A Fast Algorithm for Maximum Likelihood Estimation of Harmonic Chirp Parameters

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

Published in:
I E E E Transactions on Signal Processing

DOI (link to publication from Publisher):
10.1109/TSP.2017.2723342

Publication date:
2017

Document Version
Early version, also known as pre-print

A Fast Algorithm for Maximum Likelihood Estimation of Harmonic Chirp Parameters

Tobias Lindstrøm Jensen, Jesper Kjær Nielsen, *Member, IEEE*, Jesper Rindom Jensen, *Member, IEEE*, Mads Græsbøll Christensen, *Senior Member, IEEE*, and Søren Holdt Jensen, *Senior Member, IEEE*

***Abstract*—The analysis of (approximately) periodic signals is an important element in numerous applications. One generalization of standard periodic signals often occurring in practice are harmonic chirp signals where the instantaneous frequency increases/decreases linearly as a function of time. A statistically efficient estimator for extracting the parameters of the harmonic chirp model in additive white Gaussian noise is the maximum likelihood (ML) estimator which recently has been demonstrated to be robust to noise and accurate — even when the model order is unknown. The main drawback of the ML estimator is that only very computationally demanding algorithms for computing an estimate are known. In this paper, we give an algorithm for computing an estimate to the ML estimator for a number of candidate model orders with a much lower computational complexity than previously reported in the literature. The lower computational complexity is achieved by exploiting recursive matrix structures, including a block Toeplitz-plus-Hankel structure, the fast Fourier transform, and using a two-step approach composed of a grid and refinement step to reduce the number of required function evaluations. The proposed algorithms are assessed via Monte Carlo and timing studies. The timing studies show that the proposed algorithm is orders of magnitude faster than a recently proposed algorithm for practical sizes of the number of harmonics and the length of the signal.**

***Index Terms*—Fundamental frequency estimation, linear chirp models, Toeplitz, Hankel, fast algorithms, pitch.**

I. Introduction

THE analysis of periodic signals is an important signal processing task and problem since such signals are encountered in many applications, including music processing [1], [2], speech processing [3], [4], sonar [5], order analysis [6], and electrocardiography (ECG) [7] to name a few. Such signals can be modeled as a weighted sum of sinusoids, which is known as the harmonic model, whose frequencies are integral multiples of a common, constant fundamental frequency.

Manuscript received Month day, 2015;

T.L. Jensen and S.H. Jensen are with the Signal and Information Processing Group, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: tlj@es.aau.dk; shj@es.aau.dk). The work of T.L. Jensen was supported by the Danish Council for Independent Research | Technology and Production grant no. 4005-00122.

J.K. Nielsen is with the Signal and Information Processing Group, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark and with Bang & Olufsen A/S, Struer, Denmark (e-mail: jkn@es.aau.dk). The work of J.K. Nielsen was supported by the Danish Council for Independent Research | Technology and Production grant no. 4005-00122.

J.R. Jensen and M.G. Christensen are with the Audio Analysis Lab, Department of Architecture, Design & Media Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: jrj@create.aau.dk; mgc@create.aau.dk). The work of J.R. Jensen was supported by the Danish Council for Independent Research | Technology and Production grant no. 1337-00084.

Digital Object Identifier XX.XXX/TSP.2016.XXXXXXX

Many natural signals are not stationary, and hence the harmonic model might be unsuitable. The harmonic chirp model is a generalization of the harmonic model where the instantaneous frequency of the harmonics changes linearly as a function of time. This type of model was investigated in [8]–[11] and shown to be more accurate than the traditional harmonic model in many practical scenarios where the pitch of the signal is not stationary. The parameters of the model are the amplitudes of the harmonics, the fundamental frequency, and the fundamental chirp rate.

One method to estimate the unknown parameters of the harmonic chirp model is the non-linear least-squares (NLS) estimator for the harmonic chirp model that is identical to the maximum likelihood (ML) estimator under a white Gaussian noise assumption. Under these conditions, a number of approximate methods were investigated in [10] along with the NLS method with the conclusion that *1)* NLS is the most accurate method for any SNR, *2)* the NLS method yields the best noise variance estimation accuracy and can hence provide the best model order selection performance, *3)* the NLS estimator is the most robust estimator when the model is incorrect such as for frequency modulation. The main drawback of the NLS estimator is that it is considered "computational intensive" which gave rise to work on approximate but less accurate methods [10]. The reason is that the objective function for the NLS objective function has a very oscillating behavior like many NLS estimators for similar models which have motivated the use of the grid method (possibly combined with other methods) as in the RELAX algorithm [12], [13], weighted RELAX for time of arrival estimation [14], [15], Capon and APES spectral estimators [16], [17].

For most applications, the model order is also unknown and should be estimated as well on a segment-by-segment basis [10], [11]. However, model order selection methods require that we can compute the estimates/solutions of the estimator for a number of candidate model orders, e.g., $l = 1, \ldots, L$, see e.g. [18]. This requires that we solve the underlying optimization problem L times and the standard approach is to treat these as L individual optimization problems [10], [11].

In this paper, we will present a fast algorithm for the NLS estimator for harmonic chirp signals. A fast algorithm is achieved using three strategies. *1)* Reduce the number of objective function evaluations by decreasing the grid resolution and instead rely on a two step approach of first approximately finding the region of the global optimum and then use an iterative method for refining the solution. *2)* Formulate the problem as a single joint problem by nesting the L problems

together, such that it is possible to exploit recursive updating from candidate model order l to $l+1$ (instead of treating the overall problem as L individual problems). *3)* Exploit the matrix structures of the problem and form an algorithm with recursive in-order solvers for block Toeplitz-plus-Hankel linear systems of equations and employing the fast Fourier transform (FFT) for efficient evaluation of necessary quantities on a uniform grid. We also propose a fast recursive method for an approximate NLS method called harmonic chirp summation (HCS), denoted harmochirp-gram in [10]. The proposed algorithms have a computational complexity that is orders of magnitude lower than previously known, and both the proposed NLS and HCS algorithms will have the same asymptotic computational complexity as the less accurate harmonic separate-estimate (Harmonic-SEES) method of [10]. Furthermore, we will show that the time index selection has an impact on the Cramér-Rao lower bound (CRLB) for the harmonic chirp model similar to the non-harmonic chirp model [19]. Specifically, a symmetric time index will reduce the CRLB for the fundamental frequency by up to a factor of 16 compared to a conventional time index selection.

To the best of our knowledge, only the complex-valued harmonic chirp model has been considered in the scientific literature despite that we know of no applications where such complex-valued models appear naturally. On the contrary, many applications of the harmonic chirp model have naturally real-valued signals. To apply an estimator for complex-valued signals on real-valued signals then requires a conversion from real- to complex-valued data and this produces an estimation error that increases as any energy of the signal approaches DC or half the Nyquist rate [20]. The complex-valued model is also used in the non-harmonic linear chirp model [19], [21], [22], higher order models [23] and superimposed chirps [24]. In this paper, we instead use the real-valued model.

The remaining part of the paper is organized as follows. In Sec. II, we outline the standard algorithm for the NLS estimator. Sec. III contains analysis and a description of the implication of using a symmetric time index on the objective function and the CRLB. Sec. IV gives an analysis of the behavior of the objective function and gives some recommendations with respect to the grid resolution. We analyze the matrix structure of the problem and give a fast recursive algorithm for computing the NLS objective function in Sec. V and the HCS objective function in Sec. VI. We present a comparison with the CRLB and timings of the proposed method in Sec. VII and conclude on the results in Sec. VIII.

Notation: Let $\mathbb{R}$ denote the set of real numbers, let $\mathbb{C}$ denote the set of complex numbers, and let $\Re(x)$, $\Im(x)$ denote the real and imaginary part of $x \in \mathbb{C}$, respectively. The notation $\boldsymbol{x}^T$ denotes the transpose of a vector $\boldsymbol{x}$ and $\boldsymbol{x}^H = (\boldsymbol{x}^*)^T$ the Hermitian transpose. For a vector $\boldsymbol{x}$, $\mathrm{diag}(\boldsymbol{x})$ denotes a square matrix with the vector $\boldsymbol{x}$ on the diagonal and zeros elsewhere. We will use the notation $[\boldsymbol{x}_l]_j \in \mathbb{R}^{S\times Q}$ to denote the jth sub-block of $\boldsymbol{x}_l \in \mathbb{R}^{Sl\times Q}$. The Hadamard (element-wise) product is $\odot$ and the Kronecker product is $\otimes$.

II. Standard Algorithm

The real-valued harmonic chirp model is given by

$$x(n) = \sum_{l=1}^{\ell} a_l \cos(\psi_l(n)) - b_l \sin(\psi_l(n)) + e(n) \tag{1}$$

for time index $n = n_0, \ldots, n_0 + N - 1$, instantaneous phase $\psi_l(n) = l(\omega_0 n + \frac{1}{2}\beta_0 n^2)$ in radians/sample and (true) model order ℓ. In a vector form, this can be written as

$$\boldsymbol{x} = \boldsymbol{Z}_\ell(\omega_0, \beta_0)\boldsymbol{\alpha}_\ell + \boldsymbol{e} \tag{2}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x(n_0) & \cdots & x(n_0+N-1)\end{bmatrix}^T \tag{3}$$

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

$$\boldsymbol{Z}_l(\omega,\beta) = \begin{bmatrix}\boldsymbol{C}_l(\omega,\beta) & \boldsymbol{S}_l(\omega,\beta)\end{bmatrix} \tag{5}$$

$$\boldsymbol{C}_l(\omega,\beta) = \begin{bmatrix}\boldsymbol{c}(\omega,\beta) & \cdots & \boldsymbol{c}(l\omega,l\beta)\end{bmatrix} \tag{6}$$

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

$$\boldsymbol{c}(\omega,\beta) = \begin{bmatrix}\cos(\omega n_0 + \frac{1}{2}\beta n_0^2) & \cdots \\ \cos(\omega(n_0+N-1) + \frac{1}{2}\beta(n_0+N-1)^2)\end{bmatrix}^T \tag{8}$$

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

$$\boldsymbol{\alpha}_l = \begin{bmatrix}\boldsymbol{a}_l^T & -\boldsymbol{b}_l^T\end{bmatrix}^T \tag{10}$$

$$\boldsymbol{a}_l = \begin{bmatrix} a_1 & \cdots & a_l\end{bmatrix}^T \tag{11}$$

$$\boldsymbol{b}_l = \begin{bmatrix} b_1 & \cdots & b_l\end{bmatrix}^T . \tag{12}$$

We will assume that $\boldsymbol{Z}_\ell(\omega_0,\beta_0)$ has no linearly dependent columns such that the linear parameters in (2) are identifiable, i.e., we assume that $N \geq 2l$ and $\omega \bmod \pi \neq 0$ if $\beta \bmod 2\pi = 0$. Notice that the last two conditions pertain to the critical sampling of the signal. When the noise is white and Gaussian, the ML estimator of ω_0 and β_0 with known model order ℓ is

$$(\hat{\omega}_0, \hat{\beta}_0) = \underset{\omega_0,\beta_0}{\operatorname{argmax}}\, J_\ell(\omega_0,\beta_0) \tag{13}$$

with the non-linear objective function

$$J_l(\omega_0,\beta_0) = \boldsymbol{x}^T\boldsymbol{Z}_l(\omega_0,\beta_0)\left[\boldsymbol{Z}_l^T(\omega_0,\beta_0)\boldsymbol{Z}_l(\omega_0,\beta_0)\right]^{-1}\boldsymbol{Z}_l^T(\omega_0,\beta_0)\boldsymbol{x} \tag{14}$$

see, e.g., [9]. In other settings, (13) is simply referred to as the non-linear least-squares (NLS) estimator. In this model, we have not included any DC component since many signals such as audio and speech signals do not have a DC components. Moreover, ignoring the DC-term, makes the derivations in the following sections much simpler. In (13) we have not assumed any bounds on the parameters (ω_0,β_0). We will in the following constrain our selves to the case where the instantaneous frequency of the lth harmonic

$$\omega_l(n) = l(\omega_0 + \beta_0 n), \quad l = 1,\ldots,\ell \tag{15}$$

is positive $\omega_l(n) > 0$ and Nyquist sampled $\omega_l(n) < \pi$. In this paper we work extensively with the choice of a symmetric time

index $n_0 = -\frac{N-1}{2}$ mostly because this minimizes the CRLB for the fundamental frequency and makes the Hessian diagonal (this is discussed in detail in Sec. III). With a symmetric time index, ω_0 and β_0 must then satisfy

$$0 < \omega_0 - |\beta_0|\frac{(N-1)}{2}, \quad \omega_0 + |\beta_0|\frac{(N-1)}{2} < \frac{2\pi}{2\ell}\,. \quad (16)$$

We may add additional bounds on the form $\underline{\omega}_0 \leq \omega_0 \leq \overline{\omega}_0$ if we have additional a priori information and in essence then form a MAP estimator. The inequalities above can also be translated into bounds on the rate parameter

$$|\beta_0| < \frac{2\omega_0}{N-1}, \quad |\beta_0| < \frac{2\pi}{\ell(N-1)} - \frac{2\omega_0}{N-1}\,. \quad (17)$$

Similar to the fundamental frequency we may add additional bounds $\underline{\beta}_0 \leq \beta_0 \leq \overline{\beta}_0$ if we have additional a priori information. In the case of a Nyquist sampled signal, the feasible set without any additional bounds is,

$$\mathbb{Q}_l = \Big\{[\omega_0, \beta_0]\,|\,0 < \omega_0 - |\beta_0|\frac{(N-1)}{2}, \omega_0 + |\beta_0|\frac{(N-1)}{2} < \frac{2\pi}{2l}\Big\}, \quad l = \ell\,. \quad (18)$$

The feasible set is illustrated in Fig. 1 for various ls along with an indication of the impact of possible additional bounds using $\underline{\omega}_0, \overline{\omega}_0, \underline{\beta}_0, \overline{\beta}_0$.

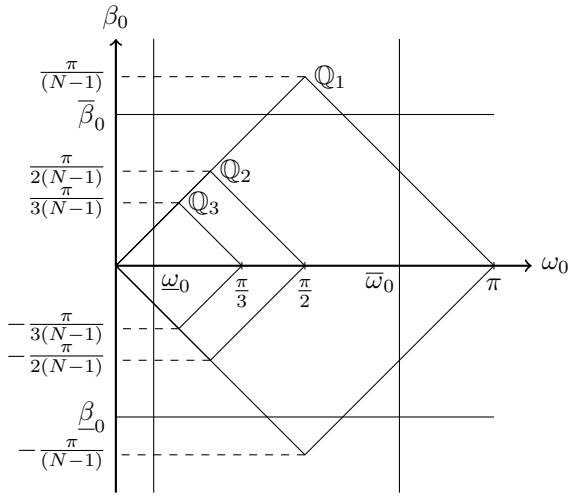

Figure 1. Example of the feasible set $\mathbb{Q}_l$ for model orders $l = 1, 2, 3$ and indication of possible additional bounds $\underline{\omega}_0, \overline{\omega}_0, \underline{\beta}_0, \overline{\beta}_0$.

Since the model order is unknown, it is necessary to use model order selection. For many standard model order selection methods, as outlined in e.g. [18], this implies that we need to compute a solution for all model orders up to a certain maximum model order $l = 1, \ldots, L$ selected such that $\ell \leq L$. Thus, we are interested in solving *all* the problems

$$\hat{\omega}_{0,l}, \hat{\beta}_{0,l} = \underset{[\omega_0,\beta_0]\in\mathbb{Q}_l}{\operatorname{argmax}}\; J_l(\omega_0, \beta_0), \qquad l = 1, \ldots, L\,. \quad (19)$$

Then many common model order selection methods will estimate the model as

$$\hat{\ell} = \underset{l=1,\ldots,L}{\operatorname{argmin}}\; \Upsilon(\hat{\omega}_{0,l}, \hat{\beta}_{0,l}) + \Phi(l) \quad (20)$$

Algorithm 1 The straightforward (naïve) algorithm for computing the NLS objective J_l for all model orders up to L, chirp rates $\mathbb{B}_l$ and fundamental frequencies $\Omega_l(\beta_0)$. One approach will be to have a uniform and equal resolution for all candidate model orders and then select $|\mathbb{B}|_l = K_l = \mathcal{O}(K_1/l) = \mathcal{O}(K/l)$ and $|\Omega_l(\beta_0)| = \mathcal{O}(F_l) = \mathcal{O}(F_1/l) = \mathcal{O}(F/l)$

```
for l ∈ {1, 2, 3, ..., L} do
 for β_0 ∈ B_l do
  for ω_0 ∈ Ω_l(β_0) do
   Form Z_l(ω_0, β_0)                                  O(lN)
   Form w_l(ω_0, β_0) = Z_l(ω_0, β_0)^T x              O(lN)
   Form A_l(ω_0, β_0) = Z_l(ω_0, β_0)^T Z_l(ω_0, β_0)  O(l^2 N)
   Solve A_l(ω_0, β_0) α_l(ω_0, β_0) = w_l(ω_0, β_0)   O(l^3)
   Calc. J_l(ω_0, β_0) = w_l^T(ω_0, β_0) α_l(ω_0, β_0) O(l)
  end for
 end for
end for
```

where $\Upsilon(\hat{\omega}_{0,l}, \hat{\beta}_{0,l})$ is a fitting measure and $\Phi(l)$ is model complexity measure. The fitting measure is typically the negative log-likelihood, which depends on J_l via

$$\Upsilon(\hat{\omega}_{0,l}, \hat{\beta}_{0,l}) = -N \ln\Big[\boldsymbol{x}^T\boldsymbol{x} - J_l(\hat{\omega}_{0,l}, \hat{\beta}_{0,l})\Big]/2 + \text{const.}\,, \quad (21)$$

whereas the model complexity measure can be, e.g., AIC, MDL [10], or MAP [11]. Solving the problem in (20) is easy if we know $J_l(\hat{\omega}_{0,l}, \hat{\beta}_{0,l})$ for all candidate model orders since L is often small. Computing $J_l(\hat{\omega}_{0,l}, \hat{\beta}_{0,l})$, however, is difficult since the problem in (19) is non-linear with an oscillating multi-modal objective function. The focus in this paper is, therefore, to solve the problem in (19) reliably and efficiently for all candidate model orders.

A reliable approach is to use the grid method [10] where we evaluate the objective function at a number of points for each model order $l = 1, \ldots, L$. Let $\mathbb{G}_l \subset \mathbb{Q}_l$ be the points on the grid where $\Omega_l(\beta_0)$ are the points on the frequency grid and $\mathbb{B}_l$ are points on the rate grid such that $\mathbb{G}_l = \{[\omega_0, \beta_0]\,|\,\beta_0 \in \mathbb{B}_l, \omega_0 \in \Omega_l(\beta_0)\}$. Then a straightforward approach to evaluate the NLS objective function on a grid is given in Algorithm 1.

Algorithm 1 was also considered in [10], [11], and found to have a high computational burden. This is also easy to see if we calculate the computational complexity of Algorithm 1. Let $K_l = |\mathbb{B}_l|$ be the number of rate points and let $|\Omega_l(\beta_0)| = \mathcal{O}(F_l)$ be the number of frequency points[1]. The computational complexity of Algorithm 1 is then

$$\mathcal{O}\left(\sum_{l=1}^{L} K_l F_l (l^2 N + l^3)\right)\,. \quad (22)$$

To shed a bit more light on the computational complexity above, consider again Fig. 1. If we apply a uniform frequency and rate grid with the same resolution for all candidate model orders $l = 1, \ldots, L$, then $K_l = K_1/l$ and $F_l = F_1/l$ since the feasible set $\mathbb{Q}_l$ shrinks as l increases. Define $K = K_1$ and

[1]here we have neglected the dependency of β_0 but in practice it is often small, e.g., from Fig. 1 we observe that the feasible set $\mathbb{Q}_l$ takes a constant $\frac{1}{2}$ of the corresponding square area and the big-O notation is still correct.

$F = 2F_1$[2], then the computational complexity of Algorithm 1 is

$$\mathcal{O}\left(\sum_{l=1}^{L} K_l F_l (l^2 N + l^3)\right) = \mathcal{O}\left(KF \sum_{l=1}^{L} \frac{1}{l^2}(l^2 N + l^3)\right) = \mathcal{O}\left(KF(LN + L^2)\right) . \quad (23)$$

Often K and F depend on L and/or N. Using the setting $F = \mathcal{O}(N^{3/2})$ and $K = \mathcal{O}(N^{5/2})$ as recommended in [10] we have an algorithm with a computational complexity of $\mathcal{O}(LN^5 + L^2N^4)$. The parameter N is typically big since we can use long(er) segments for the harmonic chirp model without (to some extend) violating the model assumptions compared to a standard harmonic model with $\beta_0 = 0$ [11]. We therefore observe that the standard method for using the NLS estimator is very costly which motivated faster but suboptimal methods [10].

The following three sections show how to improve upon the standard algorithm, reducing the computational complexity from $\mathcal{O}(LN^5 + L^2N^4)$ to $\mathcal{O}(N^2L^2\log(N))$. In Sec. III, we show the impact of the time index n_0 on the objective function and the CRLB, and Sec. IV shows an appropriate selection of the grid size when the time index is selected symmetrically $n_0 = -\frac{N-1}{2}$. In Sec V we show how to exploit the matrix structures to efficiently evaluate the objective function.

III. Consequences of Symmetric Time Index

In this section, we show two important consequences of the use of a symmetric time index $n_0 = -\frac{N-1}{2}$: *1)* the Hessian at the solution will be approximately diagonal and *2)* this will minimize the CRLB for the fundamental frequency. These two consequences are presented in the following two subsections.

A. The Hessian for the Harmonic Chirp Model

The Hessian of $J_l(\omega_0, \beta_0)$ at the global maximum is

$$\boldsymbol{\Psi}(\hat{\boldsymbol{\xi}}) = \frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} J_l(\hat{\omega}_0, \hat{\beta}_0) = \frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T} \boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{Z}}(\boldsymbol{\xi}) \boldsymbol{x} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} . \quad (24)$$

where $\boldsymbol{\xi} = \begin{bmatrix}\omega_0 & \beta_0\end{bmatrix}^T$ and

$$\boldsymbol{P}_{\boldsymbol{Z}}(\boldsymbol{\xi}) = \boldsymbol{Z}_l(\boldsymbol{\xi}) \left[\boldsymbol{Z}_l^T(\boldsymbol{\xi}) \boldsymbol{Z}_l(\boldsymbol{\xi})\right]^{-1} \boldsymbol{Z}_l^T(\boldsymbol{\xi}) . \quad (25)$$

To the best of our knowledge, an analytical expression for the Hessian does not exist. As we have detailed in App. A, however, an approximate analytical expression can be derived provided that: *1)* the data size is large enough to satisfy that the columns of $\boldsymbol{Z}_l(\hat{\boldsymbol{\xi}})$ are approximately orthogonal, and *2)* the SNR is large enough so that $\boldsymbol{x} \approx \boldsymbol{Z}_l(\hat{\boldsymbol{\xi}})\hat{\boldsymbol{\alpha}}_l$. Under these two approximations, we have shown in App. A that the approximate Hessian is given by

$$\boldsymbol{\Psi}(\hat{\omega}_0, \hat{\beta}_0) \approx -\frac{N(N^2-1)\sum_{i=1}^{l} A_i^2 i^2}{12} \cdot \begin{bmatrix} 1 & n_0 + \frac{N-1}{2} \\ n_0 + \frac{N-1}{2} & (N^2-4)/60 + \left(n_0 + \frac{N-1}{2}\right)^2 \end{bmatrix} \quad (26)$$

[2]the factor 2 is for convenience, as we will later use F as the FFT length, but will have no influence in the derived computational complexity.

where

$$A_i = \sqrt{a_i^2 + b_i^2} . \quad (27)$$

From (26), we see that the approximate Hessian of the objective function at the solution is diagonal for a symmetric time index, i.e., for $n_0 = -\frac{N-1}{2}$. If n_0 is set to the traditional choice of 0 as in [10], it is clear that the fundamental frequency and chirp rate are coupled. To illustrate the importance of this observation, we show two objective functions evaluated for the same data $\boldsymbol{x}$ using the symmetric time index $n_0 = -\frac{(N-1)}{2}$ in Fig. 2 and the non-symmetric choice $n_0 = 0$ in Fig. 3. For the symmetric case, the curvature of the objective function is approximately aligned with the standard axes which is not the case for the curvature of the non-symmetric case. We observe that, at least visually, the estimated chirp rate parameter at the maximum $\hat{\beta}_0$ will be the same. However, the fundamental frequency will not be the same but is related via $\hat{\omega}_0' \approx \hat{\omega}_0 - \hat{\beta}_0 \frac{N-1}{2}$ where $\hat{\omega}_0$ is the fundamental frequency estimate for a symmetric time index and $\hat{\omega}_0'$ is the fundamental frequency estimate for the time index $n_0 = 0$. Notice also that restricting the instantaneous frequency to be positive and having a Nyquist sampled signal give different shapes of the feasible region.

In [11], the estimator works by alternating optimization over ω_0 and β_0 (independently) starting at $\beta_0 = 0$. Such an alternating approach will work if $|\beta_0|$ is not too large but the method may get trapped in a local minima/maxima on the example presented in Fig. 2 and 3. An alternating method may be a reasonable approach in a speech application but not for a general purpose estimator.

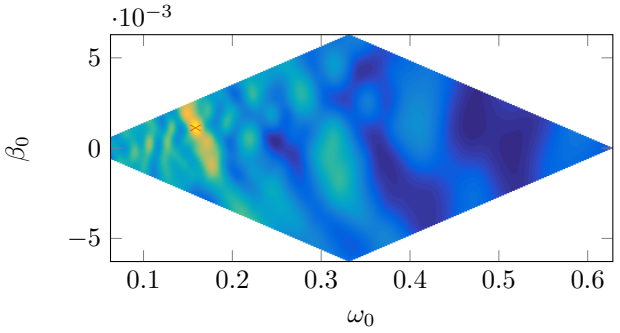

Figure 2. Example of an objective function with $N = 100$, true and candidate model order $l = \ell = 5$, symmetric time index $n_0 = -\frac{N-1}{2}$, noise free. The marker × indicates the maximum on the grid.

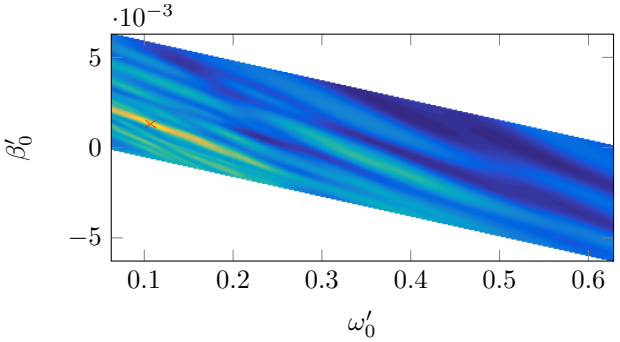

Figure 3. Example of an objective function with $N = 100$, true and candidate model order $l = \ell = 5$, non-symmetric time index $n_0 = 0$, noise free. The marker × indicates the maximum on the grid.

B. The CRLB for the Harmonic Chirp Model

The parameter vector we wish to derive the CRLB for is

$$\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\alpha}_l^T & \boldsymbol{\xi}^T\end{bmatrix}^T = \begin{bmatrix}\boldsymbol{\alpha}_l^T & \omega_0 & \beta_0\end{bmatrix}^T . \tag{28}$$

Focusing on the parameters ω_0, β_0 and using the results in Appendix B, the asymptotic inverse Fisher information matrix (FIM) is then

$$\boldsymbol{\mathcal{I}}^{-1}_{\mathrm{a},n_0}(\omega_0,\beta_0) = \frac{24\sigma^2}{N(N^2-1)(N^2-4)\sum_{i=1}^l A_i^2 i^2} \cdot \begin{bmatrix} (N^2-4)+60\left(n_0+\frac{N-1}{2}\right)^2 & -60\left(n_0+\frac{N-1}{2}\right) \\ -60\left(n_0+\frac{N-1}{2}\right) & 60 \end{bmatrix} . \tag{29}$$

Thus, the asymptotic CRLBs for any unbiased estimator of the fundamental frequency and the chirp rate are

$$\mathrm{Var}_{n_0}(\hat{\omega}_0) \geq \frac{24\sigma^2}{N(N^2-1)(N^2-4)\sum_{i=1}^l A_i^2 i^2} \cdot \left(N^2-4+60\left(n_0+\frac{N-1}{2}\right)^2\right) \tag{30}$$

$$\mathrm{Var}_{n_0}(\hat{\beta}_0) \geq \frac{1440\sigma^2}{N(N^2-1)(N^2-4)\sum_{i=1}^l A_i^2 i^2} . \tag{31}$$

It is interesting to note that with $n_0 = -\frac{N-1}{2}$, the asymptotic inverse FIM $\boldsymbol{\mathcal{I}}^{-1}_{\mathrm{a},-\frac{N-1}{2}}(\omega_0,\beta_0)$ is diagonal and minimizes the CRLB for the fundamental frequency. In fact, the asymptotic ratio of $\mathrm{Var}_{n_0}(\hat{\omega}_0)$ for the selection $n_0 = 0$ versus $n_0 = -\frac{N-1}{2}$ is

$$\lim_{N\to\infty} \frac{\mathrm{Var}_0(\hat{\omega}_0)}{\mathrm{Var}_{-\frac{N-1}{2}}(\hat{\omega}_0)} = \lim_{N\to\infty} \frac{N^2-4+15(N-1)^2}{N^2-4} = 16 . \tag{32}$$

Thus, it is possible to estimate the fundamental frequency ω_0 much more accurately by using a symmetric time index. Notice that the choice of a symmetric time index is only conceptual, will not introduce a delay in a signal processing system, also work for an even N, and that the asymptotic CRLB for the chirp rate does not depend on the start index n_0.

The argument for using a symmetric time index for complex-valued data and $\ell = 1$ (non-harmonic model) was given in [19] (see also the comment [25]). It is difficult to come with an intuitive or physical explanation of this phenomenon as also noted in [19]. Interestingly, the selection of symmetric time index also leads to minimum quantization distortion for certain sinusoidal coding schemes [26]. However, we will in the following shed some additional light on the choice of time index by considering linear transformation of the parameters of the harmonic linear chirp model.

Consider the linear transform from parameters $\boldsymbol{\xi} = \begin{bmatrix}\omega_0 & \beta_0\end{bmatrix}^T$ at the center $\boldsymbol{\xi}$ to the start point of the segment

$$\boldsymbol{\xi}' = \begin{bmatrix}\omega_0' \\ \beta_0'\end{bmatrix} = \begin{bmatrix}1 & -\frac{N-1}{2} \\ 0 & 1\end{bmatrix}\begin{bmatrix}\omega_0 \\ \beta_0\end{bmatrix} = \boldsymbol{g}(\boldsymbol{\xi}) . \tag{33}$$

Estimator efficiency is preserved over linear transforms [27]. However, the inverse FIM is different. For simplicity, consider the asymptotic inverse FIM instead of the exact FIM, but use the linear transformation rule for the inverse FIM [27, Eq. (3.30)]

$$\boldsymbol{\mathcal{I}}'^{-1}_{\mathrm{a}}(\boldsymbol{\xi}') = \frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \boldsymbol{\mathcal{I}}^{-1}_{\mathrm{a},-\frac{N-1}{2}}(\boldsymbol{\xi}) \frac{\partial \boldsymbol{g}(\boldsymbol{\xi})^T}{\partial \boldsymbol{\xi}} . \tag{34}$$

For the considered linear transform we have the Jacobian

$$\frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \begin{bmatrix}1 & -\frac{N-1}{2} \\ 0 & 1\end{bmatrix} \tag{35}$$

and then

$$\boldsymbol{\mathcal{I}}'^{-1}_{\mathrm{a}}(\boldsymbol{\xi}') = \boldsymbol{\mathcal{I}}^{-1}_{\mathrm{a},0}(\boldsymbol{\xi}) . \tag{36}$$

That is the same asymptotic inverse FIM in (29) with $n_0 = 0$. A similar derivation is possible with the reverse/inverse linear transformation

$$\begin{bmatrix}1 & -\frac{N-1}{2} \\ 0 & 1\end{bmatrix}^{-1} = \begin{bmatrix}1 & \frac{N-1}{2} \\ 0 & 1\end{bmatrix} . \tag{37}$$

The key reason for this effect is that if $n_0 \neq -(N-1)/2$, then the model dictates there will be dependency between the estimates of the parameters ω_0 and β_0 and this dependency can be removed via a linear transformation. We remark that if the estimator is not efficient, then it is not clear if the linear transform (33) is the best choice.

IV. Grid Resolution

In order to keep the computationally burden of the grid method for joint fundamental frequency and chirp rate as low as possible, we will establish how finely the objective function should be sampled in these parameters to approximately locate the global maximum of the objective function. It is then possible to rely on a two step method where *1)* the region of the global maximum is located using a grid method providing a "rough"/approximate solution *2)* the "rough"/approximate solution is then refined using an iterative algorithm. Another approach is to select the grid size fine enough to achieve the CRLB as in [10]—however, the CRLB relies on unknown parameters. We will in the simulation Sec. VII show that a two step approach can also achieve the CRLB and is computationally much more efficient.

We use the approach in [28] to identify an appropriate grid size and start with a second order Taylor approximation of a function $f : \mathbb{R} \mapsto \mathbb{R}$ around a local maximum with no active constraints $\hat{x}$ such that the derivative is $f'(\hat{x}) = 0$. The objective function $f(x)$ has approximately decreased by a factor of g from the value $f(\hat{x})$ when x is given by

$$x = \hat{x} \pm \sqrt{2\frac{1-g}{g}\frac{f(\hat{x})}{f''(\hat{x})}}, \quad f''(x) = \frac{\partial^2 f(x)}{\partial x^2} . \tag{38}$$

For the problem of joint fundamental frequency and chirp rate parameter estimation with $\boldsymbol{\xi} = \begin{bmatrix}\omega_0 & \beta_0\end{bmatrix}^T$, the objective function at the solution can be approximated using the asymptotic

result in (104) and the high-SNR approximation $\boldsymbol{x} \approx \boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})\boldsymbol{\alpha}_\ell$ as

$$J_\ell(\hat{\boldsymbol{\xi}}) = \boldsymbol{x}^T \boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}}) \left[\boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})^T \boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})\right]^{-1} \boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})^T \boldsymbol{x} \tag{39}$$

$$\approx \frac{2}{N}\|\boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})^T \boldsymbol{x}\|_2^2 \tag{40}$$

$$\approx \frac{2}{N}\|\boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})^T \boldsymbol{Z}_\ell(\hat{\boldsymbol{\xi}})\boldsymbol{\alpha}\|_2^2 \tag{41}$$

$$\approx \frac{N}{2}\|\boldsymbol{\alpha}\|_2^2 \tag{42}$$

$$\approx \frac{N}{2}\sum_{l=1}^{\ell} A_l^2, \quad A_l = \sqrt{a_l^2 + b_l^2}\,. \tag{43}$$

The Hessian is approximately given by (26). Since the approximate Hessian (26) is diagonal, we can simply apply (38) twice. Notice that this approach would be difficult to apply if the time index was not symmetric since the approximate Hessian would not be diagonal. For fixed signal energy with $f(\hat{\omega}_0) = J_\ell(\hat{\omega}_0, \hat{\beta}_0)$ or $f(\hat{\beta}_0) = J_\ell(\hat{\omega}_0, \hat{\beta}_0)$, the ratio f/f'' in (38) will be the smallest with the selection $A_l = 0$, $l = 1, \cdots, \ell - 1$ and $A_\ell = A$. We then obtain ($g \geq 1$):

$$\omega_0 \approx \hat{\omega}_0 \pm \sqrt{12\frac{g-1}{g}}\frac{1}{N\ell} \tag{44}$$

$$\beta_0 \approx \hat{\beta}_0 \pm \sqrt{12\frac{g-1}{g}}\frac{\sqrt{60}}{N^2\ell}\,. \tag{45}$$

If we set $g = 1.15$, we obtain the grid resolution for the chirp rate and the fundamental frequency as

$$\Delta_{\omega_0} = \sqrt{12\frac{g-1}{g}}\frac{1}{N\ell}\Big|_{g=1.15} \approx \frac{1.25}{N\ell} \tag{46}$$

$$\Delta_{\beta_0} = \sqrt{12\frac{g-1}{g}}\frac{\sqrt{60}}{N^2\ell}\Big|_{g=1.15} \approx \frac{9.69}{N^2\ell}\,. \tag{47}$$

Since ℓ is unknown we use the upper bound L for all candidate model orders and select the resolutions

$$\Delta_{\omega_0} = \frac{c_1}{NL} \tag{48}$$

$$\Delta_{\beta_0} = \frac{c_2}{N^2 L} \tag{49}$$

with constants c_1, c_2. On the other hand, following the constraints outlined in Fig. 1 implies that we are using

$$F = 2\pi/\frac{c_1}{NL} = \mathcal{O}(NL) \tag{50}$$

$$K = \frac{2\pi}{(N-1)}/\frac{c_2}{N^2 L} = \mathcal{O}(NL) \tag{51}$$

grid points. The selections (50)–(51) should be compared with the settings $F = \mathcal{O}(N^{3/2})$ and $K = \mathcal{O}(N^{5/2})$ suggested in [10]. Notice that the constants implied by (46)–(47) are $c_1 = 1.25$ and $c_2 = 9.69$ yields

$$F = 2\pi/c_1 NL \approx 5NL \tag{52}$$

$$K = 2\pi/c_2 NL \approx 0.65NL\,. \tag{53}$$

For the harmonic model the selection $F = 5NL$ was also discussed and justified in [28], [29]. In Sec. VII we will empirically assess a selection of the constant K with fixed $F = 5NL$ and show that $K = NL$ is sufficient to achieve the CRLB.

V. A FAST ALGORITHM FOR THE NLS ESTIMATOR

In the following we give a fast algorithm for evaluating the NLS objective on the grid $\mathbb{G}_l \,\forall\, l = 1, \ldots, L$. Specifically, we will assume that the grids are *nested* and *uniform*. A nested grid implies that $\mathbb{G}_L \subseteq \mathbb{G}_{L-1} \cdots \subseteq \mathbb{G}_1$. A nested and uniform grid then implies that if $(\omega_0, \beta_0) \in \mathbb{G}_l$ then $(l\omega_0, l\beta_0) \in \mathbb{G}_1$. The nested and uniform setting bring three key possibilities that we will exploit:

1) A uniform frequency grid $\Omega_l(\beta_0) \subseteq \{2\pi\frac{f-1}{F}\}_{f=2}^{F/2}$ ($f = 1$ is removed since this corresponds to DC) allows for computing necessary quantities efficiently via FFTs.
2) A uniform frequency and rate grid will allow us to obtain necessary quantities for model order $l = 2, \ldots, L$ based only from computed quantities from model order $l = 1$.
3) A nested grid will allow us to recursively update necessary quantities from model order $l - 1$ to l.

A key property we will exploit in the following is that the vectors $\boldsymbol{c}(\omega, \beta)$ and $\boldsymbol{s}(\omega, \beta)$ defined in (9)–(10) can be split into two vector functions

$$\boldsymbol{c}(\omega, \beta) + j\boldsymbol{s}(\omega, \beta) = \boldsymbol{z}(\omega) \odot \boldsymbol{g}(\beta) \tag{54}$$

where

$$\boldsymbol{z}(\omega) = \begin{bmatrix}\exp(j\omega n_0) & \cdots & \exp(j\omega(n_0 + N - 1))\end{bmatrix}^T \in \mathbb{C}^N \tag{55}$$

$$\boldsymbol{g}(\beta) = \begin{bmatrix}\exp(j\frac{1}{2}\beta n_0^2) & \cdots & \exp(j\frac{1}{2}\beta(n_0 + N - 1)^2)\end{bmatrix}^T \in \mathbb{C}^N. \tag{56}$$

Notice that $\boldsymbol{z}(\omega)$ can be time shifted to time index 0 using

$$\boldsymbol{z}(\omega) = \exp\left(-j\omega_0 n_0\right)\boldsymbol{z}_0(\omega)\,. \tag{57}$$

This form is of interest since $\boldsymbol{z}_0(\omega)^H \boldsymbol{x}$ is a discrete time Fourier transform (DTFT) in standard form with start index 0 at a single frequency ω of a signal $\boldsymbol{x}$.

The following 3 subsections will give fast methods for handling the following steps of Algorithm 1:

- Sec. V-C: form $\boldsymbol{w}_l(\omega_0, \beta_0) = \boldsymbol{Z}_l(\omega_0, \beta_0)^T \boldsymbol{x}$
- Sec. V-B: form $\boldsymbol{A}_l(\omega_0, \beta_0) = \boldsymbol{Z}_l(\omega_0, \beta_0)^T \boldsymbol{Z}_l(\omega_0, \beta_0)$
- Sec. V-A: solve $\boldsymbol{A}_l(\omega_0, \beta_0)\boldsymbol{\alpha}_l(\omega_0, \beta_0) = \boldsymbol{w}_l(\omega_0, \beta_0)$.

A. Solving the linear system

One key problem in Algorithm 1 is to efficiently compute the solution of the linear system

$$\boldsymbol{A}_l(\omega_0, \beta_0)\boldsymbol{\alpha}_l(\omega_0, \beta_0) = \boldsymbol{w}_l(\omega_0, \beta_0)\,. \tag{58}$$

Firstly, the coefficient matrix $\boldsymbol{A}_l(\omega_0, \beta_0) = \boldsymbol{Z}_l^T(\omega_0, \beta_0)\boldsymbol{Z}_l(\omega_0, \beta_0)$ consists of four terms given by

$$\boldsymbol{c}^T(i\omega_0, i\beta_0)\boldsymbol{c}(k\omega_0, k\beta_0) = t_{i-k}(\omega_0, \beta_0) + h_{i+k}(\omega_0, \beta_0)\ , \tag{59}$$

$$\boldsymbol{s}^T(i\omega_0, i\beta_0)\boldsymbol{s}(k\omega_0, k\beta_0) = t_{i-k}(\omega_0, \beta_0) - h_{i+k}(\omega_0, \beta_0) \tag{60}$$

$$\boldsymbol{s}^T(i\omega_0, i\beta_0)\boldsymbol{c}(k\omega_0, k\beta_0) = \tilde{t}_{i-k}(\omega_0, \beta_0) + \tilde{h}_{i+k}(\omega_0, \beta_0)\ , \tag{61}$$

$$\boldsymbol{c}^T(i\omega_0, i\beta_0)\boldsymbol{s}(k\omega_0, k\beta_0) = -\tilde{t}_{i-k}(\omega_0, \beta_0) + \tilde{h}_{i+k}(\omega_0, \beta_0)\ , \tag{62}$$

for $i,k \in \{1,2,\ldots,l\}$ where

$$t_{i-k}(\omega_0,\beta_0) = \frac{1}{2}\sum_{n=n_0}^{n_0+N-1} \cos\left[(i-k)\omega_0 n + (i-k)\beta_0 n^2/2\right] , \tag{63}$$

$$h_{i+k}(\omega_0,\beta_0) = \frac{1}{2}\sum_{n=n_0}^{n_0+N-1} \cos\left[(i+k)\omega_0 n + (i+k)\beta_0 n^2/2\right] , \tag{64}$$

$$\tilde{t}_{i-k}(\omega_0,\beta_0) = \frac{1}{2}\sum_{n=n_0}^{n_0+N-1} \sin\left[(i-k)\omega_0 n + (i-k)\beta_0 n^2/2\right] \tag{65}$$

$$\tilde{h}_{i+k}(\omega_0,\beta_0) = \frac{1}{2}\sum_{n=n_0}^{n_0+N-1} \sin\left[(i+k)\omega_0 n + (i+k)\beta_0 n^2/2\right] . \tag{66}$$

Note also the symmetries $t_{i-k}(\omega_0,\beta_0) = t_{k-i}(\omega_0,\beta_0)$ and $\tilde{t}_{i-k}(\omega_0,\beta_0) = -\tilde{t}_{k-i}(\omega_0,\beta_0)$. Using these definitions, the matrix $\boldsymbol{A}_l(\omega_0,\beta_0)$ can be partitioned as

$$\boldsymbol{A}_l(\omega_0,\beta_0) = \begin{bmatrix} \boldsymbol{T}_l(\omega_0,\beta_0) & \tilde{\boldsymbol{T}}_l^T(\omega_0,\beta_0) \\ \tilde{\boldsymbol{T}}_l(\omega_0,\beta_0) & \boldsymbol{T}_l(\omega_0,\beta_0) \end{bmatrix} + \begin{bmatrix} \boldsymbol{H}_l(\omega_0,\beta_0) & \tilde{\boldsymbol{H}}_l^T(\omega_0,\beta_0) \\ \tilde{\boldsymbol{H}}_l(\omega_0,\beta_0) & -\boldsymbol{H}_l(\omega_0,\beta_0) \end{bmatrix} \tag{67}$$

where

$$\boldsymbol{T}_l(\omega,\beta) = \begin{bmatrix} t_0(\omega,\beta) & t_1(\omega,\beta) & \cdots & t_{l-1}(\omega,\beta) \\ t_1(\omega,\beta) & t_0(\omega,\beta) & \cdots & t_{l-2}(\omega,\beta) \\ \vdots & \vdots & \ddots & \vdots \\ t_{l-1}(\omega,\beta) & t_{l-2}(\omega,\beta) & \cdots & t_0(\omega,\beta) \end{bmatrix} \tag{68}$$

$$\tilde{\boldsymbol{T}}_l(\omega,\beta) = \begin{bmatrix} \tilde{t}_0(\omega,\beta) & -\tilde{t}_1(\omega,\beta) & \cdots & -\tilde{t}_{l-1}(\omega,\beta) \\ \tilde{t}_1(\omega,\beta) & \tilde{t}_0(\omega,\beta) & \cdots & -\tilde{t}_{l-2}(\omega,\beta) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{t}_{l-1}(\omega,\beta) & \tilde{t}_{l-2}(\omega,\beta) & \cdots & \tilde{t}_0(\omega,\beta) \end{bmatrix} \tag{69}$$

$$\boldsymbol{H}_l(\omega,\beta) = \begin{bmatrix} h_2(\omega,\beta) & h_3(\omega,\beta) & \cdots & h_{l+1}(\omega,\beta) \\ h_3(\omega,\beta) & h_4(\omega,\beta) & \iddots & h_{l+2}(\omega,\beta) \\ \vdots & \iddots & \iddots & \vdots \\ h_{l+1}(\omega,\beta) & h_{l+2}(\omega,\beta) & \cdots & h_{2l}(\omega,\beta) \end{bmatrix} \tag{70}$$

$$\tilde{\boldsymbol{H}}_l(\omega,\beta) = \begin{bmatrix} \tilde{h}_2(\omega,\beta) & \tilde{h}_3(\omega,\beta) & \cdots & \tilde{h}_{l+1}(\omega,\beta) \\ \tilde{h}_3(\omega,\beta) & \tilde{h}_4(\omega,\beta) & \iddots & \tilde{h}_{l+2}(\omega,\beta) \\ \vdots & \iddots & \iddots & \vdots \\ \tilde{h}_{l+1}(\omega,\beta) & \tilde{h}_{l+2}(\omega,\beta) & \cdots & \tilde{h}_{2l}(\omega,\beta) \end{bmatrix} \tag{71}$$

and $\boldsymbol{T}_l(\omega_0,\beta_0)$, $\tilde{\boldsymbol{T}}_l(\omega_0,\beta_0)$ are Toeplitz matrices whereas $\boldsymbol{H}_l(\omega_0,\beta_0)$, $\tilde{\boldsymbol{H}}_l(\omega_0,\beta_0)$ are Hankel matrices. When $\beta_0 = 0$ and selecting $n_0 = -\frac{N-1}{2}$, then $\tilde{\boldsymbol{T}}_l(\omega_0,\beta_0) = \tilde{\boldsymbol{H}}_l(\omega_0,\beta_0) = \boldsymbol{0}$ because $\sin$ is an odd function and each term symmetric around zero cancels out in (65)–(66). Hence, we obtain two separate Toeplitz-plus-Hankel systems (see [29]) which can be solved using, e.g., the algorithm suggested in [30]. However, when $\beta_0 \neq 0$, this is no longer the case. We will instead handle the matrix structure in (67) by extending the recursive algorithm [30] to block Toeplitz-plus-Hankel in Appendix C.

We can now exploit two important structures of the problem: *1)* the linear system can be solved recursively *2)* the grids are nested $\mathbb{G}_l \subseteq \mathbb{G}_{l-1}$. These observations imply that we can obtain the solutions $\boldsymbol{\alpha}_l(\omega_0,\beta_0)\,\forall\,(\omega_0,\beta_0) \in \mathbb{G}_l$ by performing a recursive update from a subset of the "previous" solutions $\boldsymbol{\alpha}_{l-1}(\omega_0,\beta_0)\,\forall\,(\omega_0,\beta_0) \in \mathbb{G}_l \subseteq \mathbb{G}_{l-1}$. In Appendix C it is shown that one recursive update of a solution from order $l-1$ to l has a computational complexity of $\mathcal{O}(l)$ such that it is possible to compute $\boldsymbol{\alpha}_l(\omega_0,\beta_0)\,\forall\,(\omega_0,\beta_0) \in \mathbb{G}_l$ with the computational complexity $\mathcal{O}(l|\mathbb{G}_l|)$.

It is well known that when ω_0 is small then $\boldsymbol{A}_l(\omega_0,\beta_0)$ becomes ill-conditioned, see e.g. [20]. In-particular, if the frequency is $\omega_0 = 2\pi/N$ and $\beta_0 = 0$, then there will be one period in the observed signal and the condition number is $\kappa(\boldsymbol{A}_l(2\pi/N,0)) = 1$ since all the columns of $\boldsymbol{Z}_l(2\pi/N,0)$ are orthogonal. However, if there is a half period in the observed signal, then for $N = 100, \kappa(\boldsymbol{A}_6(\pi/N,0) \approx 2.5\cdot 10^6$. To this end, we add regularization in the implementations with the modified system $\boldsymbol{A}_l(\omega_0,\beta_0) + \epsilon\frac{N}{2}I_{2l}$ using $\epsilon = 10^{-8}$. Notice this can be handled by considering the modified diagonal $t_0(\omega_0,\beta_0) + \epsilon\frac{N}{2}$ in (68). Regularization can be seen as an intermediate solution between the NLS method ($\epsilon = 0$) and the harmonic chirp summation method $\epsilon = \infty$ (i.e. the harmochirp-gram in [10]) explained in Sec. VI.

B. Forming the coefficient matrix

The coefficients of the Toeplitz and Hankel matrices, e.g. (63) can also be computed efficiently. Using (54), we have

$$t_l(\omega_0,\beta_0) = \tfrac{1}{2}\sum_{n=n_0}^{n_0+N-1} \cos\left(l\omega_0 n + l\beta_0 n^2/2\right) \tag{72}$$

$$= \tfrac{1}{2}\boldsymbol{c}(l\omega_0, l\beta_0)^T\boldsymbol{1} \tag{73}$$

$$= \tfrac{1}{2}\Re\left(\boldsymbol{z}(l\omega_0)^H\boldsymbol{g}(-l\beta_0)\right) . \tag{74}$$

Notice that because the parameters ω_0 and β_0 only occurs on the form $l\omega_0$ and $l\beta_0$, then

$$t_l(\omega_0,\beta_0) = t_1(l\omega_0, l\beta_0) . \tag{75}$$

This implies that if $t_1(\omega_0,\beta_0)$ is computed on a uniform grid $(\omega_0,\beta_0) \in \mathbb{G}_1$, then $t_l(\omega_0,\beta_0)$ can be obtained by extraction from that grid using (75). The key is that only $t_1(\omega_0,\beta_0)$ needs to be computed for all $(\omega_0,\beta_0) \in \mathbb{G}_1$. Notice that Equation (74) is a DTFT (with start index n_0) at a single frequency of the chirp signal $\boldsymbol{g}(-l\beta_0)$. So for $l = 1$ and each $\beta_0 \in \mathbb{B}_1$, the values $t_1(\omega_0,\beta_0)$ can be computed for all $\omega_0 \in \Omega_l(\beta_0)$ using a single FFT of the signal $\boldsymbol{g}(\beta_0)$. Most definitions of the DFT and FFT assume a start index of 0, so for most FFTs the output would need to be phase-shifted following (57). A similar analysis is possible for $h_l(\omega_0,\beta_0), \tilde{t}_l(\omega_0,\beta_0)$ and $\tilde{h}_l(\omega_0,\beta_0)$. We also have the special case $l = 0$ with $t_0(\omega_0,\beta_0) = \frac{N}{2}$ and $\tilde{t}_0(\omega_0,\beta_0) = 0$.

C. Forming the right-hand side

The right-hand side of (58) $\boldsymbol{w}_l(\omega_0,\beta_0) = \boldsymbol{Z}_l^T(\omega_0,\beta_0)\boldsymbol{x}$ can also be updated recursively based on a simple precomputed quantity. We rewrite the right-hand side (see (5)) using (54) as

$$\boldsymbol{Z}_l^T(\omega_0,\beta_0)\boldsymbol{x} = \begin{bmatrix}\boldsymbol{C}_l^T(\omega_0,\beta_0)\boldsymbol{x}\\ \boldsymbol{S}_l^T(\omega_0,\beta_0)\boldsymbol{x}\end{bmatrix} \tag{76}$$

$$= \begin{bmatrix}\boldsymbol{C}_{l-1}^T(\omega_0,\beta_0)\boldsymbol{x}\\ \boldsymbol{c}^T(l\omega_0,l\beta_0)\boldsymbol{x}\\ \boldsymbol{S}_{l-1}^T(\omega_0,\beta_0)\boldsymbol{x}\\ \boldsymbol{s}^T(l\omega_0,l\beta_0)\boldsymbol{x}\end{bmatrix} \tag{77}$$

$$= \begin{bmatrix}\boldsymbol{C}_{l-1}^T(\omega_0,\beta_0)\boldsymbol{x}\\ \Re\left(\boldsymbol{z}^H(l\omega_0)(\boldsymbol{g}(-l\beta_0)\odot\boldsymbol{x})\right)\\ \boldsymbol{S}_{l-1}^T(\omega_0,\beta_0)\boldsymbol{x}\\ -\Im\left(\boldsymbol{z}^H(l\omega_0)(\boldsymbol{g}(-l\beta_0)\odot\boldsymbol{x})\right)\end{bmatrix}. \tag{78}$$

The interpretation of the above is that the right-hand side can be recursively updated for every l by first de-chirping the source signal $\boldsymbol{x}$ using $\boldsymbol{g}(-l\beta_0)$ followed by a DTFT at the frequency $l\omega_0$. Let

$$v(\omega_0,\beta_0) = \boldsymbol{z}^H(\omega_0)(\boldsymbol{g}(-\beta_0)\odot\boldsymbol{x}). \tag{79}$$

If $v(\omega_0,\beta_0)$ is computed on the grid $(\omega_0,\beta_0)\in\mathbb{G}_1$ then $v(l\omega_0,l\beta_0)$ is also on this grid (based on our initial assumption). Since the product with $\boldsymbol{z}(\omega_0)^H$ represents the DTFT at a single frequency, we may compute this product for all $\omega_0\in\Omega_l(\beta_0)$ using a single FFT (possible with a phase-shift following (57) depending on the definition). The computational complexity for computing $v(\omega_0,\beta_0)$ for all $(\omega_0,\beta_0)\in\mathbb{G}_1$ is then simply $K_1 = K$ FFTs of length F.

D. Computational Complexity Analysis

The proposed algorithm is presented in Algorithm 2. The linear system (67) is solved recursively in order using the recursive block Toeplitz-plus-Hankel solver outlined in Sec. C. In particular, the recursive block Toeplitz-plus-Hankel solver will have a linear computational complexity $\mathcal{O}(l)$ for updating the solution from order $l-1$ to l. Using $\Omega_l(\beta_0) = \mathcal{O}(F/l)$, the computational complexity of Algorithm 2 is then

$$\mathcal{O}\left(KF\log F + \sum_{l=1}^{L}(K/l)(F/l)(1/l+l)\right)$$
$$= \mathcal{O}(KF\log(F) + KF\log(L)) = \mathcal{O}(KF\log(F)) \tag{80}$$

where we used the inequality $\sum_{n=1}^{k}1/n < \log(k)+1$ for a harmonic series. Using the grid selection (50)–(51), Algorithm 2 then has the computational complexity $\mathcal{O}(N^2L^2\log(NL)) = \mathcal{O}(N^2L^2\log(N))$ since $N\geq 2L+3$ (number of observations larger than the number of unknown). This is orders of magnitude lower than the standard algorithm $\mathcal{O}(LN^5+L^2N^4)$ as proposed in [10] (with the notation in this paper and using the proposed feasible set). For comparison, the computational complexity of the harmonic separate-estimate method in [10] for computing estimates for all candidate model orders is $\mathcal{O}(\sum_{l=1}^{L}N^2l\log N) = \mathcal{O}(N^2L^2\log(N))$, i.e., the same computational complexity as the proposed algorithm for the NLS method.

Algorithm 2 Proposed algorithm for computing the NLS objective function $J_l(\omega_0,\beta_0)$ for $l=1,\ldots,L$, for uniform chirp rates $\mathbb{B}_l$ with $K_l = |\mathbb{B}_l| = \mathcal{O}(K/l)$ and fundamental frequencies $\Omega_l(\beta_0)$ with $F_l = |\Omega_l(\beta_0)| = \mathcal{O}(F/l)$ selected such that $\mathbb{G}_L\subseteq\mathbb{G}_{L-1}\cdots\subseteq\mathbb{G}_1$ and for all $(\omega_0,\beta_0)\in\mathbb{G}_l$ we have $(l\omega_0,l\beta_0)\in\mathbb{G}_1$.

for $\beta_0\in\mathbb{B}_1$ **do**
  Compute $v(\omega_0,\beta_0)\,\forall\omega_0\in\Omega_1(\beta_0)$ using (79) $\mathcal{O}(F\log F)$
  Compute $t_1(\omega_0,\beta_0),\tilde{t}_1(\omega_0,\beta_0),h_1(\omega_0,\beta_0),\tilde{h}_1(\omega_0,\beta_0)$ $\forall\omega_0\in\Omega_1(\beta_0)$ using e.g. (74) $\mathcal{O}(F\log F)$
end for
for $l\in\{1,2,3,\ldots,L\}$ **do**
  for $\beta_0\in\mathbb{B}_l$ **do**
    for $\omega_0\in\Omega_l(\beta_0)$ **do**
      Update right-hand size $\boldsymbol{w}_l(\omega_0,\beta_0)$ recursively from v using (78)–(79) $\mathcal{O}(1)$
      Update coefficient matrix $\boldsymbol{A}_l(\omega_0,\beta_0)$ recursively from $t_1,\tilde{t}_1,h_1,\tilde{h}_1$ using e.g. (75) and (68)–(71) $\mathcal{O}(1)$
      Solve $\boldsymbol{A}_l(\omega_0,\beta_0)\boldsymbol{\alpha}_l(\omega_0,\beta_0) = \boldsymbol{w}_l(\omega_0,\beta_0)$ recursively via the block-TH algorithm $\mathcal{O}(l)$
      Compute $J_l(\omega_0,\beta_0) = \boldsymbol{w}_l^T(\omega_0,\beta_0)\boldsymbol{\alpha}_l(\omega_0,\beta_0)$ $\mathcal{O}(l)$
    end for
  end for
end for

Notice here that the asymptotic computational complexity of forming the right-hand side and the entries in the coefficient matrix using FFTs in the initialization step is higher than solving the linear system of equations. For practical sizes of K,N,L,F and due to the large constant involved in the block-TH algorithm, this part of the algorithm may still take up a significant portion of required computations in Algorithm 2.

VI. Harmonic chirp summation

Similar to the harmonic summation algorithm [31], [32] for the estimation of the fundamental frequency for the standard harmonic model, we may form an approximate NLS estimator denoted harmonic chirp summation (HCS) (also called the harmochirp-gram [10]) using the approximation (104). In particular, if the data size N is large enough we may approximate $\boldsymbol{Z}(\omega_0,\beta_0)^T\boldsymbol{Z}(\omega_0,\beta_0)^T$ as diagonal. This estimator is then governed by the objective function

$$\hat{J}_l(\omega_0,\beta_0) = \boldsymbol{x}^T\boldsymbol{Z}_l(\omega_0,\beta_0)\boldsymbol{Z}_l(\omega_0,\beta_0)^T\boldsymbol{x} \tag{81}$$
$$= \|\boldsymbol{Z}_l(\omega_0,\beta_0)^T\boldsymbol{x}\|_2^2. \tag{82}$$

Notice this is also a scaled version of the approximation in (40). Using (5)–(7) and (54), the objective for model order $l=1$ can be written as

$$\hat{J}_1(\omega_0,\beta_0) = \|\begin{bmatrix}\boldsymbol{c}(\omega_0,\beta_0) & \boldsymbol{s}(\omega_0,\beta_0)\end{bmatrix}^T x\|_2^2 \tag{83}$$
$$= |\boldsymbol{c}(\omega_0,\beta_0) + j\boldsymbol{s}(\omega_0,\beta_0))^H x|^2 \tag{84}$$
$$= |(\boldsymbol{z}(\omega_0)\odot\boldsymbol{g}(\beta_0))^H x|^2. \tag{85}$$

The objective function for the harmonic chirp summation may then be rewritten using (5)–(7) and (83) (see also [33]):

$$\hat{J}_l(\omega_0,\beta_0) = \|\boldsymbol{Z}_l(\omega_0,\beta_0)^T\boldsymbol{x}\|_2^2 \tag{86}$$

$$= \left\| \begin{bmatrix} \boldsymbol{C}_l(\omega_0,\beta_0) & \boldsymbol{S}_l(\omega_0,\beta_0) \end{bmatrix}^T \boldsymbol{x} \right\|_2^2 \tag{87}$$

$$= \sum_{i=1}^{l} \| \begin{bmatrix} \boldsymbol{c}(i\omega_0,i\beta_0) & \boldsymbol{s}(i\omega_0,i\beta_0) \end{bmatrix}^T \boldsymbol{x}\|_2^2 \tag{88}$$

$$= \sum_{i=1}^{l} \hat{J}_1(i\omega_0,i\beta_0) \tag{89}$$

such that the objective function for candidate model order l can be computed from the objective function for model order 1. Further, it is possible to update the objective function from model order $l-1$ to l recursively

$$\hat{J}_l(\omega_0,\beta_0) = \sum_{i=1}^{l} \hat{J}_1(i\omega_0,i\beta_0) = \hat{J}_{l-1}(\omega_0,\beta_0) + \hat{J}_1(l\omega_0,l\beta_0)\,. \tag{90}$$

The computational complexity for computing $\hat{J}_1(\omega_0,\beta_0)$ using (85) for a uniform frequency and rate grid is $\mathcal{O}(KF\log F)$. Computing the objective function for the remaining model orders on a subgrid $\mathbb{Q}_l \subset \mathbb{Q}_1, l = 1,\ldots,L$ is then $\mathcal{O}(\sum_{l=2}^{L} K_l F_l) = \mathcal{O}(KF\sum_{l=2}^{L} 1/l^2) = \mathcal{O}(KF)$. The total computational complexity is then $\mathcal{O}(KF\log F)$, the same as the proposed fast NLS method (80). Using the suggested grid selection (50)–(51) ($F = \mathcal{O}(NL)$, $K = \mathcal{O}(NL)$), the computational complexity is $\mathcal{O}(N^2L^2\log(NL)) = \mathcal{O}(N^2L^2\log(N))$. From (85) we also observe that $\hat{J}_l(\omega_0,-\beta_0) = \hat{J}_l(-\omega_0,\beta_0)$ such that for symmetric rate grids, it is (almost) possible to obtain half of the objective function for free since the coefficients for negative frequencies are already computed by the FFT.

VII. Simulations

In this section, we evaluate MATLAB implementations of the presented algorithms in terms of accuracy and computational savings by Monte Carlo simulations and timing benchmarks[3]. In particular, we will investigate the behavior of the NLS method and empirically identify a usable value for K, with fixed $F = 5NL$ as suggested in [28], [29]. Notice that we are not proposing any new method but simply improving upon the computational complexity of the known NLS method so we also refer to [10], [11] for additional assessments.

A. Accuracy

We assess the accuracy of the exact NLS method and the HCS method using Monte Carlo simulations with $r = 2000$ repetitions of the model (2) using a symmetric time index and with white Gaussian noise such that the ML and NLS estimator coincide. We find the maximum on the grid and refine using the Nelder-Mead method [34] constrained into plus/minus one grid interval for both the pitch and rate parameter. For the refinement step both the NLS and HCS methods use the exact

[3]Code available via tinyurl.com/tljvbn

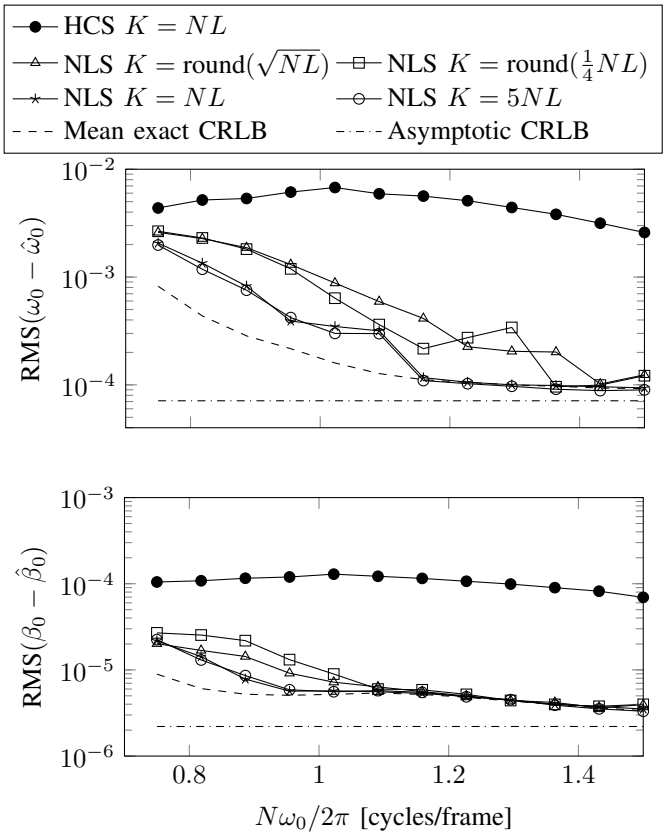

Figure 4. Estimation accuracy for different methods, the asymptotic CRLB and the mean exact CRLB for all the realizations. Varying ω. Settings: $N = 250$, $\ell = L = 6$, $\beta_0 \in [-3/N^2, 3/N^2)$ drawn uniformly, constant amplitude $\sqrt{a_l^2 + b_l^2} = 1$, $l = 1,\ldots,L$ and phase drawn from a uniform distribution in the interval $[0, 2\pi)$ for each Monte Carlo repetition. $F = 5NL$, SNR $= 10$ [dB]

NLS objective function. In the following, we compare the root-mean-squared (RMS) error result with the asymptotic CRLBs in (31)–(30). Further, we compute the exact CRLB for all realizations using the diagonal of the inverse of (112) and compute the mean exact CRLB for all the realizations, as in [35].

In the first experiment, we assess the performance at low fundamental frequencies. This is of particular interest since this is one motivation for using the real-valued data model. In particular, a mapping from real to complex-valued data will decrease estimation accuracy for low fundamental frequencies [20]. We select $N = 250$, $\ell = L = 6$, draw the rate β_0 uniformly from $[-3/N^2/12, 3/N^2)$ and vary the fundamental frequency ω_0 from $2\pi 0.75/N$ to $2\pi 1.25/N$. As shown in Fig. 4, for low fundamental frequencies, the HCS method is not accurate as also reported in [10]. We also observe that the methods cannot achieve the asymptotic CRLB but for frequencies higher than 1 cycles/frame the NLS methods can achieve the mean exact CRLB. Further, the selection $K = NL$ and $K = 5NL$ shows about the same performance, but the selection $K = \text{round}(\sqrt{NL})$ and $K = \text{round}(\frac{1}{4}NL)$ (round to nearest integer) does not provide as accurate estimates. From a computational perspective, we should then select $K = NL$.

For the next experiment we also select $N = 250$, $\ell = L = 6$ and $\beta_0 \in [-3/N^2/12, 3/N^2)$ but randomly select a higher

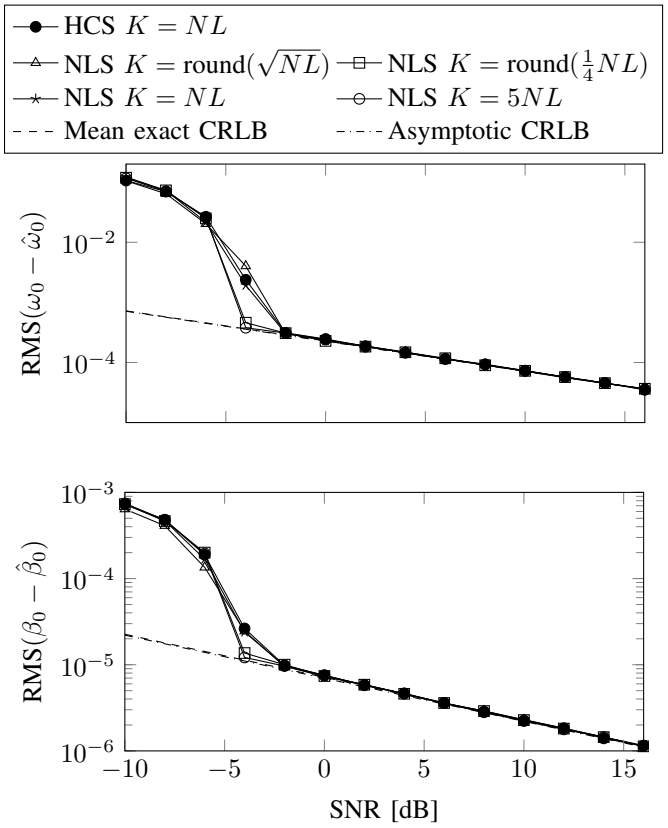

Figure 5. Estimation accuracy for different methods, the asymptotic CRLB and the mean exact CRLB for all the realizations. Varying SNR. Settings: $N = 250$, $\ell = L = 6$, $\omega_0 \in 2\pi \cdot [4/N, 8/N)$, $\beta_0 \in [-3/N^2, 3/N^2)$, constant amplitude $\sqrt{a_l^2 + b_l^2} = 1$, $l = 1, \ldots, L$ and phase drawn from a uniform distribution in the interval $[0, 2\pi)$ for each Monte Carlo repetition. $F = 5NL$.

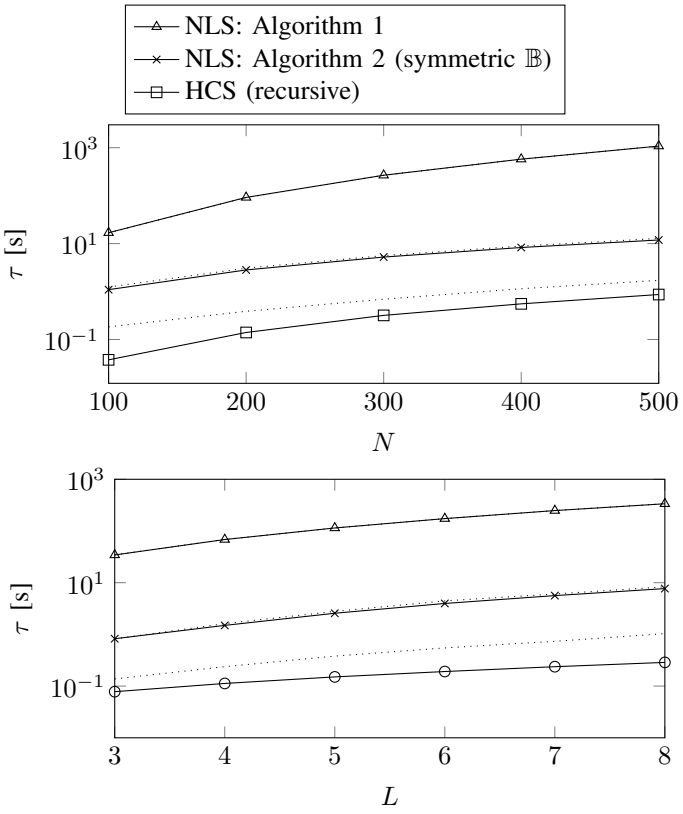

Figure 6. The computation time for four algorithms. Top: $L = 8$ Bottom: $N = 256$. The dotted lines are the corresponding timing including the Nelder-Mead refinement step for the NLS and HCS methods.

fundamental frequency $\omega_0 \in 2\pi \cdot [4/N, 8/N)$ drawn uniformly and the results being shown in Fig. 5. We observe that the mean of the exact CRLB and asymptotic CRLB are approximately equal and all the considered methods can achieve these bounds from approximately -2 [dB].

B. Timings

For comparing the computational speed at relevant dimensions, we ran different MATLAB implementations of the presented algorithms: Algorithm 1, Algorithm 2 exploiting symmetry in the rate grid as outlined in Sec. C-A, the HCS method using the recursive implementation presented in Sec. VI. The simulations were executed on an Intel(R) Core(TM) i7-5600U CPU 2.6 GHz with Ubuntu Linux kernel 3.19.0-43-generic and MATLAB R2015a (8.5.0.197613). The timings were obtained by running the algorithms three times to obtain the execution times τ_1, τ_2, τ_3. The reported execution times are then $\tau = \min(\tau_1, \tau_2, \tau_3)$. This is the same procedure used in Pythons `timeit` module [36] (when the execution time is larger than 0.2 [s]) and used as an attempt to eliminate the influence of other system processes. Fig. 6 illustrate the running time for varying N and L, respectively. The dotted lines are the corresponding timing with the Nelder-Mead refinement step with the same settings as used in the previous section. For the NLS and HCS algorithms we select $F = 5NL$ and $K = NL$ as in the previous section. From Fig. 6 it is clear that the proposed Algorithm 2 offers a much faster algorithm than the standard Algorithm 1. Notice that the standard Algorithm 1 will be many times faster than the corresponding algorithm in [10] since the number of grid evaluations is greatly reduced. The HCS algorithm in the recursive form may be of interest since it offers even faster algorithms and can be the method of choice if the HCS method is accurate enough, e.g., when the fundamental frequency is large enough compared to the frame size N.

VIII. Conclusion

In this paper, we have proposed an algorithm for computing an estimate of the non-linear least-squares estimator for the harmonic chirp model. The proposed algorithm is orders of magnitude faster than previously known algorithms. The proposed algorithm is obtained based on three strategies: *1)* Take a two-step approach using first the grid method with a coarse grid followed by a refinement step *2)* Formulate the problem such that the problem can be seen as a number of nested problems that allows for recursive updating from model order l to $l+1$. *3)* Exploit the matrix structures of the problem, in particular, a recursive block Toeplitz-plus-Hankel solver for linear systems and Fast Fourier Transforms. Furthermore, we showed the impact of the choice of a symmetric time index. In particular, a symmetric time index reduces the Cramér-Rao

lower bound for the fundamental frequency by approximately a factor of 16 compared to conventional time index selection.

APPENDIX A
APPROXIMATE HESSIAN

In this appendix, we derive an analytic, but approximate expression for the Hessian matrix $\boldsymbol{\Psi}(\hat{\boldsymbol{\xi}})$ of the objective function in (14) where $\boldsymbol{\xi} = \begin{bmatrix}\omega_0 & \beta_0\end{bmatrix}^T$. The Hessian is defined in (24), and the approximate expression is derived using the following two assumptions.

a) Asymptotic assumption: We assume that the data size N is large enough so that we can use the asymptotic result [37, App. A]

$$\lim_{N\to\infty} \frac{1}{N^{t+1}} \sum_{m=1}^{N} m^t \exp(j(\theta_1 m + \theta_2 m^2)) = 0 \tag{91}$$

for $t \geq 0$ as a good approximation. In our context, this asymptotic result can be used to show that

$$\lim_{N\to\infty} \frac{\boldsymbol{C}^T(\boldsymbol{\xi})\boldsymbol{N}^t\boldsymbol{C}(\boldsymbol{\xi})}{\mathbf{1}^T\boldsymbol{N}^t\mathbf{1}} = \lim_{N\to\infty} \frac{\boldsymbol{S}^T(\boldsymbol{\xi})\boldsymbol{N}^t\boldsymbol{S}(\boldsymbol{\xi})}{\mathbf{1}^T\boldsymbol{N}^t\mathbf{1}} = \frac{1}{2}\boldsymbol{I}_l \tag{92}$$

$$\lim_{N\to\infty} \frac{\boldsymbol{S}^T(\boldsymbol{\xi})\boldsymbol{N}^t\boldsymbol{C}(\boldsymbol{\xi})}{\mathbf{1}^T\boldsymbol{N}^t\mathbf{1}} = \mathbf{0} \tag{93}$$

where $\boldsymbol{C}(\boldsymbol{\xi})$ and $\boldsymbol{S}(\boldsymbol{\xi})$ are given in (6) and (7), respectively, and

$$\mathbf{1} = \begin{bmatrix}1 & \cdots & 1\end{bmatrix}^T \tag{94}$$

$$\boldsymbol{N} = \mathrm{diag}(\begin{bmatrix}n_0 & \cdots & n_0+N-1\end{bmatrix}) \ . \tag{95}$$

b) Large SNR assumption: We assume that the SNR is large enough so that

$$\boldsymbol{x} \approx \boldsymbol{Z}_l(\hat{\boldsymbol{\xi}})\hat{\boldsymbol{\alpha}}_l \ . \tag{96}$$

This assumption is used to simplify the expression of the Hessian matrix which, as detailed in (24), is the second order derivative of the objective. Finding this derivative involves computing the second-order differential of the projection matrix $\boldsymbol{P}_{\boldsymbol{Z}}(\hat{\boldsymbol{\xi}})$ given in (25). The expression for this differential is e.g. given in [38, App. C], and under the large SNR approximation, the expression of the (n,m)'th element of the Hessian matrix can be simplified as

$$[\boldsymbol{\Psi}(\hat{\boldsymbol{\xi}})]_{nm} \approx -2\hat{\boldsymbol{\alpha}}_l^T\boldsymbol{T}_n^T(\hat{\boldsymbol{\xi}})\left[\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{Z}}(\hat{\boldsymbol{\xi}})\right]\boldsymbol{T}_m(\hat{\boldsymbol{\xi}})\hat{\boldsymbol{\alpha}}_l \tag{97}$$

for $(n,m) \in \{1,2\}$ where $\boldsymbol{Z}_l(\boldsymbol{\xi})$ is given in (5) and

$$\boldsymbol{T}_1(\boldsymbol{\xi}) = \frac{\partial \boldsymbol{Z}_l(\boldsymbol{\xi})}{\partial\omega_0} = \begin{bmatrix}-\boldsymbol{N}\boldsymbol{S}_l(\boldsymbol{\xi})\boldsymbol{L} & \boldsymbol{N}\boldsymbol{C}_l(\boldsymbol{\xi})\boldsymbol{L}\end{bmatrix} \tag{98}$$

$$\boldsymbol{T}_2(\boldsymbol{\xi}) = \frac{\partial \boldsymbol{Z}_l(\boldsymbol{\xi})}{\partial\beta_0} = \frac{1}{2}\begin{bmatrix}-\boldsymbol{N}^2\boldsymbol{S}_l(\boldsymbol{\xi})\boldsymbol{L} & \boldsymbol{N}^2\boldsymbol{C}_l(\boldsymbol{\xi})\boldsymbol{L}\end{bmatrix} \tag{99}$$

$$\boldsymbol{L} = \mathrm{diag}(\begin{bmatrix}1 & 2 & \cdots & l\end{bmatrix}) \ . \tag{100}$$

A. Combining the two assumptions

We now use the asymptotic results in (92) and (93) to approximate the various terms in (97). First, we can derive from (92), (93), (5), (98), and (99) that

$$\lim_{N\to\infty} \frac{\boldsymbol{Z}_l^T(\boldsymbol{\xi})\boldsymbol{Z}_l(\boldsymbol{\xi})}{\mathbf{1}^T\mathbf{1}} = \frac{1}{2}\boldsymbol{I}_{2l} \tag{101}$$

$$\lim_{N\to\infty} \frac{\boldsymbol{T}_n^T(\boldsymbol{\xi})\boldsymbol{Z}_l(\boldsymbol{\xi})}{\mathbf{1}^T\boldsymbol{N}^n\mathbf{1}} = \frac{1}{2n}\begin{bmatrix}\mathbf{0} & -\boldsymbol{L}\\ \boldsymbol{L} & \mathbf{0}\end{bmatrix} \tag{102}$$

$$\lim_{N\to\infty} \frac{\boldsymbol{T}_n^T(\boldsymbol{\xi})\boldsymbol{T}_m(\boldsymbol{\xi})}{\mathbf{1}^T\boldsymbol{N}^{n+m}\mathbf{1}} = \frac{1}{2nm}\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix} \tag{103}$$

for $(n,m) \in \{1,2\}$. For a finite N, we now use these asymptotic limits as the approximations

$$\boldsymbol{Z}_l^T(\boldsymbol{\xi})\boldsymbol{Z}_l(\boldsymbol{\xi}) \approx \frac{\mathbf{1}^T\mathbf{1}}{2}\boldsymbol{I}_{2l} \tag{104}$$

$$\boldsymbol{T}_n^T(\boldsymbol{\xi})\boldsymbol{Z}_l(\boldsymbol{\xi}) \approx \frac{\mathbf{1}^T\boldsymbol{N}^n\mathbf{1}}{2n}\begin{bmatrix}\mathbf{0} & -\boldsymbol{L}\\ \boldsymbol{L} & \mathbf{0}\end{bmatrix} \tag{105}$$

$$\boldsymbol{T}_n^T(\boldsymbol{\xi})\boldsymbol{T}_m(\boldsymbol{\xi}) \approx \frac{\mathbf{1}^T\boldsymbol{N}^{n+m}\mathbf{1}}{2nm}\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix} \ . \tag{106}$$

Inserting the above into (97) yields

$$\begin{aligned}[\boldsymbol{\Psi}(\hat{\boldsymbol{\xi}})]_{nm} \approx & -2\hat{\boldsymbol{\alpha}}_l^T\frac{\mathbf{1}^T\boldsymbol{N}^{n+m}\mathbf{1}}{2nm}\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\hat{\boldsymbol{\alpha}}_l\\ & +2\hat{\boldsymbol{\alpha}}_l^T\frac{\mathbf{1}^T\boldsymbol{N}^n\mathbf{1}}{2n}\begin{bmatrix}\mathbf{0} & -\boldsymbol{L}\\ \boldsymbol{L} & \mathbf{0}\end{bmatrix}\frac{2}{\mathbf{1}^T\mathbf{1}}\frac{\mathbf{1}^T\boldsymbol{N}^m\mathbf{1}}{2m}\begin{bmatrix}\mathbf{0} & \boldsymbol{L}\\ -\boldsymbol{L} & \mathbf{0}\end{bmatrix}\hat{\boldsymbol{\alpha}}_l\\ = & -\frac{1}{nm}\left(\mathbf{1}^T\boldsymbol{N}^{n+m}\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{N}^n\mathbf{1}\mathbf{1}^T\boldsymbol{N}^m\mathbf{1}}{\mathbf{1}^T\mathbf{1}}\right)\\ & \cdot\hat{\boldsymbol{\alpha}}_l^T\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\hat{\boldsymbol{\alpha}}_l \ . \end{aligned} \tag{107}$$

The following identities can be derived

$$\mathbf{1}^T\boldsymbol{N}^2\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{N}\mathbf{1}\mathbf{1}^T\boldsymbol{N}\mathbf{1}}{\mathbf{1}^T\mathbf{1}} = \frac{N(N^2-1)}{12} \tag{108}$$

$$\begin{aligned}\mathbf{1}^T\boldsymbol{N}^3\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{N}\mathbf{1}\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}}{\mathbf{1}^T\mathbf{1}} &= \mathbf{1}^T\boldsymbol{N}^3\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}\mathbf{1}^T\boldsymbol{N}\mathbf{1}}{\mathbf{1}^T\mathbf{1}}\\ &= \frac{N(N^2-1)}{6}\left(n_0 + \frac{N-1}{2}\right)\end{aligned} \tag{109}$$

$$\begin{aligned}\mathbf{1}^T\boldsymbol{N}^4\mathbf{1} - \frac{\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}}{\mathbf{1}^T\mathbf{1}} &= \frac{N(N^2-1)}{3}\\ &\cdot\left[\frac{N^2-4}{60} + \left(n_0 + \frac{N-1}{2}\right)^2\right] \ .\end{aligned} \tag{110}$$

By inserting the above expressions in (107) and using (27), we obtain the final expression in (26) for the approximate Hessian matrix.

APPENDIX B
CRLB FOR HARMONIC CHIRP MODEL

The parameter vector we wish to derive the CRLB for is

$$\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\alpha}_l^T & \boldsymbol{\xi}^T\end{bmatrix}^T = \begin{bmatrix}\boldsymbol{\alpha}_l^T & \omega_0 & \beta_0\end{bmatrix}^T . \tag{111}$$

The (n,m)'th element of the Fisher information matrix (FIM) for this parameter vector is given by [27]

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]_{nm} = \sigma^{-2}\frac{\partial\boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial\theta_n}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta})}{\partial\theta_m} \tag{112}$$

for $(n,m) \in \{1,2,\ldots,2l+2\}$ where

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{Z}_l(\boldsymbol{\xi})\boldsymbol{\alpha}_l \ . \tag{113}$$

For the partial derivatives, we have that

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}_l} = \boldsymbol{Z}_l(\boldsymbol{\xi}) \tag{114}$$

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \omega_0} = \boldsymbol{T}_1(\boldsymbol{\xi})\boldsymbol{\alpha}_l \tag{115}$$

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \beta_0} = \boldsymbol{T}_2(\boldsymbol{\xi})\boldsymbol{\alpha}_l \tag{116}$$

where $\boldsymbol{T}_1(\boldsymbol{\xi})$ and $\boldsymbol{T}_2(\boldsymbol{\xi})$ are given in (98) and (99), respectively. Consequently, if we define the matrix

$$\boldsymbol{\Delta}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{Z}_l(\boldsymbol{\xi}) & \boldsymbol{T}_1(\boldsymbol{\xi})\boldsymbol{\alpha}_l & \boldsymbol{T}_2(\boldsymbol{\xi})\boldsymbol{\alpha}_l \end{bmatrix} \ , \tag{117}$$

the FIM is given by

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\boldsymbol{\Delta}^T(\boldsymbol{\theta})\boldsymbol{\Delta}(\boldsymbol{\theta}) \ . \tag{118}$$

Unfortunately, we cannot find an exact closed-form expression for the inverse FIM. However, if we use the asymptotic results in (101)-(103) as approximations as in (104)-(106), we obtain the approximate FIM in (119) on the top of page 13, which is also often referred to as the asymptotic FIM, and this FIM can be inverted analytically. Here, we are only interested in the CRLB for the chirp rate and the fundamental frequency parameters, and we, therefore, partition the asymptotic FIM as

$$\boldsymbol{\mathcal{I}}_{\mathrm{a}}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{b}^T & d & e \\ \boldsymbol{c}^T & e & f \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{D} \end{bmatrix} \tag{120}$$

with implicit definitions. The inverse of the latter matrix is

$$\boldsymbol{\mathcal{I}}_{\mathrm{a}}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \times & \times \\ \times & (\boldsymbol{D} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix} \tag{121}$$

where we have only written out the lower right 2×2 matrix whose diagonal elements are the CRLB for the fundamental frequency and the chirp rate, respectively. A closer inspection of this lower right matrix reveals that

$$\boldsymbol{D} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B} = -\boldsymbol{\Psi}(\boldsymbol{\xi})/2 \tag{122}$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi})$ is the Hessian matrix derived in App. A. Consequently,

$$\boldsymbol{D} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B} = \frac{N(N^2-1)\sum_{i=1}^{l} A_i^2 i^2}{24\sigma^2} \cdot \begin{bmatrix} 1 & \left(n_0 + \frac{N-1}{2}\right) \\ \left(n_0 + \frac{N-1}{2}\right) & \frac{N^2-4}{60} + \left(n_0 + \frac{N-1}{2}\right)^2 \end{bmatrix} , \tag{123}$$

and the asymptotic inverse FIM is then given in (29).

APPENDIX C
A BLOCK-TOEPLITZ PLUS BLOCK-HANKEL SOLVER

In this appendix, we will give a block Toeplitz-plus-Hankel algorithm for solving linear systems of equations with blocks of size S and Q right hand sides. We will drop the dependency on ω_0, β_0 in the following to improve readability and highlight the fact that the following deviations are more general. The derived algorithm is a block version of the algorithm presented in [30] (with blocks of size $S \times S$). A very similar algorithm was independently, but less thoroughly, derived in [39].

For the application derived in this paper, we are interested in the case with $S = 2, Q = 1$ on the form

$$\left(\begin{bmatrix} \boldsymbol{T}_l & \tilde{\boldsymbol{T}}_l^T \\ \tilde{\boldsymbol{T}}_l & \bar{\boldsymbol{T}}_l \end{bmatrix} + \begin{bmatrix} \boldsymbol{H}_l & \tilde{\boldsymbol{H}}_l \\ \tilde{\boldsymbol{H}}_l & \bar{\boldsymbol{H}}_l \end{bmatrix} \right) \begin{bmatrix} \tilde{\boldsymbol{a}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \bar{\boldsymbol{R}}_l \bar{\boldsymbol{a}}_l = \begin{bmatrix} \tilde{\boldsymbol{b}} \\ \hat{\boldsymbol{b}} \end{bmatrix} = \bar{\boldsymbol{b}}_l \tag{124}$$

where $\boldsymbol{T}_l = \boldsymbol{T}_l^T, \bar{\boldsymbol{T}}_l = \bar{\boldsymbol{T}}_l^T, \tilde{\boldsymbol{T}}_l$ indicates Toeplitz matrices, and $\boldsymbol{H}_l, \tilde{\boldsymbol{H}}_l, \bar{\boldsymbol{H}}_l$ indicates Hankel matrices, with $\boldsymbol{T}_l, \tilde{\boldsymbol{T}}_l, \bar{\boldsymbol{T}}_l, \boldsymbol{H}_l, \tilde{\boldsymbol{H}}_l, \bar{\boldsymbol{H}}_l \in \mathbb{R}^{l \times l}$ and $\bar{\boldsymbol{a}}_l \in \mathbb{R}^{Sl \times Q}, \bar{\boldsymbol{b}}_l \in \mathbb{R}^{Sl \times Q}$. With an appropriate permutation matrix $\boldsymbol{\Pi}_l$, we can obtain the symmetric matrix

$$\boldsymbol{R}_l = \begin{bmatrix} \boldsymbol{R}_{1,1} & \boldsymbol{R}_{1,2} & \ldots & \boldsymbol{R}_{1,l} \\ \boldsymbol{R}_{2,1} & \boldsymbol{R}_{2,2} & \ldots & \boldsymbol{R}_{2,l} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{l,1} & \boldsymbol{R}_{l,2} & \ldots & \boldsymbol{R}_{l,l} \end{bmatrix} = \boldsymbol{\Pi}_l \bar{\boldsymbol{R}}_l \boldsymbol{\Pi}_l^T \in \mathbb{R}^{Sl \times Sl} \tag{125}$$

with the following permuted relations

$$\boldsymbol{R}_l \boldsymbol{a}_l = \boldsymbol{b}_l, \quad \boldsymbol{\Pi}_l \bar{\boldsymbol{a}}_l = \boldsymbol{a}_l, \quad \boldsymbol{\Pi}_l \bar{\boldsymbol{b}}_l = \boldsymbol{b}_l \, . \tag{126}$$

For the case $S = 2$, the block are given as

$$\boldsymbol{R}_{i,j} = \begin{bmatrix} (\boldsymbol{T}_l)_{i,j} + (\boldsymbol{H}_l)_{i,j} & (\tilde{\boldsymbol{T}}_l)_{j,i} + (\tilde{\boldsymbol{H}}_l)_{i,j} \\ (\tilde{\boldsymbol{T}}_l)_{i,j} + (\tilde{\boldsymbol{H}}_l)_{i,j} & (\bar{\boldsymbol{T}}_l)_{i,j} + (\bar{\boldsymbol{H}}_l)_{i,j} \end{bmatrix} \in \mathbb{R}^{S \times S} \, . \tag{127}$$

The algorithm proposed for the scalar case in [30] contains a data dependent and a data independent step. The following derivation will follow this approach.

1) The data dependent step: Assume that we have computed the solutions to the following systems

$$\boldsymbol{R}_l \boldsymbol{\gamma}_l = \boldsymbol{e}_l, \quad l = 1, \ldots, L \tag{128}$$

where $\boldsymbol{e}_l = \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{I}_S \end{bmatrix}^T \in \mathbb{R}^{Sl \times S}$, $\boldsymbol{\gamma}_l \in \mathbb{R}^{Sl \times S}$ and $\boldsymbol{I}_S \in \mathbb{R}^{S \times S}$ is the identity matrix. Consider the recursive form of the matrix and the right-hand side

$$\boldsymbol{R}_l = \begin{bmatrix} \boldsymbol{R}_{l-1} & \boldsymbol{r}_{l-1} \\ \boldsymbol{r}_{l-1}^T & [\boldsymbol{r}_l]_l \end{bmatrix} \in \mathbb{R}^{Sl \times Sl}, \quad \boldsymbol{b}_l = \begin{bmatrix} \boldsymbol{b}_{l-1} \\ [\boldsymbol{b}_l]_l \end{bmatrix} \in \mathbb{R}^{Sl \times Q} \, . \tag{129}$$

Following the same analysis as in [40]

$$\boldsymbol{R}_l \begin{bmatrix} \boldsymbol{a}_{l-1} \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{l-1} \\ \boldsymbol{r}_{l-1}^T \boldsymbol{a}_{l-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{l-1} \\ [\boldsymbol{b}_l]_l - \boldsymbol{\lambda}_l \end{bmatrix} = \boldsymbol{b}_l - \boldsymbol{e_l}\boldsymbol{\lambda}_l \tag{130}$$

with

$$\boldsymbol{\lambda}_l = [\boldsymbol{b}_l]_l - \boldsymbol{r}_{l-1}^T \boldsymbol{a}_{l-1} \ . \tag{131}$$

Then

$$\begin{bmatrix} \boldsymbol{a}_{l-1} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{R}_l^{-1}\boldsymbol{b}_l - \boldsymbol{R}_l^{-1}\boldsymbol{e}_l\boldsymbol{\lambda}_l = \boldsymbol{a}_l - \boldsymbol{\gamma}_l\boldsymbol{\lambda}_l \, . \tag{132}$$

$$\boldsymbol{\mathcal{I}}_{\mathrm{a}}(\boldsymbol{\theta}) = \sigma^{-2}\begin{bmatrix} \frac{\mathbf{1}^T\mathbf{1}}{2}\boldsymbol{I}_{2l} & \frac{\mathbf{1}^T\boldsymbol{N}\mathbf{1}}{2}\begin{bmatrix}\mathbf{0} & \boldsymbol{L}\\ -\boldsymbol{L} & \mathbf{0}\end{bmatrix}\boldsymbol{\alpha}_l & \frac{\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}}{4}\begin{bmatrix}\mathbf{0} & \boldsymbol{L}\\ -\boldsymbol{L} & \mathbf{0}\end{bmatrix}\boldsymbol{\alpha}_l \\ \frac{\mathbf{1}^T\boldsymbol{N}\mathbf{1}}{2}\boldsymbol{\alpha}_l^T\begin{bmatrix}\mathbf{0} & -\boldsymbol{L}\\ \boldsymbol{L} & \mathbf{0}\end{bmatrix} & \frac{\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}}{2}\boldsymbol{\alpha}_l^T\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\boldsymbol{\alpha}_l & \frac{\mathbf{1}^T\boldsymbol{N}^3\mathbf{1}}{4}\boldsymbol{\alpha}_l^T\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\boldsymbol{\alpha}_l \\ \frac{\mathbf{1}^T\boldsymbol{N}^2\mathbf{1}}{4}\boldsymbol{\alpha}_l^T\begin{bmatrix}\mathbf{0} & -\boldsymbol{L}\\ \boldsymbol{L} & \mathbf{0}\end{bmatrix} & \frac{\mathbf{1}^T\boldsymbol{N}^3\mathbf{1}}{4}\boldsymbol{\alpha}_l^T\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\boldsymbol{\alpha}_l & \frac{\mathbf{1}^T\boldsymbol{N}^4\mathbf{1}}{8}\boldsymbol{\alpha}_l^T\begin{bmatrix}\boldsymbol{L}^2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}^2\end{bmatrix}\boldsymbol{\alpha}_l \end{bmatrix} \tag{119}$$

Using the above, an algorithm can then compute the solution to $\boldsymbol{R}_l\boldsymbol{a}_l = \boldsymbol{b}_l$ using the recursive updates

$$\boldsymbol{\lambda}_l = [\boldsymbol{b}_L]_l - \boldsymbol{r}_{l-1}^T\boldsymbol{a}_{l-1} \tag{133}$$

$$\boldsymbol{a}_l = \begin{bmatrix}\boldsymbol{a}_{l-1}\\ \mathbf{0}\end{bmatrix} + \boldsymbol{\gamma}_l\boldsymbol{\lambda}_l\,. \tag{134}$$

2) The data independent step: In the data independent step, we focus on solving

$$\boldsymbol{R}_{l+1}\boldsymbol{\gamma}_{l+1} = \boldsymbol{e}_{l+1} \tag{135}$$

efficiently given the solution $\boldsymbol{\gamma}_l$ to $\boldsymbol{R}_l\boldsymbol{\gamma}_l = \boldsymbol{e}_l$. Due to the block Toeplitz-plus-Hankel structure of $\boldsymbol{R}_l$, we have

$$\boldsymbol{R}_l = \begin{bmatrix}\boldsymbol{t}_0 & \boldsymbol{t}_1^T & \dots & \boldsymbol{t}_{l-1}^T\\ \boldsymbol{t}_1 & \boldsymbol{t}_0 & \ddots & \boldsymbol{t}_{l-2}^T\\ \vdots & \ddots & \ddots & \vdots\\ \boldsymbol{t}_{l-1} & \boldsymbol{t}_{l-2} & \dots & \boldsymbol{t}_0\end{bmatrix} + \begin{bmatrix}\boldsymbol{h}_2 & \boldsymbol{h}_3 & \dots & \boldsymbol{h}_{l+1}\\ \boldsymbol{h}_3 & \boldsymbol{h}_4 & \iddots & \boldsymbol{h}_{l+2}\\ \vdots & \iddots & \iddots & \vdots\\ \boldsymbol{h}_{l+1} & \boldsymbol{h}_{l+2} & \dots & \boldsymbol{h}_{2l}\end{bmatrix} \tag{136}$$

where $\boldsymbol{t}_i \in \mathbb{R}^{S\times S}, i = 0,\dots,l-1$, $\boldsymbol{h}_i \in \mathbb{R}^{S\times S}, i = 2,\dots,2l$. From this, it follows that

$$\begin{aligned}(\boldsymbol{L}_l + \boldsymbol{L}_l^T)\boldsymbol{R}_l - \boldsymbol{R}_l(\boldsymbol{L}_l + \boldsymbol{L}_l^T) = \\ \begin{bmatrix}\boldsymbol{t}_1 - \boldsymbol{t}_1^T & -\boldsymbol{t}_2^T & \dots & -\boldsymbol{t}_{l-1} & \mathbf{0}\\ \boldsymbol{t}_2 & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{t}_{l-1}^T\\ \vdots & \vdots & & \vdots & \vdots\\ \boldsymbol{t}_{l-1} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{t}_2^T\\ \mathbf{0} & -\boldsymbol{t}_{l-1} & \dots & -\boldsymbol{t}_2 & \boldsymbol{t}_1^T - \boldsymbol{t}_1\end{bmatrix}\\ + \begin{bmatrix}\mathbf{0} & -\boldsymbol{h}_2 & \dots & -\boldsymbol{h}_{l-1} & -\boldsymbol{h}_l + \boldsymbol{h}_{l+2}\\ \boldsymbol{h}_2 & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{h}_{l+3}\\ \vdots & \vdots & & \vdots & \vdots\\ \boldsymbol{h}_{l-1} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{h}_{2l}\\ \boldsymbol{h}_l - \boldsymbol{h}_{l+2} & -\boldsymbol{h}_{l+1} & \dots & -\boldsymbol{h}_{2l} & \mathbf{0}\end{bmatrix}\end{aligned} \tag{137}$$

$$= \boldsymbol{q}_l\boldsymbol{e}_1^T - \boldsymbol{e}_1\boldsymbol{q}_l^T + \boldsymbol{r}_l\boldsymbol{e}_l^T - \boldsymbol{e}_l\boldsymbol{r}_l^T \tag{138}$$

where $\boldsymbol{L}_l$ is a lower triangular shift matrix of size $Sl \times Sl$:

$$\boldsymbol{L}_l = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \boldsymbol{I}_S & \mathbf{0} & \mathbf{0} & \ddots & \vdots\\ \mathbf{0} & \boldsymbol{I}_S & \mathbf{0} & & \mathbf{0}\\ \vdots & \ddots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{I}_S & \mathbf{0}\end{bmatrix}, \tag{139}$$

and

$$\boldsymbol{q}_l = \begin{bmatrix}\boldsymbol{t}_0\\ \boldsymbol{t}_2 + \boldsymbol{h}_2\\ \vdots\\ \boldsymbol{t}_l + \boldsymbol{h}_l\end{bmatrix}, \quad \boldsymbol{r}_l = \begin{bmatrix}\boldsymbol{t}_l^T + \boldsymbol{h}_{l+2}\\ \boldsymbol{t}_{l-1}^T + \boldsymbol{h}_{l+1}\\ \vdots\\ \boldsymbol{t}_1^T + \boldsymbol{h}_{2l+1}\end{bmatrix}, \quad l = 1,\dots,L-1. \tag{140}$$

Notice that in the scalar case $S = 1$ we have the form presented in [30]. The vector $\boldsymbol{r}_l$ is also shown in (129). Multiplying (138) with the symmetric matrix $\boldsymbol{R}_l^{-1}$ from both the right and left yields

$$\boldsymbol{R}_l^{-1}(\boldsymbol{L}_l + \boldsymbol{L}_l^T) - (\boldsymbol{L}_l + \boldsymbol{L}_l^T)\boldsymbol{R}_l^{-1} = \boldsymbol{\phi}_l\boldsymbol{\psi}_l^T - \boldsymbol{\psi}_l\boldsymbol{\phi}_l^T - \boldsymbol{\rho}_l\boldsymbol{\gamma}_l^T + \boldsymbol{\gamma}_l\boldsymbol{\rho}_l^T \tag{141}$$

where $\boldsymbol{\phi}_l$, $\boldsymbol{\psi}_l$, and $\boldsymbol{\rho}_l$ are the solutions to

$$\boldsymbol{R}_l\boldsymbol{\phi}_l = \boldsymbol{q}_l, \quad l = 1,\dots,L-1 \tag{142}$$

$$\boldsymbol{R}_l\boldsymbol{\psi}_l = \boldsymbol{e}_1, \quad l = 1,\dots,L-1 \tag{143}$$

$$\boldsymbol{R}_l\boldsymbol{\rho}_l = -\boldsymbol{r}_l, \quad l = 1,\dots,L-1. \tag{144}$$

By multiplying $\boldsymbol{e}_l$ on the right of (141) and solve for the second last term on the right hand side, we have

$$\begin{aligned}\boldsymbol{\rho}_l[\boldsymbol{\gamma}_l]_l^T = (\boldsymbol{L}_l + \boldsymbol{L}_l^T)\boldsymbol{\gamma}_l - \boldsymbol{R}_l^{-1}\begin{bmatrix}\boldsymbol{e}_{l-1}\\ \mathbf{0}\end{bmatrix}\\ + \boldsymbol{\gamma}_l[\boldsymbol{\rho}_l]_l^T + \boldsymbol{\phi}_l[\boldsymbol{\psi}_l]_l^T - \boldsymbol{\psi}_l[\boldsymbol{\phi}_l]_l^T \triangleq \boldsymbol{\beta}_l\,.\end{aligned} \tag{145}$$

With this definition, $\boldsymbol{\beta}_l$ can be related to update formulas for $\boldsymbol{\gamma}_{l+1}$. To derive this, we write (135) using (129) as

$$\begin{bmatrix}\boldsymbol{R}_l & \boldsymbol{r}_l\\ \boldsymbol{r}_l^T & [\boldsymbol{r}_{l+1}]_{l+1}\end{bmatrix}\begin{bmatrix}[\boldsymbol{\gamma}_{l+1}]_{1:l}\\ [\boldsymbol{\gamma}_{l+1}]_{l+1}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \boldsymbol{I}_S\end{bmatrix}\,. \tag{146}$$

Using (144), the above can be restated as

$$\boldsymbol{R}_l^{-1}\boldsymbol{r}_l = -\boldsymbol{\rho}_l = -[\boldsymbol{\gamma}_{l+1}]_{1:l}[\boldsymbol{\gamma}_{l+1}]_{l+1}^{-1} \tag{147}$$

$$[\boldsymbol{\gamma}_{l+1}]_{l+1} = \left([\boldsymbol{r}_{l+1}]_{l+1} + \boldsymbol{r}_l^T\boldsymbol{\rho}_l\right)^{-1}\,. \tag{148}$$

By combining (145), (147), and (148), we can get linear complexity updates of $\boldsymbol{\gamma}_{l+1}$

$$[\boldsymbol{\gamma}_{l+1}]_{l+1} = \left([\boldsymbol{r}_{l+1}] + \boldsymbol{r}_l^T\boldsymbol{\beta}_l[\boldsymbol{\gamma}_l]_l^{-T}\right)^{-1} \tag{149}$$

$$[\boldsymbol{\gamma}_{l+1}]_{1:l} = \boldsymbol{\beta}_l[\boldsymbol{\gamma}_l]_l^{-T}[\boldsymbol{\gamma}_{l+1}]_{l+1}\,. \tag{150}$$

To calculate $\boldsymbol{\beta}_l$ in (145) efficiently, we consider the second, fourth, and fifth term.

2. Since

$$\boldsymbol{R}_l\begin{bmatrix}\boldsymbol{\gamma}_{l-1}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{R}_{l-1} & \boldsymbol{r}_{l-1}\\ \boldsymbol{r}_{l-1}^T & [\boldsymbol{r}_l]_l\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}_{l-1}\\ \mathbf{0}\end{bmatrix} \tag{151}$$

$$= \begin{bmatrix}\boldsymbol{e}_{l-1}\\ \mathbf{0}\end{bmatrix} + \boldsymbol{e}_l(\boldsymbol{r}_{l-1}^T\boldsymbol{\gamma}_{l-1})\,, \tag{152}$$

it follows that

$$\boldsymbol{R}_l^{-1}\begin{bmatrix}\boldsymbol{e}_{l-1}\\ \boldsymbol{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\gamma}_{l-1}\\ \boldsymbol{0}\end{bmatrix} - \boldsymbol{\gamma}_l(\boldsymbol{r}_{l-1}^T\boldsymbol{\gamma}_{l-1})\ . \tag{153}$$

The last factor is simply

$$\boldsymbol{r}_{l-1}^T\boldsymbol{\gamma}_{l-1} = \boldsymbol{r}_{l-1}^T\boldsymbol{R}_{l-1}^{-1}\boldsymbol{e}_{l-1} = -\boldsymbol{\rho}_{l-1}^T\boldsymbol{e}_{l-1} = -[\boldsymbol{\rho}_{l-1}]_{l-1}^T\ . \tag{154}$$

4.-5. By using (142) and (143) with a similar analysis as in (130)–(132) we obtain that these vectors can be updated with linear computational complexity as

$$\boldsymbol{\phi}_l = \begin{bmatrix}\boldsymbol{\phi}_{l-1}\\ \boldsymbol{0}\end{bmatrix} + \boldsymbol{\gamma}_l([\boldsymbol{q}_l]_l - \boldsymbol{r}_{l-1}^T\boldsymbol{\phi}_{l-1}) \tag{155}$$

$$\boldsymbol{\psi}_l = \begin{bmatrix}\boldsymbol{\psi}_{l-1}\\ \boldsymbol{0}\end{bmatrix} - \boldsymbol{\gamma}_l(\boldsymbol{r}_{l-1}^T\boldsymbol{\psi}_{l-1})\ . \tag{156}$$

Then $\boldsymbol{\beta}_l$ can now be written as

$$\boldsymbol{\beta}_l = \left(\boldsymbol{L}_l + \boldsymbol{L}_l^T\right)\boldsymbol{\gamma}_l + \boldsymbol{\gamma}_l([\boldsymbol{\rho}_l]_l^T - [\boldsymbol{\rho}_{l-1}]_{l-1}^T) - \begin{bmatrix}\boldsymbol{\gamma}_{l-1}\\ \boldsymbol{0}\end{bmatrix} + \boldsymbol{\phi}_l[\boldsymbol{\psi}_l]_l^T - \boldsymbol{\psi}_l[\boldsymbol{\phi}_l]_l^T\ . \tag{157}$$

To summarize the data-independent step: given $\boldsymbol{\gamma}_l$, $\boldsymbol{r}_l$, $[\boldsymbol{r}_{l+1}]_{l+1}$, $\boldsymbol{q}_l$, and $[\boldsymbol{\rho}_{l-1}]_{l-1}$, we can compute $\boldsymbol{\gamma}_{l+1}$ with linear complexity using the following algorithm:

1) Compute $\boldsymbol{\phi}_l$ and $\boldsymbol{\psi}_l$ using (155) and (156).
2) Compute $[\boldsymbol{\rho}_l]_l$ using (154).
3) Compute $\boldsymbol{\beta}_l$ using (157).
4) Compute $\boldsymbol{\gamma}_{l+1}$ using (149) and (150).

The data-dependent step can then be added, such that given $\boldsymbol{\gamma}_{l+1}$ and $\boldsymbol{a}_l$:

5) Compute $\boldsymbol{a}_{l+1}$ using (133)–(134).

This will yield a recursive algorithm with the computational complexity dominated by matrix-matrix products (block-vector times block as in e.g. (155)), yielding $\mathcal{O}(S^3l)$ operations per recursion, and $\mathcal{O}(S^3L^2)$ for all $l = 1,\ldots,L-1$. Applying (133)–(134) costs $\mathcal{O}(S^2L^2Q)$, to a total of $\mathcal{O}(L^2(S^3+S^2Q)$. This is the same as classic block-Levinson algorithms with the computational complexity $\mathcal{O}(S^3L^2)$ for $Q = 1$ [41].

A. Exploiting symmetry

With symmetric time index $n_0 = -\frac{N-1}{2}$ and using (63)–(66), we also have

$$t_l(\omega_0,\beta_0) = t_l(\omega_0,-\beta_0) \tag{158}$$

$$h_l(\omega_0,\beta_0) = h_l(\omega_0,-\beta_0) \tag{159}$$

$$\tilde{t}_l(\omega_0,\beta_0) = -t_l(\omega_0,-\beta_0) \tag{160}$$

$$\tilde{h}_l(\omega_0,\beta_0) = -h_l(\omega_0,-\beta_0)\,. \tag{161}$$

With the permutation used in (125), then

$$\boldsymbol{R}_l(\omega_0,\beta_0) = \boldsymbol{\Pi}\boldsymbol{A}_l(\omega_0,\beta_0)\boldsymbol{\Pi}^T = \begin{bmatrix}\boldsymbol{R}_{1,1}(\omega_0,\beta_0) & \boldsymbol{R}_{1,2}(\omega_0,\beta_0) & \ldots & \boldsymbol{R}_{1,l}(\omega_0,\beta_0)\\ \boldsymbol{R}_{2,1}(\omega_0,\beta_0) & \boldsymbol{R}_{2,2}(\omega_0,\beta_0) & \ldots & \boldsymbol{R}_{2,l-1}(\omega_0,\beta_0)\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{R}_{l,1}(\omega_0,\beta_0) & \boldsymbol{R}_{l,2}(\omega_0,\beta_0) & \ldots & \boldsymbol{R}_{l,l}(\omega_0,\beta_0)\end{bmatrix} \tag{162}$$

where

$$\boldsymbol{R}_{i,j}(\omega_0,\beta_0) = \begin{bmatrix}[\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{1,1} & [\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{1,2}\\ [\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{2,1} & [\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{2,2}\end{bmatrix} \tag{163}$$

$$= \begin{bmatrix}t_{i-j}(\omega_0,\beta_0)+h_{i+j}(\omega_0,\beta_0) & -\tilde{t}_{i-j}(\omega_0,\beta_0)+\tilde{h}_{i+j}(\omega_0,\beta_0)\\ \tilde{t}_{i-j}(\omega_0,\beta_0)+\tilde{h}_{i+j}(\omega_0,\beta_0) & t_{i-j}(\omega_0,\beta_0)-h_{i+j}(\omega_0,\beta_0))\end{bmatrix} \tag{164}$$

Then

$$\boldsymbol{R}_{i,j}(\omega_0,-\beta_0) = \begin{bmatrix}[\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{1,1} & -[\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{1,2}\\ -[\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{2,1} & [\boldsymbol{R}_{i,j}(\omega_0,\beta_0)]_{2,2}\end{bmatrix} \tag{165}$$

$$= \boldsymbol{R}_{i,j}(\omega_0,\beta_0)\odot\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}. \tag{166}$$

Let

$$\boldsymbol{R}_l(\omega_0,\beta_0)\boldsymbol{\gamma}_l(\omega_0,\beta_0) = \boldsymbol{e}_l \tag{167}$$

then

$$\boldsymbol{R}_{i,j}(\omega_0,-\beta_0)\left([\boldsymbol{\gamma}_l]_j\odot\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\right) = (\boldsymbol{R}_{i,j}(\omega_0,\beta_0)[\boldsymbol{\gamma}_l]_j)\odot\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}. \tag{168}$$

This implies that

$$\boldsymbol{R}_l(\omega_0,-\beta_0)\left(\boldsymbol{\gamma}_l(\omega_0,\beta_0)\odot\left(\boldsymbol{1}_l\otimes\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\right)\right) = \boldsymbol{e}_l \tag{169}$$

where $\boldsymbol{1}_l = [1,1,\ldots,1]^T \in \mathbb{R}^{l\times 1}$. The interpretation is that all off-diagonal elements of each 2×2 block of $\boldsymbol{\gamma}_l(\omega_0,\beta_0)$ just need a sign-change to obtain the solution of the data-independent step $\boldsymbol{\gamma}_l(\omega_0,-\beta_0)$. This amounts to a 50% saving in the data-independent step of the algorithm if $\mathbb{B}_l$ is symmetric.

REFERENCES

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

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

[3] H. Dudley, "The carrier nature of speech," *AT&T Tech J.*, vol. 19, no. 4, pp. 495–515, Oct. 1940.

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

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

[6] S. Gade, H. Herlufsen, H. Konstantin-Hansen, and N. J. Wismer, "Order tracking analysis," Brüel & Kjær A/S, Technical Review No. 2, 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," *Math. Biosci.*, vol. 12, no. 1–2, pp. 41–51, Oct. 1971.

[8] 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.

[9] 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.

[10] 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.

[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. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.
[13] ——, "Angle and waveform estimation via RELAX," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 1077–1087, 1997.
[14] R. Wu and J. Li, "Time-delay estimation via optimizing highly oscillatory cost functions," *IEEE J. Ocean. Eng.*, vol. 23, no. 3, pp. 235–244, 1998.
[15] R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via MODE-WRELAX," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 294–307, 1999.
[16] E. G. Larsson and P. Stoica, "Fast implementation of two-dimensional APES and CAPON spectral estimators," *Multidimension. Syst. Signal Process.*, vol. 13, no. 1, pp. 35–53, 2002.
[17] G.-O. Glentis, "A fast algorithm for APES and Capon spectral estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4207–4220, Sep. 2008.
[18] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[19] 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.
[20] M. G. Christensen, "Accurate estimation of low fundamental frequencies from real-valued measurements," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 10, pp. 2042–2056, 2013.
[21] T. Abatzoglou, "Fast maximum likelihood joint estimation of frequency and frequency rate," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 708–715, Nov 1986.
[22] B. Volcker and B. Ottersten, "Chirp parameter estimation using rank reduction," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, vol. 2, Nov 1998, pp. 1443–1446.
[23] P. O'Shea, "A fast algorithm for estimating the parameters of a quadratic FM signal," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 385–393, Feb 2004.
[24] S. Saha and S. Kay, "Maximum likelihood parameter estimation of superimposed chirps using Monte Carlo importance sampling," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 224–230, Feb 2002.
[25] B. Ristic and B. Boashash, "Comments on "the Cramer-Rao lower bounds for signals with constant amplitude and polynomial phase," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1708–1709, June 1998.
[26] P. Korten, J. Jensen, and R. Heusdens, "High-resolution spherical quantization of sinusoidal parameters," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 966–981, March 2007.
[27] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice Hall PTR, Mar. 1993.
[28] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Grid size selection for nonliner least-squares optimisation in spectral estimation and array processing," in *Proc. European Signal Processing Conf.*, Aug. 2016, pp. 1653–1657.
[29] ——, "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient," *Elsevier Signal Processing*, vol. 135, pp. 188–197, 2017.
[30] I. Gohberg and I. Koltracht, "Efficient algorithm for Toeplitz plus Hankel matrices," *Integr. Equat. Oper. Th.*, vol. 12, no. 1, pp. 136–142, 1989.
[31] A. M. Noll, "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate," in *Proc. of the symposium on computer process. commun.*, vol. 779, 1969.
[32] D. J. Hermes, "Measurement of pitch by subharmonic summation," *J. Acoust. Soc. Am.*, vol. 83, no. 1, pp. 257–264, 1988.
[33] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast harmonic chirp summation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2017.
[34] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer J.*, vol. 7, no. 4, pp. 308–313, 1965.
[35] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1070–1087, 1991.
[36] G. van Rossum and the Python development team, *The Python Library Reference*, Python Software Foundation, 2016, v. 2.7.11, https://docs.python.org/2/index.html.
[37] A. Lahiri, D. Kundu, and A. Mitra, "Estimating the parameters of multiple chirp signals," *J. Multivariate Analysis*, vol. 139, pp. 189–206, 2015.
[38] 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, 2014.
[39] A. Ringh and J. Karlsson, "A fast solver for the circulant rational covariance extension problem," in *Proc. European Control Conf. (ECC)*, July 2015, pp. 727–733.
[40] I. Gohberg, T. Kailath, and I. Koltracht, "Efficient solution of linear systems of equations with recursive structure," *Linear Algebra App.*, vol. 80, pp. 81 – 113, 1986.
[41] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, vol. 70, no. 8, pp. 1885–1891, 1965.

Tobias Lindstrøm Jensen received his M.Sc. in Electrical Engineering from Aalborg University in 2007 and Ph.D. degree in 2011. In 2007, he was an intern at Wipro-NewLogic Technologies in Sophia-Antipolis, France. In 2009 and 2016 he was a visiting researcher at University of California, Los Angeles (UCLA).

TLJ is currently an associate professor at the Department of Electronic Systems, Aalborg University, funded by the Danish Council for Independent Research under an individual post-doc grant with Sapere Aude (former Danish Council for Independent Research Young Researcher's Award). His research interests includes optimization algorithms, inverse problems and estimation, signal and image processing, signal processing for communications, numerical algorithms and computing. TLJ is a reviewer for e.g. IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, IEEE Transactions on Communications, IEEE Transactions on Wireless Communications, IEEE Transactions on Audio, Speech and Language Processing, EURASIP Journal on Advanced Signal Processing.

Jesper Kjær Nielsen (S'12–M'13) received the B.Sc, M.Sc (Cum Laude), and Ph.D. degrees in electrical engineering with a specialisation in signal processing from Aalborg University, Denmark, in 2007, 2009, and 2012, respectively. From 2012 to 2016, he was with the Department of Electronic Systems, Aalborg University, as an industrial postdoctoral researcher (12-15) and as a fixed-term associate professor (15-16). Bang & Olufsen A/S (B&O) was the industrial partner in these four years, and Jesper spend approximately 1/3 of his time at B&O in Struer. Jesper is currently with the Audio Analysis Lab, Aalborg University, in a three year position as an assistant professor. He is part-time employed by B&O and part time employed on a research project with the Danish hearing aid company GN ReSound.

Jesper has been a Visiting Scholar in the Signal Processing and Communications Laboratory, University of Cambridge in 2009 and at the Department of Computer Science, University of Illinois at Urbana-Champaign in 2011. Moreover, he has been a guest researcher in the Signal & Information Processing Lab at TU Delft in 2014. His research interests include spectral estimation, (sinusoidal) parameter estimation, microphone array processing, as well as statistical and Bayesian methods for signal processing.

Jesper Rindom Jensen (S'09–M'12) was born in Ringkøbing, Denmark in August 1984. He received the M.Sc. degree *cum laude* for completing the elite candidate education in 2009 from Aalborg University in Denmark. In 2012, he received the Ph.D. degree from Aalborg University. Currently, he is a Assistant Professor at the Department of Architecture, Design & Media Technology (AD:MT) at Aalborg University in Denmark, where he is also a member of the Audio Analysis Lab. Before this, he has held two Postdoc positions at AD:MT at Aalborg University. He has been a Visiting Researcher at the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, at the Friedrich-Alexander Universität Erlangen-Nürnberg in Erlangen, Germany, and at the University of Surrey, UK.

His research interests include signal processing theory and methods for, 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. He has published more than 60 papers on these topics in top-tier, peer-reviewed conference proceedings and journals. Moreover, he is the co-author of two books, namely, "Speech Enhancement – A Signal Subspace Perspective" and "Signal Enhancement with Variable Span Linear Filters".

He has received a highly competitive postdoc grant from the Danish Independent Research Council, as well as several travel grants from private foundations. Furthermore, he is an affiliate member of the IEEE Signal Processing Theory and Methods Technical Committee, and is Member of the IEEE.

Mads Græsbøll Christensen (S'00–M'05–SM'11)) received the M.Sc. and Ph.D. degrees in 2002 and 2005, respectively, from Aalborg University (AAU) in Denmark, where he is also currently employed at the Dept. of Architecture, Design & Media Technology as Professor in Audio Processing and is head and founder of the Audio Analysis Lab.

He was formerly with the Dept. of Electronic Systems at AAU and has been held visiting positions at Philips Research Labs, ENST, UCSB, and Columbia University. He has published 3 books and more than 180 papers in peer-reviewed conference proceedings and journals, and he has given tutorials at EUSIPCO, SMC, and INTERSPEECH and a keynote talk at IWAENC. His research interests lie in 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 has received several awards, including the Spar Nord Foundation's Research Prize, a Danish Independent Research Council Young Researcher's Award, the Statoil Prize, and the EURASIP Early Career Award. He is a beneficiary of major grants from the Danish Independent Research Council, the Villum Foundation, and Innovation Fund Denmark. He is an Associate Editor for IEEE/ACM Trans. on Audio, Speech, and Language Processing, a former Associate Editor of IEEE Signal Processing Letters, 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 Senior Member of the IEEE, Member of EURASIP, and Member of the Danish Academy of Technical Sciences.

Søren Holdt Jensen S'87–M'88–SM'00) received the M.Sc. degree in electrical engineering from Aalborg University (AAU), Aalborg, Denmark, in 1988, and the Ph.D. degree (signal processing) from the Technical University of Denmark (DTU), Lyngby, Denmark, in 1995. He is Full Professor at Aalborg University in Speech and Signal Processing. Before joining the Department of Electronic Systems, Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Taastrup (Copenhagen), Denmark; the Electronics Institute of Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNI•C), Lyngby; the Electrical Engineering Department (ESAT-SISTA) of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. His current research interest are in statistical signal processing, numerical algorithms, optimization engineering, machine learning, and digital processing of acoustic, audio, communication, image, multimedia, speech, and video signals. He is co-author of the textbook *Software-Defined GPS and Galileo Receiver—A Single-Frequency Approach*, Birkhäuser, Boston, USA, also translated to Chinese: National Defence Industry Press, China. Prof. Jensen has been Associate Editor for the IEEE Transactions on Signal Processing, IEEE/ACM Transactions on Audio, Speech and Language Processing, Elsevier Signal Processing, and EURASIP Journal on Advances in Signal Processing. He is a recipient of an individual European Community Marie Curie (HCM: Human Capital and Mobility) Fellowship, former Chairman of the IEEE Denmark Section and the IEEE Denmark Section's Signal Processing Chapter (founder and first chairman). He is member of the Danish Academy of Technical Sciences (ATV) and has been member of the Danish Council for Independent Research (2011–2016) appointed by two different Danish Ministers of Science.