\kappa_k^2$ which shows that if the Frobenius norm of the product is zero, then all the non-trivial angles are $\frac{\pi}{2}$, i.e., the two subspaces are orthogonal. This expression can be used to find the fundamental frequency as

$$\hat{\omega}_0 = \arg\min_{\omega_0}\|\boldsymbol{\Pi}_Z\boldsymbol{\Pi}_G\|_F^2, \quad (42)$$

and the estimate can be seen to be the value for which the sum of cosine to the angles squared is the least. Finally, (42) can be expressed as

$$\hat{\omega}_0 = \arg\min_{\omega_0}\mathrm{Tr}\left\{\mathbf{Z}\left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{G}\mathbf{G}^H\right\}, \quad (43)$$

which is asymptotically equivalent to the fundamental frequency estimator in [18] but different for finite M and N in that it takes the non-orthogonality of the sinusoids for low M and ω_0 into account. Hence, it can be expected to yield superior estimates for low fundamental frequencies. This estimator requires that a number of quantities are computed in the initialization, i.e., only once, namely the EVD of $\mathbf{R}$ and the projection matrix for the noise subspace, which results in a complexity of $\mathcal{O}((M - L)M^2) + \mathcal{O}(M^3)$ (which is obviously only valid for $L < M$). For each candidate fundamental frequency, operations having complexity $\mathcal{O}(L^2M) + \mathcal{O}(M^2L) + \mathcal{O}(L^3)$ are computed.

As for the covariance matrix, it has to be estimated and its dimensions chosen. For this method, this is done as described in (30), only with a different definition of $\mathbf{x}(n)$ as described earlier in this section. Unlike the optimal filtering method, it is not required for this method that the estimated matrix has full rank. It must, however, allow for the estimation of a basis for the signal subspace, which requires that $M \leq N - 2L + 1$. Additionally, for the orthogonal complement to the signal subspace to be non-empty, $M \geq 2L + 1$, which means that we obtain the following inequality for M:

$$2L + 1 \leq M \leq N - 2L + 1. \tag{44}$$

E. Shift-Invariance

The final estimator is also a subspace method and thus builds on the same matrix covariance model as in Section III-D. The last method was based on the noise subspace eigenvectors, while the present one is based on the signal subspace eigenvectors. More specifically, it is based on the principle used in [19]. The signal subspace is given by $\mathcal{S} = \mathcal{R}(\mathbf{S})$ with the matrix $\mathbf{S}$ being defined as in (36). As established earlier, the columns of $\mathbf{S}$ span the same space as the columns of $\mathbf{Z}$, i.e., $\mathcal{R}(\mathbf{S}) = \mathcal{R}(\mathbf{Z})$. Therefore, we may express the relation between these matrices as $\mathbf{S} = \mathbf{Z}\mathbf{B}$ where

$$\mathbf{B} = \mathbf{P}\mathbf{Z}^H\mathbf{S}\mathbf{\Psi}^{-1}, \tag{45}$$

with $\mathbf{B}$ being a square and full rank matrix as both $\mathbf{S}$ and $\mathbf{Z}$ do, and it is hence invertible, something that we will make use of later. The matrix $\mathbf{Z}$ exhibits a particular structure, known as shift-invariance. This property can be expressed in the following way. Define the matrices $\underline{\mathbf{Z}}$ and $\overline{\mathbf{Z}}$ by removing the last and first rows of $\mathbf{Z}$, i.e., $\underline{\mathbf{Z}} = [\ \mathbf{I}\ \mathbf{0}\ ]\mathbf{Z}$ and $\overline{\mathbf{Z}} = [\ \mathbf{0}\ \mathbf{I}\ ]\mathbf{Z}$ where it follows that $\mathbf{I}$ is $(M-1) \times (M-1)$. Now, doing the same for $\mathbf{S}$ we obtain $\underline{\mathbf{S}} = [\ \mathbf{I}\ \mathbf{0}\ ]\mathbf{S}$ and $\overline{\mathbf{S}} = [\ \mathbf{0}\ \mathbf{I}\ ]\mathbf{S}$. From these definitions and (45), it can easily be seen that $\underline{\mathbf{S}}$ and $\underline{\mathbf{Z}}$ are related as $\underline{\mathbf{S}} = \underline{\mathbf{Z}}\mathbf{B}$. More importantly, however, due to the particular structure of the model, the matrices $\overline{\mathbf{Z}}$ and $\underline{\mathbf{Z}}$ can be related as $\overline{\mathbf{Z}} = \underline{\mathbf{Z}}\mathbf{D}$ where

$$\mathbf{D} = \mathrm{diag}\left(\left[\ e^{j\omega_0}\ e^{-j\omega_0}\ \cdots\ e^{j\omega_0 L}\ e^{-j\omega_0 L}\ \right]\right). \tag{46}$$

This property is known as shift-invariance. However, since we are interested in finding the parameters that characterize $\mathbf{Z}$, this is of little use by itself. From the above it also follows that $\overline{\mathbf{S}} = \underline{\mathbf{S}}\mathbf{\Sigma}$ and the matrix relating $\underline{\mathbf{S}}$ to $\overline{\mathbf{S}}$ can be shown to be (see, e.g., [41])

$$\mathbf{\Sigma} = \mathbf{B}^{-1}\mathbf{D}\mathbf{B}, \tag{47}$$

i.e., the matrix $\mathbf{\Sigma}$ has the frequencies of the harmonics as the arguments of its eigenvalues. Since $\mathbf{S}$ and hence $\overline{\mathbf{S}}$ and $\underline{\mathbf{S}}$ are known from the EVD of the sample covariance matrix, this is useful in the following way: Given $\underline{\mathbf{S}}$ to $\overline{\mathbf{S}}$, we can solve for $\mathbf{\Sigma}$, from which we can find the frequencies via the EVD. Since the sample covariance will be corrupted by noise in practice, so will $\underline{\mathbf{S}}$ and $\overline{\mathbf{S}}$, and, consequently the above relations will only hold approximately, i.e., $\overline{\mathbf{S}} \approx \underline{\mathbf{S}}\mathbf{\Sigma}$, which means we have to introduce some way of finding $\mathbf{\Sigma}$. Here, we proceed by

estimating $\mathbf{\Sigma}$ using total least-squares (TLS) as follows. Define $\underline{\mathbf{\Delta}}$ and $\overline{\mathbf{\Delta}}$ as the minimal perturbations of $\underline{\mathbf{S}}$ to $\overline{\mathbf{S}}$, respectively:

$$\min\left\|\left[\overline{\mathbf{\Delta}}\ \underline{\mathbf{\Delta}}\right]\right\|_F \quad \text{s.t.} \quad \overline{\mathbf{S}} + \overline{\mathbf{\Delta}} = (\underline{\mathbf{S}} + \underline{\mathbf{\Delta}})\mathbf{\Sigma}. \tag{48}$$

An estimate $\widehat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ is then obtained as the solution to $\overline{\mathbf{S}} + \overline{\mathbf{\Delta}} = (\underline{\mathbf{S}} + \underline{\mathbf{\Delta}})\mathbf{\Sigma}$ for the perturbations solving (48) (see [41] for further details).

The frequencies obtained from the eigenvalues of $\widehat{\mathbf{\Sigma}}$ are not constrained to being integer multiples of a fundamental frequency, i.e., they are unconstrained frequencies, and, hence, cannot be used directly for estimating the fundamental frequency. Much like for the WLS method in Section III-B, we must fit a fundamental frequency to these frequencies. We now proceed to express $\widehat{\mathbf{\Sigma}}$ in terms of the empirical EVD as

$$\widehat{\mathbf{\Sigma}} = \mathbf{C}\widehat{\mathbf{D}}\mathbf{C}^{-1} \tag{49}$$

with $\mathbf{C}$ containing the empirical eigenvectors of $\widehat{\mathbf{\Sigma}}$ and

$$\widehat{\mathbf{D}} = \mathrm{diag}\left(\left[\ e^{j\hat{\Omega}_1^+}\ e^{j\hat{\Omega}_1^-}\ \cdots\ e^{j\hat{\Omega}_L^+}\ e^{j\hat{\Omega}_{2L}^-}\ \right]\right). \tag{50}$$

We here denote the estimated frequencies as $\{\hat{\Omega}_l^+ \in (0, \pi)\}_{l=1}^L$ and $\{\hat{\Omega}_l^- \in (-\pi, 0)\}_{l=1}^L$. Moreover, we assume that they are ordered $\hat{\Omega}_1^+ < \ldots < \hat{\Omega}_L^+$ and $\hat{\Omega}_1^- > \ldots > \hat{\Omega}_L^-$ and that the corresponding eigenvectors in $\mathbf{C}$ are ordered accordingly.

Recall that $\overline{\mathbf{S}} = \underline{\mathbf{S}}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}$, and thus $\overline{\mathbf{S}}\mathbf{C} \approx \underline{\mathbf{S}}\mathbf{C}\mathbf{D}$, where $\mathbf{D}$ depends on the unknown fundamental frequency ω_0. We can now introduce a metric that measures the extent to which the right and left side resemble each other as $\|\overline{\mathbf{S}}\mathbf{C} - \underline{\mathbf{S}}\mathbf{C}\mathbf{D}\|_F^2$. This expression can be expanded as

$$\|\overline{\mathbf{S}}\mathbf{C} - \underline{\mathbf{S}}\mathbf{C}\mathbf{D}\|_F^2 = -2\,\mathrm{Re}\left(\mathrm{Tr}\left\{\overline{\mathbf{S}}\mathbf{C}\mathbf{D}^H\mathbf{C}^H\underline{\mathbf{S}}^H\right\}\right) \tag{51}$$

$$+ \mathrm{Tr}\left\{\overline{\mathbf{S}}\mathbf{C}\mathbf{C}^H\overline{\mathbf{S}}^H\right\} + \mathrm{Tr}\left\{\underline{\mathbf{S}}\mathbf{C}\mathbf{C}^H\underline{\mathbf{S}}^H\right\}. \tag{52}$$

Noting that the last two terms do not depend on ω_0 and introducing $\delta_l = \left[\mathbf{C}^H\underline{\mathbf{S}}^H\overline{\mathbf{S}}\mathbf{C}\right]_{ll}$, we finally obtain the estimator

$$\hat{\omega}_0 = \arg\max_{\omega_0} 2\,\mathrm{Re}\left\{\sum_{l=1}^L \delta_{2l-1}e^{-j\omega_0 l} + \delta_{2l}e^{j\omega_0 l}\right\}. \tag{53}$$

As can be seen, the resulting estimator is extremely simple having complexity $\mathcal{O}(L)$ for each fundamental frequency candidate, albeit the initialization, i.e., the computation of δ_l, is somewhat complex. More specifically, it requires computations of complexity $\mathcal{O}(M^3) + \mathcal{O}(L^3) + \mathcal{O}(M^2 L) + \mathcal{O}(L^2 M)$. We also note that the involved cost function is generally smooth and well-behaved. Regarding the size of the covariance matrix, M should be chosen according to (44) for obtaining a rank $2L$ estimate of $\mathbf{S}$ and for $\mathbf{\Sigma}$ to be unique.

IV. EXPERIMENTAL RESULTS

A. Exact vs. Asymptotic Bounds

We will start out the experimental part of this paper by exploring the difference between the exact and asymptotic CRLBs for the problem of estimating the fundamental frequency and the dependency of this difference on various parameters. This is interesting for a number of reasons. Many of the estimators derived based on complex models are based on
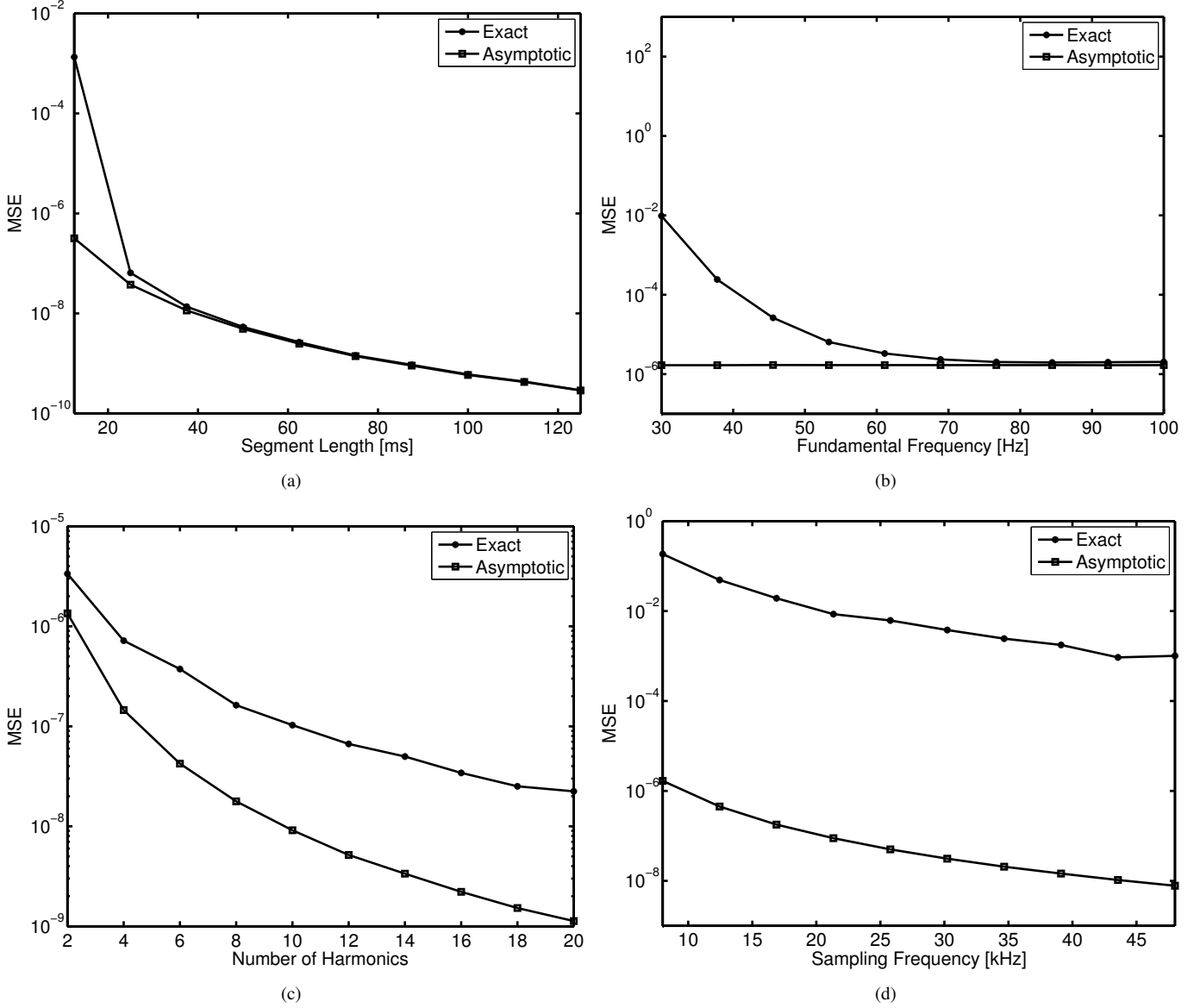
Fig. 1. Exact and asymptotic Cramér-Rao lower bounds as functions of various parameters, namely (a) the segment length (in ms), (b) the fundamental frequency (in Hz), (c) the number of harmonics L, and (d) the sampling frequency (in kHz). Each point on the curves is obtained over 1000 realizations of the involved parameters.

the same asymptotic approximation that the asymptotic CRLB is based on. Hence, if the asymptotic approximation is accurate for the CRLB, it is also likely to be accurate for the various estimators. Moreover, we can also learn something about the conditions under which the approximation will hold and learn if anything can be done about it. To make it easier to interpret the results, we will do this assuming typical physical values encountered in speech and audio applications. In the first experiment, a low fundamental frequency of 50 Hz is assumed along with a sampling frequency of 8 kHz. Moreover, the noise variance is kept fixed at one throughout these experiments. The remaining parameters were uniformly distributed phases and Rayleigh distributed amplitudes with five harmonics. Based on these values, the exact CRLB based on (9) and the asymptotic approximation in (10) were computed as a function of the segment length (in ms) for 1000 realizations of the parameters

for each experimental condition. The results, in the form of the averages over these realizations, are shown in Figure 1(a). As can be seen, there is a huge discrepancy between the two bounds for short segments, and this discrepancy vanishes for long segments. This clearly shows that the claim that the problem of estimating low fundamental frequencies is difficult is indeed true. It also shows that it is entirely unrealistic to expect estimators to perform close to the asymptotic CRLB under these circumstances, and, hence, an estimator may be falsely deemed suboptimal if its performance is compared to the wrong bound.

In the next experiment, the segment length is kept fixed at 20 ms while the fundamental frequency is varied with the remaining parameters and experimental conditions being as before. The results are shown in Figure 1(b). The same observations as for the varying segment length can be observed

here, namely that as the fundamental frequency is lowered, relative to N, the discrepancy between the asymptotic and exact CRLBs grow. Beyond a certain frequency, here 80 Hz, there is basically no difference between the two bounds and asymptotic approximations must therefore be valid from this frequency and beyond. It should be noted that depending on the physics of the observed phenomenon, a low fundamental frequency may also have more harmonics, as they can in principle extend up to half the sampling frequency. This is not reflected in this experiment. It can be seen from (10) that, in theory, the more harmonics that are present, the more accurately the underlying fundamental frequency also can be estimated, at least for a sufficiently high N. For this reason, the next experiment focuses on the dependency on the number of harmonics, L. In this experiment, a fundamental frequency of 50 Hz is used for different L while other experimental settings were as before. The results can be seen in Figure 1(c). From the figure, it can be seen that the discrepancy between the two bounds actually increases as a function of L, meaning that the more harmonics are in the signal, it becomes relatively more difficult to determine the fundamental frequency, due to it being so low. On the other hand, the bound does decrease as a function of L even if the gap increases, so it is still beneficial to incorporate the additional harmonics in the model. Part of the reason that the bounds decrease as a function of L is that it effectively leads to an increase in SNR, as defined in (11) when the noise variance is kept fixed.

The final experiment involving the differences between the CRLBs is one where all the prior parameters are kept fixed while the sampling frequency is changed, and this is motivated as follows: since the highest possible segment length (in ms) is dictated by the stationarity of the observed signal, it is not possible to mitigate the problems associated with low fundamental frequencies by simply changing the segment length beyond a certain point. However, the sampling frequency can of course be changed in many situations, and raising the sampling frequency while keeping the segment length in ms fixed, of course leads to a higher number of samples N. Here, the behavior of the asymptotic and exact CRLBs is observed for a 20 ms segment and a fundamental frequency of 50 Hz with five harmonics. In Figure 1(d), the resulting curves can be seen. The figure shows that simply changing the sampling frequency does not alleviate the discrepancy between the two CRLBs, and the explanation is that while raising the fundamental does lead to a higher N, it also leads to a lower ω_0. But it is also interesting to note that both bounds do decrease as a function of the sampling rate, meaning that we are able to estimate the fundamental frequency more accurately by increasing the sampling frequency. An explanation for this is that while increasing the sampling frequency results in a proportionally higher N and lower ω_0 the effect of the noise on the ability to estimate the parameters is nonlinear. That this is the case can be seen from (10), from which it can be observed that the bound is inversely proportional to N^3.

B. Tested Methods

In the following experiments, we will compare the performance of the presented estimators to the previously published methods based on a complex signal model and/or asymptotic approximations. We will denote the methods for real signals by prefix "r" and their complex counterparts by prefix "c". To summarize, the following methods will be compared:

- rWLS is the harmonic fitting method based on WLS as presented in Section III-B. It requires that unconstrained frequencies and their amplitudes are found. This is done using ESPRIT and LS, respectively.
- rFILT is the optimal filtering method presented in Section III-C.
- rNLS is the NLS method of Section III-A.
- rABS is the subspace method based on measuring the angles between subspaces as described in Section III-D.
- rSHIFT is another subspace method, but based on the shift-invariance property, as presented in Section III-E.

We will compare the performance of these methods to a number of reference methods, namely the following:

- cWLS is the harmonic fitting method as originally proposed in [6]. It uses asymptotic approximations of the weighting matrix to obtain a simple expression for the fundamental frequency. Like its real counterpart it requires unconstrained frequency and their amplitude estimates. Here, the same as for rWLS are used.
- cFILT is the optimal filtering method proposed in [8]. It differs from rFILT in that it does not take the existence of complex conjugate pairs of harmonics into account.
- cNLS is the approximate NLS method as described in [8]. It is similar to the methods of [4], [5]. It differs from rNLS in the following way: it is based on the asymptotic orthogonality of complex sinusoids and, hence, takes neither the existence of complex conjugate pairs nor the interaction between the harmonics into account.
- cABS is the MUSIC-based method of [18], except that the model order is assumed known. Unlike rABS, it uses an approximation of the angles between the subspaces.
- cSHIFT is the method proposed in [19], which is based on the shift-invariance property of the signal subspace. It differs from rSHIFT in that it does not take the existence of complex conjugate pairs of complex sinusoids into account. Unlike [19] it uses TLS rather than LS.

All estimators are implemented in a two-step fashion where a coarse fundamental frequency estimate is first found using a grid search after which a simple dichotomous search is used to obtain a refined estimate. The same grid size and dichotomous search algorithm is used for all the methods. For most of the methods, a covariance matrix size/filter length of $M = N/2$ is used, except for the optimal filtering methods where $M = N/4$ have been used (the reason for this will become clear later). For the estimators relying on a complex model, the real signal is mapped to a complex one via the Hilbert transform. The optimal filtering methods require an invertible covariance matrix for which reason the down-sampled analytic signal is used for cFILT. To address the numerical issue associated with very low fundamental frequencies, which may cause the involved matrices to be rank deficient numerically but not on paper, the Moore-Penrose pseudo-inverse [40] is used whenever appropriate.
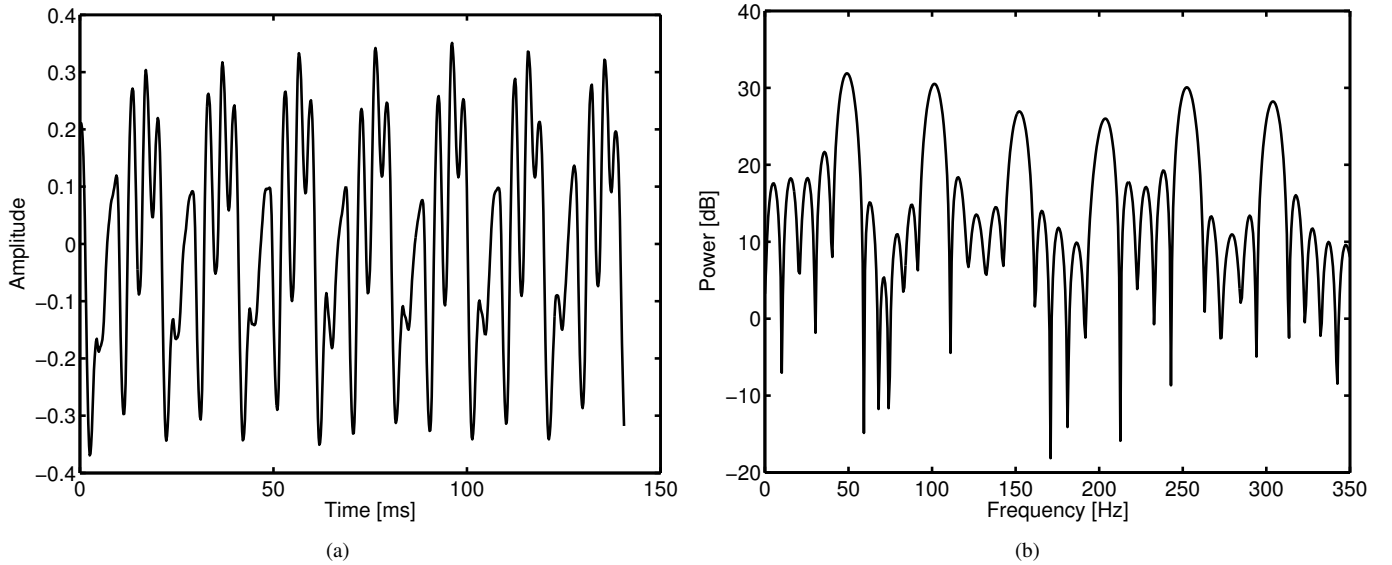
Fig. 2. Example of a signal having a low frequency, here a tone played by a contrabassoon. Shown are (a) the time-domain signal, and (b) part of its spectrum, namely the low frequencies, estimated using the periodogram.
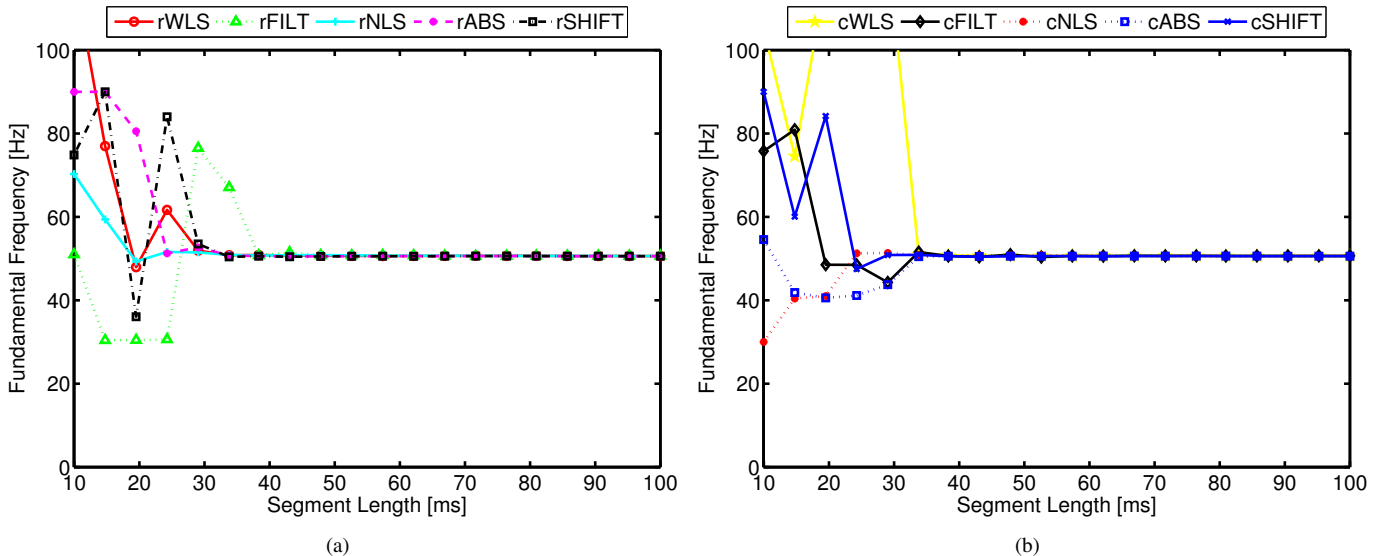

Fig. 3. Fundamental frequency estimates obtained for the signal in Figure 2 as a function of the segment length (in ms) for (a) the estimators for real-valued signals and (b) their complex counterparts.

C. A Signal Example

Next, we will illustrate the problems associated with low fundamental frequencies using a recorded signal, namely a tone played by a contrabassoon. The signal is shown in Figure 2(a) along with its spectrum in Figure 2(b), here estimated using the periodogram computed using a 8192 point FFT and a rectangular window. Note that a sampling frequency of 8820 Hz is used. In studying the effect of the low fundamental frequency on the ability to obtain accurate estimates, the segment length will be varied from 10 ms to 100 ms (with all segments beginning at the start of the signal shown in Figure 2(a)). The various estimators are then run on these segments. The number of harmonics was determined by visual inspection of the spectrum. The results are shown in Figure 3 for (a) the presented estimators, and (b) the estimators based

on asymptotic approximations and complex signal models. A number of interesting observations can be made from the figures. Firstly, all estimators, both the real ones and their complex counterparts, converge to the same result when the segment length is increased. It can also be seen that all the methods break down when the segment length gets extremely short. Moreover, for this particular example, the methods for real signals generally outperform the complex ones, but it should also be noted that other factors may play a role due to the complex nature of real-life signals.

D. Monte Carlo Simulations

The methods are compared using Monte Carlo simulations by generating signals according to the model in (2) and then applying the various estimators to the resulting signal. The
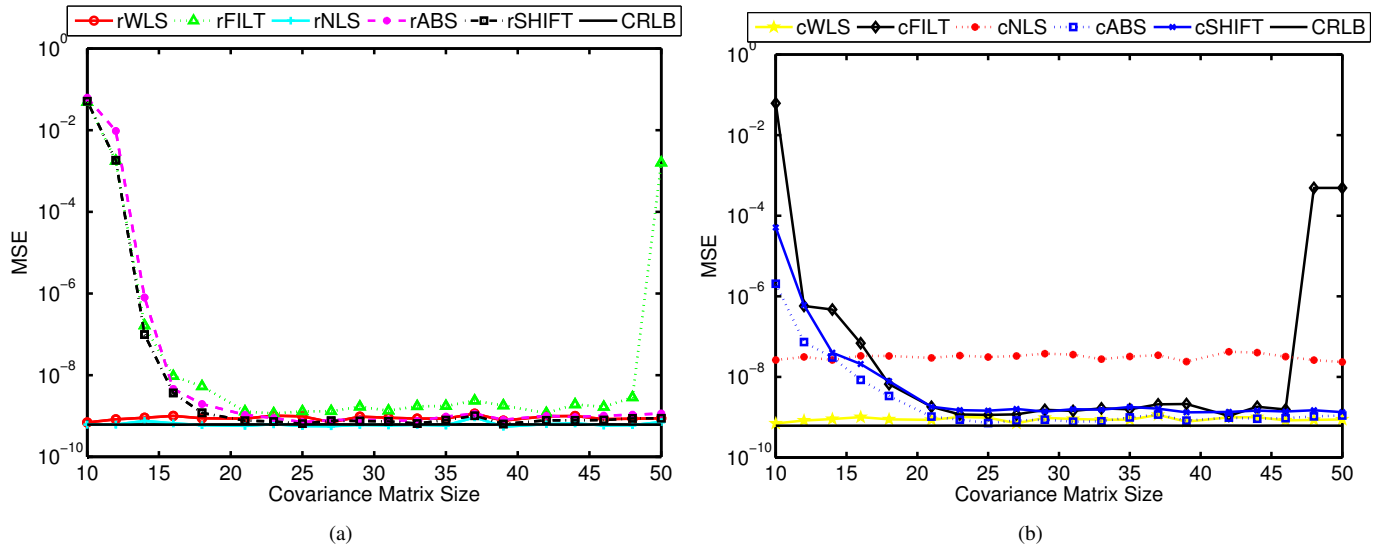
Fig. 4. Performance measured in terms of the Mean Square estimation Error (MSE) as a function of the covariance matrix size, M, for (a) the real estimators and (b) their complex counterparts based on asymptotic approximations.

so-obtained parameter estimates are then compared to the true parameters and the estimation error is measured in terms of the mean square error (MSE). For each set of experimental conditions, 100 realizations are used and the CRLB shown in the figures to follow is the average over the exact CRLB. The signals were generated with the following parameters, except when otherwise stated (e.g., when a certain parameter is varied): a fundamental frequency with $\omega_0 = 0.3129$ is used with five harmonics, each having unit amplitude and phases uniformly distributed between $-\pi$ and π. Segments of $N = 100$ samples were used with white Gaussian noise added at an SNR of 40 dB, according to the definition of the SNR in (11).

First, the influence of the covariance matrix size, which is also the filter length for the filtering methods, on the performance of the various estimators is investigated. This is done by simply varying M while keeping all other parameters fixed. The results are shown in Figure 4 for the real estimators (a) and the complex ones (b). Note that neither the NLS nor the WLS class methods make use of the covariance matrix an their performance hence does not depend on M. It can generally be observed that as long as the covariance matrix size is not chosen too low or too high, the methods perform well. In fact, the only class of methods that are sensitive to M being close to $M/2$ appears to be the optimal filtering methods (we remind the reader that $N = 100$ is used here). All methods, except one, perform close to the CRLB. For the cNLS method, a gap between its MSE and the CRLB can be seen. This demonstrates the clear sub-optimality of this method for the problem at hand and illustrates the importance of avoiding asymptotic approximations. It should be noted that the cNLS method performs extremely well for sufficiently high N and ω_0, being statistically efficient. Moreover, it has also been confirmed experimentally that the poor performance reported (and in the experiments to follow) here is not due to the suboptimality of the Hilbert transform used but rather, as stated, the asymptotic approximation.

We will now proceed to investigate the dependency of the performance for the various estimators on the number of samples N. For the methods requiring a covariance matrix, it was stated that M should be chosen proportionally to N; otherwise, the estimator would not be consistent. So, in varying N, the covariance matrix size will also be varied with $M = N/2$ for all methods, except the optimal filtering methods for which $M = N/4$ is used. The results are shown in Figure 5(a) and Figure 5(b) for the two classes of methods. It can be seen that all the methods appear to be consistent in that the MSE decreases as a function of N. It can also be seen that the filtering methods, rFILT and cFILT perform poorly for low N, and that cNLS is clearly sup-optimal performing far from the CRLB, unlike rNLS, for the entire range of N shown here. Similarly, the cSHIFT methods perform poorly. Other than that, it appears that the remaining methods, aside from rNLS, break down below 40 samples.

In the next experiment, the performance of the various methods is investigated as a function of the SNR. From the asymptotic SNR in (10), one would perhaps expect this to be a trivial experiment as the noise variance is a linear parameter. However, due to the estimation problem being nonlinear, it is difficult to predict exactly how the performance of estimators will depend on the SNR. Moreover, it is well-known that, for nonlinear problems, estimators will exhibit so-called threshold behavior, which means that below a certain point, the estimators will break down producing what is essentially useless results. The MSE as a function of the SNR is depicted in Figures 6(a) and 6(b) for the real and complex estimators, respectively. A number of interesting observations can be made from these figures. For most of the methods, except cNLS, it can be seen that the performance increases as a function of the SNR, as can be expected from good estimators. The cNLS method can be seen to hit a floor for high SNRs. This is likely to be due to the approximations used in that method being inaccurate. For low, SNRs, however, this appear to not matter much as the error is dominated by the noise,
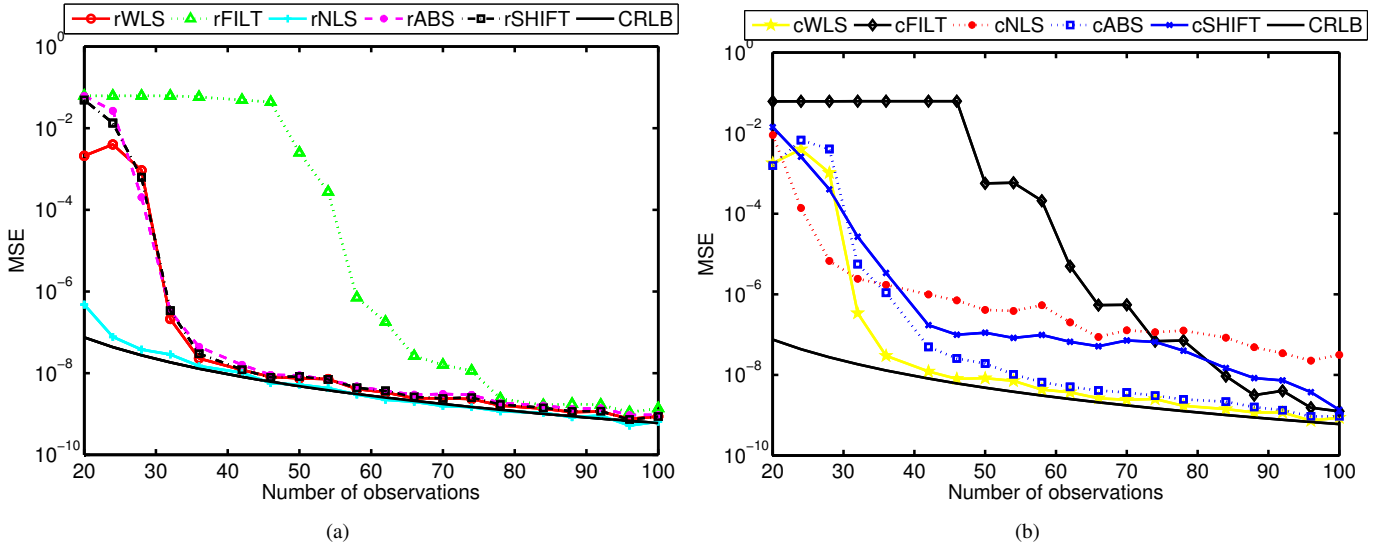
Fig. 5. Performance measured in terms of the Mean Square estimation Error (MSE) as a function of the number of observations, N, for (a) the real estimators and (b) their complex counterparts based on asymptotic approximations.
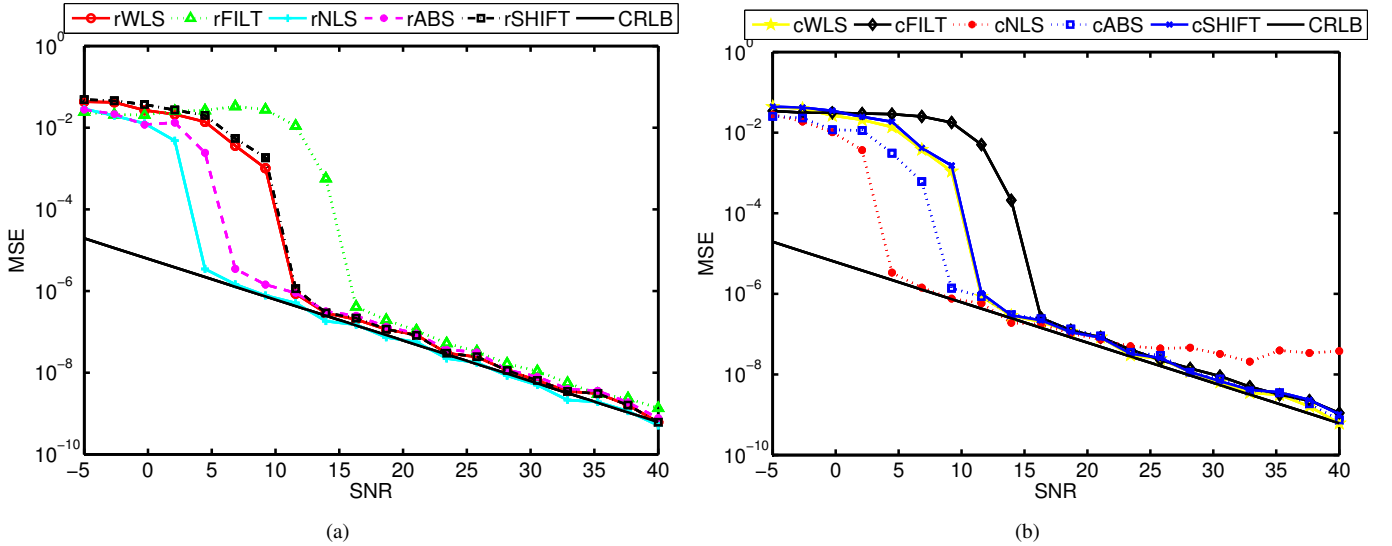

Fig. 6. Performance measured in terms of the Mean Square estimation Error (MSE) as a function of the SNR for (a) the real estimators and (b) their complex counterparts based on asymptotic approximations.

with the MSE following the CRLB. It even appears that the cNLS method breaks down later than the cWLS, cSHIFT and cFILT methods with also the cABS method performing quite well for low SNRs. The rNLS can be observed to mitigate the problems of the cNLS as it follows the CRLB even for high SNRs. In fact, it can be seen to be statistically efficient above SNRs of 5 dB. Curiously, the rABS and cABS appear to perform almost equally well, being fairly robust against low SNRs, although it is not statistically efficient. The rWLS, rFILT and rSHIFT methods appear to perform similarly to their complex counterparts in this experiment, with the optimal filtering method performing the worst.

In the final and most important experiment, the role of the fundamental frequency will be investigated. More specifically, the fundamental frequency is varied from a value for which it is expected that all methods work to a low value close to

zero, and it is expected they eventually will exhibit threshold behavior. The results are shown in Figures 7(a) and 7(b) for the two classes of methods. Starting with the complex methods, a number of interesting points can be made. Firstly, all except the cWLS perform poorly with the resulting MSEs differing substantially from the CRLB. The cWLS method performs well, following the CRLB, until about a fundamental frequency of 0.06. The cABS method also performs quite well, but performs further from the CRLB as the fundamental frequency is lowered. The cNLS, cFILT and cSHIFT methods can be seen to generally not perform well at all. For the real methods, it can be observed that the rNLS method performs the best, followed by the rWLS, rABS, and rSHIFT methods with the rFILT method performing quite poorly and worst of the methods. Comparing the two figures an important observation can be made: it can clearly be seen that all methods, except the
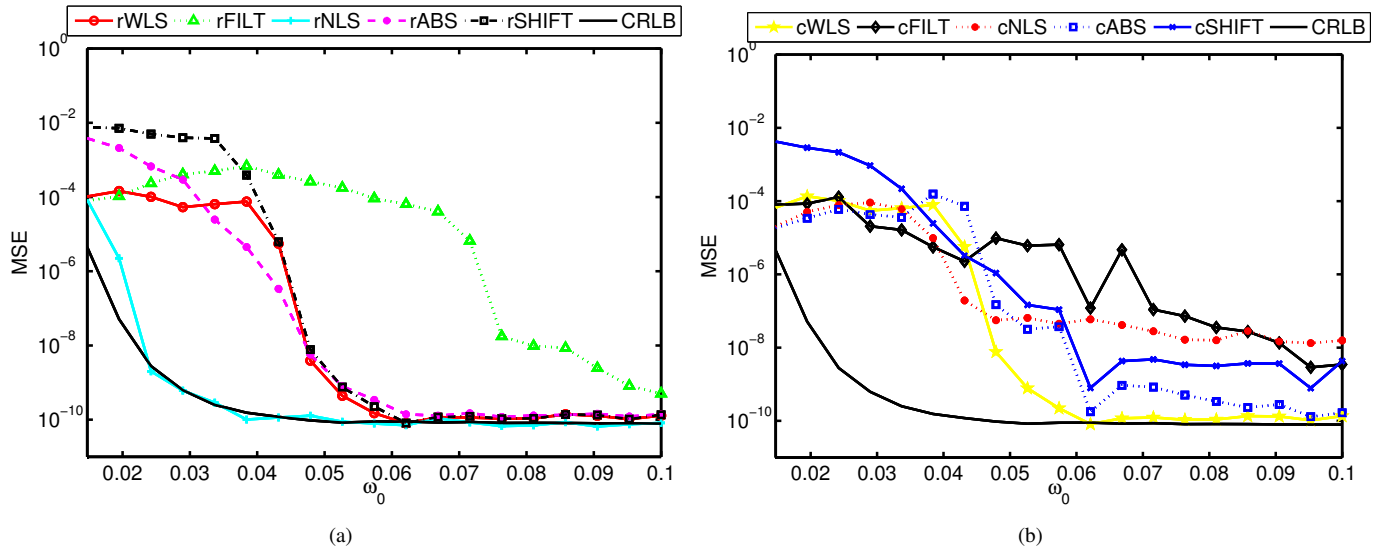
Fig. 7. Performance measured in terms of the Mean Square estimation Error (MSE) as a function of the fundamental frequency, ω_0, for (a) the real estimators and (b) their complex counterparts based on asymptotic approximations.

rWLS method, are improved by the modifications presented in this paper. This clearly demonstrates that the commonly used approximations are not suitable for low fundamental frequencies and that it is possible to avoid them. Regarding the rWLS method, from the experiments, it appears that the approximations used in the weighting matrix in the cWLS method is not the reason for threshold behavior as the rWLS method behaves in the same way, rather the dominant error source is most likely the unconstrained frequencies. The reader should be aware that the rWLS method, like the cWLS method, is dependent on the unconstrained frequencies being accurate, and it can of course be expected that this will not be the case when the fundamental frequency is low. Note that the high sensitivity of this method to spurious frequency estimates was also demonstrated in [18], albeit under different circumstances.

V. Conclusion

In this paper, the problem of estimating low fundamental frequencies from real-valued measurements has been considered. The problem has been analyzed via comparisons of the asymptotic and approximate Cramér-Rao lower bounds. These comparisons show that the asymptotic approximations frequently used in estimators and in the computation of estimation bounds are not accurate under these circumstances. To mitigate this, a number of estimators have been presented in which such approximations are avoided, and these estimators can therefore be said to be exact. The estimators are based on the methodologies of maximum likelihood, leading to a nonlinear least-squares method and a harmonic fitting algorithm that fits individual frequencies to a fundamental frequency estimate, optimal filtering as known from Capon's classical beam-former, and subspace methods, herein one based on subspace orthogonality and one based on subspace shift-invariance. All of the methods, except the harmonic fitting one, which makes use of an set of intermediate parameters,

have cubic complexity in the number of samples and/or the number of harmonics. In Monte Carlo simulations, the performance of the various estimators has been investigated and compared to methods employing asymptotic approximations. These simulations showed that, among the considered methods, the nonlinear least-squares method performed the best, the optimal filtering method performed the worst, and the remaining methods in-between. More importantly, however, the simulations showed that for all the considered methods, except the harmonic fitting one, it is possible to achieve improved performance by using the exact estimators. Moreover, it can be seen that not only do the proposed methods perform closer to the Cramér-Rao lower bound, but their threshold behavior is also improved for low fundamental frequencies.

References

[1] E. Conte, A. Filippi, and S. Tomasin, "ML period estimation with application to vital sign monitoring," *IEEE Signal Process. Lett.*, vol. 17, no. 11, pp. 905–908, 2010.

[2] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34(5), pp. 1124–1138, Oct. 1986.

[3] G. Ogden, L. Zurk, M. Siderius, E. Sorensen, J. Meyers ad S. Matzner, and M. Jones, "Frequency domain trackin of passive vessel harmonics," *J. Acoust. Soc. Am.*, vol. 126, pp. 2249, 2009.

[4] M. Noll, "Pitch determination of human speech by harmonic product spectrum, the harmonic sum, and a maximum likelihood estimate," in *Proc. Symposium on Computer Processing Communications*, 1969, pp. 779–797.

[5] B. G. Quinn and P. J. Thomson, "Estimating the frequency of a periodic function," *Biometrika*, vol. 78(1), pp. 65–74, 1991.

[6] H. Li, P. Stoica, and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[7] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a posteriori probability pitch tracking in noisy environments using harmonic model," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 12(1), pp. 76–87, 2004.

[8] M. G. Christensen, P. Stoica, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88(4), pp. 972–983, Apr. 2008.

[9] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, pp. 353–362, Oct. 1974.

[10] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41(2), pp. 293–309, 1967.

[11] Y. Medan, E. Yair, and D. Chazan, "Super resolution pitch determination of speech signals," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 40–48, Jan. 1991.

[12] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111(4), pp. 1917–1930, Apr. 2002.

[13] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., chapter 5, pp. 495–518. Elsevier Science B.V., 1995.

[14] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11(7), pp. 609–612, July 2004.

[15] J. Moorer, "The optimum comb method of pitch period analysis of continuous digitized speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, pp. 330–338, Oct 1974.

[16] M. G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Processing*, vol. 58(12), pp. 5969–5983, Dec. 2010.

[17] D. Chazan, Y. Stettiner, and D. Malah, "Optimal multi-pitch estimation using the EM algorithm for co-channel speech separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 27–30 April 1993, vol. 2, pp. 728–731.

[18] M. G. Christensen, A. Jakobsson and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15(5), pp. 1635–1644, July 2007.

[19] M. G. Christensen, A. Jakobsson and S. H. Jensen, "Fundamental frequency estimation using the shift-invariance property," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2007, pp. 631–635.

[20] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, vol. 5 of *Synthesis Lectures on Speech & Audio Processing*, Morgan & Claypool Publishers, 2009.

[21] E. J. Hannan and B. G. Quinn, "The resolution of closely adjacent spectral lines," *J. of Time Series Analysis*, vol. 10, pp. 13–31, 1989.

[22] D. Huang, "On low and high frequency estimation," *J. of Time Series Analysis*, vol. 17(4), pp. 351–365, 1996.

[23] P. Stoica and A. Eriksson, "MUSIC estimation of real-valued sine-wave frequencies," *Signal Processing*, vol. 42, pp. 139–146, 1995.

[24] K. Mahata, "Subspace fitting approaches for frequency estimation using real-valued data," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3099–3110, Aug. 2005.

[25] A. Jakobsson, T. Ekman, and P. Stoica, "Capon and APES spectrum estimation for real-valued signals," in *Eighth IEEE Digital Signal Processing Workshop*, 1998.

[26] H.C. So, K. W. Chan, Y.T. Chan, and K.C. Ho, "Linear prediction approach for efficient frequency estimation of multiple real sinusoids: algorithms and analyses," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2290–2305, July 2005.

[27] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34(3), pp. 276–280, Mar. 1986.

[28] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. on Advances in Signal Processing*, vol. 2009, pp. 1–11, 2009, Article ID 948756.

[29] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results and an application," *IEEE Trans. Signal Process.*, vol. 48(2), pp. 338–352, Feb. 2000.

[30] S. Godsill and M. Davy, "Bayesian computational models for inharmonicity in musical instruments," in *Proc. IEEE Workshop on Appl. of Signal Process. to Aud. and Acoust.*, 2005, pp. 283–286.

[31] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2001.

[32] G. L. Bretthorst, "An introduction to parameter estimation using Bayesian probability theory," in *Max. Entropy and Bayesian Methods*, P. Fougere, Ed., pp. 53–79. 1990.

[33] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31(5), pp. 1235–1248, Oct. 1983.

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[35] P. Stoica, A. Jakobsson, and J. Li, "Cisoid Parameter Estimation in the Colored Noise Case: Asymptotic Cramér-Rao Bound, Maximum Likelihood and Nonlinear Least-Squares," *IEEE Trans. Signal Process.*, vol. 45, pp. 2048–2059, August 1997.

[36] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Elsevier Signal Processing*, vol. 17, pp. 383–387, 1989.

[37] A.L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, no. 2, pp. 421–441, 1998.

[38] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21(4), pp. 36–47, July 2004.

[39] G. Bienvenu, "Influence of the spatial coherence of the background noise on high resolution passive methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1979, pp. 306–309.

[40] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, 3rd edition, 1996.

[41] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.

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 in 2002 and 2005, respectively, from Aalborg University (AAU) in Denmark, where he is also currently employed at the Dept. of Architecture, Design & Media Technology as Associate Professor. At AAU, he is head of the Audio Analysis Lab which conducts research in audio signal processing.

He was formerly with the Dept. of Electronic Systems, Aalborg University and has been a Visiting Researcher at Philips Research Labs, ENST, UCSB, and Columbia University. He has published more than 100 papers in peer-reviewed conference proceedings and journals as well as 1 research monograph. 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, including an ICASSP Student Paper Award, the Spar Nord Foundation's Research Prize for his Ph.D. thesis, a Danish Independent Research Council Young Researcher's Award, and the Statoil Prize 2013, as well as prestigious grants from the Danish Independent Research Council and the Villum Foundation's Young Investigator Programme. He has served as an Associate Editor for the IEEE Signal Processing Letters.