Fast Harmonic Chirp Summation

Nielsen, Jesper Kjær; Jensen, Tobias Lindstrøm; Jensen, Jesper Rindom; Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

DOI (link to publication from Publisher):
10.1109/ICASSP.2017.7952189

Publication date:
2017

Document Version
Other version

FAST HARMONIC CHIRP SUMMATION

Jesper Kjær Nielsen[1,2], *Tobias Lindstrøm Jensen*[3], *Jesper Rindom Jensen*[1], *Mads Græsbøll Christensen*[1], *and Søren Holdt Jensen*[3]

[1]Audio Analysis Lab, AD:MT
Aalborg University, Denmark
{jkn,jrj,mgc}@create.aau.dk

[2]Acoustic Research
Bang & Olufsen A/S
Struer, Denmark

[3]Dept. of Electronic Systems
Aalborg University, Denmark
{tlj,shj}@es.aau.dk

ABSTRACT

The harmonic chirp signal model has only very recently been introduced for modelling approximately periodic signals with a time-varying fundamental frequency. A number of estimators for the parameters of this model have already been proposed, but they are either inaccurate, non-robust to noise, or very computationally intensive. In this paper, we propose a fast algorithm for the harmonic chirp summation method which has been demonstrated in the literature to be accurate and robust to noise. The proposed algorithm is orders of magnitudes faster than previous algorithms which is also demonstrated via timing studies.

Index Terms— Harmonic chirp model, harmonic chirp summation, fast algorithm

1. INTRODUCTION

Many real-world signals are approximately periodic on a short time-scale. These signals are encountered in many applications such as music processing [1, 2], speech processing [3, 4], sonar [5], order analysis [6], and electrocardiography [7]. Since the underlying parameters of most real-world signals are time-varying, the periodic assumption is no longer accurate when the signal is viewed on a larger time-scale. An example of this is illustrated in Fig. 1 where the spectrogram of an accelerating car engine is shown. When viewed on a short time scale, the frequency components in the signal are approximately constant, but the spectrogram clearly reveals that they are increasing. Thus, we get a more accurate model of the signal by replacing the constant frequency components in the harmonic model with time-varying ones. The harmonic chirp model (HCM) is such a generalisation of the harmonic model where the frequency components are modelled as varying linearly in time, and we here discuss how the harmonic chirp summation (HCS) method can be implemented efficiently for the estimation of the intercept, i.e., the fundamental frequency, and the slope, i.e., the fundamental chirp rate, of such linearly varying frequency components.

The HCM has only very recently been used in [8] as an alternative to modelling non-stationary speech using amplitude modulation models [9]. The HCM was also used in the context of speech processing in [10, 11], but was considered in a more general framework in [12, 13] in which animal sound signals were analysed. In all of these papers, the complex-valued HCM was used although we know of no application where such signals naturally occur. Since the HCM model has only been used for a couple of years, only a few estimators of the fundamental frequency and chirp rate have been proposed, and efficient algorithms for them have not received much attention. Unfortunately, the estimation of the fundamental frequency and chirp rate involves solving a multi-modal 2D optimisation problem which is computationally expensive. Recently in [13], four different estimators were considered, and the results showed that a non-linear least squares (NLS) estimator and the HCS method (called the harmochirp-gram in [13]) had the best estimation accuracy and robustness to noise, but were also the most costly to compute. On the other hand, the authors proposed an estimator called harmonic-SEES which is much faster, but breaks down in noisy conditions for both parameter estimation and model order selection. In [8, 11], the HCS method was implemented in an iterative fashion to keep the computational complexity low, but this approach breaks down if the chirp rate is too high relative to the segment length.

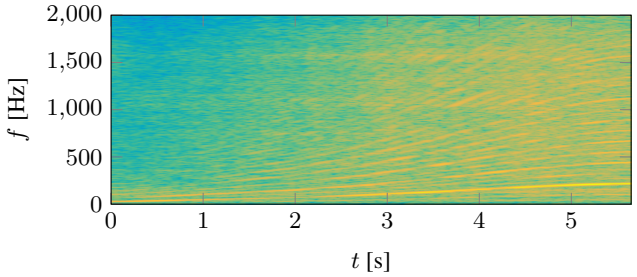

Fig. 1. A spectrogram of an accelerating car engine. The recording is downloaded from `http://tinyurl.com/prosigorder`.

The HCS method is an approximate NLS estimator and, therefore, inherits many of its desirable properties such as noise robustness and asymptotic efficiency. Contrary to the NLS estimator, however, the HCS method does not work well for low fundamental frequencies (measured in cycles/segment), but this decrease in time-frequency resolution is the price to pay to avoid the matrix inversion in the NLS estimator. We think that the HCS method is a very good trade-off between estimation accuracy and computational complexity, and in this paper we focus on reducing the time complexity of the HCS method. Compared to the HCS algorithm in [13], we reduce the complexity by several orders of magnitudes. This gain is obtained by 1) computing a part of the HCS objective with an FFT, 2) exploiting the feasible region of the parameter space, 3) making the algorithm recursive over the model order, and 4) selecting the start index of the time series so that the HCS objective can be evaluated on the coarsest possible grid. Moreover, this start index is shown to lower the Cramér-Rao Lower Bound (CRLB) for any unbiased estimator of the fundamental frequency by up to a factor of 16 compared to the standard start index $n_0 = 0$ which was used in [13].

The work by J.K. Nielsen, T.L. Jensen, and J.R. Jensen was partly supported by the Independent Research Council for Technology and Production grant no. 4005-00122 and 1337-00084, respectively. The work by M.G. Christensen was supported by the Villum Foundation

2. THE HARMONIC CHIRP MODEL

The real-valued harmonic chirp model (HCM) for a data set $\{x(n)\}_{n=n_0}^{n_0+N-1}$ is the real part of the complex-valued HCM introduced in [8] and given by

$$x(n) = \sum_{i=1}^{l} A_i \cos(\theta_i(n)) + e(n) \tag{1}$$

where n_0 is a start index, $A_i \geq 0$ is an amplitude, $l \in \{1, \cdots, L\}$ is the model order, $e(n) \in \mathbb{R}$ is additive noise, and $\theta_i(n) \in [0, 2\pi)$ is the instantaneous phase. In the HCM, this phase is modelled with a second order polynomial as

$$\theta_i(n) = \phi_i + i\omega_0 n + i\beta_0 n^2/2 \tag{2}$$

where ϕ_i, ω_0, and β_0 are the phase at $n = 0$, the fundamental frequency, and the fundamental chirp rate, respectively. The signal model in (1) can be rewritten into a vector form given by

$$\boldsymbol{x} = \boldsymbol{Z}_l(\omega_0, \beta_0)\boldsymbol{\alpha}_l + \boldsymbol{e} \tag{3}$$

with the following definitions

$$\boldsymbol{x} = \begin{bmatrix} x(n_0) & \cdots & x(n_0+N-1) \end{bmatrix}^T \tag{4}$$

$$\boldsymbol{e} = \begin{bmatrix} e(n_0) & \cdots & e(n_0+N-1) \end{bmatrix}^T \tag{5}$$

$$\boldsymbol{Z}_l(\omega, \beta) = \begin{bmatrix} \boldsymbol{C}_l(\omega, \beta) & \boldsymbol{S}_l(\omega, \beta) \end{bmatrix} \tag{6}$$

$$\boldsymbol{C}_l(\omega, \beta) = \begin{bmatrix} \boldsymbol{c}(\omega, \beta) & \cdots & \boldsymbol{c}(l\omega, l\beta) \end{bmatrix} \tag{7}$$

$$\boldsymbol{S}_l(\omega, \beta) = \begin{bmatrix} \boldsymbol{s}(\omega, \beta) & \cdots & \boldsymbol{s}(l\omega, l\beta) \end{bmatrix} \tag{8}$$

$$\boldsymbol{c}(\omega, \beta) = \begin{bmatrix} \cos(\omega n_0 + \frac{1}{2}\beta n_0^2) & \cdots \\ \cos(\omega(n_0+N-1) + \frac{1}{2}\beta(n_0+N-1)^2) \end{bmatrix}^T \tag{9}$$

$$\boldsymbol{s}(\omega, \beta) = \begin{bmatrix} \sin(\omega n_0 + \frac{1}{2}\beta n_0^2) & \cdots \\ \sin(\omega(n_0+N-1) + \frac{1}{2}\beta(n_0+N-1)^2) \end{bmatrix}^T \tag{10}$$

$$\boldsymbol{\alpha}_l = \begin{bmatrix} a_1 & \cdots & a_l & -b_1 & \cdots & -b_l \end{bmatrix}^T \tag{11}$$

where $a_i = A_i \cos(\phi_i)$, and $b_i = A_i \sin(\phi_i)$. We assume that the signal has been Nyquist sampled so that the instantaneous frequency $\omega_i(n) = \theta_i'(n) = i(\omega_0 + \beta_0 n)$ satisfies

$$0 < \omega_i(n) < \pi\,, \quad \text{for } i \in \{1, 2, \ldots, l\}\,. \tag{12}$$

If this constraint is satisfied by $\omega_l(n)$, then it is also satisfied for all lower model orders, so we here focus on $\omega_l(n)$. As we argue in Sec. 3, the best choice of the start index is $n_0 = -(N-1)/2$. For this start index, the feasible region $\mathbb{Q}_l$ for the fundamental frequency ω_0 and chirp rate β_0 can be found by inserting the minimum and maximum values of $\omega_l(n)$ into (12). Thus,

$$\mathbb{Q}_l = \Big\{(\omega_0, \beta_0) \,|\, 0 < \omega_0 - |\beta_0|(N-1)/2 \\ \wedge\ \omega_0 + |\beta_0|(N-1)/2 < \pi/l\Big\}\,. \tag{13}$$

Often, the feasible region is not only determined by Nyquist sampling, but also by prior knowledge on the parameters. Here, we assume that such prior knowledge is represented by upper and lower bounds ($\underline{\omega}_0, \overline{\omega}_0, \underline{\beta}_0$, and $\overline{\beta}_0$) on the fundamental frequency and chirp rate. Thus, the total feasible region is the intersection between $\mathbb{Q}_l$ and these bounds, i.e.,

$$\mathbb{P}_l = \Big\{(\omega_0, \beta_0) \in \mathbb{Q}_l \,|\, \underline{\omega}_0 \leq \omega_0 \leq \overline{\omega}_0 \ \wedge\ \underline{\beta}_0 \leq \beta_0 \leq \overline{\beta}_0\Big\}\,. \tag{14}$$

The feasible sets for a number of different model orders are illustrated in Fig. 2.

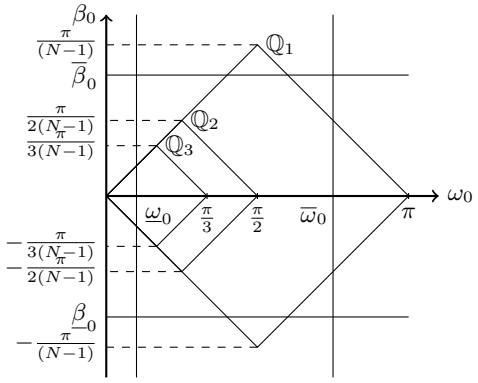

Fig. 2. Example of the feasible set for model orders $l = 1, 2, 3$.

2.1. Harmonic Chirp Summation

As shown in [8, 13], the nonlinear least squares (NLS) estimator of the fundamental frequency and chirp rate for a given l is

$$(\hat{\omega}_0, \hat{\beta}_0) = \underset{(\omega_0, \beta_0) \in \mathbb{P}_l}{\operatorname{argmax}}\ C_l(\omega_0, \beta_0) \tag{15}$$

where the NLS objective is given by

$$C_l(\omega_0, \beta_0) = \boldsymbol{x}^T \boldsymbol{Z}_l(\omega_0, \beta_0) \\ \times \Big[\boldsymbol{Z}_l^T(\omega_0, \beta_0)\boldsymbol{Z}_l(\omega_0, \beta_0)\Big]^{-1} \boldsymbol{Z}_l^T(\omega_0, \beta_0)\boldsymbol{x}\,. \tag{16}$$

As alluded to in the introduction, the NLS objective has a very oscillatory behaviour, but the maximiser can be found reliably using a grid search over $\mathbb{P}_l$ as in [13]. However, this is very costly since we must invert a $2l \times 2l$ matrix for all points on the grid (or solve a linear system of the same size) for all candidate model orders $l \in \{1, \ldots, L\}$. In the harmonic chirp summation (HCS) method, this inversion is avoided by replacing the matrix with its asymptotic expression (in N). Specifically, we have from the results in [14] that

$$\lim_{N\to\infty} 2N^{-1} \boldsymbol{Z}_l^T(\omega_0, \beta_0)\boldsymbol{Z}_l(\omega_0, \beta_0) = \boldsymbol{I}_{2l} \tag{17}$$

where $\boldsymbol{I}_{2l}$ is the $2l \times 2l$ identity matrix. Using this as an approximation for a finite N, we obtain the approximate NLS objective

$$J_l(\omega_0, \beta_0) = \boldsymbol{x}^T \boldsymbol{Z}_l(\omega_0, \beta_0)\boldsymbol{Z}_l^T(\omega_0, \beta_0)\boldsymbol{x} \tag{18}$$

which is the objective of the HCS method. The cost of making the approximation based on (17) is that the HCS method breaks down when the fundamental frequency is low [13]. Low here refers to the number of periods in a segment of data. Thus, we can get a better time-frequency resolution by using the NLS estimator since we can work with shorter segment of data before the estimator breaks down.

2.1.1. Efficient implementation

A closer inspection of (18) reveals that it can be rewritten as

$$J_l(\omega_0, \beta_0) = \sum_{i=1}^{l} \Big|\boldsymbol{f}^H(i\omega_0)\,(\boldsymbol{g}^*(i\beta_0) \odot \boldsymbol{x})\Big|^2 = \sum_{i=1}^{l} J_1(i\omega_0, i\beta_0)$$

where $\odot$ denotes element-wise multiplication and

$$\boldsymbol{f}(\omega) = \begin{bmatrix} 1 & \exp(j\omega 1) & \cdots & \exp(j\omega(N-1)) \end{bmatrix}^T \tag{19}$$

$$\boldsymbol{g}(\beta) = \begin{bmatrix} \exp(j\beta n_0^2) & \cdots & \exp(j\beta(n_0+N-1)^2) \end{bmatrix}^T\,. \tag{20}$$

Thus, for each candidate β_0, we can efficiently compute the objective on a uniform grid using an FFT algorithm. Moreover, the objective $J_1(i\omega_0, i\beta_0)$ evaluated on a uniform grid covering $\mathbb{Q}_1$ acts

as a mother objective for the objectives of higher model orders since each of these can be computed by recursively summing the appropriate elements of $J_1(\omega_0, \beta_0)$. Finally, $J_l(\omega_0, -\beta_0) = J_l(-\omega_0, \beta_0)$ so if we use a symmetric rate grid around 0, we basically get one half of the objective evaluations for free since the FFT algorithm already computes the objective for negative candidate fundamental frequencies. In total, the time complexity is $\mathcal{O}(KF \log F)$ for calculating $J_1(\omega_0, \beta_0)$ and $\mathcal{O}(\sum_{i=2}^{L}(K/i)(F/i)) = \mathcal{O}(KF)$ for calculating $\{J_l(\omega_0, \beta_0)\}_{l=2}^{L}$ where K and F are the number of candidate fundamental chirp rates for $l = 1$ and the number of FFT grid points, respectively. Thus, the total time complexity of our proposed algorithm for the HCS method is dominated by $\mathcal{O}(KF \log F)$ which means that, in terms of the order of time complexity, we get the objectives $\{J_l(\omega_0, \beta_0)\}_{l=2}^{L}$ for free. In Sec. 4, we establish how K and F depend on N and L.

3. THE START INDEX, THE HESSIAN, AND THE CRLB

Traditionally, the start index n_0 is set to either 0 or 1 without much thought. This was also the case in [13]. In many applications, however, even a non-integer choice for this index can be made since the important information lies in the distance between the sampling times, not an arbitrary starting time. This was also noted in [15] where it was shown that selecting the start index as $n_0 = -(N-1)/2$ minimised the CRLB of the phase of a chirp signal (i.e., the HCM with $l = 1$). This indexing was also adopted in [8, 11], but the consequences of this choice for the general HCM with $l > 1$ were not analysed, and we, therefore, do this here. Specifically, we show (asymptotically) that a start index of $n_0 = -(N-1)/2$ *1)* diagonalises the Hessian of $J_l(\omega_0, \beta_0)$, and *2)* minimises the CRLB of any unbiased estimator of the fundamental frequency.

3.1. The Hessian

Unfortunately, the derivation of the Hessian is lengthy, so we skip it here, but if we make use of the asymptotic result in (17) and assume that the SNR is high so that $\boldsymbol{x} \approx \boldsymbol{Z}(\omega_0, \beta_0)\boldsymbol{\alpha}$, it can be shown that

$$\boldsymbol{H}_l(\hat{\boldsymbol{\xi}}) = \frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} J_l(\boldsymbol{\xi})\Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \approx -\frac{N^2(N^2-1)\sum_{i=1}^{l} A_i^2 i^2}{24} \times \begin{bmatrix} 1 & n_0 + \frac{N-1}{2} \\ n_0 + \frac{N-1}{2} & \frac{N^2-4}{60} + \left(n_0 + \frac{N-1}{2}\right)^2 \end{bmatrix} . \quad (21)$$

where $\boldsymbol{\xi} = \begin{bmatrix} \omega_0 & \beta_0 \end{bmatrix}^T$. Clearly, setting $n_0 = -(N-1)/2$ makes the Hessian diagonal, and this can make simple iterative procedures for refining a coarse grid-based estimate converge much faster to the maximum. In Fig. 3, an example of the objective for $n_0 = -(N-1)/2$ is shown. For comparison, an example for $n_0 = 0$ can be found in [13]. Another advantage of having a diagonal Hessian is that the grid size can be designed for the fundamental frequency and chirp rate independently. We discuss this more thoroughly in Sec. 4.

3.2. The CRLB

The asymptotic CRLB for the parameter vector $\boldsymbol{\eta}_l = \begin{bmatrix} \boldsymbol{\alpha}_l^T & \omega_0 & \beta_0 \end{bmatrix}^T$ can be derived under the same approximations as made in the derivation of the Hessian. Assuming that the noise is white and Gaussian with variance σ^2, we obtain the Fisher information matrix (FIM) in (22) on the top of page 4 where $\boldsymbol{n} = \begin{bmatrix} n_0 & \cdots & n_0 + N - 1 \end{bmatrix}^T$ and $\boldsymbol{l} = \begin{bmatrix} 1 & \cdots & l \end{bmatrix}^T$. As we are here only interested in the bound

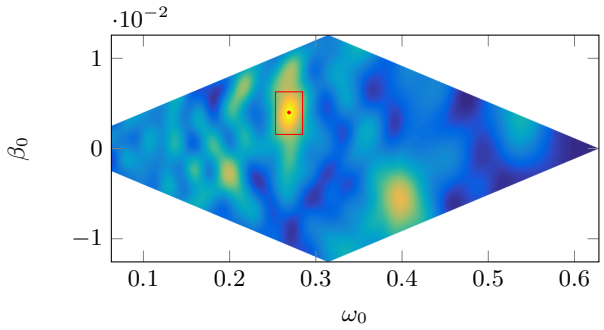

Fig. 3. Example of an objective function for $N = 50$, $l = 5$, $n_0 = -(N-1)/2$, and no noise. The red dot indicates the maximum, and the red rectangle shows a scaled grid cell.

on the fundamental frequency and chirp rate, we only focus on the lower 2×2 matrix of the asymptotic inverse FIM which is given by

$$\boldsymbol{\mathcal{I}}_l^{-1}(\omega_0, \beta_0) = \frac{24\sigma^2}{N(N^2-1)(N^2-4)\sum_{i=1}^{l} A_i^2 i^2} \times \begin{bmatrix} (N^2-4) + 60\left(n_0 + \frac{N-1}{2}\right)^2 & -60\left(n_0 + \frac{N-1}{2}\right) \\ -60\left(n_0 + \frac{N-1}{2}\right) & 60 \end{bmatrix} . \quad (23)$$

The CRLB for β_0 does not depend on the start index, but the CRLB for ω_0 does and is minimised by $n_0 = -(N-1)/2$. In fact, the ratio between the CRLB for ω_0 for $n_0 = 0$ and $n_0 = -(N-1)/2$ approaches 16 for $N \to \infty$. Thus, we get a much more accurate estimate of the fundamental frequency for $n_0 = -(N-1)/2$.

4. GRID SIZE SELECTION

As we have alluded to earlier, we recommend that the maximiser of the HCS objective in (18) is found by evaluating the objective over a coarse grid followed by a refinement procedure such as the Nelder-Mead method [16]. This ensures that the number of required objective evaluations is low. In [13], the authors did not use a refinement step and recommended that the number of grid points was based on the CRLB. Specifically, they suggested that $F = \mathcal{O}(N^{3/2})$ and $K = \mathcal{O}(N^{5/2})$ grid points were used for the fundamental frequency and chirp rate, respectively. By following the procedure suggested in [17] instead, we show here how a much lower value can be selected by combining a coarse grid search with a refinement step.

Suppose that we set the grid sizes in $\Delta\boldsymbol{\xi} = \begin{bmatrix} \Delta\omega_0 & \Delta\beta_0 \end{bmatrix}^T$ so that the objective has decreased by a factor of $g > 1$ when we move $\pm\Delta\boldsymbol{\xi}$ away from the maximiser $\hat{\boldsymbol{\xi}}$. That is,

$$J_l(\hat{\boldsymbol{\xi}} \pm \Delta\boldsymbol{\xi}) = J_l(\hat{\boldsymbol{\xi}})/g \, . \quad (24)$$

Unfortunately, this equation cannot be solved for $\Delta\boldsymbol{\xi}$. However, we can obtain a approximate solution by replacing $J_l(\hat{\boldsymbol{\xi}} \pm \Delta\boldsymbol{\xi})$ by its second order Taylor approximation around $\hat{\boldsymbol{\xi}}$. Since $J_l(\hat{\boldsymbol{\xi}}) \approx (N^2/4)\sum_{i=1}^{l} A_i^2$ and the Hessian in (21) is diagonal for $n_0 = -(N-1)/2$, we obtain that

$$\Delta\omega_0 \approx \sqrt{2\frac{1-g}{g}\frac{J_l(\hat{\boldsymbol{\xi}})}{[\boldsymbol{H}_l(\hat{\boldsymbol{\xi}})]_{11}}} > \sqrt{12\frac{g-1}{g}}\frac{1}{NL} \quad (25)$$

$$\Delta\beta_0 \approx \sqrt{2\frac{1-g}{g}\frac{J_l(\hat{\boldsymbol{\xi}})}{[\boldsymbol{H}_l(\hat{\boldsymbol{\xi}})]_{22}}} > \sqrt{12\frac{g-1}{g}}\frac{\sqrt{60}}{N^2L} \quad (26)$$

where the lower bounds are obtained by selecting the worst case value of the amplitudes, i.e., $A_i = 0$ for $i = 1, \ldots, L-1$ and

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\eta}_l) = \sigma^{-2} \begin{bmatrix} N & \mathbf{0}^T & 0 & 0 \\ \mathbf{0} & \frac{N}{2}\boldsymbol{I}_{2l} & -\frac{\boldsymbol{1}^T\boldsymbol{n}}{2}\begin{bmatrix}\boldsymbol{l}\odot\boldsymbol{b}\\ \boldsymbol{l}\odot\boldsymbol{a}\end{bmatrix} & -\frac{\boldsymbol{n}^T\boldsymbol{n}}{4}\begin{bmatrix}\boldsymbol{l}\odot\boldsymbol{b}\\ \boldsymbol{l}\odot\boldsymbol{a}\end{bmatrix} \\ 0 & -\frac{\boldsymbol{n}^T\boldsymbol{1}}{2}\begin{bmatrix}(\boldsymbol{l}\odot\boldsymbol{b})^T & \boldsymbol{l}\odot\boldsymbol{a})^T\end{bmatrix} & \frac{\boldsymbol{n}^T\boldsymbol{n}}{2}\sum_{i=1}^{l} A_i^2 i^2 & \frac{(\boldsymbol{n}\odot\boldsymbol{n})^T\boldsymbol{n}}{4}\sum_{i=1}^{l} A_i^2 i^2 \\ 0 & -\frac{\boldsymbol{n}^T\boldsymbol{n}}{4}\begin{bmatrix}(\boldsymbol{l}\odot\boldsymbol{b})^T & (\boldsymbol{l}\odot\boldsymbol{a})^T\end{bmatrix} & \frac{\boldsymbol{n}^T(\boldsymbol{n}\odot\boldsymbol{n})}{4}\sum_{i=1}^{l} A_i^2 i^2 & \frac{(\boldsymbol{n}\odot\boldsymbol{n})^T(\boldsymbol{n}\odot\boldsymbol{n})}{8}\sum_{i=1}^{l} A_i^2 i^2 \end{bmatrix} \tag{22}$$

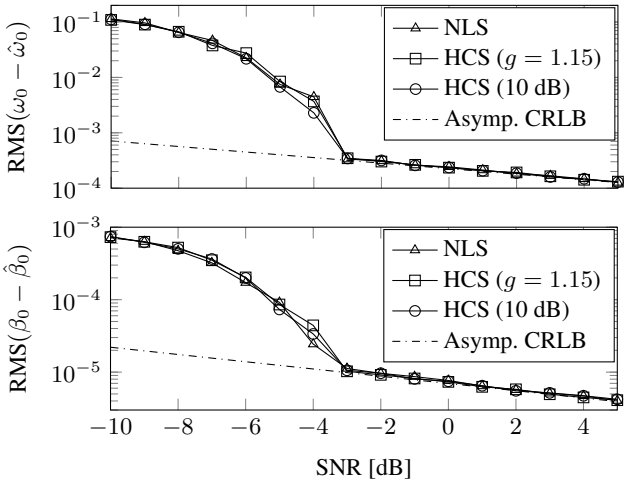

Fig. 4. Estimation accuracy as a function of the SNR. Settings: $N = 250$, $l = L = 6$, $\omega_0 \sim \mathcal{U}(8\pi/N, 16\pi/N)$, $\beta_0 \sim \mathcal{U}(-3/N^2, 3/N^2)$, $A_i = 1$, $i = 1, \ldots, l$, and $\phi_i \sim \mathcal{U}(0, 2\pi)$.

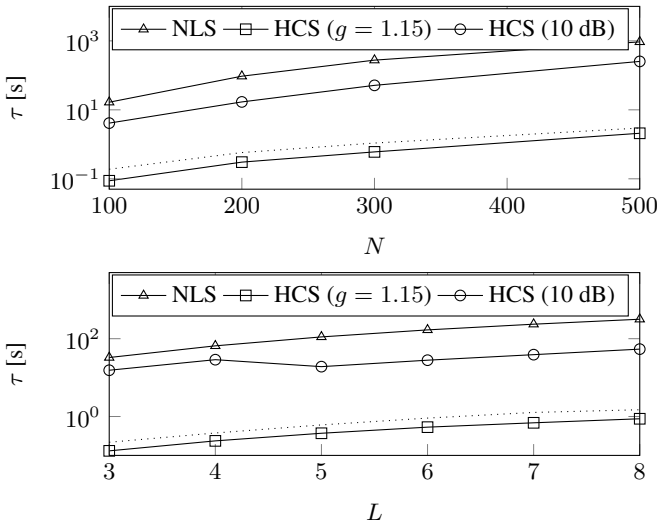

Fig. 5. The computation times for computing the objectives on the grid with $F = 5NL$ and $K = NL$. Top: $L = 6$ Bottom: $N = 256$. The dotted lines are the corresponding timing of the HCS ($g = 1.15$) with the Nelder-Mead refinement step.

$A_L = A > 0$. Note that we use the maximum candidate model order L instead of the true order l since the latter is unknown. In the evaluation section, we find a suitable value for g, but the above can be used to find the relative size of the grid cell. In Fig. 3, an example of such a grid cell is shown, and it clearly captures the shape of the main peak of the objective.

From the grid sizes and $\mathbb{Q}_1$, we can now determine how the number of grid points F and K scale with N and L. Specifically,

$$F = 2\pi/\Delta\omega_0 = \mathcal{O}(NL) \tag{27}$$

$$K = 2\pi/((N-1)\Delta\beta_0) = \mathcal{O}(NL) \tag{28}$$

so that the time complexity for the algorithm in Sec. 2.1.1 is equal to $\mathcal{O}(N^2L^2 \log NL)$. This is significantly lower than the complexity $\mathcal{O}(N^5L^2)$ of the algorithm for the HCS method proposed in [13] (when the objective is evaluated for all candidate model orders), and only slightly larger than the complexity $\mathcal{O}(N^2L^2 \log N)$ of the Harmonic-SEES, also proposed in [13]. The Harmonic-SEES, however, breaks down in noisy conditions.

5. NUMERICAL EVALUATIONS

Due to the limited space and that we here propose a fast algorithm of an already known estimator, we neither evaluate the accuracy of the estimator nor apply it to real world data. For such evaluations, we refer the interested reader to [8, 11, 13]. Instead, we here only find a suitable value for the grid cells and evaluate the computation times of our algorithm for the HCS method. We compare the proposed fast HCS algorithm to our own implementation of the NLS method and a fine-grid version (as suggested in [13]) of our fast HCS algorithm since it, unfortunately, has not been possible to get access to the code in [13][1]. To find a suitable grid size or, equivalently, value of g for our proposed HCS algorithm, we ran a Monte Carlo simulation consisting of 1000 runs at each SNR from -10 dB to 5 dB in steps of 1 dB. In each run, all the model parameters except the amplitudes were generated at random as described in the caption of Fig. 4.

This figure shows the estimation accuracy of the NLS and the HCS methods for a grid size given by setting $g = 1.15$ as well as for the HCS method for a grid size given by the CLRB at an SNR of 10 dB. All grid estimates were refined using the NLS objective. We see no major difference in estimation accuracy between the three methods suggesting that using a combination of a coarse grid followed by a refinement works as well as using a much finer grid with refinement. The former approach, however, dramatically reduces the computation time as illustrated in Fig. 5 for various data length and model orders[2]. The fine-grid version of our HCS algorithm is roughly 50 times slower than the version with the coarser grid. Since the only difference between the two HCS algorithms is the coarseness of the grid we are confident that our HCS algorithm is much more than a factor of 50 times faster than the naïve HCS algorithm in [13].

6. CONCLUSION

In this paper, we have developed a fast algorithm for the harmonic chirp summation method which has been demonstrated in the literature to be an accurate and robust estimator of the parameters in the harmonic chirp model. The proposed estimator is orders of magnitudes faster than a recently suggested algorithm. The speed up was obtained by using an FFT algorithm in the evaluation of the cost function and by using a combination of a coarse grid search and a local refinement instead of a fine grid search. Numerical evaluations showed that using the former approach is as accurate as the latter, but leads to a significantly lower computation time. We believe that our fast algorithm for the HCS method now makes the method practically useful in a large number of applications.

[1] Our code will be available at `http://tinyurl.com/jknvbn`.

[2] For orders 3 and 4, the HCS (10 dB) method is costly compared to higher orders. This is due to MATLAB's FFT implementation whose time complexity can be quite high when the FFT size is prime or has large prime factors.

7. REFERENCES

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

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

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

[4] R. J. Sluijter, *The Development of Speech Coding and the First Standard Coder for Public Mobile Telephony*, Ph.D. thesis, Techniche Universiteit Eindhoven, 2005.

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

[6] S. Gade, H. Herlufsen, H. Konstantin-Hansen, and N. J. Wismer, "Order tracking analysis," Technical review no. 2, Brüel & Kjær A/S, 1995.

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

[8] M. G. Christensen and J. R. Jensen, "Pitch estimation for non-stationary speech," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2014, pp. 1400–1404.

[9] Y. Pantazis, O. Rosec, and Y. Stylianou, "Chirp rate estimation of speech based on a time-varying quasi-harmonic model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3985–3988.

[10] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Enhancement and noise statistics estimation for non-stationary voiced speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 645–658, 2016.

[11] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Instantaneous pitch estimation with optimal segmentation for non-stationary voiced speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 12, pp. 2354–2367, 2016.

[12] J. Swärd, J. Brynolfsson, A. Jakobsson, and M. Hansson-Sandsten, "Sparse semi-parametric estimation of harmonic chirp signals," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1798–1807, 2015.

[13] Y. Doweck, A. Amar, and I. Cohen, "Joint model order selection and parameter estimation of chirps with harmonic components," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1765–1778, 2015.

[14] A. Lahiri, D. Kundu, and A. Mitra, "Estimating the parameters of multiple chirp signals," *J. Multivariate Analysis*, vol. 139, pp. 189–206, 2015.

[15] P. M. Djuric and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2118–2126, Dec. 1990.

[16] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer J.*, vol. 7, no. 4, pp. 308–313, 1965.

[17] J. K Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Grid size selection for nonlinear least-squares optimisation in spectral estimation and array processing," in *Proc. European Signal Processing Conf.*, 2016, pp. 1653–1657.