An Iterative Subspace-based Multi-Pitch Estimation Algorithm

Zhang, Johan Xi; Christensen, Mads Græsbøll; Jensen, Søren Holdt; Moonen, Marc

Published in:
Signal Processing

DOI (link to publication from Publisher):
10.1016/j.sigpro.2010.06.010

Publication date:
2011

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

Signal Processing 91 (2011) 150–154

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

An iterative subspace-based multi-pitch estimation algorithm

Johan Xi Zhang [a,*,1], Mads Græsbøll Christensen [b], Søren Holdt Jensen [a,♣], Marc Moonen [c,♣]

[a] Department of Electronic Systems, Aalborg University, Aalborg, Denmark
[b] Department of Media Technology, Aalborg University, Aalborg, Denmark
[c] Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium

ARTICLE INFO

Article history:
Received 30 January 2010
Received in revised form
19 April 2010
Accepted 7 June 2010
Available online 12 June 2010

Keywords:
Multi-pitch estimation
Noise subspace
Orthogonality
Cyclic minimizer

ABSTRACT

In this paper, we present an iterative method for estimation of pitches from signals containing multiple sources using subspace techniques. The resulting estimator is termed Iterative Harmonic MUltiple SIgnal Classification (I-HMUSIC). Different modifications of I-HMUSIC are proposed that improve upon the classical MUSIC algorithm, including a computationally efficient method for noise subspace updating I-HMUSIC and its modifications are evaluated and compared with both the Cramér–Rao lower bound (CRLB) and non-iterative HMUSIC; good statistical performances have been obtained.

1. Introduction

The problem of estimating the fundamental frequency or pitch of a periodic waveform has been of interest to the signal processing community for many years. Fundamental frequency estimators are important for many applications such as automatic note transcription, audio coding, and classification of music. Numerous algorithms have been proposed see, e.g. [1–7]. In real recorded signals, however, multi-pitch scenarios occur more frequently than single-pitch scenarios. A number of multi-pitch algorithm are described in [6,3,4,1], to which we refer the interested reader.

A model using complex exponentials for the multi-pitch estimation problem can be defined as follows: consider a signal consisting of K sources of harmonically related complex exponentials with fundamental frequencies ω_k embedded in additive noise, i.e.,

$$x(n)=\sum_{k=1}^{K}\sum_{l=1}^{L_k}\beta_{l,k}e^{j\omega_k ln}+e(n),\quad \beta_{l,k}=A_{l,k}e^{j\theta_{l,k}} \tag{1}$$

for $n=0,\ldots,N-1$, where $A_{l,k}$ is the real-valued amplitude, ω_k is the fundamental frequency, L_k is the model order, $\theta_{l,k}$ is the phase, and $e(n)$ is the complex symmetric white Gaussian noise. The problem is to estimate ω_k from $x(n)$ with N measured samples. The estimation problem associated with the real case can be cast as (1) by the use of analytic signals, which is valid when there is little or no spectral content of interest near 0 and π. In this paper, we assume that the number of sources K and the model orders L_k of the individual sources are known, order estimation can be achieved with multi-dimensional search of the extended cost function [1,5].

Recently, subspace-based fundamental frequency estimators have shown good estimation performance both for single and multi-pitch cases [5,1,8]. This type of methods is forming a cost function by exploiting the orthogonality properties between the noise and signal subspaces decomposed from a covariance matrix. In subspace-based multi-pitch estimators, however, the cost function is usually multi-modal with many local extrema.

* Corresponding author. Tel.: +45 31598878.
E-mail addresses: jxz@es.aau.dk (J.X. Zhang), mgc@imi.aau.dk (M.G. Christensen), shj@es.aau.dk (S.H. Jensen), marc.moonen@esat.kuleuven.be (M. Moonen).
♣ EURASIP member.

[1] The work is supported by the Marie Curie Fellowship, Contract no. MEST-CT-2005-021175.

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.06.010

Those extrema are the ambiguities between the signal subspace on source of interest and part of the signal subspace of other sources, especially when there are large variations between the model orders L_k. This gives rise to estimation errors such as sub-octave or octave errors on the fundamental frequency for sources other than the source with the highest model order. To avoid local extrema we present an iterative HMUSIC (I-HMUSIC) for multi-pitch estimation. In I-HMUSIC, we use the deflation technique to sequentially estimate the fundamental frequencies starting with the source containing the highest model order. This exploits the property that the estimation of the signal subspace for the source with highest model order is only orthogonal to the noise subspace with the fundamental frequency of interest which gives a cost function without local extrema. The estimated source is then removed from the mixture $x(n)$, and then the next source with the second highest model order is estimated. This procedure simplifies the multi-pitch estimation problem into K sequentially related single-pitch problems. For further improvement of the estimation accuracies, iterative re-estimation of the ω_k's can be performed, based on previously estimated parameters, this type of iteration is normally referred as a cyclic minimizer (CM) [9]. Within this framework of I-HMUSIC, different modifications of the algorithm were also proposed.

Iteratively decomposing a covariance matrix into its subspaces is usually computationally heavy. In this paper, a fast updating method of the noise subspace is also proposed. This method is based on the approximative orthogonality between Vandermonde vectors with distinct frequencies. Our method is asymptotically exact and computationally simpler compared to subspace decomposition algorithms based on eigenvalue decomposition (EVD) or singular value decomposition (SVD).

2. Preliminaries

In this section, we present the fundamentals of the so-called covariance matrix model and introduce some useful vector and matrix notations. The signal sub-vector containing m samples of the observed complex signal (1) is defined as

$$\mathbf{x}(n)=[x(n)\ x(n-1)\ \cdots\ x(n-m+1)]^T, \tag{2}$$

with $(\cdot)^T$ denoting vector transpose, and when m is a user parameter.

The covariance matrix $\mathbf{R}\in\mathbb{C}^{m\times m}$ of $\mathbf{x}(n)$ can then be written as

$$\mathbf{R}=\mathbf{E}\{\mathbf{x}(n)\mathbf{x}^H(n)\}=\sum_{k=1}^{K}\mathbf{Z}_k\mathbf{P}_k\mathbf{Z}_k^H+\sigma^2\mathbf{I}, \tag{3}$$

where $\mathbf{E}\{\cdot\}$ and $(\cdot)^H$ denote the statistical expectation and the Hermitian transpose, respectively. Furthermore, $\mathbf{Z}_k$ is a Vandermonde matrix of source k, which is defined as

$$\mathbf{Z}_k=[\mathbf{z}(\omega_k)\ \cdots\ \mathbf{z}(\omega_k L_k)], \tag{4}$$

where $\mathbf{z}(\omega)=[1\ e^{j\omega}\ \cdots\ e^{j\omega(m-1)}]^T$. The matrix $\mathbf{P}_k=\mathrm{diag}([A_{1,k}^2\ \cdots\ A_{L_k,k}^2])$ contains signal amplitudes, σ^2 denotes the noise variance, and $\mathbf{I}$ is the identity matrix.

Let $\mathbf{R}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be the EVD or SVD of the covariance matrix. Then, $\mathbf{U}$ contains the m orthonormal eigenvectors of $\mathbf{R}$, i.e., $\mathbf{U}=[\mathbf{u}_1\ \ldots\ \mathbf{u}_m]$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues. Lets $\mathbf{G}$ be formed from the $m-Q$ least significant eigenvalues where $Q=\sum_{k=1}^{K}L_k$, with $\mathbf{S}=[\mathbf{u}_1\ \ldots\ \mathbf{u}_Q]$, and $\mathbf{G}=[\mathbf{u}_{Q+1}\ \ldots\ \mathbf{u}_m]$. The noise subspace $\mathbf{G}$ is orthogonal to $\mathbf{Z}_k$, i.e., $\mathbf{Z}_k^H\mathbf{G}=\mathbf{0}$, for $k=1,\ldots,K$.

3. I-HMUSIC algorithm

Having defined the covariance matrix model and useful notations, we now present I-HMUSIC. This method can be summarized into two nested loops: the inner loop uses the deflation technique to sequentially estimate the fundamental frequencies, while the outer loop uses CM to increase the accuracies by repeating the inner loop until some convergence criterion is reached [4,9]. The outer and the inner loops have iteration indices i, k, respectively.

We start with the inner loop, where the deflation technique is adopted to avoid local extrema by estimating the individual sources in a sequence based on the model orders sorted as $L_1\geq L_2\geq\cdots\geq L_K$. The deflated covariance matrix with previously estimated sources removed is defined as

$$\mathbf{R}_{k,i}(\boldsymbol{\varphi}_{k,i})=\mathbf{R}-\sum_{k'=1}^{k-1}\hat{\mathbf{Z}}_{k',i}\hat{\mathbf{P}}_{k',i}\hat{\mathbf{Z}}_{k',i}^H=\mathbf{U}_{k,i}\mathbf{\Lambda}_{k,i}\mathbf{U}_{k,i}^H, \tag{5}$$

where $\boldsymbol{\varphi}_{k,i}=[\hat{\omega}_{1,i}\ \cdots\ \hat{\omega}_{K,i}\ \hat{\mathbf{b}}_{1,i}^T\ \cdots\ \hat{\mathbf{b}}_{K,i}^T]$ is a parameter vector that is updated after each iteration and initially $\boldsymbol{\varphi}_{k,1}=\mathbf{0}$. Without loss of generality, indices i of the parameter vector can be dropped, e.g. $\boldsymbol{\varphi}_k$. The amplitude vector is defined as $\hat{\mathbf{b}}_k=[\hat{\beta}_{1,k}\ \cdots\ \hat{\beta}_{L_k,k}]^T$. The eigenvectors of $\mathbf{U}_k$ are used to form the noise and signal subspace notated as

$$\mathbf{S}_{k,i}=[\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_{L_k}], \tag{6}$$

$$\mathbf{G}_{k,i}=[\mathbf{u}_{L_k+1}\ \mathbf{u}_{L_k+2}\ \ldots\ \mathbf{u}_m], \tag{7}$$

where L_k is the number of principal eigenvalues associated with the remaining harmonics in $\mathbf{R}_k$, given as

$$L_k=Q-\sum_{k'=1}^{k-1}L_{k'} \tag{8}$$

for $k=1,\ldots,K$. Then, we proceed with the estimation of the fundamental frequency of source k as

$$\hat{\omega}_k=\underset{\omega_k}{\mathrm{argmax}}\frac{1}{\|\mathbf{Z}_k^H\mathbf{G}_{k,i}\|_F^2}, \tag{9}$$

where $\|\cdot\|_F$ is the Frobenius norm. With the fundamental frequency, the signal amplitudes can easily be estimated using a least squares estimate, e.g.

$$\hat{\mathbf{b}}_k=(\hat{\mathbf{Z}}_k^H\hat{\mathbf{Z}}_k)^{-1}\hat{\mathbf{Z}}_k^H\left(\mathbf{x}(n)-\sum_{k'=1}^{k-1}\hat{\mathbf{Z}}_{k'}\hat{\mathbf{b}}_{k'}\right). \tag{10}$$

The estimated amplitudes and fundamental frequency are then substituted into (5) to estimate the covariance

matrix for the next source $k+1$. The estimation of fundamental frequency is repeated until $\{\omega_k\}_{k=1}^{K}$ is completely estimated.

In the outer loop, we iteratively re-estimate $\mathbf{R}_{k,i}(\boldsymbol{\varphi}_k)$ based on previous estimates of $\boldsymbol{\varphi}_k$. Increased accuracy of ω_k then results from an improved estimate of the decomposed covariance matrix $\mathbf{R}_{k,i}$. In many applications, the required accuracy for the estimated parameters is satisfied without using CM, i.e., for $i=1$.

Algorithm outline.

Loop 1 $i=1,\ldots,convergence$:
Loop 2 $k=1,\ldots,K$:
$\mathbf{U}_{k,i}\boldsymbol{\Lambda}_{k,i}\mathbf{U}_{k,i}^H = \mathrm{SVD}\{\mathbf{R}_{k,i}(\boldsymbol{\varphi}_k)\}$
$\mathbf{G}_{k,i} = [\mathbf{u}_{L_k+1}\ \ \mathbf{u}_{L_k+2}\ \ \cdots\ \ \mathbf{u}_m]$
$\hat{\omega}_k = \underset{\omega_k}{\mathrm{argmax}} \dfrac{1}{\|\mathbf{Z}_{k,i}^H\mathbf{G}_{k,i}\|_F^2}$
$\hat{\mathbf{b}}_k = (\hat{\mathbf{Z}}_{k,i}^H\hat{\mathbf{Z}}_{k,i})^{-1}\hat{\mathbf{Z}}_{k,i}^H\left(\mathbf{x}(n) - \sum_{k'=1}^{k-1}\hat{\mathbf{Z}}_{k',i}\hat{\mathbf{b}}_{k',i}\right)$
Update $\boldsymbol{\varphi}_k$
End loop 2
End loop 1

3.1. Fast noise subspace updating method

Generally, the EVD or SVD computation used in (5) is computationally heavy. Here, we will present an approximate method to efficiently update the noise subspace for source k. The relationship between the projection matrix of the signal subspace (6) and the projection matrix of the Vandermonde matrix is defined as

$$\boldsymbol{\Pi}_{\mathbf{S}} = \boldsymbol{\Pi}_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H, \tag{11}$$

where $\mathbf{Z}=[\mathbf{Z}_1\ \ldots\ \mathbf{Z}_K]$, and $\boldsymbol{\Pi}_{\mathbf{S}} = \mathbf{S}\mathbf{S}^H$. The columns of Vandermonde matrix $\mathbf{Z}$ are asymptotically orthogonal for any set of distinct frequencies when $m\to\infty$, i.e., [1]

$$\lim_{m\to\infty}\boldsymbol{\Pi}_{\mathbf{Z}} = \lim_{m\to\infty} m\mathbf{Z}(\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H = \sum_{k=1}^{K}\mathbf{Z}_k\mathbf{Z}_k^H. \tag{12}$$

Using (11), an approximately estimate of the noise subspace projection matrix which is also equivalent to the true one is given as

$$\boldsymbol{\Pi}_{\mathbf{G}_k} \approx \boldsymbol{\Pi}_{\mathbf{G}} + \sum_{k'=1}^{k-1}\mathbf{Z}_{k'}(\mathbf{Z}_{k'}^H\mathbf{Z}_{k'})^{-1}\mathbf{Z}_{k'}^H, \tag{13}$$

where $\mathrm{Trace}\{\mathbf{Z}_k^H\boldsymbol{\Pi}_{\mathbf{G}_k}\mathbf{Z}_k\} = 0$. The expression in (13) can then be substituted into (9) for fundamental frequency estimation. Our proposed updating method is computationally simpler because we only need to estimate the true noise subspace once and because it works without estimating the signal amplitudes.

In summary, three modifications of I-HMUSIC are proposed. The first one uses the nested loop system to first sequentially estimate fundamental frequencies and adopts CM to refine the results. This is referred to as I-HMUSIC (SVD, CM). The second one only uses the inner loop to estimate the estimates, referred to as I-HMUSIC (SVD). The third one uses the fast noise subspace updating method I-HMUSIC (FAST).

4. Experimental results

We start the simulation with a simple demonstration of the differences between the cost function in non-iterative HMUSIC used in [10] and the proposed I-HMUSIC(SVD,CM). Signal parameters used in this example consist of two sources with fundamental frequencies ω_1 and ω_2, and harmonic orders are $L_1=8$ and $L_2=4$. The resulting cost functions with non-iterative HMUSIC and I-HMUSIC are shown in Fig. 1, we can see that the cost function has a spurious peak at the octave error of ω_1 while I-HMUSIC shows a cleared cost function. Sometimes, this type of spurious peak is hard to distinguish from real peaks when sources consisting of different model order appear.

Before we start statistical evaluation, we will plot RMSE between the projection matrix of the noise subspace estimated using fast noise subspace updating method and the projection matrix estimated on the noise subspace using SVD. The resulting plot is shown in Fig. 2,

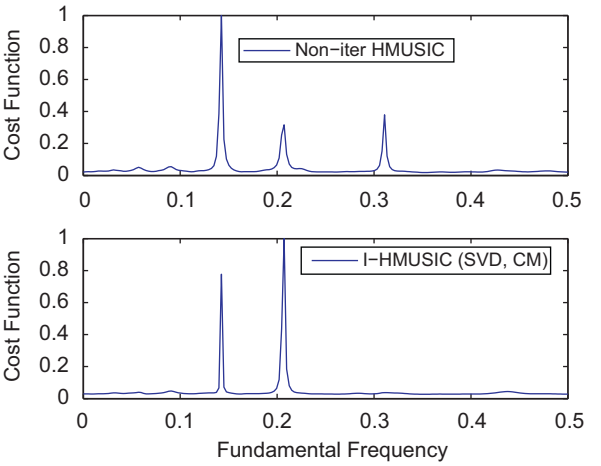

Fig. 1. (Top) Cost function of the non-iterative HMUSIC evaluated on $N=512$. (Bottom) Cost function of the proposed I-HMUSIC.

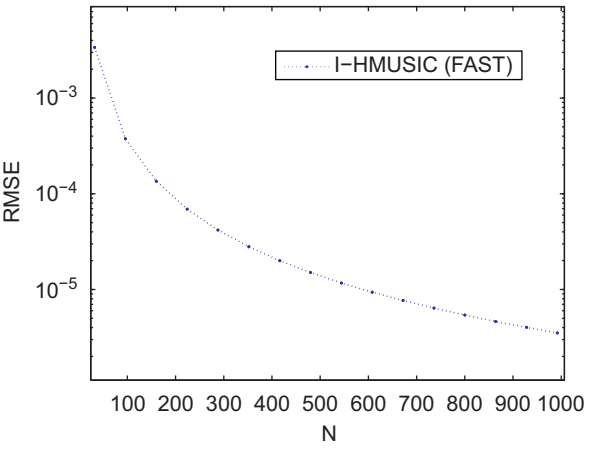

Fig. 2. Error between the true projection matrix and the approximated using our proposed fast method.

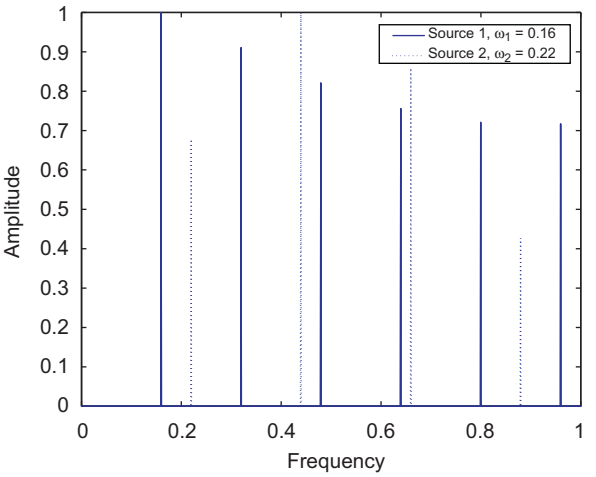

Fig. 3. Source amplitudes used in Monte Carlo simulations.

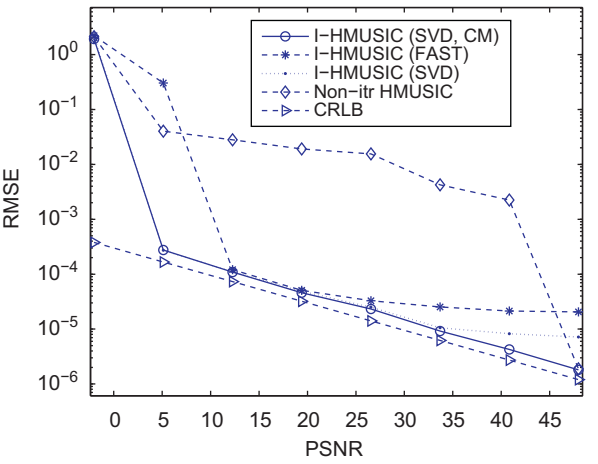

Fig. 4. Estimated RMSE as a function of PSNR for N=512.

as expected the errors of the approximative noise subspace decrease with increased N.

Next, we evaluate the proposed estimators statistical properties using Monte Carlo simulations by generating signals according to the model (1) with the phase and the noise being randomized over each realization, 100 trials are run. The estimators are evaluated for K=2, $\omega_1 = 0.16$ with L_1=8 and $\omega_2 = 0.22$ with L_2=4. The amplitude for each source is generated with a 3th-order AR-filter, as could be expected for natural spectra, the corresponding amplitudes are shown in Fig. 3. Signal frame is evaluated for N=512 samples, and the user parameter set to $m = \lfloor N/2 \rfloor$. Here, the root mean square error is defined as

$$RMSE = \sqrt{\frac{1}{DK}\sum_{d=1}^{D}\sum_{k=1}^{K}(\hat{\omega}_{k,d}-\omega_k)^2}, \tag{14}$$

with $\hat{\omega}_{k,d}$ and ω_k being the estimate and the true fundamental frequency, respectively, and with D being the number of Monte Carlo trials. The asymptotic Cramér–Rao lower bound (CRLB) and the pseudo signal-to-noise ratio (PSNR) for k'th source are defined in [10]. The cost function of (9) is evaluated first using FFT-based method to obtain a coarse estimate of ω_k. Then, these coarse estimates are used to initialize the gradient-based methods to achieve a refined estimate [5]. Furthermore, our methods are compared with CRLB and non-iterative HMUSIC [10].

In this example, we will consider the case where RMSE is a function of PSNR. The evaluated results are shown in Fig. 4, I-HMUSIC (SVD, CM) performs best which closely following CRLB, while I-HMUSIC (FAST) and I-HMUSIC (SVD) perform a bit worse. The main contribution of error in I-HMUSIC (SVD) is during the estimation of the amplitudes; the problem encountered here is that we estimate the amplitudes for one of the sources, while the noise term is white noise plus harmonic interferences from the other sources. It is well-known that least square estimate is optimal only when the noise term is white

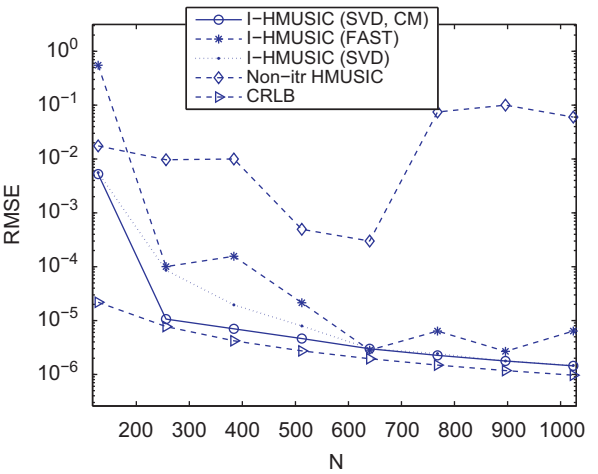

Fig. 5. Estimated RMSE as a function of N with PSNR fixed at 43 dB.

Gaussian noise which is not the case in this problem. Therefore, an iterative estimate of the signal amplitudes with additional knowledge of parameters on harmonic interferences will increase the performance and also reduce RMSE of I-HMUSIC (SVD, CM). The I-HMUSIC (FAST) algorithm will not follow CRLB with increased PSNR, because of errors in the approximated projection matrix, this is shown in Fig. 2. It is interesting to note that CM operation will not increase the performance of I-HMUSIC (FAST), because here we only use basis vectors of the subspaces of interest and no estimation of amplitudes is required.

Next, we proceed to evaluate the proposed estimators RMSE as a function of window length N, PSNR fixed at 43 dB. The results are shown in Fig 5, on the evaluated window length I-HMUSIC (SVD, CM) performs best where the RMSE closely follow CRLB. As expected, the performance of I-HMUSIC (FAST) approaches CRLB with increased N.

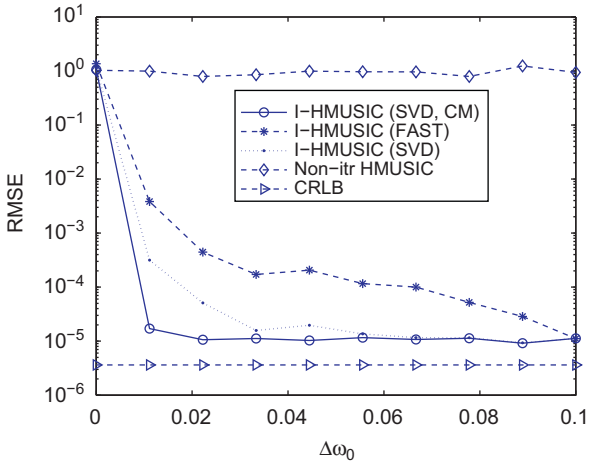

Fig. 6. Estimated RMSE as a function of $\Delta\omega_0$ for N=512 with PSNR fixed at 48.6 dB.

In this final experiment, we will evaluate the minimum resolution between two sources, i.e., $\Delta\omega_0 = |\omega_1 - \omega_2|$, for PSNR is fixed at 48.6 dB and N=512. In order to avoid frequency overlaps between higher harmonic orders the evaluated signals are slightly different from previous examples, here L_1=8 and L_2=1. The results are shown in Fig. 6. It can be seen that I-HMUSIC (SVD, CM) performs best for closely spaced harmonics. Methods such as I-HMUSIC (SVD) and I-HMUSIC (FAST) only give a slightly worse performance but it can be far enough in practical speech and audio applications.

In all examples, the performance of non-iterative HMUSIC is not promising which is explained by spurious peaks in the cost function when one of the source model orders is less than the dominant model order.

5. Conclusion

In this paper, we have proposed an I-HMUSIC for multi-pitch estimation problem, several modifications of the method have also been introduced. Overall, I-HMUSIC gives a cost function that contains less local minimum than non-iterative multi-pitch HMUSIC. We have evaluated our methods with Monte Carlo simulations under different scenarios, in all the cases good statistical properties have been shown. The performance of the various modifications is concluded to be determined by the trade-off between estimation accuracy and computational complexity, where highest complexity gives best accuracy.

References

[1] M.G. Christensen, A. Jakobsson, Multi-Pitch Estimation, Synthesis Lectures on Speech and Audio Processing, 2009.
[2] A. de Cheveigne, H. Kawahara, YIN, a fundamental frequency estimator for speech and music, J. Acoust. Soc. Am. 111 (4) (2002) 1917–1930.
[3] A. Klapuri, Multiple fundamental frequency estimation based on harmonicity and spectral smoothness, IEEE Trans. Signal Process. 47 (2000) 338–352.
[4] H. Li, P. Stoica, J. Li, Computationally efficient parameter estimation for harmonic sinusoidal signals, Signal Processing 80 (2000) 1937–1944.
[5] M.G. Christensen, A. Jakobsson, S.H. Jensen, Joint high-resolution fundamental frequency and order estimation, IEEE Trans. Acoust. Speech Signal Process. 15 (5) (2007) 1635–1644.
[6] R. Badeau, V. Emiya, B. David, Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2009.
[7] V. Emiya, B. David, R. Badeau, A parametric method for pitch estimation of piano tones, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.
[8] J.X. Zhang, M.G. Christensen, J. Dahl, S.H. Jensen, M. Moonen, A robust and computationally efficient subspace-based fundamental frequency estimator, IEEE Trans. Audio Speech Language Process. 18 (3) (2010).
[9] P. Stoica, Y. Selen, Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher, IEEE Signal Process. Mag. 21 (1) (2004) 112–114.
[10] M.G. Christensen, P. Stoica, A. Jakobsson, S.H. Jensen, Multi-pitch estimation, Signal Processing 88 (4) (2008) 972–983.