GRID SIZE SELECTION FOR NONLINEAR LEAST-SQUARES OPTIMISATION IN SPECTRAL ESTIMATION AND ARRAY PROCESSING

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

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

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

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

ABSTRACT

In many spectral estimation and array processing problems, the process of finding estimates of model parameters often involves the optimisation of a cost function containing multiple peaks and dips. Such non-convex problems are hard to solve using traditional optimisation algorithms developed for convex problems, and computationally intensive grid searches are therefore often used instead. In this paper, we establish an analytical connection between the grid size and the parametrisation of the cost function so that the grid size can be selected as coarsely as possible to lower the computation time. Additionally, we show via three common examples how the grid size depends on parameters such as the number of data points or the number of sensors in DOA estimation. We also demonstrate that the computation time can potentially be lowered by several orders of magnitude by combining a coarse grid search with a local refinement step.

Index Terms— Optimisation, DOA estimation, fundamental frequency estimation, periodogram.

1. INTRODUCTION

A classical problem in signal processing is the search for a global extremum of a cost function containing numerous and closely spaced peaks and dips. An example of such a cost function is shown in Fig. 1, and, as we detail in Sec. 4, the maximiser of this cost function is the estimate of the fundamental frequency of a periodic signal. From a global perspective, any iterative (zero order or derivative based) method is unfortunately likely to converge to a local maximiser instead of the global one, and a grid search is therefore often used instead. However, such a grid search can often be computationally intensive, especially in multi-dimensional problems, so the selection of a proper grid size for the problem at hand is very important.

The problem of finding the fundamental frequency estimate from the cost function in Fig. 1 is an example of a nonlinear least-squares (NLS) optimisation problem which is non-convex and considered to be much more difficult than the class of convex optimisation problems [1]. Unfortunately, NLS optimisation problems frequently occur in spectral estimation and array processing [2, 3]. The reason that we here focus on spectral estimation and array processing is that the NLS optimisation problems within these fields are often solved using the following approach.

1. If the problem is multi-dimensional, decouple it into a series of lower dimensional problems so a grid search is feasible.

The work by T.L. Jensen and J.R. Jensen was partly supported by the Independent Research Council for Technology and Production 4005-00122 and 1337-00084, respectively. The work by M.G. Christensen was supported by the Villum Foundation

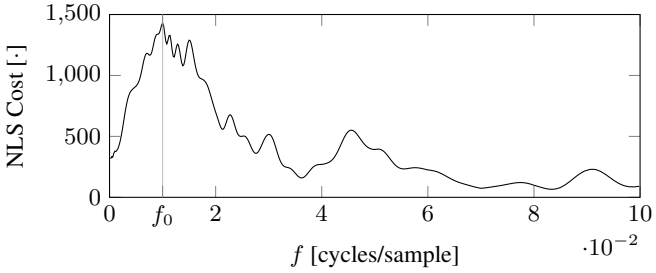

Fig. 1. An example of the cost function of the non-linear least squares (NLS) estimator for fundamental frequency estimation. The true fundamental frequency was set to $f_0 = 0.01$ cycles/sample.

2. Compute the cost function over a uniform grid.
3. Find the optimum on the grid and optionally refine the estimate by using a line search or a derivative based method.

The main motivations for using such a grid-based approach in these applications over alternative and more advanced approaches such as interval analysis combined with branch-and-bound methods (see, e.g., the overview work [4]), are the conceptual simplicity and that the cost function (or a part of it) can often be evaluated using an FFT algorithm. A good example of this is found in [5] and [6] where the three step procedure above was used iteratively, called the RELAX algorithm, and used for the estimation of the frequencies of independent sinusoids and the direction of arrivals of both narrowband and wideband sources, respectively. Other examples include the (approximate) NLS and harmonic MUSIC methods for fundamental frequency estimation [7, 8], the weighted RELAX algorithm for time of arrival estimation [9, 10], the Capon and APES spectral estimators [11, 12], and, recently, the maximum likelihood (ML) estimator for joint fundamental frequency and chirp-rate estimation [13, 14].

A fundamental question in the above procedure is how fine the uniform grid should be? One possibility is to base the grid spacing on the Cramér-Rao lower bound (CRLB) of the signal model as suggested in, e.g., [14], but evaluating the cost function over such a fine grid might be very time consuming, even though it (partly) can be evaluated by an FFT algorithm and eliminates the need for a refinement step. Instead, as suggested in [10], a better approach is to compute the cost function on a relatively coarse grid and then use the refinement step to get the desired accuracy of the estimators. The disadvantage of the latter approach is that we might miss the true peak of the cost function if the grid is too coarse. Therefore, it is desirable to establish a connection between the model parameters of a signal model and the width of the main peak of the cost function at the global optimum. For example, it is well known that the 3 dB half-power width of the main peak of the periodogram is approxi-

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

mately N^{-1} where N is the number of data points [2], but does this also apply to the cost function in Fig. 1? Do we have to decrease the grid spacing for direction of arrival (DOA) estimation if we increase the number of sensors in a uniform circular array (UCA)? And how does the grid-size depend on the array radius?

In the rest of this paper, we answer these and several other questions. First, we introduce the general framework in Sec. 2 and then examine three common problems which are the periodogram in Sec. 3, the NLS and harmonic summation estimators for fundamental frequency estimation in Sec. 4, and the estimator of the range and DOA for a UCA in Sec. 5.

2. FINDING THE GRID SPACING OF A COST FUNCTION

Assume that we have a cost function $f : S \mapsto \mathbb{R}$ which has continuous second-order partial derivatives and a global maximum[1] at $\hat{\boldsymbol{\theta}} \in S \subseteq \mathbb{R}^P$ where S is a subset of the P-dimensional set of real numbers. We assume that the maximum is not on the boundary of S so that the gradient is 0 at $\hat{\boldsymbol{\theta}}$. Consequently, the second order Taylor expansion around this extremum is

$$f(\boldsymbol{\theta}) \approx f(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{H}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{1}$$

where $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{\nabla}^2 f(\boldsymbol{\theta})$ is the Hessian matrix. For this approximation, we would like to find the values of $\boldsymbol{\theta}$ close to $\hat{\boldsymbol{\theta}}$ which satisfies

$$f(\boldsymbol{\theta}) = f(\hat{\boldsymbol{\theta}})/g \tag{2}$$

where $g > 1$ is some constant. For example, we should set $g = 2$ to find the approximate 3 dB bandwidth when $\hat{\boldsymbol{\theta}}$ is a maximiser. Inserting (1) in (2) gives that

$$2f(\hat{\boldsymbol{\theta}})(1-g)g^{-1} \approx (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{H}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \,. \tag{3}$$

From the above, we see that the set of solutions lie on a P-dimensional ellipsoid centred at the extremum. In order to establish how we can use this observation to design the size of the grid cells, we consider two special cases of (3) and then the general case.

2.1. The Scalar Case

In the scalar case $P = 1$, we can easily solve (3) for θ and obtain

$$\theta = \hat{\theta} \pm \Delta\theta = \hat{\theta} \pm \sqrt{2\frac{1-g}{g}\frac{f(\hat{\theta})}{H(\hat{\theta})}} \,. \tag{4}$$

That is, if we increase or decrease the global extremum $\hat{\theta}$ by $\Delta\theta$, then the value of the cost function has approximately changed by a factor of g^{-1}. If we return to the example in Fig. 1, we see that setting $g = 2$ is probably too big. In fact, one can easily construct a case where a $g > 1$ is too big (consider the case where two peaks have almost the same height). Selecting a g that works well in most situations for a particular application is therefore largely a heuristic choice by the user that could be evaluated using Monte Carlo simulations. We will look more into that in the examples later in this paper.

2.2. The Diagonal Case

When the Hessian is a diagonal matrix, we can rewrite (3) into

$$2f(\hat{\boldsymbol{\theta}})(1-g)g^{-1} = \sum_{p=1}^{P} \left[\boldsymbol{H}(\hat{\boldsymbol{\theta}})\right]_{pp} \left(\theta_p - \hat{\theta}_p\right)^2 \tag{5}$$

[1] This is without loss of generality since any minimisation problem can be formulated as a maximisation problem.

where $[\cdot]_{pp}$ denotes the (p,p)th element. We can use (5) to find the relative size of a grid cell and then experiment with the value of g to scale the cell to a suitable size. This approach ensures that over- or undersampling in some of the dimensions is avoided. Thus,

$$\theta_p = \hat{\theta}_p \pm \Delta\theta_p = \hat{\theta}_p \pm \sqrt{2\frac{1-g}{g}\frac{f(\hat{\boldsymbol{\theta}})}{\left[\boldsymbol{H}(\hat{\boldsymbol{\theta}})\right]_{pp}}} \tag{6}$$

for $p = 1, \ldots, P$ by solving (5) for θ_p with $\theta_q = \hat{\theta}_q$ for $q \neq p$.

2.3. The General Case

The general case is more difficult. If the parametrisation of the cost function allows it, one approach is to reparametrise the cost function in the transformed variables $\boldsymbol{\phi} = \boldsymbol{U}^T\boldsymbol{\theta}$ and $\hat{\boldsymbol{\phi}} = \boldsymbol{U}^T\hat{\boldsymbol{\theta}}$ where $\boldsymbol{U}$ contains the eigenvectors of $\boldsymbol{H}(\hat{\boldsymbol{\theta}})$. This allows us to rewrite (3) as

$$2f(\hat{\boldsymbol{\theta}})(1-g)g^{-1} = (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^T \boldsymbol{\Lambda} (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}) \tag{7}$$

so that $\boldsymbol{\Lambda}$ is now a diagonal matrix containing the eigenvalues of the Hessian $\boldsymbol{H}(\hat{\theta})$. Unfortunately, this might be difficult since the Hessian depends on the parameter $\hat{\boldsymbol{\theta}}$ we are trying to find.

3. EXAMPLE 1: THE PERIODOGRAM

The periodogram is the cost function of the ML estimator of the frequency of a single complex sinusoid in white Gaussian noise. For N uniformly sampled observations $\{x(n)\}_{n=0}^{N-1}$, such a signal can be written as

$$\boldsymbol{x} = \boldsymbol{z}(\omega)\alpha + \boldsymbol{e} \tag{8}$$

where α is a complex-valued weight,

$$\boldsymbol{x} = \begin{bmatrix} x(0) & x(1) \cdots & x(N-1) \end{bmatrix}^T \tag{9}$$

$$\boldsymbol{z}(\omega) = \begin{bmatrix} 1 & \exp(j\omega) & \cdots & \exp(j\omega(N-1)) \end{bmatrix}^T \,, \tag{10}$$

and $\boldsymbol{e}$ is defined analogously to $\boldsymbol{x}$. The ML estimator of the frequency $\omega \in [0, 2\pi)$ is given by the solution to [2, pp. 157–162]

$$\hat{\omega} = \underset{\omega \in [0, 2\pi)}{\operatorname{argmax}} \, \phi(\omega) \tag{11}$$

where the cost function $\phi(\omega)$ is the periodogram

$$\phi(\omega) = N^{-1} \left| \boldsymbol{z}^H(\omega)\boldsymbol{x} \right|^2 = N \left| N^{-1} \boldsymbol{z}^H(\omega)\boldsymbol{x} \right|^2 \,. \tag{12}$$

At the maximiser $\hat{\omega}$, the periodogram has the value

$$\phi(\hat{\omega}) = N \left|\hat{\alpha}\right|^2 \tag{13}$$

where $\hat{\alpha}$ is the ML and least squares estimate of α. The second order derivative of the periodogram is approximately given by

$$H(\hat{\omega}) \approx \, - \left|\hat{\alpha}\right|^2 \frac{N(N^2-1)}{6} \tag{14}$$

where the approximation follows by replacing $\boldsymbol{x}$ by $\boldsymbol{z}(\hat{\omega})\hat{\alpha}$. Thus,

$$\Delta\omega = \sqrt{2\frac{1-g}{g}\frac{\phi(\hat{\omega})}{H(\hat{\omega})}} \approx \sqrt{12\frac{g-1}{g}\frac{1}{N}} \tag{15}$$

for N being sufficiently large. The result is hardly surprising and was also found in [2, p. 33] by analysing the width of the main lobe of a rectangular window function. In the case of real-valued data, we obtain the same approximate result for $\Delta\omega$.

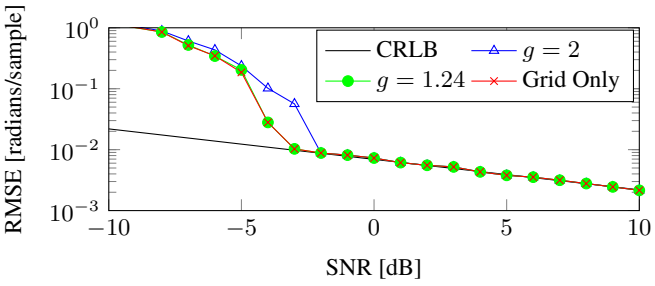

Fig. 2. The RMSE of different implementations of the ML estimator of a single frequency in white Gaussian noise.

g	2	1.24	1.0024	1.0024	1.00024	Grid Only
F	65	100	301	946	2989	28673
τ	1.27	1.25	1.21	1.19	1.27	3.00

Table 1. The computation times τ in ms for different implementations of the ML estimator of a single frequency in white Gaussian noise. The scalar F is the number of FFT-points.

To evaluate the reduction in computation time and the possible loss in estimation accuracy, we conducted a Monte Carlo simulation in which the ML estimate of the fundamental frequency and the computation time were computed over 1000 different noise and phase realisations for various signal-to-noise ratios (SNR) and gains. The simulations were run on a 1st generation Intel Core i7 processor desktop computer running Ubuntu Linux 14.04.3 and MATLAB 8.5, and the code is available online at `http://kom.aau.dk/~jkn/publications/publications.php`. We required that the estimates were computed to within the tolerance predicted by the CRLB at an SNR of 30 dB. For the signal model in (8), the CRLB guarantees that

$$\mathrm{Var}(\hat{\omega}) \geq \frac{6\sigma^2}{|\alpha|^2 N(N^2-1)} \tag{16}$$

where σ^2 is the noise variance. We used $N = 50$ observations in the simulation. From Fig. 2, we see that setting $g = 2$ gave a slightly worse performance than the grid only approach (i.e. FFT without refinement). For $g = 1.24$, however, we basically have the same estimation accuracy between using only an FFT and using an FFT combined with a Fibonacci refinement search [15, pp. 85–92]. Unfortunately, these observations cannot support that a $g = 1.24$ works well in all cases since coloured noise will result in larger spurious peaks in the cost function. Table 1 lists the FFT length and the average of the minimum computation time in each Monte Carlo run for a number of different values of g. The computation times reveal two things. First, that the FFT only approach is more expensive than using a combined approach. Second, the lowest computation time should be selected as a compromise between the FFT length and the number of refinement steps. Finding this optimum, however, is difficult to do in general since it depends on the number of data points and the particular implementation of the FFT and the refinement step. However, our simulations suggest that the combined approach is always faster than the FFT only approach. Finally, the computational gain varies strongly with the data length. For example, the gain is approximately 35 instead of 3, if we increase the data length to $N = 500$.

4. EXAMPLE 2: FUNDAMENTAL FREQUENCY ESTIMATION

The periodic signal model is an extension of the single frequency case considered in Sec. 3. Specifically, it is given by

$$\boldsymbol{x} = \sum_{i=1}^{l} \boldsymbol{z}(i\omega_0)\alpha_i + \boldsymbol{e} = \boldsymbol{Z}(\omega_0)\boldsymbol{\alpha} + \boldsymbol{e} \tag{17}$$

where $\omega_0 \in (0, 2\pi/l)$ is the fundamental frequency, l is the number of harmonic components, and

$$\boldsymbol{Z}(\omega_0) = \begin{bmatrix} \boldsymbol{z}(\omega_0) & \cdots & \boldsymbol{z}(l\omega_0) \end{bmatrix} \tag{18}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_l \end{bmatrix}^T . \tag{19}$$

When $\boldsymbol{e}$ is again a white Gaussian noise vector, the ML estimator of the fundamental frequency is [7]

$$\hat{\omega}_0 = \underset{\omega_0 \in (0, 2\pi/l)}{\operatorname{argmax}} \; C(\omega_0) \tag{20}$$

$$C(\omega_0) = \boldsymbol{x}^H \boldsymbol{Z}(\omega_0) \left[\boldsymbol{Z}^H(\omega_0)\boldsymbol{Z}(\omega_0) \right]^{-1} \boldsymbol{Z}^H(\omega_0)\boldsymbol{x} \tag{21}$$

$$\approx N^{-1} \| \boldsymbol{Z}^H(\omega_0)\boldsymbol{x} \|_2^2 . \tag{22}$$

The approximate cost function is used in the harmonic summation (HS) method [16, 17] and follows from the asymptotic result that

$$\lim_{N\to\infty} N \left[\boldsymbol{Z}^H(\omega_0)\boldsymbol{Z}(\omega_0) \right]^{-1} = \boldsymbol{I}_l \tag{23}$$

where $\boldsymbol{I}_l$ is the $l \times l$ identity matrix. Using this limit as an approximation, we can write the cost function at $\omega_0 = \hat{\omega}_0$ as

$$C(\hat{\omega}_0) \approx N \hat{\boldsymbol{\alpha}}^H \hat{\boldsymbol{\alpha}} \tag{24}$$

where $\hat{\boldsymbol{\alpha}}$ is the least squares estimate of $\boldsymbol{\alpha}$ at $\omega_0 = \hat{\omega}_0$. The second order derivative of the cost function is a lot harder to compute, but under the approximation following from (23), it can be found to [18]

$$H(\hat{\omega}_0) \approx -\frac{N(N^2-1)}{6} \sum_{i=1}^{l} |\hat{\alpha}_i|^2 i^2 \tag{25}$$

which reduces to (14) for $l = 1$. In contrast to the previous example, $\Delta\omega_0$ will in general depend on the complex weights and therefore the data. To avoid this, we select the weights so that $\Delta\omega_0$ is as small as possible, thus producing a lower bound on $\Delta\omega_0$. This lower bound is achieved by selecting $\boldsymbol{\alpha}$ so that

$$\tilde{\boldsymbol{\alpha}} = \begin{array}{ll} \operatorname{argmax} & \sum_{i=1}^{l} |\alpha_i|^2 i^2 \\ \text{subject to} & \boldsymbol{\alpha}^H\boldsymbol{\alpha} = C,\ \boldsymbol{\alpha} \in \mathbb{C}^{l\times 1} \end{array} . \tag{26}$$

By differentiating the Lagrangian of the problem w.r.t. $\boldsymbol{\alpha}^*$, we see that $\tilde{\boldsymbol{\alpha}}$ is related to the eigenvector pertaining to the maximum eigenvalue of the matrix $\mathrm{diag}(1, \ldots, l)^2$ by any complex-valued factor γ lying on a circle with radius $\sqrt{C}$. Thus, we get that

$$\tilde{\boldsymbol{\alpha}} = \begin{bmatrix} 0 & \cdots & 0 & \gamma \end{bmatrix}^T , \quad |\gamma| = \sqrt{C} . \tag{27}$$

Inserting this worst case value for $\hat{\boldsymbol{\alpha}}$ now readily gives

$$\Delta\omega_0 \approx \sqrt{12\frac{g-1}{g}\frac{\hat{\boldsymbol{\alpha}}^H\hat{\boldsymbol{\alpha}}}{\sum_{i=1}^{l}|\hat{\alpha}_i|^2 i^2}} > \sqrt{12\frac{g-1}{g}}\frac{1}{Nl} \tag{28}$$

for N being sufficiently large. Thus, we see that the grid size becomes up to a factor of l smaller compared to the case of a single complex sinusoid. The grid size is the same in the real-valued case.

To evaluate the reduction in computation time and the possible loss in estimation accuracy, we conducted a Monte Carlo simulation consisting of 250 runs for every g and SNR. The data length and the

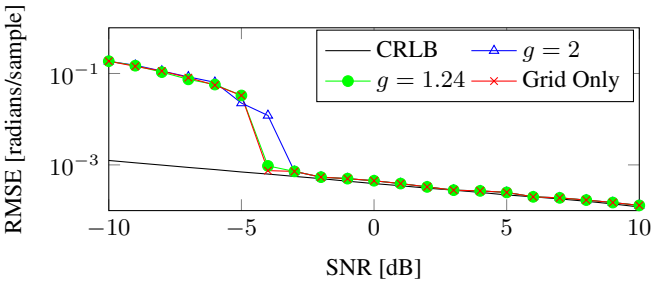

Fig. 3. The RMSE of the ML estimator of the fundamental frequency of a periodic signal in white Gaussian noise.

g	2	1.24	1.024	1.0024	1.00024	Grid Only
F	1285	2000	6011	18911	59770	811116
τ	0.03	0.03	0.03	0.05	0.12	1.42

Table 2. The computation times τ in seconds for different implementations of the ML fundamental frequency estimator. The scalar F is the number of FFT-points.

number of harmonics were set to $N = 100$ and $l = 10$, respectively. In most practical situations, the number of harmonics is unknown and must also be estimated. To do this, a fundamental frequency for all candidate models $l \in \{1, \ldots, L\}$ must be found, and an efficient algorithm for obtaining this is given in [19] which reduces the computational complexity of the ML estimator to the same order as that of the HS method. We therefore evaluate the former here and report the total computation time of computing estimates for all orders up to $L = 10$. We required that the estimates were computed to within the tolerance predicted by the best-case CRLB at an SNR of 30 dB. For the signal model in (17), the CRLB guarantees that [8]

$$\mathrm{Var}(\hat{\omega}_0) \geq \frac{6\sigma^2}{N(N^2-1)\sum_{i=1}^{l}|\alpha_i|^2 i^2} \tag{29}$$

and the best case is obtained by using $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$. In Fig. 3 and Table 2, we have depicted the estimation accuracy and the computation times for various values of g and SNRs. We see that a $g = 1.24$ basically gives the same estimation accuracy as a full grid search, but is approximately a factor of 50 faster.

5. EXAMPLE 3: TOA AND DOA ESTIMATION

In the third and final example, we consider joint time of arrival (TOA) and DOA estimation with a uniform circular array (UCA) and a known source signal in the far-field. As a special case of this problem, we also consider far-field DOA estimation for an unknown source signal. For the data generated by the kth sensor, we use the free-field signal model given by

$$\boldsymbol{x}_k = \beta_k \boldsymbol{s}(\eta_k) + \boldsymbol{e}_k\,, \quad k = 1, \ldots, K \tag{30}$$

where $\beta_k \in \mathbb{R}$ is a gain parameter, $\eta_k \in [0, N)$ is the delay from the source to sensor k in samples, and $\boldsymbol{s}(0) \in \mathbb{R}^N$ is a source signal vector. For a UCA with radius r in meters and a source placed in the far-field at a range of d meters and a DOA of θ radians, the gains are the same (i.e., $\beta_k = \beta$ for $k = 1, \ldots, K$), and the delays are

$$\eta_k = \frac{f_\mathrm{s}}{c}\left(d - r\cos\left(\theta - 2\pi\frac{k-1}{K}\right)\right) \tag{31}$$

where f_s is the sampling frequency in Hz and c is the propagation speed in meters/second. As we have recently detailed in [20], an often implicit assumption on a broadband source signal is that it is periodic in N so that a time-shift becomes a phase-shift in the frequency domain. That is,

$$\boldsymbol{s}(\eta_k) = \sum_{i=-l}^{l} \boldsymbol{z}(i\omega_0)\alpha_i \exp(-ji\omega_0\eta_k) \tag{32}$$

where $l = \lfloor N/2 \rfloor$ is the maximum number of harmonics and $\omega_0 = 2\pi/N$ is the fundamental frequency in radians/sample of the broadband signal model. For notational convenience, we define the angles

$$\xi = \omega_0 d f_\mathrm{s}/c \in [0, 2\pi) \tag{33}$$

$$\rho = \omega_0 r f_\mathrm{s}/c \in [0, 2\pi) \tag{34}$$

which are representations of the range d and radius r in radians. If we concatenate the K data vectors $\{\boldsymbol{x}_k\}_{k=1}^K$ into $\boldsymbol{x}$, we obtain

$$\boldsymbol{x} = \boldsymbol{h}(\xi, \theta)\beta + \boldsymbol{e} \tag{35}$$

$$\boldsymbol{h}(\xi, \theta) = \begin{bmatrix}\boldsymbol{s}^T(\eta_1) & \cdots & \boldsymbol{s}^T(\eta_K)\end{bmatrix}^T\,. \tag{36}$$

If we assume that the noise is white and Gaussian, the ML estimator of the range η and the DOA θ is

$$(\hat{\xi}, \hat{\theta}) = \underset{\xi\in[0,2\pi),\theta\in[0,2\pi)}{\operatorname{argmax}} J(\xi, \theta) \tag{37}$$

$$J(\xi, \theta) = \left|\boldsymbol{h}^T(\xi, \theta)\boldsymbol{x}\right|^2 \tag{38}$$

$$= \left|\boldsymbol{f}^H(\xi)\boldsymbol{A}^H\sum_{k=1}^{K}\boldsymbol{D}_k^H(\theta)\boldsymbol{F}^H\boldsymbol{x}_k\right|^2 \tag{39}$$

where we have defined

$$\boldsymbol{s}(\eta_k) = \boldsymbol{F}\boldsymbol{A}\boldsymbol{D}_k(\theta)\boldsymbol{f}(\xi) \tag{40}$$

$$\boldsymbol{F} = \begin{bmatrix}\boldsymbol{z}(-2\pi l/N) & \cdots & \boldsymbol{z}(2\pi l/N)\end{bmatrix} \tag{41}$$

$$\boldsymbol{f}(\xi) = \begin{bmatrix}\exp(jl\xi) & \cdots & \exp(-jl\xi)\end{bmatrix}^T \tag{42}$$

$$\boldsymbol{D}_k(\theta) = \mathrm{diag}\left(\boldsymbol{f}\left(-\rho\cos\left(\theta - 2\pi\frac{k-1}{K}\right)\right)\right) \tag{43}$$

$$\boldsymbol{A} = \mathrm{diag}(\alpha_{-l}, \ldots, \alpha_l) = N^{-1}\,\mathrm{diag}\left(\boldsymbol{F}^H\boldsymbol{s}(0)\right)\,. \tag{44}$$

Note that $\boldsymbol{F}^H\boldsymbol{x}_k$ and $\boldsymbol{A}$ can both be computed efficiently using an FFT, and that $\boldsymbol{f}(\xi)$ is a shifted inverse DTFT vector so that the cost function can be computed efficiently using an inverse FFT for every candidate θ. At the maximiser $(\hat{\xi}, \hat{\theta})$, we have

$$J(\hat{\xi}, \hat{\theta}) = K^2\hat{\beta}^2 P_\mathrm{s}^2 \tag{45}$$

where P_s is the power of the source signal and given by

$$P_\mathrm{s} = \boldsymbol{s}^T(0)\boldsymbol{s}(0) = N\sum_{i=-l}^{l}|\alpha_i|^2\,. \tag{46}$$

After a lot of tedious algebra, we obtain a diagonal Hessian given by

$$\boldsymbol{H}(\hat{\xi}, \hat{\theta}) \approx -2K^2\hat{\beta}^2 N P_\mathrm{s}\sum_{i=-l}^{l}|\alpha_i|^2 i^2\begin{bmatrix}1 & 0\\ 0 & \rho^2/2\end{bmatrix}\,. \tag{47}$$

For a known source signal, we know $\{\alpha_i\}_{i=-l}^l$. Therefore, the grid sizes for the range and DOA are

$$\Delta\xi \approx \sqrt{\frac{g-1}{g}\frac{N^{-1}P_\mathrm{s}}{\sum_{i=-l}^{l}|\alpha_i|^2 i^2}} > \sqrt{2\frac{g-1}{g}}\frac{1}{N} \tag{48}$$

$$\Delta\theta \approx \sqrt{2\frac{g-1}{g}\frac{N^{-1}P_\mathrm{s}}{\rho^2\sum_{i=-l}^{l}|\alpha_i|^2 i^2}} > \sqrt{4\frac{g-1}{g}}\frac{1}{\rho N} \tag{49}$$

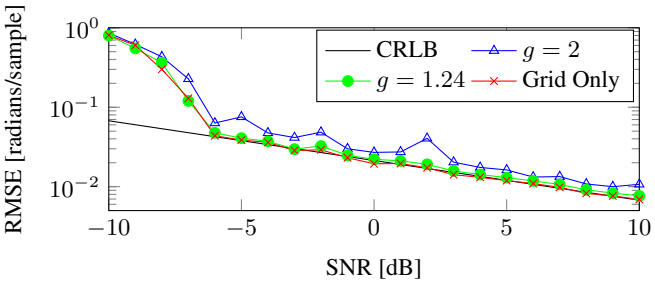

Fig. 4. The RMSE of different implementations of the ML estimator of the DOA.

g	10	2	1.24	1.024	1.0024	Grid Only
F_ξ	192	258	402	1208	3798	16126
F_θ	15	19	30	89	279	1671
τ	0.004	0.004	0.006	0.018	0.105	1.985

Table 3. The computation times τ in seconds for different implementations of the joint TOA and DOA ML estimator. The scalar F is the number of FFT-points.

where the lower bounds follow from inserting $\alpha_{-l}^* = \alpha_l = \gamma$ and $\alpha_i = 0$ for $|i| < l$. From the above grid sizes, we see that neither of them depends on the number K of sensors. This might be somewhat surprising since the CRLB

$$\text{cov}(\xi, \theta) \geq \frac{\sigma^2}{K\beta^2 N \left[\sum_{i=-l}^{l} |\alpha_i|^2 i^2\right]} \begin{bmatrix} 1 & 0 \\ 0 & 2/\rho^2 \end{bmatrix} \qquad (50)$$

depends on it.

For an unknown source signal, we can write the signal model as

$$\boldsymbol{x}_k = \boldsymbol{Z}\boldsymbol{D}_k(\theta)\boldsymbol{A}\boldsymbol{f}(\xi)\beta + \boldsymbol{e}_k = \boldsymbol{Z}\boldsymbol{D}_k(\theta)\boldsymbol{g} + \boldsymbol{e}_k \qquad (51)$$

where $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}(\xi)\beta$ is unknown. By following the same procedure as above, we obtain the exact same grid size $\Delta\theta$ and CRLB for the DOA θ. Thus, the lower bound on the variance of any DOA estimator does not depend on whether we know the source signal or not when the array is a UCA.

To evaluate the reduction in computation time and the possible loss in estimation accuracy, we again conducted a Monte Carlo simulation consisting of 250 runs for every g and SNR. The setup was $N = 100$ data points, $K = 3$ sensors, and an array radius of $r = 0.1$ m. The known source signal was a realisation from a white and Gaussian process, and the TOA and DOA were selected at random in each run. The refinement was performed by first refining the DOA and then the TOA to within a tolerance predicted by the CRLB at 15 dB[2]. In Fig. 4, we have shown the estimation accuracy of the DOA estimator for various values of g and SNRs. Again, we see that a combination of a coarse grid search and a refinement step essentially gives the same estimation accuracy as a full grid search. The computation times, however, were vastly different as detailed in Table 3. For example, a full grid search is more than 350 times slower than a combined search with a $g = 1.24$.

6. CONCLUSION

The optimisation of NLS cost functions is largely an engineering task which requires a careful design to get a good trade-off between the computation time and the probability of finding the global optimum. The results in this paper can help in this design by establishing how the shape of the cost function is related to the model parameters. This is important in order to perform the optimisation as a combination of a coarse grid search and a refinement step. As we demonstrated in the three examples in the paper, the computation time of the combined approach is potentially several orders in magnitude lower than that of a full grid search.

[2]At 30 dB, which we have used in the previous simulations, the memory requirements were too big for the grid only approach.

7. REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.

[2] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice Hall, May 2005.

[3] H. L. V. Trees, *Optimum Array Processing*. Wiley-Interscience, Mar. 2002.

[4] E. Hansen and G. W. Walster, *Global optimization using interval analysis: revised and expanded*. CRC Press, 2003, vol. 264.

[5] 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.

[6] ——, "Angle and waveform estimation via relax," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 1077–1087, 1997.

[7] B. G. Quinn and P. J. Thomson, "Estimating the frequency of a periodic function," *Biometrika*, vol. 78, no. 1, pp. 65–74, Mar. 1991.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Springer, Mar. 2007.

[16] 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.

[17] D. J. Hermes, "Measurement of pitch by subharmonic summation," *J. Acoust. Soc. Am.*, vol. 83, no. 1, pp. 257–264, 1988.

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

[19] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "A fast algorithm for maximum likelihood-based fundamental frequency estimation," in *Proc. European Signal Processing Conf.*, 2015, pp. 589–593.

[20] J. R. Jensen, J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "On frequency domain models for TDOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 11–15.