ML Estimation and Detection of Multiple Frequencies through Periodogram Estimate Refinement

J. Selva

Abstract—This letter presents a method to detect and estimate multiple frequencies based on the maximum likelihood (ML) principle. The method addresses the three main difficulties in this kind of computation, which are the detection of the number of frequencies, the coarse localization of the cost function's global maximum, and the iterative refinement of an initial estimate. Fundamentally, it consists of first detecting and estimating single frequencies or frequency clusters using the periodogram, and then refining this last estimate through a Newton-type method. This second step is fast because its complexity is independent of the number of samples, once a single FFT has been computed. These two steps are iteratively repeated until no mode frequency is above a fixed detection threshold. The main advantage of the proposed method is its low complexity, given that its computational burden is just that of a few FFTs in typical scenarios. The method is assessed in a numerical example.

I. INTRODUCTION

In many signal processing applications, it is necessary to estimate multiple frequencies and there is a wide variety of estimators for this problem in the literature. Among them, the maximum likelihood (ML) estimator is valued for its asymptotical efficiency and its good performance even if the noise is colored [1]–[4]. However, in practice its computation is deemed too complex [5, p. 148], and other sub-optimal estimators that exploit the problem's specific structure are usually preferred, like MUSIC and ESPRIT, [6], [7]. These alternative estimators achieve a complexity reduction but have poorer statistical performance. Also, there exist iterative methods like SAGE [8], that compute the actual ML estimate by decoupling the problem into several one-dimensional minimizations in each iteration. However, they require a high number of iterations to converge.

In the literature, the computation of the ML estimator through standard techniques is deemed too complex for the following reasons. First, the minimization of its cost function must be performed iteratively, through a method like the Modified Variable Projection (MVP) [9], that involves first and second order differentials. Second, these iterative methods require an initial iterate close to the actual cost function's global maximum. And third, it is also necessary to detect

the number of frequencies. In [10], it was shown that each iteration in the MVP method can be performed with low complexity, thus allowing the fast computation of the ML estimate, once an initial iterate with a given number of frequencies is available. Fundamentally, it was shown that the ML estimator can be obtained in two steps. First, one FFT is computed with complexity $O(N \log N)$, where N is the number of samples. And second, each MVP iteration is performed with complexity $O(K^3)$ for K estimated frequencies. In order to assess the relevance of this result, note that computing one MVP iteration through conventional methods has complexity $O(NK^3)$, N is very large (on the thousands), and K small. So, the method in [10] decouples the complexity from $O(NK^3)$ to $O(N \log N) + O(K^3)$. In practice, this decoupling involves a complexity reduction by a large factor.

The purpose of this letter is to present a joint solution for the detection and initial iterate problems just commented, and then combine it with an improved version of the method in [10]. The outcome is a fast method to detect and estimate multiple frequencies based on the ML principle. We propose to employ as initial iterate the set of abscissas corresponding to the periodogram peaks, that lie above a fixed detection threshold. In doing so, we are viewing the periodogram as a frequency cluster estimator. Then, additional frequencies are added if necessary and the method in [10] applied again, until all the modes have been detected and estimated.

The letter has been organized as follows. In the next section, we introduce the multiple frequency estimation problem, and recall the ML estimator and the MVP method. Then, in Sec. III we shortly describe the fast computation of the MVP iteration and related functions. Afterward, Sec. IV presents the proposed estimation and detection method. Finally, the statistical and computational performances are assessed in Sec. V.

A. Notation

In the paper, we adopt the following notation:

- Vectors and matrices are written in lower and upper bold face respectively, (vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$). $\boldsymbol{I}$ denotes an identity matrix of proper size, and '⊙' is the element-by-element product of two equal-size matrices or vectors.
- $[\boldsymbol{x}]_k$ and $[\boldsymbol{X}]_{p,q}$ are the kth and (p, q) elements of $\boldsymbol{x}$ and $\boldsymbol{X}$ respectively.
- $\boldsymbol{A}^\dagger$ denotes the pseudo-inverse of a matrix $\boldsymbol{A}$. If $\boldsymbol{A}$ has full-column rank, then $\boldsymbol{A}^\dagger = (\boldsymbol{A}^H \boldsymbol{A})^{-1} \boldsymbol{A}^H$.
- '≡' is used for introducing new symbols or functions.

II. Signal Model for Multiple Frequency Estimation, ML Estimator, and MVP Method

Consider a signal $x(t)$ consisting of K undamped exponentials with complex amplitudes a_k and distinct frequencies f_k, which is observed in the presence of a complex white Gaussian noise process $\epsilon(t)$ of variance σ^2. The model for the samples $x_n \equiv x(n)$ at instants $n = 0, 1, \ldots, N - 1$ is

$$x_n = \sum_{k=1}^{K} a_k e^{j2\pi n f_k} + \epsilon_n, \tag{1}$$

where $\epsilon_n \equiv \epsilon(n)$. The noise process characteristics imply that the samples ϵ_n are independent with equal variance σ^2 and follow a complex circularly-symmetric Gaussian distribution. Stacking the last equation column-wise we obtain the well-known model

$$\boldsymbol{x} = \boldsymbol{\Phi}(\boldsymbol{f})\boldsymbol{a} + \boldsymbol{\epsilon}, \tag{2}$$

where $(k = 1, 2, \ldots, K)$

$$[\boldsymbol{x}]_{n+1} \equiv x_n, \qquad [\boldsymbol{f}]_k \equiv f_k, \quad [\boldsymbol{\phi}(f)]_{n+1} \equiv e^{j2\pi nf},$$
$$[\boldsymbol{\Phi}(\boldsymbol{f})]_{\cdot,k} \equiv \boldsymbol{\phi}(f_k), \quad [\boldsymbol{\epsilon}]_{n+1} \equiv \epsilon_n.$$

Since $\boldsymbol{\epsilon}$ has zero mean and covariance $\boldsymbol{I}\sigma^2$, the ML estimator of $\boldsymbol{f}$ and $\boldsymbol{a}$ is identical to the least squares estimator [5, p. 147],

$$\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{f}}\} = \arg\min_{\boldsymbol{a}, \boldsymbol{f}} \|\boldsymbol{x} - \boldsymbol{\Phi}(\boldsymbol{f})\boldsymbol{a}\|^2. \tag{3}$$

Besides, this cost function can be concentrated in the coefficients $\boldsymbol{a}$ using the pseudo-inverse $\boldsymbol{\Phi}^{\dagger}(\boldsymbol{f})$. If this is done, it turns out that the ML estimate of the frequencies is also given by

$$\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} L(\boldsymbol{f}), \tag{4}$$

where

$$L(\boldsymbol{f}) \equiv \|\boldsymbol{\Phi}(\boldsymbol{f})\boldsymbol{\Phi}^{\dagger}(\boldsymbol{f})\boldsymbol{x}\|^2. \tag{5}$$

The MVP method optimizes (4) iteratively from a given initial estimate of $\boldsymbol{f}$ using a Newton-type iteration, in which the gradient and the approximation to the Hessian are given by

$$\boldsymbol{g}(\boldsymbol{f}) \equiv -2\mathrm{Re}\big\{(\boldsymbol{\Phi}^{\dagger}\boldsymbol{x}) \odot \big(\boldsymbol{D}^H(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x}\big)^*\big\}, \tag{6}$$
$$\boldsymbol{H}(\boldsymbol{f}) \equiv 2\mathrm{Re}\big\{\big(\boldsymbol{D}^H(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{D}\big) \odot \big((\boldsymbol{\Phi}^{\dagger}\boldsymbol{x}) \cdot (\boldsymbol{\Phi}^{\dagger}\boldsymbol{x})^H\big)^*\big\},$$

where we have written $\boldsymbol{\Phi}$ rather than $\boldsymbol{\Phi}(\boldsymbol{f})$ for simplicity and $[\boldsymbol{D}]_{\cdot,k} \equiv \boldsymbol{\phi}'(f_k)$, [10].

III. Fast Evaluation of the Cost and Related Functions

As already commented in the introduction, the optimization of (4) is deemed too complex in the literature, due to the evaluation of $L(\boldsymbol{f})$ and its differentials, which is viewed as too costly. However, in [10] a technique was presented that allows one to perform these evaluations inexpensively. The techinque starts by computing the zero padded FFT of $\boldsymbol{x}$ of size $M > N$ (usually $M = 2N$). Then, if $\tilde{\boldsymbol{x}}$ denotes this FFT output, the technique makes it possible to interpolate $L(\boldsymbol{f})$ and various functions related with this cost function from $\tilde{\boldsymbol{x}}$ in a small number of operations. Actually, the complexity of these interpolations is independent of the data size N. For what we

require in this letter, we may interpolate from $\tilde{\boldsymbol{x}}$: $L(\boldsymbol{f})$, $\boldsymbol{g}(\boldsymbol{f})$, $\boldsymbol{H}(\boldsymbol{f})$, and the periodogram introduced in the sequel.

To see how this kind of interpolation works for $L(\boldsymbol{f})$, let us substitute the expression of the pseudo-inverse into (5),

$$L(\boldsymbol{f}) = \boldsymbol{x}^H \boldsymbol{\Phi} (\boldsymbol{\Phi}^H \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^H \boldsymbol{x}. \tag{7}$$

Note that in this expression $\boldsymbol{\Phi}$ is a very "tall" matrix, given that its number of rows is equal to the number of data samples N (on the thousands), but its number of columns is equal to the number of frequencies K in the estimation problem, a small number. This implies that most of the computational burden in evaluating (7) is taken up by two products, namely $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^H \boldsymbol{x}$. However, the product $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ must not be performed at all, given that there is a closed-form formula for its components. Specifically, we have $[\boldsymbol{\Phi}^H \boldsymbol{\Phi}]_{k,r} \equiv \gamma(f_r - f_k)$, $(k, r = 1, 2, \ldots, K)$, where $\gamma(f)$ is the complex spectral window defined by

$$\sum_{n=0}^{N-1} e^{j2\pi nf} = \gamma(f) \equiv \begin{cases} N & \text{if } f = 0 \\ \dfrac{e^{j2\pi Nf} - 1}{e^{j2\pi f} - 1} & \text{otherwise.} \end{cases} \tag{8}$$

As to the second product $\boldsymbol{\Phi}^H \boldsymbol{x}$, its components are samples of the spectrum of $\boldsymbol{x}$, i.e, $[\boldsymbol{\Phi}^H \boldsymbol{x}]_k = X(f_k)$, defined by

$$X(f) \equiv \sum_{n=0}^{N-1} x_n e^{-j2\pi nf}. \tag{9}$$

Thus, the evaluation of $\boldsymbol{\phi}(f)^H \boldsymbol{x}$ comes down to evaluating $X(f)$ at arbitrary frequencies. At this point, the relevant fact is that there exist very accurate barycentric formulas for interpolating $X(f)$, [11], [12]. More precisely, for an integer $M > N$, we may interpolate $X(f)$ from the $2P + 1$ samples of the form m/M closer to f in the following steps:

1) If $f = k/M + u$ is a modulo-$(1/M)$ decomposition of f, (integer k and $-1 \leq 2Mu < 1$), compute the coefficients

$$a_p = \frac{w_p}{u - p/M}, \quad -P \leq p \leq P, \tag{10}$$

where w_p is a set of constants.

2) Interpolate $X(f)$ using the barycentric formula

$$X(f) \approx \sum_{p=-P}^{P} X((k - p)/M)a_p \bigg/ \sum_{p=-P}^{P} a_p. \tag{11}$$

Note that (10) and (11) just involve a few arithmetic operations for each input sample $X((k - p)/M)$. Following the last references, one may choose w_p so that the interpolation error in (11) decreases exponentially with P. So, in practice (11) is able to interpolate $X(f)$ with high accuracy for small P. An additional point is the computation of the samples $X((k - p)/M)$ in (11). But note that these samples are just components of $\tilde{\boldsymbol{x}}$, the M-length FFT of $\boldsymbol{x}$.

In summary, for an arbitrary frequency vector $\boldsymbol{f}$, $L(\boldsymbol{f})$ can be evaluated in the following steps,

1) Compute $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ using (8). [Complexity $O(K^2)$.]
2) Compute $\boldsymbol{\Phi}^H \boldsymbol{x}$ using (11). [Complexity $O(KP)$.]
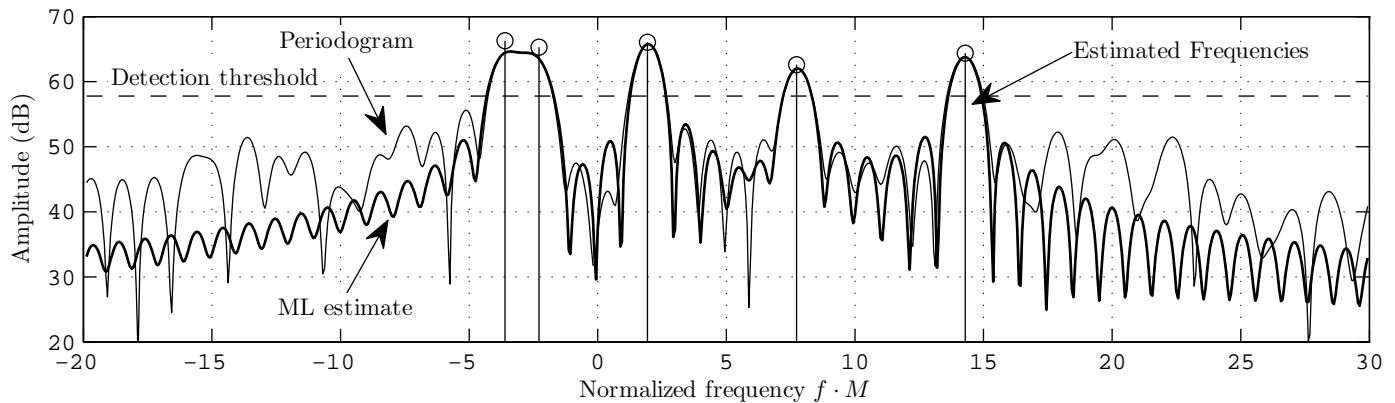3) Compute the remaining operations in (7). [$O(K^3)$.]

Fig. 1. Periodogram realization (5 frequencies) and corresponding ML estimate. The stems indicate the module and position of the estimated modes. Note that the periodogram estimate would place a single frequency in the first local maximum above the threshold.

As shown in [10], this interpolation technique can be applied as well to the cost function's gradient and approximate Hessian in (6), $g(f)$ and $H(f)$. This is so because these functions only depend on $\Phi^H \Phi$ and $\Phi^H x$ and their derivatives. For details, see [10].

Finally, the fast evaluation can also be applied to an extended version of the Schuster periodogram, that will be a key function in the detection method in the next section. We will call "residual periodogram" to the Schuster periodogram of x once the contribution of a set of modes, specified by a vector of frequencies η, has been subtracted. Its definition is

$$L_1(f, x; \eta) \equiv |\phi(f)^H (I - \Phi(\eta)\Phi^\dagger(\eta))x|^2. \quad (12)$$

Besides, if η is empty then we will denote this periodogram simply as $L_1(f, x)$ and it will coincide with the usual periodogram, $L_1(f, x) = |\phi(f)^H x|^2$. The sampled residual periodogram will be (12) but restricted to the frequencies m/M, $m = 0, 1, \ldots, M-1$, and is defined by

$$\begin{aligned} L_{1,s}(m, x; \gamma) &\equiv L_1(m/M, x; \gamma), \\ L_{1,s}(m, x) &\equiv L_1(m/M, x). \end{aligned} \quad (13)$$

IV. PROPOSED DETECTION METHOD

The detection method proposed in this letter is a combination of two procedures that improve on a given estimate specified by a vector of frequencies f_o. The method starts with an empty f_o and applies the procedures successively until there is no change in this vector.

The first procedure works as follows. If the modes specified by f_o approximately match modes actually present in x, then the residual $x - \Phi(f_o)\Phi(f_o)^\dagger x$ is approximately composed by the remaining modes plus noise. This composition is reflected in the periodogram $L_{1,s}(m, x; f_o)$ in which the local peaks with high amplitude are mainly produced by one or more of the remaining modes, while the low-amplitude peaks are produced by the noise. Fig. 1 is a typical periodogram of this kind. The procedure first detects the significant peaks and then appends the frequency m_1/M to f_o for each of them, where m_1 is the peak's abscissa. For assessing whether a given peak is significant, the procedure uses a test for unlikely high noise

peaks. More precisely, if $x - \Phi(f_o)\Phi(f_o)^\dagger x$ just contains noise, then the procedure declares a given peak as significant if its amplitude is above a threshold A. This threshold is pre-computed from the test

$$Prob(\max_m L_{1,s}(m, \epsilon) > A) = P_{FA}, \quad (14)$$

where $Prob(\cdot)$ denotes probability, ϵ is a noise vector as in (2), and P_{FA} a given false-alarm probability. The threshold A is pre-computed through Monte Carlo simulation. In (14), we implicitly assume that there is available an estimate of the noise variance σ^2.

The second procedure fundamentally consists of refining f_o using the MVP iterative process in which $L(f)$, $g(f)$, and $H(f)$ are computed through the interpolation method in the previous section. The MVP process is stopped when the cost function increase becomes negligible. During the MVP iterations, the components of f_o are checked for possible frequency pairs whose difference is too small, before the cost function or any of its differentials is evaluated. If one such pair is found, then one of the two frequencies is eliminated from f_o. Note that the mode corresponding to the eliminated frequency is not estimable, because the small difference produces a large value of the Cramer-Rao bound of both frequencies. The minimum frequency difference is the one ensuring that the logarithm of the condition number of $\Phi(f'_o)^H \Phi(f'_o)$ is below half the numerical precision, where f'_o is a vector containing only the frequency pair. For example, if working with double precision then the condition number must be smaller than $10^{15.65/2}$.

A. Comments

The threshold in (14) detects unusual noise peaks, but is employed to detect several modes that may interfere with each other. Thus, the significance of the detected modes is not jointly tested, and this is a weak point of the proposed detector. Whether this is a drawback or not depends on the specific application but, if necessary, the computed frequencies and amplitudes can be used to evaluate with low complexity a more robust detector like the Generalized Likelihood Ratio Test, [13, Ch. 6]. Besides, the interpolation techniques in Sec.
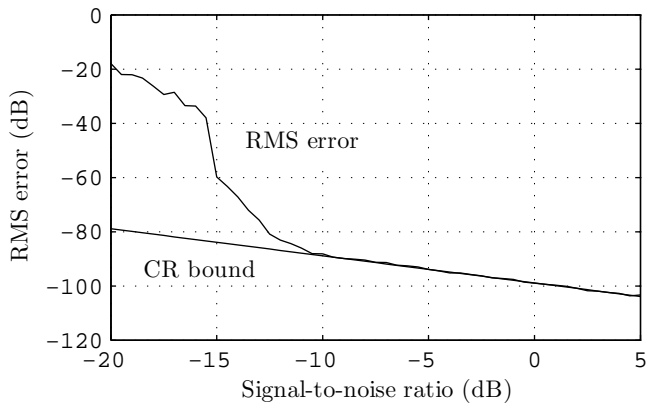
(a) RMS error of first frequency and corresponding Cramer-Rao bound.
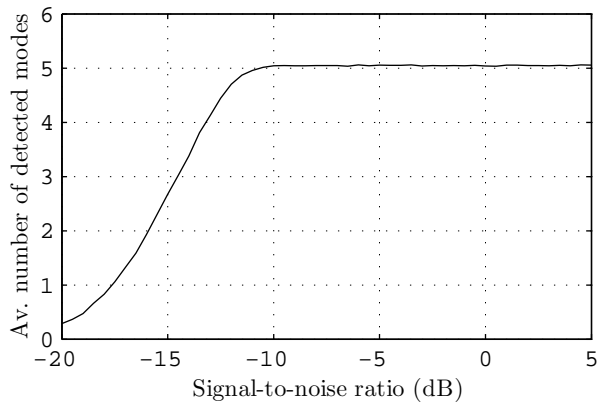
(b) Average number of detected frequencies.
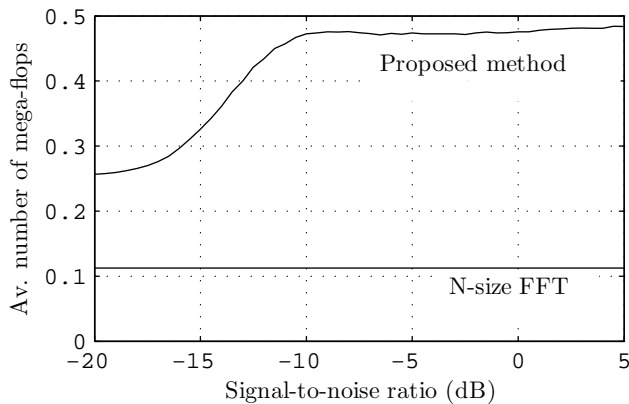
(c) Computational cost in millions of floating-point operations (Mega flops).

Fig. 2. RMS error (dB), average number of detected frequencies, and computational cost (mega-flops) of proposed estimator/detector.

III can be used to eliminate non-significant modes and re-adjust the ML estimation in such a test. It is worth mentioning that there exist analytical thresholds for real periodograms that seem adaptable to (14); see [14].

V. NUMERICAL EXAMPLE

The proposed method has been assessed in a scenario with five modes ($K = 5$) and noise samples following the distribution specified in Sec. II. The modes' frequencies and amplitudes were the following:

Frequencies ($f \cdot N$):	-3.61, -2.31, 1.92, 7.65, 14.24.
Modules $|a|$:	1, 0.803, 0.958, 0.687, 0.678.
Phases $\arg\{a\}$ (rad):	0.489, 0.599, 1.43, 2.24, -1.66.

Fig. 1 is a typical snapshot in this scenario. The number of data and periodogram samples were $N = 2048$ and $M = 4096$, and the interpolation order was $P = 6$. The false alarm probability was set to $P_{FA} = 0.05$, and the corresponding threshold was obtained through a Monte Carlo simulation and its value was $A = 2.305 \cdot 10^4 \sigma^2$. Note that the separation between the first two frequencies was only $(f_2 - f_1)N = 1.3$ samples.

Fig. 2(a) shows the root-mean-square (RMS) error of the first frequency together with the corresponding Cramer-Rao bound. As expected for the ML estimator, the RMS error converges to this last bound for sufficiently high signal-to-noise (SNR) ratio.

Fig. 2(b) shows the average number of detected frequencies. We can see that the method was able to make out the five frequencies at approximately the RMS threshold region. In this example, at most two iterations were necessary. In the first, the four lobes in Fig. 1 were detected, and in the second an additional frequency was detected in the first lobe. Note that in Fig. 2(b) the number of detected frequencies is slightly above 5 for high SNRs. Actually, it is roughly equal to 5.05 because the false alarm probability was set to 0.05, i.e, occasionally there was a small noisy mode.

Finally, Fig. 2(c) presents the computational cost in millions of floating-point operations. Approximately 0.48 MFlops sufficed to detect and estimate the five frequencies. This cost is very small. Just for comparison, note that the cost of one size-N FFT is $5N \log_2(N) = 0.112$ MFlops. So, we have that the computational cost of the whole estimation and detection is roughly equal to that of $0.48/0.112 = 4.29$ N-size FFTs. For a comparison with other methods in the literature, see [10, Sec. VII.E], where a less efficient interpolation method already provided roughly a factor-six improvement in computational burden over the existing methods.

VI. CONCLUSIONS

We have presented a method to estimate and detect multiple frequencies based on the ML principle. The method employs the periodogram to detect frequency clusters, and the MVP algorithm to refine a given initial estimate. It works iteratively and in each iteration adds to the set of detected frequencies those given by the peaks of the residual periodogram. The iterative procedure stops when no peaks are left. The method is very efficient computationally due to the interpolation method, that makes it possible to compute elaborate detection schemes with a complexity that is independent of the data size. The only operation with a complexity proportional to this last size is one initial FFT.

REFERENCES

[1] P. Stoica and A. Nehorai, "Statistical analysis of two non-linear least-squares estimators of sine waves parameters in the colored noise,"

in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr 1988, pp. 2408–2411.

[2] Petre Stoica, Randolph L. Moses, Benjamin Friedlander, and Torsten Söderström, "Maximum likelihood estimation of the paramaters of multiple sinusoids from noisy measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392, Mar. 1989.

[3] F. Andersson, M. Carlsson, J. Y. Tourneret, and H. Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5761–5774, Nov 2014.

[4] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5066–5081, Oct 2016.

[5] P. Stoica and R. Moses, *Introduction to spectral analysis*, Prentice Hall, 1997.

[6] Ralph O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on and Antennas Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[7] R Roy, A. Paulraj, et al., "ESPRIT–a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 1340–1342, Oct. 1986.

[8] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 434–450, Mar 1999.

[9] L. Kaufman, "A variable projection method for solving separable nonlinear least squares problems," *BIT*, vol. 15, pp. 49–57, 1975.

[10] J. Selva, "An efficient Newton-type method for the computation of ML estimators in a Uniform Linear Array," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2036–2045, June 2005.

[11] J. Selva, "Design of barycentric interpolators for uniform and nonuniform sampling grids," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1618 –1627, Mar 2010.

[12] J. Selva, "Efficient maximum likelihood estimation of a 2-d complex sinusoidal based on barycentric interpolation," in *International Conference on Acoustics Speech, and Signal Processing*, May 2011, pp. 4212–4215.

[13] Steven M. Kay, *Fundamentals of Statistical Signal Processing. Detection Theory*, vol. 2 of *Signal Processing*, Prentice Hall, first edition, 1993.

[14] R. V. Baluev, "Assessing statistical significance of periodogram peaks," *Monthly Notices of the Royal Astronomical Society*, , no. 385, pp. 1279–1285, Dec 2008.