Parametric Modeling for Two-Dimensional Harmonic Signals With Missing Harmonics

Zhou, Zhenhua; Christensen, Mads Græsbøll; Jensen, Jesper Rindom; Zhang, Shengli

Published in:
IEEE Access

DOI (link to publication from Publisher):
10.1109/ACCESS.2019.2907456

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 26, 2019, accepted March 6, 2019, date of current version April 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907456

Parametric Modeling for Two-Dimensional Harmonic Signals With Missing Harmonics

ZHENHUA ZHOU[1], MADS G. CHRISTENSEN[2], (Senior Member, IEEE), JESPER R. JENSEN[2], (Member, IEEE), AND SHENGLI ZHANG[1], (Member, IEEE)

[1]College of Information Engineering, Shenzhen University, Shenzhen 518060, China
[2]Audio Analysis Lab, AD:MT, Aalborg University, 9220 Aalborg, Denmark

Corresponding author: Shengli Zhang (zsl@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801130 and Grant 61771315, in part by the Villum Foundation of Young Investigator Programme (''Spatio-Temporal Filtering Methods for Enhancement and Separation of Speech Signals''), and in part by the Shenzhen Natural Science Foundation Projects under Grant JCYJ20160226192223251, Grant JCYJ20170412104656685, and Grant JCYJ20170302142312688.

ABSTRACT The problem of parametric modeling for the two-dimensional (2-D) harmonic signals with missing harmonics is addressed. The modeling process consists of two parts. First, we devise the joint spatial-temporal linearly constrained minimum variance beamformer and perform the joint estimation of the spatial and temporal fundamental frequencies for each pitch of the signal based on the maximum harmonic model. Second, we differentiate the competing signal models by the number of harmonic components and derive the maximum *a posteriori* criterion for the 2-D harmonic signals. As a result, the harmonic components of each pitch are detected according to their spectral powers. The simulation and experimental results are provided to show the superiority of the proposed signal modeling methodology over other existing schemes.

INDEX TERMS Fundamental frequency estimation, harmonic detection, two-dimensional harmonic signal, linearly constrained minimum variance beamformer, maximum *a posteriori* criterion, maximum harmonic model.

I. INTRODUCTION

Parametric modeling for the harmonic signal, whose frequencies are integer multiples of the fundamental ones, is a classical but still open problem in the spectral analysis research due to its application in a wide range of areas such as music and voiced speech signal processing [1], [2], biomedical engineering [3], sonar [4] and so on.

In this work, we focus on the area of music and voiced speech signal processing, where the signal of interest (SOI) is modeled as the multi-pitch harmonic signal [5]. Correspondingly, the parametric modeling consists of two parts - harmonic detection and parameter estimation for each source (or pitch), by which we mean the detection of the indexes of the individual harmonic components and the estimation of their characteristic parameters. For the parameter estimation, the fundamental frequencies are of most interest. Once their estimates are obtained, the remaining linear parameters such as amplitudes and initial phases, can be computed as a linear least-squares (LLS) solution [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Jing Liang.

Conventionally, the fundamental frequency estimation is performed in the single-channel setup [7]–[13]. Consequently, when the frequencies of the harmonic components from different pitches are close to each other relative to the length of the observation, that is, the problem of spectrally overlapping harmonics occurs, it is difficult to differentiate these pitches. To achieve better source separation and improve the estimation accuracy, the multiple-channel setup of microphone arrays has been adopted in the recent years [14]–[16], where the music and voiced speech signals are measured from the multiple channels, that is microphones, successively at some time delay. As a result, the measured signals constitute a two-dimensional (2-D) harmonic signal, which is characterized by the spatial-temporal fundamental frequencies. Accordingly, several joint estimation methods for the 2-D fundamental frequencies have been proposed in the relevant works [14]–[18], in order to enhance the source separation. Since more data are observed from the multiple channels than those from the single channel, the parameter estimation refinement is also expected [19]. However, most of these methods are parametric, and do not perform well if the information about the harmonic indexes is

not complete. The common nonparametric estimation methods based on the fast Fourier transform (FFT) can handle this case, but subjected to the limited ability to resolve the closely-spaced frequencies relative to the length of the observation signal [20]. Hence, a kind of high-resolution nonparametric estimation method is needed in the case of the incomplete signal model.

When it comes to the harmonic detection, we note that it often happens in practice that the frequencies of the harmonic components consist of just a subset of the consecutive integer multiples of the fundamental quantities. This means that some integer-order harmonics are not present in the observed signal. For example, in the public switched telephone network (PSTN), the speech signals are commonly high-pass filtered at about 300 Hz, meaning that the lower harmonics for speakers with the frequencies below 300 Hz are missing from the observed signal. Similarly, the harmonic signal may consist of only odd-order or even-order harmonics [5], or some harmonic components may be so weak that they are buried in the background noise. In these situations, the appearance or disappearance of the harmonic components incurs the exponential growth of the number of the potential signal models with respect to the maximum possible order of the harmonic components of the signal. Therefore, it is not appropriate to utilize such model comparison methods as [21]–[23] to determine which harmonic components exist or miss due to the high computational burden. To overcome the above difficulty, it is proposed to characterize the candidate signal models only in terms of the number of harmonics, and we aim to estimate the signal order, that is the number of the harmonic components existing in the observed signal. As a result, the harmonic components are detected as the ones with the largest spectral powers.

In the signal order estimation, one class of the methods are based on the rank determination of the data matrix, including the order estimators of MUltiple SIgnal Classification (MUSIC) [24], ESTimation ERror (ESTER) [25] and Subspace-based Automatic Model Order Selection (SAMOS) [26], etc. Although it is easy to conduct rank determination following their respective criteria, these methods do not utilize the statistical model of the observed data. In [27], various information theoretic criteria have been reviewed to select the order of the sinusoidal signals in white Gaussian noise, such as the minimum description length (MDL) criterion, direct Kullback-Leibler (KL) approach, cross-validatory KL approach based on the Akaike information criterion (AIC), generalized cross-validatory KL approach based on the generalized information criterion (GIC) and Bayesian approach based on the Bayesian information criterion (BIC). In [28], the authors try to estimate the order of the one-dimensional (1-D) sinusoidal signals from the Bayesian viewpoint, and derive the maximum *a posteriori* (MAP) criterion, which is regarded as optimal in the sense of maximizing the average probability of correct detection. In [29], the MAP criterion is extended to the 2-D sinusoidal signals. Nevertheless, the authors consider the general model of the 2-D sinusoidal signal, and do not take the harmonic structure into account when deriving the 2-D MAP criterion. What is more important is that, in [29], the asymptotic assumptions of the lengths of the two dimensional data are both used, which is impractical for the number of the signal channels.

In this paper, we model the 2-D harmonic signals by combining the parameter estimation and harmonic detection together. Firstly, we propose the maximum harmonic model, which encompasses all the candidate models of the 2-D harmonic signals. Then, we develop a high-resolution nonparametric estimator, that is the joint spatial-temporal linearly constrained minimum variance (JST-LCMV) beamformer, to perform the joint estimation of the 2-D fundamental frequencies based on the maximum harmonic model. Accordingly, the spectral powers of all the estimated harmonic components in the maximum harmonic model are obtained. Secondly, the harmonic detection procedure is proposed, where we estimate the signal order by extending the MAP criterion to the 2-D harmonic signal, and detect the harmonic components according to their spectral power estimates.

The rest of this paper is organized as follows. The problem of parametric modeling for the 2-D harmonic signals with missing harmonics is formulated in Section II. The proposed parametric modeling methodology for the 2-D harmonic signal is presented in Section III, which includes the joint estimation of the 2-D fundamental frequencies with the JST-LCMV beamformer and the detection of the existing harmonic components based on the MAP criterion and the spectral powers of the harmonic components of the maximum harmonic model. Simulation and experimental results are presented in Section IV to evaluate the performance of the proposed parametric modeling framework by comparing with other parameter estimation and harmonic detection methods. Finally, conclusions are drawn in Section V.

II. PROBLEM FORMULATION

In this paper, we consider the model of the 2-D harmonic signal without multi-path components, which is generated by the multiple recordings of a 1-D harmonic signal $s(n)$. Here, the signal $s(n)$ consists of K pitches:

$$s(n) = \sum_{k=1}^{K} s^{(k)}(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \rho_{l,k} e^{j\omega_k p_{l,k} n}, \tag{1}$$

for $n = 1, 2, \cdots, N$, where $j = \sqrt{-1}$ stands for the unit imaginary number, N denotes the data length, ω_k and $p_{l,k}$ ($p_{1,k} < \cdots < p_{L_k,k}$) are the temporal fundamental frequency and the harmonic indexes of the k-th pitch, respectively, and $\rho_{l,k}$ is the complex-valued amplitude of the l-th harmonic component of the k-th pitch. Since $p_{l,k}$ ($l = 1, \cdots, L_k$) are not necessarily consecutive integers, there maybe exist missing harmonics in $s(n)$.

Assume that each pitch of $s(n)$, that is $s^{(k)}(n)$, is recorded through the channels $1, 2, \cdots, I$, successively, and the time delay between any two successive recordings is τ_k.

As a result, the signal measured from the i-th channel is given by:

$$x_i(n) = s_i(n) + q_i(n), \tag{2}$$

$$s_i(n) = \beta_i \sum_{k=1}^{K} s^{(k)}(n - f_s \tau_k i), \tag{3}$$

for $n = 1, 2, \cdots, N$ and $i = 1, 2, \cdots, I$, where $s_i(n)$ and $q_i(n)$ represent the SOI and noise parts of $x_i(n)$, respectively, β_i is the gain of the i-th channel, and f_s is the sampling frequency. Here, $N \gg I$, which is the normal case in practice. Without loss of generality, β_i $(i = 1, \cdots, I)$ are set as 1 uniformly here. The noises $q_i(n)$ are assumed to be uncorrelated white Gaussian with the unknown variance σ^2. According to (1), $s_i(n)$ of (3) is further expressed as:

$$\begin{aligned} s_i(n) &= \sum_{k=1}^{K} \sum_{l=1}^{L_k} \rho_{l,k} e^{j\omega_k p_{l,k}(n - f_s \tau_k i)} \\ &= \sum_{k=1}^{K} \sum_{l=1}^{L_k} \rho_{l,k} e^{j(\omega_k p_{l,k} n + \theta_k p_{l,k} i)}, \end{aligned} \tag{4}$$

with $\theta_k \triangleq -\omega_k f_s \tau_k$ being the so-called spatial fundamental frequency. Consequently, $s_i(n)$ $(n = 1, \cdots, N; i = 1, \cdots, I)$ constitute a 2-D harmonic signal.

The model of the 2-D harmonic signal, that is $s_i(n)$ of (4), is well-suited for the application of processing the voiced speech signals and many musical instrumental signals at microphone arrays [14]–[17]. When the microphones are arranged as the uniform linear array (ULA), for example, we have that $\tau_k = d \sin \alpha_k / c$, and

$$\theta_k = -\omega_k f_s d \sin \alpha_k / c, \tag{5}$$

with α_k, d and c denoting the direction-of-arrival (DOA), inter-element spacing and propagation speed of the sound wave, respectively. Note that when the multi-path propagation occurs, the signal model of (4) is still valid but with $\omega_{k_1} = \omega_{k_2}$, for some $k_1 \neq k_2$.

It should be addressed that in this paper, we assume that the multiple recordings of each pitch of the 1-D harmonic signal are spaced with the equal time delay. Based on this point, we develop the joint fundamental frequency estimation method in Section III. When the 1-D harmonic signal is measured spatially at a non-uniform rate, we can utilize the interpolation-based techniques such as [30]–[32] to recover the multiple recordings spaced with the equal time delay as (2).

In this paper, we aim to devise a high-resolution nonparametric method to estimate the 2-D fundamental frequencies of (ω_k, θ_k) jointly under the assumption of the unknown harmonic indexes, and then to determine the harmonic indexes of each pitch, that is $p_{l,k}$, based on the parameter estimates. Afterwards, the parametric modeling of the 2-D harmonic signal $s_i(n)$ is completed. Here, K is assumed known *a priori*. Nevertheless, as shown in Section IV.E, when K is unknown and relaxed as $K \leq K_{max}$, our methodology is robust to such difficulty and can still select the correct model as when K is known. As a result, the number of sources is determined automatically.

III. PARAMETRIC MODELING METHODOLOGY

The work of parametrically modeling the 2-D harmonic signal consists of two parts: i) the joint estimation of the temporal and spatial fundamental frequencies with the JST-LCMV beamformer; ii) the detection of the existing harmonics with the MAP criterion. In this section, the detail of the development of the JST-LCMV beamformer and the MAP criterion is illustrated.

A. 2-D FUNDAMENTAL FREQUENCY ESTIMATION WITH THE JST-LCMV BEAMFORMER

Since the information about the harmonic indexes is unavailable, the parametric methods are not appropriate in the 2-D fundamental frequency estimation here. To overcome such difficulty, we turn to the nonparametric methods. Traditionally, the nonparametric parameter estimation methods are classified into two categories - signal independent and signal dependent [20]. For the former, the FFT-based method is extensively used, where we search for the frequency estimates according to the locations of the peaks of the signal's periodogram. This kind of method is essentially a filtering approach, where a bandpass filter is utilized with the finite impulse response (FIR) given by the standard Fourier transform vector (e.g., for the 1-D signal, $\left[1 \; e^{-j\tilde{\omega}} \; \cdots \; e^{-j(N-1)\tilde{\omega}}\right]^T$). However, the FFT-based approach is subjected to the well-known frequency resolution limit of the FFT (see Section 2.4 of [20]).

To alleviate this problem, several signal dependent beamformers have been developed (see [33] for an overview), among which the linearly constrained minimum variance (LCMV) beamformer is popular due to its simplicity and effectiveness. Since the LCMV beamformer is designed through passing the SOI undistorted while suppressing the signal out of interest as much as possible, it becomes more SOI selective and bears higher frequency resolution than the signal independent methods. The standard version of the LCMV beamformer considers the spectral analysis only in one-dimension (temporal domain or spatial domain). To achieve the joint estimation of 2-D fundamental frequencies, here we extend the LCMV beamformer to the two-dimensional scenario of the temporal and spatial domains, and develop the JST-LCMV beamformer. To improve the estimation accuracy, we make use of the harmonic relation of the signal components in devising the JST-LCMV beamformer.

Due to the lack of the information about the harmonic indexes, we define the maximum harmonic model for the observed signal at the first step:

$$x_i(n) = s_i(n) + q_i(n), \tag{6}$$

$$s_i(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_{max,k}} \rho'_{l,k} e^{j(\omega_k l n + \theta_k l i)}, \tag{7}$$

where $L_{max,k}$ is the maximum possible order of the harmonics for the k-th pitch (which means $p_{L_k,k} \leq L_{max,k}$), and $\rho'_{l,k}$ is the complex-valued amplitude of the l-th harmonic component of the k-th pitch in the maximum harmonic model. Comparing (4) and (7), it is seen that the maximum harmonic model covers all the possible signal models of $s_i(n)$ of (4) with respect to different L_k and $p_{l,k}$; and $\rho'_{l,k} = 0$, for $l \neq p_{1,k}, \cdots, p_{L_k,k}$.

Now considering a segment of the signal $x_i(n)$ sampled at N_s temporal instants and from I_s channels, we construct the $N_s \times I_s$ data matrix $\mathbf{X}_{i_s}(n_s)$ as: (see (8), as shown at the bottom of the next page), for $n_s = 1, \cdots, N-N_s+1$, $i_s = 1, \cdots, I-I_s+1$, and express the FIR of the 2-D filter at the temporal and spatial frequencies $(\tilde{\omega}, \tilde{\theta})$, as [34]: (see (9), as shown at the bottom of the next page). Furthermore, we convert the 2-D data matrix and filter response into the 1-D ones by means of vectorization as:

$$\mathbf{x}_{i_s}(n_s) = vec\left\{\mathbf{X}_{i_s}(n_s)\right\}, \tag{10}$$

$$\mathbf{h}_{\tilde{\omega},\tilde{\theta}} = vec\left\{\mathbf{H}_{\tilde{\omega},\tilde{\theta}}\right\}, \tag{11}$$

with $vec\{\cdot\}$ denoting the column-wise stacking operator.

Our proposed signal dependent joint spatial-temporal filtering method for the joint estimation of (ω_k, θ_k) is based on the LCMV beamformer [35]. Different from the FFT-based method, we utilize the single spatial-temporal filter with multiple harmonic constraints. The JST-LCMV beamformer aims to minimize the output power subject to a distortionless response at certain frequency pairs. The output of the spatial-temporal filter is given by:

$$y_{i_s}(n_s) = \mathbf{h}^H_{\tilde{\omega},\tilde{\theta}} \mathbf{x}_{i_s}(n_s), \tag{12}$$

with the superscript of H denoting the Hermitian transpose. Accordingly, the output power is:

$$\begin{aligned} E\left\{y_{i_s}(n_s) y^H_{i_s}(n_s)\right\} &= \mathbf{h}^H_{\tilde{\omega},\tilde{\theta}} E\left\{\mathbf{x}_{i_s}(n_s) \cdot \mathbf{x}^H_{i_s}(n_s)\right\} \mathbf{h}_{\tilde{\omega},\tilde{\theta}} \\ &= \mathbf{h}^H_{\tilde{\omega},\tilde{\theta}} \mathbf{R} \mathbf{h}_{\tilde{\omega},\tilde{\theta}}, \end{aligned} \tag{13}$$

where $\mathbf{R}$ is the covariance matrix of the signal $\mathbf{x}_{i_s}(n_s)$. With the maximum possible harmonic order for the k-th pitch, $L_{max,k}$, and its corresponding 2-D fundamental frequency (ω_k, θ_k), the LCMV beamformer design problem for the k-th pitch is formulated as the following optimization problem:

$$\begin{aligned} \mathbf{h}_{\omega_k,\theta_k} = \arg\min_{\tilde{\mathbf{h}}} \tilde{\mathbf{h}}^H \mathbf{R} \tilde{\mathbf{h}}, \\ s.t.\ \tilde{\mathbf{h}}^H \mathbf{a}(l\omega_k, l\theta_k) = 1, \\ for\ l = 1, 2, \cdots, L_{max,k}, \end{aligned} \tag{14}$$

where $\mathbf{a}(\tilde{\omega}, \tilde{\theta}) \in \mathbb{C}^{(I_s N_s)\times 1}$ is the spatial-temporal Fourier transform vector, and is defined as:

$$\mathbf{a}(\tilde{\omega}, \tilde{\theta}) = \mathbf{a}_2(\tilde{\theta}) \otimes \mathbf{a}_1(\tilde{\omega}), \tag{15}$$

$$\mathbf{a}_1(\tilde{\omega}) = \left[1\ e^{-j\tilde{\omega}}\ \cdots\ e^{-j(N_s-1)\tilde{\omega}}\right]^T, \tag{16}$$

$$\mathbf{a}_2(\tilde{\theta}) = \left[1\ e^{-j\tilde{\theta}}\ \cdots\ e^{-j(I_s-1)\tilde{\theta}}\right]^T, \tag{17}$$

with $\otimes$ denoting the Kronecker product. The above optimization problem is solved by using the Lagrange multiplier method [36], which gives the following expression for the optimal stacked spatial-temporal filter response:

$$\begin{aligned} \mathbf{h}_{\omega_k,\theta_k} = \mathbf{R}^{-1}\mathbf{A}_k(\omega_k, \theta_k) \cdot \\ (\mathbf{A}^H_k(\omega_k, \theta_k)\mathbf{R}^{-1}\mathbf{A}_k(\omega_k, \theta_k))^{-1}\mathbf{1}_{L_{max,k}}, \end{aligned} \tag{18}$$

with $\mathbf{1}_{L_{max,k}}$ denoting an $L_{max,k} \times 1$ vector with all the elements being one, and $\mathbf{A}_k(\tilde{\omega}, \tilde{\theta})$ defined as:

$$\mathbf{A}_k(\tilde{\omega}, \tilde{\theta}) = \left[\mathbf{a}(\tilde{\omega}, \tilde{\theta})\ \cdots\ \mathbf{a}(L_{max,k}\tilde{\omega}, L_{max,k}\tilde{\theta})\right]. \tag{19}$$

By inserting the optimal filter response (18) into the output power expression (13), the 2-D fundamental frequencies of the harmonic signal $s_i(n)$ of (4), are estimated in a joint way by searching for the peaks of the output power $J(\tilde{\omega}, \tilde{\theta})$ over the set of the 2-D fundamental frequency candidate $\mathbf{\Omega} \times \mathbf{\Theta}$, with

$$J(\tilde{\omega}, \tilde{\theta}) = \mathbf{1}^H_{L_{max,k}} \left(\mathbf{A}^H_k(\tilde{\omega}, \tilde{\theta})\mathbf{R}^{-1}\mathbf{A}_k(\tilde{\omega}, \tilde{\theta})\right)^{-1} \mathbf{1}_{L_{max,k}}, \tag{20}$$

for $\tilde{\omega} \in \Omega = (0, 2\pi)$ and $\tilde{\theta} \in \Theta = (0, 2\pi)$.

In the real-life setting, it is impossible to access the covariance matrix $\mathbf{R}$, which is usually estimated in the forward way as:

$$\begin{aligned} \hat{\mathbf{R}}_F = \frac{1}{(N-N_s+1)(I-I_s+1)} \\ \cdot \sum_{n_s=1}^{N-N_s+1} \sum_{i_s=1}^{I-I_s+1} \mathbf{x}_{i_s}(n_s) \cdot \mathbf{x}^H_{i_s}(n_s). \end{aligned} \tag{21}$$

It is seen from the signal model of (2) that $\mathbf{x}_{i_s}(n_s)$ of (10) is second-order stationary. Thus, $\mathbf{R}$ can also be estimated in the backward way as (see Section 4.8 of [20]):

$$\hat{\mathbf{R}}_B = \mathbf{J}_0 \hat{\mathbf{R}}^T_F \mathbf{J}_0, \tag{22}$$

with $\mathbf{J}_0$ denoting the exchange matrix with the anti-diagonal elements being 1 and the rest being 0. To enhance the estimation accuracy, here $\mathbf{R}$ is estimated by averaging $\hat{\mathbf{R}}_F$ and $\hat{\mathbf{R}}_B$ as follows:

$$\hat{\mathbf{R}} = \frac{1}{2}\left(\hat{\mathbf{R}}_F + \hat{\mathbf{R}}_B\right). \tag{23}$$

In the practical procedure, we estimate the K pairs of 2-D fundamental frequencies of $s_i(n)$ of (4) one by one. Firstly, we obtain the coarse estimate of each pair of the 2-D fundamental frequencies, denoted by $(\hat{\omega}^{(0)}_k, \hat{\theta}^{(0)}_k)$, from the grid of its admissible range. Then, we conduct the estimation refinement by searching for the locally minimum point of $J(\tilde{\omega}, \tilde{\theta})$ of (20) from $(\hat{\omega}^{(0)}_k, \hat{\theta}^{(0)}_k)$ based on the gradient-based method. In detail, we calculate the estimate for the k-th pitch iteratively as:

$$\begin{bmatrix} \hat{\omega}^{(i+1)}_k \\ \hat{\theta}^{(i+1)}_k \end{bmatrix} = \begin{bmatrix} \hat{\omega}^{(i)}_k \\ \hat{\theta}^{(i)}_k \end{bmatrix} + \delta \cdot \nabla J(\tilde{\omega}, \tilde{\theta})\Big|_{\tilde{\omega}=\hat{\omega}^{(i)}_k, \tilde{\theta}=\hat{\theta}^{(i)}_k}, \tag{24}$$

where i and $i+1$ are the iteration indexes, $\delta > 0$ is a small constant which is found using a line search

algorithm [37], and $\nabla J(\tilde{\omega},\tilde{\theta}) = \left[\frac{\partial J(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\omega}} \ \frac{\partial J(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\theta}}\right]^T$ is the gradient of $J(\tilde{\omega},\tilde{\theta})$. According to the rules of matrix derivative [38], $\nabla J(\tilde{\omega},\tilde{\theta})$ is: (see (25), as shown at the bottom of this page), where $\Re\mathfrak{e}\{\cdot\}$ stands for the operation of taking real part, and $\mathbf{Q}_k(\tilde{\omega},\tilde{\theta}) = \left(\mathbf{A}_k^H(\tilde{\omega},\tilde{\theta})\hat{\mathbf{R}}^{-1}\mathbf{A}_k(\tilde{\omega},\tilde{\theta})\right)^{-1}$, $\mathbf{B}_{1,k} = \frac{\partial\mathbf{A}_k(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\omega}}$, $\mathbf{B}_{2,k} = \frac{\partial\mathbf{A}_k(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\theta}}$.

It should be noted that in the fundamental frequency estimation of the harmonic signal with missing harmonics, there maybe exist relatively lower but still high peaks of $J(\tilde{\omega},\tilde{\theta})$ of (20) at the frequency pairs corresponding to the integer multiples or divisors of the true fundamental frequencies [5]. For the multi-pitch signal, such peaks are probably higher than the peak at another fundamental frequency pair and thus, their locations may be mistaken as the fundamental frequency estimates. To overcome this difficulty, it is proposed to cancel the estimated harmonic components corresponding to the k-th pitch from the observation once $(\hat{\omega}_k,\hat{\theta}_k)$ is solved, and then, to select the initial estimate of the next pair of 2-D fundamental frequency, that is $(\hat{\omega}_{k+1}^{(0)},\hat{\theta}_{k+1}^{(0)})$, based on such preprocessed data. Define the observation as:

$$\mathbf{x} = \left[\mathbf{x}_1^T \ \mathbf{x}_2^T \ \cdots \ \mathbf{x}_I^T\right]^T, \tag{26}$$

$$\mathbf{x}_i = \left[x_i(1) \ x_i(2) \ \cdots \ x_i(N)\right]^T, \tag{27}$$

and we cancel the harmonic components corresponding to the k-th pitch by the projection operation [20]:

$$\begin{aligned}\mathbf{x}_{\backslash k} &= \Big(\mathbf{I}_{IN} - \mathbf{P}_k(\hat{\omega}_k,\hat{\theta}_k)\left(\mathbf{P}_k^H(\hat{\omega}_k,\hat{\theta}_k)\mathbf{P}_k(\hat{\omega}_k,\hat{\theta}_k)\right)^{-1} \\ &\quad \mathbf{P}_k^H(\hat{\omega}_k,\hat{\theta}_k)\Big)\mathbf{x} \\ &\triangleq \mathbf{P}_k^{\perp}\mathbf{x},\end{aligned} \tag{28}$$

where $\mathbf{I}_{IN}$ is an $(IN)\times(IN)$ identity matrix, and

$$\mathbf{P}_k(\hat{\omega}_k,\hat{\theta}_k) = \left[\mathbf{p}(\hat{\omega}_k,\hat{\theta}_k) \ \cdots \ \mathbf{p}(L_{max,k}\hat{\omega}_k, L_{max,k}\hat{\theta}_k)\right],$$

with

$$\begin{aligned}\mathbf{p}(\tilde{\omega},\tilde{\theta}) &\triangleq \mathbf{p}_2(\tilde{\theta})\otimes\mathbf{p}_1(\tilde{\omega}), \\ \mathbf{p}_1(\tilde{\omega}) &\triangleq \left[e^{j\tilde{\omega}} \ e^{j2\tilde{\omega}} \ \cdots \ e^{jN\tilde{\omega}}\right]^T, \\ \mathbf{p}_2(\tilde{\theta}) &\triangleq \left[e^{j\tilde{\theta}} \ e^{j2\tilde{\theta}} \ \cdots \ e^{jI\tilde{\theta}}\right]^T.\end{aligned}$$

Up to now, the procedure of the 2-D fundamental frequency estimation is summarized in Algorithm 1.

Algorithm 1 Procedure of the 2-D Fundamental Frequency Estimation

Input: The signal observation $x_i(n)$, $n = 1,\cdots,N$, $i = 1,\cdots,I$, and the maximum possible orders $L_{max,k}$, $k = 1,\cdots,K$. Stack $x_i(n)$ as $\mathbf{x}$, and set $\mathbf{x}^\circ = \mathbf{x}$.

Output: The 2-D fundamental frequency estimates $(\hat{\omega}_k,\hat{\theta}_k)$, $k = 1,\cdots,K$.

for $k = 1; k \le K; k++$ **do**
- Estimate the covariance matrix $\mathbf{R}$ with $\mathbf{x}^\circ$ as (23);
- Construct the cost function $J(\tilde{\omega},\tilde{\theta})$ of (20) based on the maximum harmonic model of (6) with $L_{max,k}$;
- Obtain the coarse estimate of the fundamental frequency of the k-th pitch, $(\hat{\omega}_k^{(0)},\hat{\theta}_k^{(0)})$, from the grid of $\mathbf{\Omega}\times\mathbf{\Theta}$;
- Solve the refined estimate $(\hat{\omega}_k,\hat{\theta}_k)$ by searching for the peak of $J(\tilde{\omega},\tilde{\theta})$ of (20) near $(\hat{\omega}_k^{(0)},\hat{\theta}_k^{(0)})$;
- Cancel the harmonic components corresponding to the k-th pitch from the observation $x_i(n)$ by (28) as
$$\mathbf{x}^\circ = \mathbf{P}_k^{\perp}\mathbf{x}^\circ;$$

end

Return $(\hat{\omega}_k,\hat{\theta}_k)$, $k = 1,\cdots,K$.

According to (12)-(19), we can estimate the spectral power at some 2-D frequency pair $(\tilde{\omega},\tilde{\theta})$ by the JST-LCMV

$$\mathbf{X}_{i_s}(n_s) = \begin{bmatrix} x_{i_s}(n_s) & x_{i_s+1}(n_s) & \cdots & x_{i_s+I_s-1}(n_s) \\ x_{i_s}(n_s+1) & x_{i_s+1}(n_s+1) & \cdots & x_{i_s+I_s-1}(n_s+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_{i_s}(n_s+N_s-1) & x_{i_s+1}(n_s+N_s-1) & \cdots & x_{i_s+I_s-1}(n_s+N_s-1)\end{bmatrix} \tag{8}$$

$$\mathbf{H}_{\tilde{\omega},\tilde{\theta}} = \begin{bmatrix} \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(0,0) & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(0,1) & \cdots & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(0,I_s-1) \\ \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(1,0) & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(1,1) & \cdots & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(1,I_s-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(N_s-1,0) & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(N_s-1,1) & \cdots & \mathbf{H}_{\tilde{\omega},\tilde{\theta}}(N_s-1,I_s-1)\end{bmatrix} \tag{9}$$

$$\begin{aligned}\nabla J(\tilde{\omega},\tilde{\theta}) &= \left[\frac{\partial J(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\omega}} \ \frac{\partial J(\tilde{\omega},\tilde{\theta})}{\partial\tilde{\theta}}\right]^T \\ &= -2\Re\mathfrak{e}\left\{\begin{bmatrix}\mathbf{1}_{L_{max,k}}^H\mathbf{Q}_k(\tilde{\omega},\tilde{\theta})\mathbf{A}_k^H(\tilde{\omega},\tilde{\theta})\hat{\mathbf{R}}^{-1}\mathbf{B}_{1,k}(\tilde{\omega},\tilde{\theta})\mathbf{Q}_k(\tilde{\omega},\tilde{\theta})\mathbf{1}_{L_{max,k}} \\ \mathbf{1}_{L_{max,k}}^H\mathbf{Q}_k(\tilde{\omega},\tilde{\theta})\mathbf{A}_k^H(\tilde{\omega},\tilde{\theta})\hat{\mathbf{R}}^{-1}\mathbf{B}_{2,k}(\tilde{\omega},\tilde{\theta})\mathbf{Q}_k(\tilde{\omega},\tilde{\theta})\mathbf{1}_{L_{max,k}}\end{bmatrix}\right\}\end{aligned} \tag{25}$$

beamformer as:

$$s_{LCMV}(\tilde{\omega},\tilde{\theta}) = \frac{1}{\mathbf{a}^H(\tilde{\omega},\tilde{\theta})\hat{\mathbf{R}}^{-1}\mathbf{a}(\tilde{\omega},\tilde{\theta})}. \quad (29)$$

Consequently, the spectral power of the l-th harmonic component of the k-th pitch in the maximum harmonic model, is estimated as:

$$s_{LCMV}(l\hat{\omega}_k, l\hat{\theta}_k) = \frac{1}{\mathbf{a}^H(l\hat{\omega}_k, l\hat{\theta}_k)\hat{\mathbf{R}}^{-1}\mathbf{a}(l\hat{\omega}_k, l\hat{\theta}_k)}. \quad (30)$$

The above spectral power estimate for the maximum harmonic model will be utilized in the harmonic detection as detailed in the next part.

B. HARMONIC DETECTION WITH THE MAP CRITERION

In this part, we devise the harmonic detection scheme via extending the well-known MAP criterion [28] into the case of the 2-D harmonic signal and with the estimated 2-D fundamental frequencies. In the MAP model selection criterion, it is assumed that there exist a series of candidate models, and the correct model is determined with the largest posterior probability given the observation data. The posterior probability of each model is calculated according to the Bayesian rule and depends on the prior probability of the candidate models and unknown parameters.

According to (4), (2) is rewritten in the form of matrix-vector multiplication as follows:

$$\mathbf{x} = \mathbf{\Pi}\boldsymbol{\rho} + \mathbf{q}, \quad (31)$$

where $\mathbf{q}$ is the noise part of $\mathbf{x}$,

$$\boldsymbol{\rho} = \left[\boldsymbol{\rho}_1^T \ \boldsymbol{\rho}_2^T \ \cdots \ \boldsymbol{\rho}_K^T\right]^T,$$
$$\boldsymbol{\rho}_k = \left[\rho_{1,k} \ \rho_{2,k} \ \cdots \ \rho_{L_k,k}\right]^T,$$

and

$$\mathbf{\Pi} = \left[\mathbf{\Pi}_1 \ \mathbf{\Pi}_2 \ \cdots \ \mathbf{\Pi}_K\right],$$
$$\mathbf{\Pi}_k = \left[\mathbf{p}(p_{1,k}\omega_k, p_{1,k}\theta_k) \ \cdots \ \mathbf{p}(p_{L_k,k}\omega_k, p_{L_k,k}\theta_k)\right],$$

for $k = 1, 2, \cdots, K$.

Now we assume that there exist M candidate models, which are differentiated in terms of the number of harmonics, and are indexed as $m = 1, 2, \cdots, M$, respectively. For the m-th model, there exist m harmonic components across the K_m present pitches. This means that these m harmonic components come from K_m different pitches. Correspondingly, the m-th model includes the unknown parameters of the K_m pairs of 2-D fundamental frequencies, the m complex-valued amplitudes and the noise level, which are denoted by $\boldsymbol{\psi}_{K_m} \in \mathbb{R}^{2K_m\times 1}$, $\boldsymbol{\rho}_m \in \mathbb{C}^{m\times 1}$ and $\sigma > 0$, respectively. Here, $\boldsymbol{\psi}_{K_m}$ is further represented by $\boldsymbol{\psi}_{K_m} = \left[\boldsymbol{\omega}_{K_m}^T \ \boldsymbol{\theta}_{K_m}^T\right]^T$, with $\boldsymbol{\omega}_{K_m} \triangleq \left[\omega_1 \ \cdots \ \omega_{K_m}\right]^T$ and $\boldsymbol{\theta}_{K_m} \triangleq \left[\theta_1 \ \cdots \ \theta_{K_m}\right]^T$ denoting the temporal and spatial fundamental frequencies in the m-th model, respectively.

To calculate the posterior probability of the model m given the observed data $\mathbf{x}$, denoted by $p(m|\mathbf{x})$, the prior probability of the candidate models and the unknown parameters should be selected first. It is assumed *a priori* that the M competing models are equiprobable. That is, the prior probability of the m-th model:

$$p(m) = \frac{1}{M}, \ m = 1, 2, \cdots, M. \quad (32)$$

As a result, to maximize $p(m|\mathbf{x})$ is equivalent to the maximization of $p(\mathbf{x}|m)$, that is the conditional probability distribution function (PDF) of $\mathbf{x}$ given the model m [39]. The goal of harmonic detection is to derive a model selection rule based on the noninformative prior of the unknown parameters. In other words, we should select such priors that they can represent the lack of the prior knowledge of the values of the unknown parameters before the data are observed [28]. In addition, we assume that all the unknown parameters, that is $\boldsymbol{\psi}_{K_m}$, $\boldsymbol{\rho}_m$ and σ, are independent of each other.

According to the Bayesian rule, we have that the prior PDF of the unknown parameters given the m-th model is expressed as:

$$p(\boldsymbol{\psi}_{K_m}, \boldsymbol{\rho}_m, \sigma|m) = p(\boldsymbol{\rho}_m, \sigma|\boldsymbol{\psi}_{K_m}, m) \cdot p(\boldsymbol{\psi}_{K_m}|m). \quad (33)$$

Assuming that the fundamental frequencies are independent of each other, the lack of the prior knowledge of these frequencies means that $\boldsymbol{\psi}_{K_m}$ are uniformly distributed in $\mathbb{D}_{K_m} \triangleq [0, \ 2\pi]^{2K_m}$. Thus,

$$p(\boldsymbol{\psi}_{K_m}|m) = \frac{1}{(2\pi)^{2K_m}}. \quad (34)$$

Due to the independence of $\boldsymbol{\rho}_m$ and σ, it holds that

$$p(\boldsymbol{\rho}_m, \sigma|\boldsymbol{\psi}_{K_m}, m) = p(\boldsymbol{\rho}_m|\boldsymbol{\psi}_{K_m}, m) \cdot p(\sigma|\boldsymbol{\psi}_{K_m}, m). \quad (35)$$

From the assumption that little is known *a priori* relative to the information contained in the observed data, the prior PDF of $\boldsymbol{\rho}_m$ and σ are locally uniform [39], or equivalently,

$$p(\boldsymbol{\rho}_m|\boldsymbol{\psi}_{K_m}, m) = \gamma(m), \quad (36)$$

and

$$p(\sigma|\boldsymbol{\psi}_{K_m}, m) = c\sigma^{-1}, \quad (37)$$

where $c > 0$ is a constant, and $\gamma(m)$ is positively constant for any given m. Note that, (36) and (37) result in an improper prior distribution of (35) if defined in $\mathbb{C}^{m\times 1} \times \mathbb{R}^+$ or $\mathbb{R}^{2m\times 1} \times \mathbb{R}^+$. Hence, (36) and (37) are defined only locally, and not over the entire definition domain. This is consistent with the boundedness of the amplitudes $\boldsymbol{\rho}_m$ and the noise level σ. In detail, (36) and (37) define the priors of $\boldsymbol{\rho}_m$ and σ only over the range where the likelihood functions of the corresponding parameters are close to or at the maximum, whereas the priors decay to zero outside this range to ensure that they represent proper PDFs.

Note that the noise $q_i(n)$ is white Gaussian. On the assumption of the true model m and the true parameter values $\boldsymbol{\psi}_{K_m}$, $\boldsymbol{\rho}_m$ and σ, the PDF of the observed data is:

$$p(\mathbf{x}|m, \boldsymbol{\psi}_{K_m}, \boldsymbol{\rho}_m, \sigma) = \frac{1}{\pi^{NI}\sigma^{2NI}} \cdot e^{-\frac{1}{\sigma^2}(\mathbf{x}-\mathbf{D}_m\boldsymbol{\rho}_m)^H(\mathbf{x}-\mathbf{D}_m\boldsymbol{\rho}_m)}, \quad (38)$$

where

$$\mathbf{D}_m = \left[\mathbf{d}_1\ \mathbf{d}_2\ \cdots\ \mathbf{d}_m\right],$$
$$\mathbf{d}_i = \mathbf{p}_2(\bar{\theta}_i) \otimes \mathbf{p}_1(\bar{\omega}_i),$$

for $1 \le i \le m$. For $\bar{\omega}_i$ and $\bar{\theta}_i$, there exist $1 \le k \le K_m$ and $1 \le l \le L_{max,k}$, so that $\bar{\omega}_i = l\omega_k$ and $\bar{\theta}_i = l\theta_k$.

In Appendix A, it is derived that when $N \to \infty$, the conditional PDF of the observed data $\mathbf{x}$ based on the model m,

$$p(\mathbf{x}|m) = C(m)(IN)^{-2m}(\mathbf{x}^H\hat{\mathbf{D}}_m^{\perp}\mathbf{x})^{-IN} \cdot (2\pi)^{K_m}(IN)^{-K_m}|\hat{\mathbf{H}}_m|^{-1/2}, \quad (39)$$

where $C(m) = c2^{-(2K_m+1)}\pi^{-(2K_m-m+IN)}\Gamma(IN)\gamma(m)$, the projection matrix $\mathbf{D}_m^{\perp}$ is defined as follows:

$$\mathbf{D}_m^{\perp} \triangleq \mathbf{I}_{IN} - \mathbf{D}_m(\mathbf{D}_m^H\mathbf{D}_m)^{-1}\mathbf{D}_m^H, \quad (40)$$

the Hessian matrix of $\ln \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}$,

$$\mathbf{H}_m = \frac{\partial^2 \ln \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}}{\partial\boldsymbol{\psi}_{K_m}\partial\boldsymbol{\psi}_{K_m}^T}, \quad (41)$$

$\hat{\mathbf{D}}_m^{\perp}$ and $\hat{\mathbf{H}}_m$ are the versions of $\mathbf{D}_m^{\perp}$ and $\mathbf{H}_m$, respectively, with $\boldsymbol{\psi}_{K_m}$ replaced by $\hat{\boldsymbol{\psi}}_{K_m}$, the maximum likelihood (ML) estimate of $\boldsymbol{\psi}_{K_m}$ for the observation $\mathbf{x}$ and based on the m-th data model.

The signal order estimate of the MAP criterion, denoted by $\hat{M}_{MAP}$, is taken as the candidate order which maximizes the conditional PDF, $p(\mathbf{x}|m)$. Taking the negative logarithm of (39), we have that for the large data length N, the order estimate

$$\begin{aligned}\hat{M}_{MAP} &= \arg\min_{m\in\mathbb{Z}^+}\{-\ln p(\mathbf{x}|m)\}\\ &\approx \arg\min_{m\in\mathbb{Z}^+}(IN)\ln \mathbf{x}^H\hat{\mathbf{D}}_m^{\perp}\mathbf{x}\\ &\quad + (2m+K_m)\ln IN + \frac{1}{2}\ln|\hat{\mathbf{H}}_m|. \end{aligned} \quad (42)$$

To find the determinant of $\hat{\mathbf{H}}_m$, we compute the elements of $\mathbf{H}_m$ one by one, and evaluate their values at the ML estimate of $\boldsymbol{\psi}_{K_m}$. In detail, the first- and second-order derivatives of $\ln \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}$ are:

$$\frac{\partial \ln \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}}{\partial\boldsymbol{\psi}_{K_m}(i_1)} = (\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})^{-1} \cdot \frac{\partial(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}(i_1)} \quad (43)$$

and (see (44), as shown at the bottom of the next page), respectively, with $\boldsymbol{\psi}_{K_m}(i_1)$ and $\boldsymbol{\psi}_{K_m}(i_2)$ denoting the i_1-th and i_2-th elements of $\boldsymbol{\psi}_{K_m}$, respectively, for $i_1, i_2 = 1, 2, \cdots, 2K_m$. Since the ML estimate of $\boldsymbol{\psi}_{K_m}$, that is $\hat{\boldsymbol{\psi}}_{K_m}$, is an extreme point of $\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}$, we have that

$$\left.\frac{\partial \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}}{\partial\boldsymbol{\psi}_{K_m}(i_1)}\right|_{\boldsymbol{\psi}_{K_m}=\hat{\boldsymbol{\psi}}_{K_m}} = 0. \quad (45)$$

As a result, $\hat{\mathbf{H}}_m$ has the following form:

$$\hat{\mathbf{H}}_m = (\mathbf{x}^H\hat{\mathbf{D}}_m^{\perp}\mathbf{x})^{-1} \cdot \hat{\mathbf{H}}_{0,m}, \quad (46)$$

where

$$\mathbf{H}_{0,m} = \frac{\partial^2(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}\partial\boldsymbol{\psi}_{K_m}^T} \quad (47)$$

is the Hessian matrix of $\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}$, and

$$\hat{\mathbf{H}}_{0,m} = \mathbf{H}_{0,m}\big|_{\boldsymbol{\psi}_{K_m}=\hat{\boldsymbol{\psi}}_{K_m}}. \quad (48)$$

Furthermore, (42) is written as

$$\begin{aligned}\hat{M}_{MAP} &\approx \arg\min_{m\in\mathbb{Z}^+}(IN)\ln \mathbf{x}^H\hat{\mathbf{D}}_m^{\perp}\mathbf{x}\\ &\quad + (2m+K_m)\ln IN + \frac{1}{2}\ln|\hat{\mathbf{H}}_{0,m}|. \end{aligned} \quad (49)$$

According to (A.11)-(A.13), $\mathbf{D}_m^{\perp}$ is simplified as

$$\begin{aligned}\mathbf{D}_m^{\perp} &= \mathbf{I}_{IN} - \mathbf{D}_m(\mathbf{D}_m^H\mathbf{D}_m)^{-1}\mathbf{D}_m^H\\ &\approx \mathbf{I}_{IN} - \frac{1}{IN}\mathbf{D}_m\mathbf{D}_m^H, \end{aligned} \quad (50)$$

where $\mathbf{D}_m^H\mathbf{D}_m$ is approximated as $\mathbf{D}_m^H\mathbf{D}_m \approx (IN)\mathbf{I}_m$. Thus, the elements of $\mathbf{H}_{0,m}$ are calculated approximately as: (see (51), as shown at the bottom of the next page).

As a result, the MAP order estimate becomes:

$$\begin{aligned}\hat{M}_{MAP} &\approx \arg\min_{m\in\mathbb{Z}^+}(IN)\ln \mathbf{x}^H\hat{\mathbf{D}}_m^{\perp}\mathbf{x}\\ &\quad + 2m\ln IN + \frac{1}{2}\ln|\hat{\mathcal{H}}_{0,m}|\\ &\triangleq \arg\min_{m\in\mathbb{Z}^+} Q(m), \end{aligned} \quad (52)$$

where the (i_1, i_2)-th element of $\mathcal{H}_{0,m}$ $(i_1, i_2 = 1, 2, \cdots, 2K_m)$ is computed as (see (53), as shown at the bottom of the next page), and

$$\hat{\mathcal{H}}_{0,m} = \mathcal{H}_{0,m}\big|_{\boldsymbol{\psi}_{K_m}=\hat{\boldsymbol{\psi}}_{K_m}}. \quad (54)$$

Correspondingly, the harmonic components of the $\hat{M}_{MAP}$-th model are detected as existing in the 2-D harmonic signal $s_i(n)$ of (4).

Remark 1: Here, we select the m harmonic components with the largest spectral powers from those of the maximum harmonic model, to constitute the m-th model and to calculate the MAP criterion of (52). Instead of calculating the MAP criterion for all the combinations of the candidate harmonic components, we compare it only for $m_{max} = \sum_{k=1}^{K} L_{max,k}$ times, which saves a great deal of computation.

Remark 2: It is seen from (52) that the MAP criterion consists of three terms. The first term is about data fitting. Note that the M candidate models are nested

(referring to Remark 1). Thus, the first term decreases with increasing m. Meanwhile, the second term increases, which penalizes overfitting to a larger degree for a more complex model. It is difficult to analyze the third term directly. But extensive numerical tests show that its variation with respect to m is usually small compared with that of the second term, and it acts as finely tuning the MAP criterion.

Remark 3: In practice, it is computationally prohibitive to solve the ML estimate of $\boldsymbol{\psi}_{K_m}$, that is $\hat{\boldsymbol{\psi}}_{K_m}$ in (52), especially for each m. Here, we obtain the approximate value of $\hat{\boldsymbol{\psi}}_{K_m}$ from Algorithm 1, which is based on the maximum harmonic model and the nonparametric JST-LCMV beamformer.

In summary, the complete procedure of the harmonic detection is listed in Algorithm 2.

Algorithm 2 Procedure of the harmonic detection

Input: The signal observation $x_i(n)$, $n = 1, \cdots, N$, $i = 1, \cdots, I$, the 2-D fundamental frequency estimates $(\hat{\omega}_k, \hat{\theta}_k)$, and the maximum possible orders $L_{max,k}$, $k = 1, \cdots, K$.

Output: The harmonic indexes $p_{l,k}$, $k = 1, \cdots, K$.

1 Set $m_{max} = \sum_{k=1}^{K} L_{max,k}$;
2 Estimate the spectral power $s_{LCMV}(l\hat{\omega}_k, l\hat{\theta}_k)$ $(k = 1, \cdots, K,\ l = 1, \cdots, L_{max,k})$ of $x_i(n)$ using (30);
3 Sort $s_{LCMV}(l\hat{\omega}_k, l\hat{\theta}_k)$ in the descending order as $\lambda_1 > \lambda_2 > \cdots > \lambda_{m_{max}}$;
4 **for** $m = 1; m \le m_{max}; m++$ **do**
5 Constitute the m-th model with the harmonic components corresponding to $\lambda_1, \lambda_2, \cdots, \lambda_m$;
6 Calculate $Q(m)$ of (52) for the m-th model;
7 **end**
8 Determine $m = \hat{M}_{MAP}$ which minimizes $Q(m)$;
9 Return the harmonic indexes corresponding to λ_m, $m = 1, 2, \cdots, \hat{M}_{MAP}$.

IV. SIMULATION AND EXPERIMENT RESULTS

A. SIMULATION SETTING

In this section, we firstly compare the spectrograms of the joint spatial-temporal FFT (JST-FFT) and the JST-LCMV beamformer, from which it is seen that the latter bears higher frequency resolution.

Then, we evaluate the estimation and detection performance of the proposed JST-LCMV beamformer and MAP criterion, respectively, by comparing with other methods. The mean square errors (MSE), which account for the combination of the bias and standard error, are adopted to evaluate the performance of the 2-D fundamental frequency estimation. They are defined as follows:

$$MSE_{\omega} = \frac{1}{SK}\sum_{k=1}^{K}\sum_{s=1}^{S}(\hat{\omega}_k^{(s)} - \omega_k)^2 \tag{55}$$

and

$$MSE_{\theta} = \frac{1}{SK}\sum_{k=1}^{K}\sum_{s=1}^{S}(\hat{\theta}_k^{(s)} - \theta_k)^2, \tag{56}$$

with ω_k, θ_k and $\hat{\omega}_k^{(s)}$, $\hat{\theta}_k^{(s)}$ being the true parameter values and their estimates at the s-th trial, respectively; and S being the number of trials.

The performance of the harmonic detection is evaluated in terms of the probability of correct detection (PCD): $PCD = S_0/S$, with S_0 and S being the number of correct-detection trials and the total number of trials, respectively.

For the performance evaluation, the results provided are the averages of 1000 runs and they are produced with respect to signal-to-noise ratio (SNR), which is defined as: $SNR = \sigma_s^2/\sigma^2$ with the signal power $\sigma_s^2 = \sum_{k=1}^{K}\sum_{l=1}^{L_k}|\rho_{l,k}|^2/\sum_{k=1}^{K}L_k$.

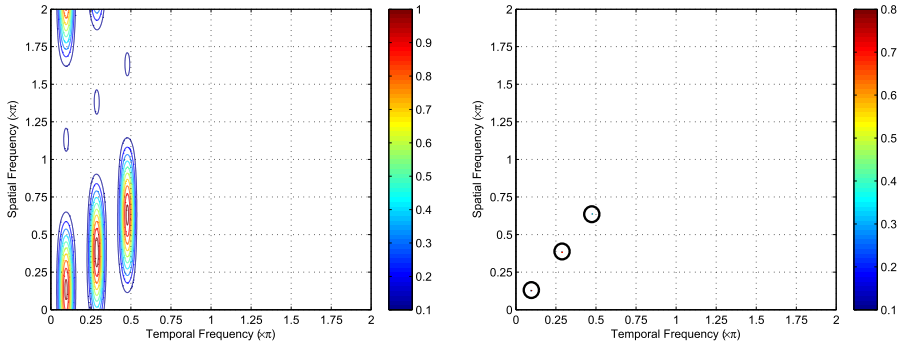
In the simulation, we consider the 2-D harmonic signals of the two cases: single-pitch and two-pitch, whose data lengths are set as $(N, I) = (50, 5)$ and $(N, I) = (100, 10)$, respectively. Their parameter settings are shown in Tables 1 and 2, respectively. It is seen from these two tables that there exist only the odd-order harmonic components in the signals for simulation.

It is shown in [40] that the best shift-invariance characteristic of the signal subspace is achieved when the data matrix is as square as possible. Thus, for the estimate of the covariance matrix, $\hat{\mathbf{R}}$ of (23), I_s and N_s are set as: $I_s = \lfloor 0.5\ I \rfloor$ and $N_s = \lfloor 0.5\ N \rfloor$ ($\lfloor u \rfloor$ denoting the largest integer smaller than u), respectively, which is found empirically to result in good performance.

$$\frac{\partial^2 \ln \mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}}{\partial\boldsymbol{\psi}_{K_m}(i_1)\partial\boldsymbol{\psi}_{K_m}(i_2)} = -(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})^{-2}\frac{\partial(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}(i_1)}\cdot\frac{\partial(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}(i_2)} + (\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})^{-1}\frac{\partial^2(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}(i_1)\partial\boldsymbol{\psi}_{K_m}(i_2)} \tag{44}$$

$$\frac{\partial^2(\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x})}{\partial\boldsymbol{\psi}_{K_m}(i_1)\partial\boldsymbol{\psi}_{K_m}(i_2)} \approx -\frac{2}{IN}\mathfrak{Re}\left\{\mathbf{x}^H\left(\frac{\partial^2\mathbf{D}_m}{\partial\boldsymbol{\psi}_{K_m}(i_1)\partial\boldsymbol{\psi}_{K_m}(i_2)}\cdot\mathbf{D}_m^H + \frac{\partial\mathbf{D}_m}{\partial\boldsymbol{\psi}_{K_m}(i_1)}\cdot\frac{\partial\mathbf{D}_m^H}{\partial\boldsymbol{\psi}_{K_m}(i_2)}\right)\mathbf{x}\right\} \tag{51}$$

$$\mathcal{H}_{0,m}(i_1, i_2) = \mathfrak{Re}\left\{\mathbf{x}^H\left(\frac{\partial^2\mathbf{D}_m}{\partial\boldsymbol{\psi}_{K_m}(i_1)\partial\boldsymbol{\psi}_{K_m}(i_2)}\cdot\mathbf{D}_m^H + \frac{\partial\mathbf{D}_m}{\partial\boldsymbol{\psi}_{K_m}(i_1)}\cdot\frac{\partial\mathbf{D}_m^H}{\partial\boldsymbol{\psi}_{K_m}(i_2)}\right)\mathbf{x}\right\} \tag{53}$$



(a) (b)

FIGURE 1. Spectrograms of the single-pitch harmonic signal from the: (a) JST-FFT and (b) JST-LCMV beamformer.

TABLE 1. Simulation Setting of Single-Pitch Harmonic Signal.

ω_1	θ_1	$p_{l,1}$	$\rho_{l,1}$
		1	e^{j}
0.3	0.4	3	e^{2j}
		5	e^{3j}

TABLE 2. Simulation Setting of Two-Pitch Harmonic Signal.

k	ω_k	θ_k	$p_{l,k}$	$\rho_{l,k}$
			1	e^{j}
1	0.3	0.4	3	e^{2j}
			5	e^{3j}
			1	e^{4j}
2	0.5	0.6	3	e^{5j}
			5	e^{6j}

B. COMPARISON IN FREQUENCY DOMAIN

Before investigating the estimation and detection performance of the JST-LCMV beamformer, we first observe the spectrograms from the JST-FFT and the JST-LCMV beamformer, which can help us have a better understanding of the advantage of the JST-LCMV beamformer over the standard tool for spectral analysis. As shown in Figs. 1 and 2, the spectrograms are represented in the form of contour plots, with the height denoting the spectral power. For the spectrogram from the JST-FFT, the filter's impulse response vector is the Fourier transform vector, and the spectral power at the 2-D frequency $(\tilde{\omega}, \tilde{\theta})$ is estimated as:

$$\begin{aligned} s_{FFT}(\tilde{\omega},\tilde{\theta}) &= E\left\{y_{i_s}(n_s)y_{i_s}^H(n_s)\right\} \\ &= \mathbf{h}_{\tilde{\omega},\tilde{\theta}}^H E\left\{\mathbf{x}_{i_s}(n_s)\mathbf{x}_{i_s}^H(n_s)\right\}\mathbf{h}_{\tilde{\omega},\tilde{\theta}} \\ &\approx \mathbf{h}_{\tilde{\omega},\tilde{\theta}}^H \hat{\mathbf{R}}\mathbf{h}_{\tilde{\omega},\tilde{\theta}} \\ &= \frac{1}{(I_sN_s)^2}\mathbf{a}^H(\tilde{\omega},\tilde{\theta})\hat{\mathbf{R}}\mathbf{a}(\tilde{\omega},\tilde{\theta}), \end{aligned} \tag{57}$$

with $\mathbf{a}(\tilde{\omega},\tilde{\theta})$ and $\hat{\mathbf{R}}$ defined in (15) and (23), respectively.

Fig. 1 shows the spectrograms of the single-pitch harmonic signal from the JST-FFT and the JST-LCMV beamformer at $SNR = 40\ dB$. From this figure, it is seen that for the single-pitch signal, both of the JST-FFT and JST-LCMV beamformer can differentiate all the three harmonic components, while the former has wider main lobe. Fig. 2 shows the spectrograms of the two-pitch harmonic signal. For the two-pitch signal, the JST-FFT method is unable to differentiate the two harmonic components with the same temporal frequency of 1.5, while the latter can. From Figs. 1 and 2, it is concluded that the JST-LCMV beamformer bears better frequency resolution, and it is reasonable to utilize the JST-LCMV beamformer to perform the parameter estimation and spectral analysis instead of the standard FFT-based method.

C. JOINT ESTIMATION OF 2-D FUNDAMENTAL FREQUENCIES

Now we evaluate the estimation performance of the JST-LCMV beamformer based on the maximum harmonic model and with $L_{max,k}$ set as $3, 5, 7$ for each pitch. For comparison, the results of the JST-FFT, harmonic MUSIC (HMUSIC) [14] and nonlinear least squares (NLS) [15] methods are also included. For fairness, these methods are also applied to the maximum harmonic model. Here, the Cramér-Rao lower bound (CRLB) [41] is provided as a benchmark, which demonstrates the best results theoretically achievable.

Fig. 3 shows the MSE results of the 2-D fundamental frequency estimation for the single-pitch signal, where the different columns of the subfigures correspond to the different values of $L_{max,k}$, and the first and second rows of the subfigures show the estimation results of the temporal and spatial fundamental frequencies, respectively.

Let us focus on the case that $L_{max,k} = 7$ at first. It is seen that the MSEs of the JST-LCMV beamformer decrease approximately linearly with respect to the SNR, which means that the 2-D fundamental frequency estimation with the JST-LCMV beamformer is consistent with respect to the SNR. Furthermore, the MSE curves of the JST-LCMV

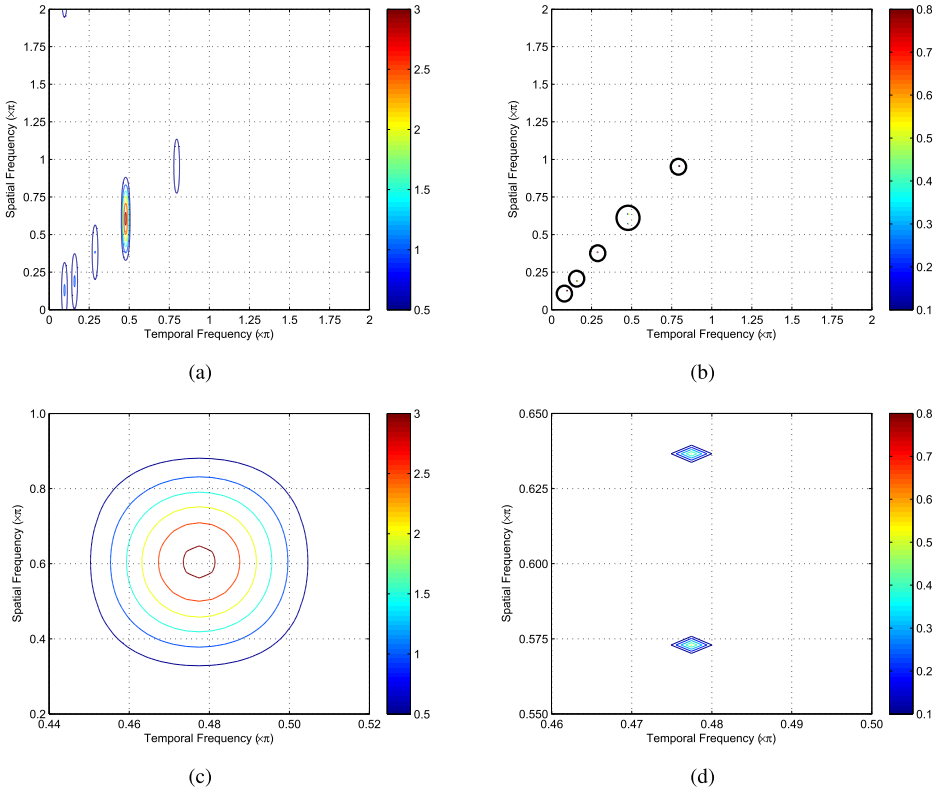


FIGURE 2. Spectrograms of the two-pitch harmonic signal from the: (a) JST-FFT, (b) JST-LCMV beamformer, (c) JST-FFT (zoomed), and (d) JST-LCMV beamformer (zoomed).

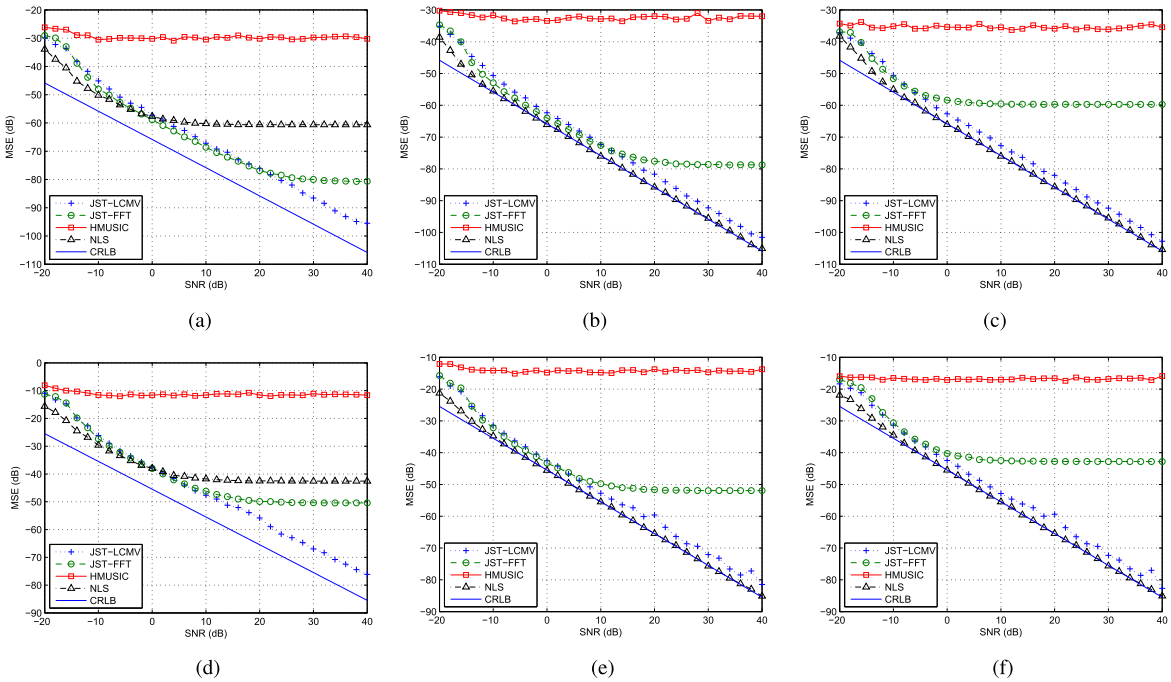


FIGURE 3. MSEs of the single-pitch parameter estimation versus SNR for: (a)–(c) temporal fundamental frequency and (d)–(f) spatial fundamental frequency.

beamformer keep parallel with the CRLBs when the SNR is sufficiently large. In detail, when $SNR \geq 0\ dB$, the gaps between the MSEs of the LCMV beamformer and their corresponding CRLBs are about 3 dB for the temporal and spatial fundamental frequencies, respectively. Meanwhile, the MSEs of the JST-FFT method decrease more and more

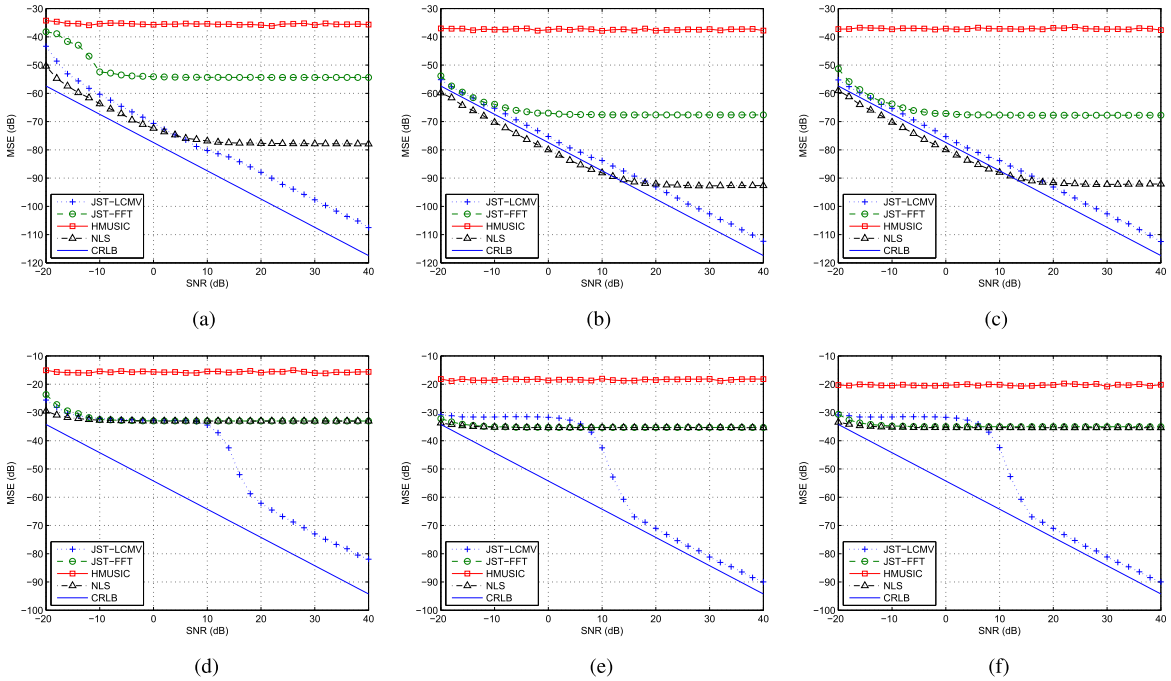


FIGURE 4. MSEs of the two-pitch parameter estimation versus SNR for: (a)–(c) temporal fundamental frequency and (d)–(f) spatial fundamental frequency.

slowly with the increase of the SNR. When $SNR \geq 10\ dB$, they keep nearly constant. This is because of the limited frequency resolution of the JST-FFT method aforementioned, which causes the estimation bias. The MSEs of the NLS method agree well with the CRLBs when $SNR \geq -8\ dB$, which is consistent with the fact that the NLS method is statistically efficient under the assumption of white Gaussian noise [41]. The MSEs of the HMUSIC keep nearly constant in the whole SNR range. This is because the HMUSIC is a parametric frequency estimation method, and does not work when the information of the signal model is not complete.

When $L_{max,k} = 5$, the layout of the MSE curves of the JST-LCMV beamformer, JST-FFT, NLS and HMUSIC methods are similar to those when $L_{max,k} = 7$. This is because the maximum harmonic model covers the true signal model for both of these two cases.

When $L_{max,k} = 3$, it is seen that the gaps between the MSEs of the LCMV beamformer and their corresponding CRLBs become obviously larger. In detail, when $SNR \geq 0\ dB$, the gaps are about 8 dB for the temporal and spatial fundamental frequencies, respectively. This demonstrates that the LCMV beamformer still works when $L_{max,k}$ is smaller than the largest harmonic index of the true signal model, but with a less satisfying performance. Note that when $L_{max,k} = 3$, the MSEs of the NLS method keep nearly constant when $SNR \geq 10\ dB$. This is because the maximum harmonic model with $L_{max,k} = 3$ excludes the true signal model, and the NLS method cannot minimize the signal-fitting error with the 2-D fundamental frequency estimates effectively.

Fig. 4 shows the MSE results of the 2-D fundamental frequency estimation for the two-pitch harmonic signal with the same subfigure layout as Fig. 3. We see that when $L_{max,k} = 5$ and 7, for the JST-LCMV beamformer, the MSEs of the temporal and spatial fundamental frequencies keep the gaps of $3-5\ dB$ from the CRLBs when $SNR \geq 16\ dB$. Meanwhile, the MSEs of the JST-FFT method keep nearly constant when $SNR \geq 0\ dB$. When $L_{max,k} = 3$, the gaps between the MSEs of the JST-LCMV beamformer and the CRLBs are enlarged to $10-12\ dB$.

We extend the NLS method to the two-pitch harmonic signal following [7]. It is seen that for the NLS method, the MSEs of the temporal and spatial fundamental frequencies keep nearly constant when $SNR \geq 20\ dB$, which is related to the decoupling difficulty in the extension of the NLS method to the multi-pitch harmonic signal [9]. Note that for $L_{max,k} = 5$ and 7, the MSEs of the temporal fundamental frequency are smaller than the CRLB when $SNR \leq 10\ dB$, which demonstrates that the NLS method provides the biased estimates when the SNR is small enough. In addition, the MSEs of the spatial fundamental frequency are always larger than the CRLB in the whole SNR range. The HMUSIC algorithm still does not work due to the incomplete model information.

By comparing the fundamental frequency estimation performance of the above methods for the single-pitch and two-pitch harmonic signals, we adopt the estimation results of

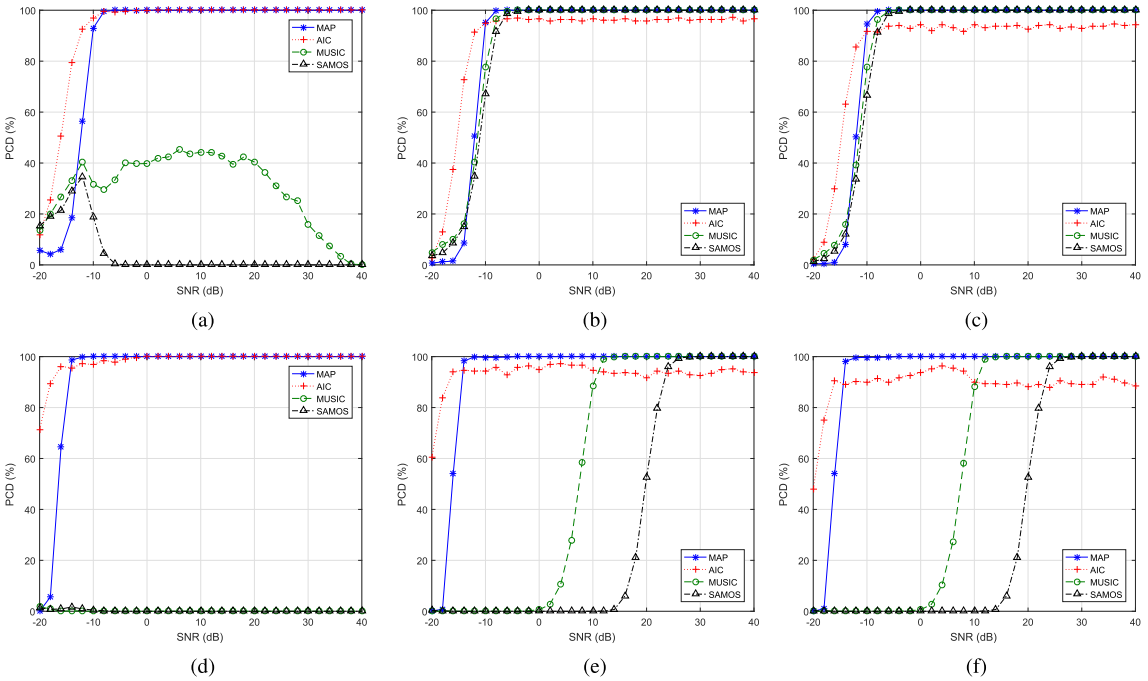


FIGURE 5. PCDs of the harmonic detection versus SNR for the: (a)–(c) single-pitch signal and (d)–(f) two-pitch signal.

the JST-LCMV beamformer for the harmonic detection in the next subsection.

D. DETECTION OF THE HARMONICS

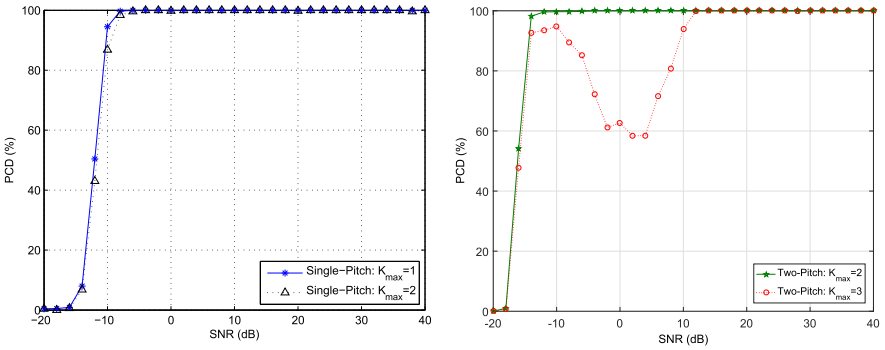
Having obtained the estimates of the 2-D fundamental frequencies, it is possible to detect the existing harmonic components of the 2-D harmonic signal based on the MAP criterion and following the procedure in Section III.B. Fig. 5 shows the harmonic detection performance of the MAP criterion, where the different columns of the subfigures correspond to the different values of $L_{max,k}$; and the first and second rows of the subfigures show the results for the single-pitch and two-pitch signals, respectively. For comparison, we also employ the MUSIC [24], SAMOS [26] and AIC [27] methods to estimate the orders of the 2-D harmonic signals under consideration, and provide their corresponding results of harmonic detection. Here, we extend the SAMOS order estimator to the 2-D harmonic signal by utilizing the shift invariance in both of the temporal and spatial dimensions. To make the performance evaluation of Figs. 5(a) and (d) feasible, we regard only the harmonic components with the indexes not larger than 3 as existing in the true signal model when $L_{max,k}$ is set as 3.

When $L_{max,k} = 7$, the proposed MAP criterion together with the MUSIC and SAMOS methods are consistent asymptotically with respect to the SNR. For the single-pitch harmonic signal, the detection with the MAP criterion achieves a 100% success rate above the threshold $SNR = -8\ dB$. By comparison, the MUSIC and SAMOS methods perform perfectly when $SNR \geq -4\ dB$ and $SNR \geq -2\ dB$, respectively. For the two-pitch signal, the PCD of the MAP criterion achieves the value of 100% when $SNR \geq -4\ dB$, while the PCDs of the MUSIC and SAMOS methods become 100% when $SNR \geq 16\ dB$ and $SNR \geq 30\ dB$, respectively. This means that the MAP criterion bears the threshold SNR advantage of 2 dB and 20 dB over the MUSIC and SAMOS methods for the signal-pitch and two-pitch signals, respectively. It is also noted that in the whole SNR range and for both the single-pitch and two-pitch harmonic signals, the PCD of the AIC criterion is always lower than 100%, which means that the AIC criterion is unable to provide the consistent harmonic detection asymptotically with respect to the SNR.

By comparing Figs. 5(b), (e) and Figs. 5(c), (f), it is seen that the harmonic detection performance with $L_{max,k} = 5$ is similar to that with $L_{max,k} = 7$. This is reasonable, for the maximum harmonic models with $L_{max,k} = 5$ and 7 both cover the true signal model.

It is seen from Figs. 5(a) and (d) that, when $L_{max,k} = 3$, the methods based on the rank determination of the data matrix, that is the MUSIC and SAMOS methods, do not work in the whole SNR range, while the MAP criterion provides the perfect harmonic detection results when $SNR \geq -6\ dB$ and $SNR \geq -10\ dB$ for the single-pitch and two-pitch harmonic signals, respectively. Interestingly, the AIC criterion performs perfectly when $SNR \geq 2\ dB$ and $SNR \geq 0\ dB$ for the single-pitch and two-pitch harmonic signals, respectively, although it is inconsistent with respect to the SNR when $L_{max,k} = 5$ and 7. This suggests that the AIC criterion tends to underestimate the orders of the 2-D harmonic signals.

Remark 4: Based on the simulation results of Section IV. C-D, it is suggested to set $L_{max,k}$ of the maximum harmonic



(a) (b)

FIGURE 6. PCDs of the harmonic detection with the MAP criterion versus SNR for the: (a) single-pitch signal and (b) two-pitch signal.

model as a large value so that the parametric modeling methodology for the 2-D harmonic signals can provide satisfying performance.

E. ABOUT SOURCE ENUMERATION

As mentioned in Section II, the source number K is assumed known *a priori* here. When K is unknown, the problem of the determination of K, that is source enumeration, occurs. In this subsection, we will explore the source enumeration ability of our proposed signal modeling methodology. Here, it is assumed that $K \leq K_{max}$, and we estimate the K_{max} pairs of 2-D fundamental frequencies following Algorithm 1. Based on the estimated parameters and the maximum harmonic model:

$$x_i(n) = s_i(n) + q_i(n),$$
$$s_i(n) = \sum_{k=1}^{K_{max}} \sum_{l=1}^{L_{max,k}} \rho'_{l,k} e^{j(\omega_k ln + \theta_k li)},$$

we constitute the candidate models, and conduct the harmonic detection with the developed MAP criterion.

Fig. 6 shows the harmonic detection performance of the MAP criterion with $K_{max} = K$ (the same as K being known) and $K_{max} = K + 1$. The $L_{max,k}$ is set as 7 as suggested in the last subsection. It is seen from Fig. 6 that the proposed signal modeling methodology can give the perfect performance for $K_{max} = K$ and $K + 1$ when the SNR is sufficiently high. This is reasonable, since according to the MAP criterion, only the candidate model with the largest posterior probability is selected. When the SNR is sufficiently high, the true signal model usually satisfies this condition.

It is noted that for the single-pitch signal, when K_{max} is set as K and $K + 1$, the harmonic detection performances are similar, while for the two-pitch signal, there is a threshold SNR loss of 18 dB when K_{max} is set as $K+1$. Combining with the results of Figs. 3 and 4, it is inferred that the performance of the harmonic detection is also subjected to the quality of the parameter estimation. If the accurate estimates are unavailable, the signal modeling is less robust to a more complex set of candidate models.

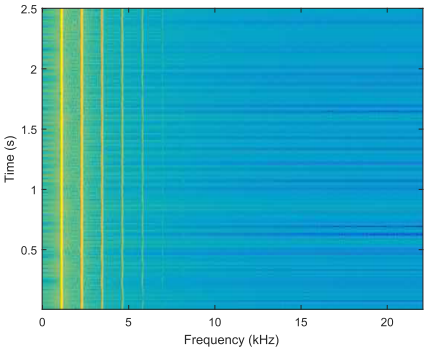


FIGURE 7. Spectrogram of the anechoic trumpet signal for the experiment.

F. EXPERIMENT ON REAL-LIFE DATA

In this subsection, we evaluate the performance of the proposed JST-LCMV beamformer and MAP criterion through an experiment on the real-life data, that is the anechoic trumpet signal.[1] Firstly, we analyze its temporal fundamental frequency and localize the sound source in a joint way. Afterwards, the harmonic detection is conducted. The DOA of the sound source is set as 50°. Due to the limited laboratory conditions, here we utilize this anechoic trumpet signal and its temporally delayed version to mimic the measured signals by two microphones spaced with $d = 5\ cm$. Thus, β_1 and β_2 of (3) are equal to 1. The sampling frequency is $f_s = 44.1\ kHz$, and the propagation speed of the sound wave is measured as $c = 341.7\ m/s$. Accordingly, the time delay between the two sets of experimental signals is $\tau \approx 5/f_s$.

Fig. 7 shows the spectrogram of the anechoic trumpet signal using the short-time Fourier transform,[2] where the yellower strips correspond to the stronger harmonic components. From this spectrogram, it is seen that there exists one pitch with the temporal fundamental frequency around 1170 Hz and

[1]The anechoic trumpet signal was download from http://theremin.music.uiowa.edu/MIS.html

[2]The spectrogram is produced with the function "spectrogram" of MATLAB.

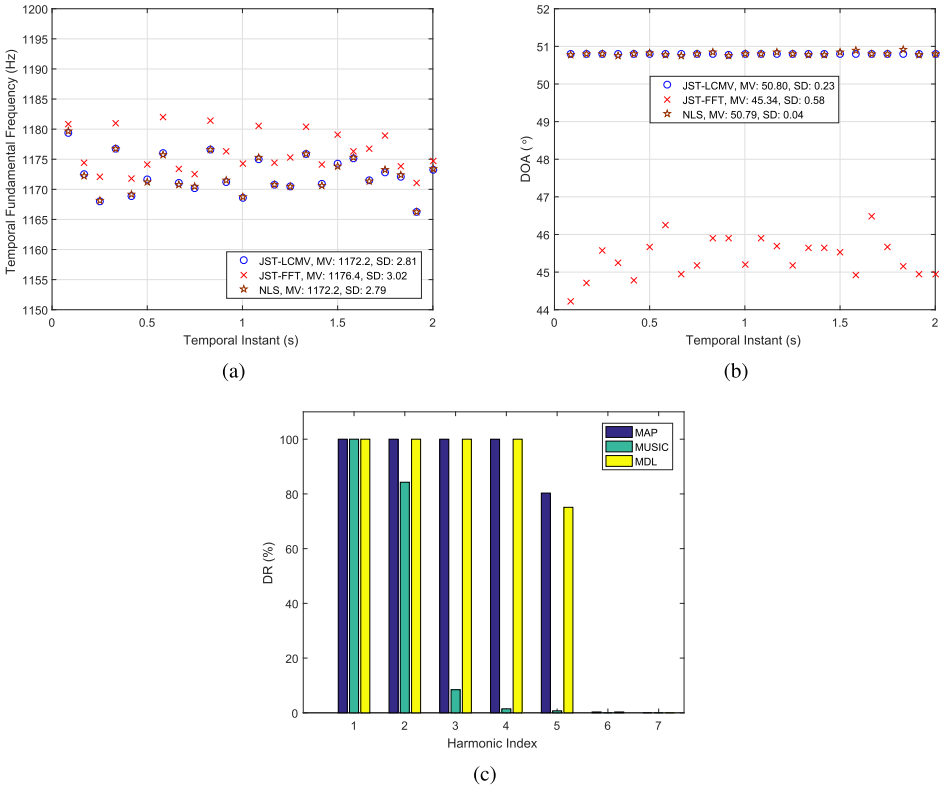


FIGURE 8. Experimental results of the signal modeling: (a) temporal fundamental frequency estimation, (b) DOA estimation, and (c) harmonic detection.

the 7 consecutive harmonic components; and the higher-order components have the smaller spectral powers.

To evaluate the statistical properties of the proposed signal modeling method, we take 1000 equally spaced segments of the sampled signals during $0 - 2.5\ s$, each segment with the data length of $N = 100$, and conduct the signal modeling for each segment. For the parameter estimation, since the exact true value is unavailable, we calculate the mean value (MV) and standard deviation (SD) of the estimates instead; for the harmonic detection, we calculate the detection rate (DR) for the harmonic components indexed by $l = 1, 2, \cdots, 7$. Here, the DR of the l-th component is defined as: $DR_l = S_l/S$, where S_l is the number of the trials where the l-th component is detected, and S is the total number of trials, that is 1000. In view of the small number of channels, N_s and I_s are set as 50 and 2 for the JST-LCMV beamformer, respectively. For the harmonic detection, the order of the maximum harmonic model is set as $L_{max,1} = 7$.

Figs. 8(a) and (b) show the MVs and SDs of the JST-LCMV beamformer, JST-FFT, and NLS methods, together with their respective estimation results at some temporal instants. It is seen that the JST-LCMV beamformer bears the comparable performance with the theoretically optimal NLS method. The MV of the DOA estimates of the JST-FFT method deviates from the true location by larger than 4°, while it performs less stably than the JST-LCMV beamformer. Fig. 8(c) shows the DRs of the MAP criterion together with the MDL [27] and MUSIC methods. Since the AIC criterion is not consistent and the SAMOS method is not so good as the MUSIC according to Section IV.D, they are not compared here. From Fig. 8(c), it is seen that the MAP and MDL criteria detect the harmonic components $l = 1, 2, 3$ and 4 with the DR of 100%. Although the MUSIC method detects the first harmonic component with the DR of 100%, its performance degrades quickly for the higher-order components. For the weaker 5-th component, the MAP criterion bears the DR higher than that of the MDL by 5%. Additionally, all the three methods can hardly perceive the 6-th and 7-th components, which are so weak that they are almost buried in the background noise.

V. CONCLUSION

In this paper, a complete framework is proposed to conduct the parametric modeling for the 2-D harmonic signals with missing harmonics. This includes the estimation of the 2-D fundamental frequencies and the detection of the harmonic components. To achieve this objective, we develop the JST-LCMV beamformer for the joint 2-D parameter estimation based on the maximum harmonic model; and derive the maximum *a posteriori* criterion to estimate the signal order. The corresponding number of harmonic components from the maximum harmonic model with the largest spectral powers are selected as existing in the 2-D harmonic signals. Simulation and experimental results demonstrate the performance advantages of the proposed signal modeling methodology

by comparing with other parameter estimation and harmonic detection methods.

This paper only focuses on the 2-D harmonic signal. In practice, it is also necessary to consider other kinds of high-dimensional sinusoidal signals. In the future, we will go on with the topic of parametric modeling for the high-dimensional sinusoidal signals, and explore more about the application of our methodology of signal modeling to the practical scenarios including localization, biomedical signal analysis, music and speech signal processing, and so on.

APPENDIX A
DERIVATION OF (39)

Let

$$\hat{\boldsymbol{\rho}}_m \triangleq (\mathbf{D}_m^H \mathbf{D}_m)^{-1} \mathbf{D}_m^H \mathbf{x}, \tag{A.1}$$

and we have

$$\begin{aligned}(\mathbf{x} - \mathbf{D}_m \boldsymbol{\rho}_m)^H (\mathbf{x} - \mathbf{D}_m \boldsymbol{\rho}_m) &= \mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x} \\ &+ (\boldsymbol{\rho}_m - \hat{\boldsymbol{\rho}}_m)^H \mathbf{D}_m^H \mathbf{D}_m (\boldsymbol{\rho}_m - \hat{\boldsymbol{\rho}}_m).\end{aligned} \tag{A.2}$$

Applying the priors and evaluating the marginal distribution, we have that (see (A.3), as shown at the top of the next page). Following the similar principle in [29], $p(\mathbf{x}, \boldsymbol{\psi}_{K_m}, \sigma|m)$ is approximated by

$$\begin{aligned}p(\mathbf{x}, \boldsymbol{\psi}_{K_m}, \sigma|m) &\approx p(\sigma|\boldsymbol{\psi}_{K_m}, m) p(\boldsymbol{\psi}_{K_m}|m) \gamma(m) \\ &\cdot \frac{1}{(\pi\sigma^2)^{IN}} e^{-\frac{1}{\sigma^2}\mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x}} \frac{(\pi\sigma^2)^m}{|\mathbf{D}_m^H \mathbf{D}_m|}.\end{aligned} \tag{A.4}$$

Substituting (34) and (37) into (A.4), it is derived that:

$$\begin{aligned}&p(\mathbf{x}, \boldsymbol{\psi}_{K_m}, \sigma|m) \\ &= c 2^{-2K_m} \pi^{-2K_m + m - IN} \gamma(m) \\ &\cdot |\mathbf{D}_m^H \mathbf{D}_m|^{-1} \sigma^{-2IN+2m-1} e^{-\frac{1}{\sigma^2}\mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x}}.\end{aligned} \tag{A.5}$$

Then, $p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m)$ is evaluated as [42]:

$$\begin{aligned}&p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m) \\ &= \int_{\mathbb{R}^+} p(\mathbf{x}, \boldsymbol{\psi}_{K_m}, \sigma|m) d\sigma \\ &\approx c 2^{-(2K_m+1)} \pi^{-(2K_m - m + IN)} \gamma(m) |\mathbf{D}_m^H \mathbf{D}_m|^{-1} \\ &\cdot (\mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x})^{-(IN-m)} \cdot \Gamma(IN - m),\end{aligned} \tag{A.6}$$

where $\Gamma(\cdot)$ denotes the standard Gamma function. Furthermore, when $IN \gg m$, $\Gamma(IN - m) \approx (IN)^{-m}\Gamma(IN)$, and (A.6) is rewritten as:

$$\begin{aligned}&p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m) \\ &\approx c 2^{-(2K_m+1)} \pi^{-(2K_m - m + IN)} \Gamma(IN) \gamma(m) \\ &\cdot (IN)^{-m} |\mathbf{D}_m^H \mathbf{D}_m|^{-1} (\mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x})^{-(IN-m)}.\end{aligned} \tag{A.7}$$

Finally, to obtain the expression of the conditional PDF $p(\mathbf{x}|m) = \int_{\mathbb{D}_{K_m}} p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m) d\boldsymbol{\psi}_{K_m}$, we make use of the Laplace integration method [43], which considers the integral of the form

$$\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{t}) e^{\alpha h(\mathbf{t})} d\mathbf{t}, \tag{A.8}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{t}$ are vectors, α is a large positive parameter, $g(\mathbf{t})$ and $h(\mathbf{t})$ are real functions of $\mathbf{t}$. To align with the form of the Laplace integration method, we rewrite (A.7) as:

$$\begin{aligned}p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m) &= C(m)(IN)^{-m} \\ &\cdot |\mathbf{D}_m^H \mathbf{D}_m|^{-1} e^{-(IN-m)\ln \mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x}}.\end{aligned} \tag{A.9}$$

Assume that the number of the harmonic components of the signal, that is the signal order, is equal to m. Define the matrices $\hat{\mathbf{D}}_m$ and $\hat{\mathbf{D}}_m^{\perp}$ in the same way as $\mathbf{D}_m$ and $\mathbf{D}_m^{\perp}$, respectively, with $\boldsymbol{\psi}_{K_m}$ replaced by its ML estimate $\hat{\boldsymbol{\psi}}_{K_m}$. Then, according to the Laplace integration method, the conditional PDF $p(\mathbf{x}|m)$ is derived as:

$$\begin{aligned}p(\mathbf{x}|m) &= \int_{\mathbb{D}_{K_m}} p(\mathbf{x}, \boldsymbol{\psi}_{K_m}|m) d\boldsymbol{\psi}_{K_m} \\ &= \int_{\mathbb{D}_{K_m}} C(m)(IN)^{-m} |\mathbf{D}_m^H \mathbf{D}_m|^{-1} \\ &\times e^{-(IN-m)\ln \mathbf{x}^H \mathbf{D}_m^{\perp} \mathbf{x}} d\boldsymbol{\psi}_{K_m} \\ &\approx C(m)(IN)^{-m} |\hat{\mathbf{D}}_m^H \hat{\mathbf{D}}_m|^{-1} e^{-(IN-m)\ln \mathbf{x}^H \hat{\mathbf{D}}_m^{\perp} \mathbf{x}} \\ &\cdot (2\pi)^{K_m} |\hat{\mathbf{H}}_m|^{-1/2} (IN - m)^{-K_m}.\end{aligned} \tag{A.10}$$

When $N \to \infty$, there holds that for $\mu \neq 2\pi k$ ($k \in \mathbb{Z}$),

$$\sum_{n=0}^{N-1} e^{jn\mu} = \mathcal{O}(1), \tag{A.11}$$

$$\sum_{n=0}^{N-1} n e^{jn\mu} = \mathcal{O}(N), \tag{A.12}$$

$$\sum_{n=0}^{N-1} n^2 e^{jn\mu} = \mathcal{O}(N^2). \tag{A.13}$$

As a result, when the data length N is large, $|\hat{\mathbf{D}}_m^H \hat{\mathbf{D}}_m|^{-1}$ is simplified as

$$|\hat{\mathbf{D}}_m^H \hat{\mathbf{D}}_m|^{-1} \approx (IN)^{-m}, \tag{A.14}$$

and

$$\begin{aligned}p(\mathbf{x}|m) &\approx C(m)(IN)^{-2m} (\mathbf{x}^H \hat{\mathbf{D}}_m^{\perp} \mathbf{x})^{-(IN-m)} \\ &\cdot (2\pi)^{K_m} (IN - m)^{-K_m} |\hat{\mathbf{H}}_m|^{-1/2} \\ &\approx C(m)(IN)^{-2m} (\mathbf{x}^H \hat{\mathbf{D}}_m^{\perp} \mathbf{x})^{-IN} \\ &\cdot (2\pi)^{K_m} (IN)^{-K_m} |\hat{\mathbf{H}}_m|^{-1/2}.\end{aligned} \tag{A.15}$$

APPENDIX B
DERIVATION OF THE CRLB FOR 2-D FUNDAMENTAL FREQUENCY ESTIMATION

Here, we derive the CRLB for the 2-D fundamental frequency estimation problem. Firstly, suppose that at the i-th ($i = 1, 2, \cdots, I$) channel we observe the signal $\mathbf{s}_i(\boldsymbol{\mu})$ corrupted by the noise $\mathbf{q}_i$ as follows:

$$\mathbf{x}_i = \mathbf{s}_i(\boldsymbol{\mu}) + \mathbf{q}_i, \tag{B.1}$$

$$
\begin{aligned}
&p(\mathbf{x}, \boldsymbol{\psi}_{K_m}, \sigma|m) \\
&= \int_{\mathbb{C}^m} p(\mathbf{x}|m, \boldsymbol{\psi}_{K_m}, \boldsymbol{\rho}_m, \sigma) p(\boldsymbol{\rho}_m, \boldsymbol{\psi}_{K_m}, \sigma|m) d\boldsymbol{\rho}_m \\
&= p(\sigma|\boldsymbol{\psi}_{K_m}, m) p(\boldsymbol{\psi}_{K_m}|m) \frac{1}{(\pi\sigma^2)^{IN}} \int_{\mathbb{C}^m} e^{-\frac{1}{\sigma^2}(\mathbf{x}-\mathbf{D}_m\boldsymbol{\rho}_m)^H(\mathbf{x}-\mathbf{D}_m\boldsymbol{\rho}_m)} p(\boldsymbol{\rho}_m|\boldsymbol{\psi}_{K_m}, m) d\boldsymbol{\rho}_m \\
&= p(\sigma|\boldsymbol{\psi}_{K_m}, m) p(\boldsymbol{\psi}_{K_m}|m) \frac{1}{(\pi\sigma^2)^{IN}} e^{-\frac{1}{\sigma^2}\mathbf{x}^H\mathbf{D}_m^{\perp}\mathbf{x}} \int_{\mathbb{C}^m} e^{-\frac{1}{\sigma^2}(\boldsymbol{\rho}_m-\hat{\boldsymbol{\rho}}_m)^H\mathbf{D}_m^H\mathbf{D}_m(\boldsymbol{\rho}_m-\hat{\boldsymbol{\rho}}_m)} p(\boldsymbol{\rho}_m|\boldsymbol{\psi}_{K_m}, m) d\boldsymbol{\rho}_m
\end{aligned} \tag{A.3}
$$

where $\mathbf{x}_i \in \mathbb{C}^{N\times 1}$, $\mathbf{s}_i(\boldsymbol{\mu}) \in \mathbb{C}^{N\times 1}$, $\mathbf{q}_i \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{\mu} \in \mathbb{R}^{(2K+2\sum_{k=1}^{K} L_k)\times 1}$ are defined as:

$$\mathbf{x}_i \triangleq \left[x_i(1)\ x_i(2)\ \cdots\ x_i(N)\right]^T, \tag{B.2}$$

$$\mathbf{s}_i(\boldsymbol{\mu}) \triangleq \left[s_i(1)\ s_i(2)\ \cdots\ s_i(N)\right]^T, \tag{B.3}$$

$$\mathbf{q}_i \triangleq \left[q_i(1)\ q_i(2)\ \cdots\ q_i(N)\right]^T, \tag{B.4}$$

and

$$
\begin{aligned}
\boldsymbol{\mu} \triangleq \big[&\omega_1,\ \omega_2,\ \cdots,\ \omega_K,\ \theta_1,\ \theta_2,\ \cdots,\ \theta_K,\ \kappa_{1,1},\ \kappa_{2,1}, \\
&\cdots,\ \kappa_{L_K,K},\ \phi_{1,1},\ \phi_{2,1},\ \cdots,\ \phi_{L_K,K}\big]^T,
\end{aligned} \tag{B.5}
$$

respectively, with $\kappa_{l,k}$ and $\phi_{l,k}$ being the module and phase of $\rho_{l,k}$ ($k = 1, 2, \cdots, K$, $l = 1, 2, \cdots, L_k$), respectively.

We derive the CRLB under the assumption that $\mathbf{q}_i$ is complex-valued white Gaussian noise with the variance σ^2. Under this assumption, the log-likelihood function of the observed signal $\mathbf{x}_i$ is expressed as:

$$
\begin{aligned}
&\ln p(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_I) \\
&= -(NI)\ln \pi\sigma^2 \\
&\quad - \frac{1}{\sigma^2}\sum_{i=1}^{I}(\mathbf{x}_i - \mathbf{s}_i(\boldsymbol{\mu}))^H \cdot (\mathbf{x}_i - \mathbf{s}_i(\boldsymbol{\mu})),
\end{aligned} \tag{B.6}
$$

and the Fisher information matrix (FIM) of $\mathbf{x}_i$ is [41]:

$$\mathbf{I}(\boldsymbol{\mu}) = -E\left\{\frac{\partial^2 \ln p(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_I)}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T}\right\}. \tag{B.7}$$

Since the covariance matrices of the observation signals do not depend on the parameter vector $\boldsymbol{\mu}$, we express the FIM related to the estimation problem at hand as:

$$\mathbf{I}(\boldsymbol{\mu}) = \frac{2}{\sigma^2}\mathfrak{Re}\left\{\sum_{i=1}^{I}\mathbf{U}_i^H(\boldsymbol{\mu})\cdot\mathbf{U}_i(\boldsymbol{\mu})\right\}, \tag{B.8}$$

where $\mathbf{U}_i(\boldsymbol{\mu}) \in \mathbb{C}^{N\times(2K+\sum_{k=1}^{K} L_k)}$ is the gradient matrix defined as:

$$\mathbf{U}_i(\boldsymbol{\mu}) \triangleq \left[\mathbf{U}_i(\boldsymbol{\omega})\ \mathbf{U}_i(\boldsymbol{\theta})\ \mathbf{U}_i(\boldsymbol{\kappa})\ \mathbf{U}_i(\boldsymbol{\phi})\right], \tag{B.9}$$

with

$$\mathbf{U}_i(\boldsymbol{\omega}) \triangleq \frac{\partial\mathbf{s}_i(\boldsymbol{\mu})}{\partial\boldsymbol{\omega}^T}, \tag{B.10}$$

$$\mathbf{U}_i(\boldsymbol{\theta}) \triangleq \frac{\partial\mathbf{s}_i(\boldsymbol{\mu})}{\partial\boldsymbol{\theta}^T}, \tag{B.11}$$

$$\mathbf{U}_i(\boldsymbol{\kappa}) \triangleq \frac{\partial\mathbf{s}_i(\boldsymbol{\mu})}{\partial\boldsymbol{\kappa}^T}, \tag{B.12}$$

$$\mathbf{U}_i(\boldsymbol{\phi}) \triangleq \frac{\partial\mathbf{s}_i(\boldsymbol{\mu})}{\partial\boldsymbol{\phi}^T}. \tag{B.13}$$

Here, the parameter vectors $\boldsymbol{\omega}$, $\boldsymbol{\theta}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\phi}$ are defined as:

$$\boldsymbol{\omega} \triangleq \left[\omega_1\ \omega_2\ \cdots\ \omega_K\right]^T, \tag{B.14}$$

$$\boldsymbol{\theta} \triangleq \left[\theta_1\ \theta_2\ \cdots\ \theta_K\right]^T, \tag{B.15}$$

$$\boldsymbol{\kappa} \triangleq \left[\kappa_{1,1}\ \kappa_{2,1}\ \cdots\ \kappa_{L_K,K}\right]^T, \tag{B.16}$$

$$\boldsymbol{\phi} \triangleq \left[\phi_{1,1}\ \phi_{2,1}\ \cdots\ \phi_{L_K,K}\right]^T. \tag{B.17}$$

Accordingly, the CRLB of $\boldsymbol{\mu}(i)$ is derived as:

$$CRLB\left(\boldsymbol{\mu}(i)\right) = \boldsymbol{\rho}(i), \tag{B.18}$$

$$\boldsymbol{\rho} = diag(\mathbf{I}^{-1}(\boldsymbol{\mu})), \tag{B.19}$$

where $diag(\mathbf{I}^{-1}(\boldsymbol{\mu}))$ stands for the diagonal elements of the inverse matrix of $\mathbf{I}(\boldsymbol{\mu})$, and $\boldsymbol{\mu}(i)$ and $\boldsymbol{\rho}(i)$ denote the i-th elements of $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$, respectively.

REFERENCES

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. New York, NY, USA: Springer, 1998.

[2] H. Dudley, "The carrier nature of speech," *Bell Syst. Tech. J.*, vol. 19, no. 4, pp. 495–515, Oct. 1940.

[3] V. K. Murthy *et al.*, "Analysis of power spectral densities of electrocardiograms," *Math. Biosci.*, vol. 12, nos. 1–2, pp. 41–51, Oct. 1971.

[4] 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. Amer.*, vol. 129, no. 6, pp. 3768–3776, Jun. 2011.

[5] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*. San Rafael, CA, USA: Morgan & Claypool, 2009.

[6] Z. Zhou and H. C. So, "Linear prediction approach to oversampling parameter estimation for multiple complex sinusoids," *Signal Process.*, vol. 92, no. 6, pp. 1458–1466, Jun. 2012.

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

[8] Z. Duan, B. Pardo, and C. Zhang, "Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 2121–2133, Nov. 2010.

[9] Z. Zhou, H. C. So, and F. K. W. Chan, "Optimally weighted music algorithm for frequency estimation of real harmonic sinusoids," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3537–3540.

[10] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Default Bayesian estimation of the fundamental frequency," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 3, pp. 598–610, Mar. 2013.

[11] M. G. Christensen, "Accurate estimation of low fundamental frequencies from real-valued measurements," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 10, pp. 2042–2056, Oct. 2013.

[12] G. Zhang and S. Godsill, "Fundamental frequency estimation in speech signals with variable rate particle filters," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 5, pp. 890–900, May 2016.

[13] J. Swärd, H. Li, and A. Jakobsson, "Off-grid fundamental frequency estimation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 2, pp. 296–303, Feb. 2018.
[14] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and multi-pitch estimation based on subspace techniques," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–11, 2012.
[15] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 5, pp. 923–933, May 2013.
[16] S. Karimian-Azari, J. R. Jensen, and M. G. Christensen, "Computationally efficient and noise robust DOA and pitch estimation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1613–1625, Sep. 2016.
[17] Y. Wu, A. Leshem, J. R. Jensen, and G. Liao, "Joint pitch and DOA estimation using the ESPRIT method," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 32–45, Jan. 2015.
[18] S. I. Adalbjörnsson, T. Kronvall, S. Burgess, K. Åström, and A. Jakobsson, "Sparse localization of harmonic audio sources," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 1, pp. 117–129, Jan. 2016.
[19] Z. Zhou, H. C. So, and M. G. Christensen, "Parametric modeling for damped sinusoids from multiple channels," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3895–3907, Aug. 2013.
[20] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA : Prentice-Hall, 2005.
[21] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *J. Amer. Stat. Assoc.*, vol. 103, no. 481, pp. 410–423, Jan. 2008.
[22] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Bayesian model comparison and the BIC for regression models," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 6362–6366.
[23] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, and S. H. Jensen, "Bayesian model comparison with the g-prior," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 225–238, Jan. 2014.
[24] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 62, pp. 1–11, 2009.
[25] 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.
[26] J. M. Papy, L. D. Lathauwer, and S. V. 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.
[27] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[28] P. M. Djurić, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.
[29] M. Kliger and J. M. Francos, "MAP model order selection rule for 2-D sinusoids in white noise," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2563–2575, Jul. 2005.
[30] P. Hyberg, M. Jansson, and B. Ottersten, "Array interpolation and DOA MSE reduction," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4464–4471, Dec. 2005.
[31] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, Nov. 2018.
[32] C. Zhou, Y. Gu, Z. Shi, and Y. D. Zhang, "Off-grid direction-of-arrival estimation using coprime array interpolation," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1710–1714, Nov. 2018.
[33] H. L. Van Trees, *Optimum Array Processing*. New York, NY, USA: Wiley, 2002.
[34] A. Jakobsson, S. L. Marple, and P. Stoica, "Computationally efficient two-dimensional Capon spectrum analysis," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2651–2661, Sep. 2000.
[35] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Joint DOA and fundamental frequency estimation methods based on 2-D filtering," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 2091–2095.
[36] A. Antoniou and W. S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York, NY, USA: Springer, 2007.
[37] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[38] A. Jennings and J. J. McKeown, *Matrix Computation*. Chichester, West Sussex: Wiley, 1992.
[39] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. New York, NY, USA: Wiley, 1992.
[40] S. Vanhuffel, H. Chen, C. Decanniere, and P. Vanhecke, "Algorithm for time-domain NMR data fitting based on total least squares," *J. Magn. Reson., Ser. A*, vol. 110, no. 2, pp. 228–237, Oct. 1994.
[41] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: PTR Prentice-Hall, 1993.
[42] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York, NY, USA: Academic, 1980.
[43] D. V. Widder, *The Laplace Transform*. Princeton, NJ, USA: Princeton Univ. Press, 1941.

ZHENHUA ZHOU was born in Shanghai, China, in 1985. He received the bachelor's degree from the University of Shanghai for Science and Technology, in 2007, the master's degree from Shanghai Jiao Tong University, in 2010, and the Ph.D. degree from the City University of Hong Kong, in 2013, all in electronic engineering.

He was a Research Associate with The Hong Kong University of Science and Technology, from 2016 to 2017. He is currently a Research Fellow with the College of Information Engineering, Shenzhen University. His research interests include spectral analysis, optimization, statistical signal processing, machine learning and their application to speech and audio signal processing, radar signal processing, and wireless communication.

MADS G. CHRISTENSEN (S'00–M'05–SM'11) received the M.Sc. and Ph.D. degrees from Aalborg University (AAU), Aalborg, Denmark, in 2002 and 2005, respectively.

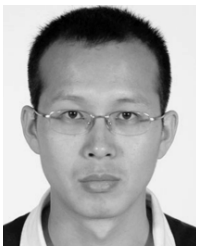
He was with the Department of Electronic Systems, AAU, and held visiting positions with Philips Research Laboratories, ENST, UCSB, and Columbia University. He is currently with the Department of Architecture, Design and Media Technology, AAU, as a Professor of audio processing, and is also the Head and the Founder of the Audio Analysis Laboratory. He is a beneficiary of major grants from the Danish Independent Research Council, the Villum Foundation, and the Innovation Fund Denmark. He has published three books and more than 180 papers in peer-reviewed conference proceedings and journals, and he has given tutorials with the EUSIPCO, SMC, and INTERSPEECH and a keynote talk at IWAENC. His research interests include audio and acoustic signal processing, where he has worked on topics, such as microphone arrays, noise reduction, signal modeling, speech analysis, audio classification, and audio coding.

Dr. Christensen is a member of the IEEE Audio and Acoustic Signal Processing Technical Committee, and a Founding Member of the EURASIP Special Area Team in Acoustic, Sound and Music Signal Processing. He is also a member of the EURASIP and the Danish Academy of Technical Sciences. He received several awards, including the Spar Nord Foundation's Research Prize, the Danish Independent Research Council Young Researcher's Award, the Statoil Prize, and the EURASIP Early Career Award. He is an Associate Editor of the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, and a former Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

JESPER R. JENSEN (S'09–M'12) was born in Ringkøbing, Denmark, in 1984. He received the M.Sc. degree for completing the elite candidate education and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 2009 and 2012, respectively.

He held two Postdoctoral positions with the Department of Architecture, Design and Media Technology (AD:MT), Aalborg University, where he is currently an Assistant Professor and also a member of the Audio Analysis Laboratory. He was a Visiting Researcher with the University of Quebec, INRS-EMT, Montreal, QC, Canada, with the Friedrich–Alexander Universitat Erlangen–Nurnberg, Erlangen, Germany, and with the University of Surrey, U.K. He has received a highly competitive postdoctoral grant from the Danish Independent Research Council, as well as several travel grants from private foundations. He has published more than 60 papers on these topics in top-tier, peer-reviewed conference proceedings and journals. Moreover, he has coauthored two books, namely *Speech Enhancement: A Signal Subspace Perspective* (Academic Press, 2014) and *Signal Enhancement With Variable Span Linear Filters* (Springer, 2016). His research interests include signal processing theory and methods, e.g., microphone array and joint audio–visual signal processing. Examples of more specific research interests within this scope are enhancement, separation, localization, tracking, parametric analysis, and modeling.

Dr. Jensen is an Affiliate Member of the IEEE Signal Processing Theory and Methods Technical Committee.



SHENGLI ZHANG (M'08) received the B.Eng. degree in electronic engineering and the M.Eng. degree in communication and information engineering from the University of Science and Technology of China, Hefei, China, in 2002 and 2005, respectively, and the Ph.D. degree in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2008. From 2014 to 2015, he was a Visiting Professor with Stanford University. Since 2008, he has been a Faculty Member with Shenzhen University, Shenzhen, China, where he is currently a Professor. His current research interests include wireless networks and wireless communications, especially physical layer network coding and cooperative wireless networks.