Preprints of the
15th IFAC Symposium on System Identification
Saint-Malo, France, July 6-8, 2009

Identification of real sinusoids in noise, the global matched filter approach.

Jean Jacques Fuchs *

* *Irisa-Université de Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France.(e-mail: fuchs@irisa.fr).*

Abstract: We consider the basic identification of real sinusoids in Gaussian noise problem. Since the maximum likelihood function is ill-conditioned and possesses numerous local maxima, we propose to somehow approximate it by a function that is convex and easy to optimize. The procedure amounts to apply the Global Matched Filter to the present problem using a complex redundant basis. It performs simultaneously the detection of the sinusoids and the identification of their characteristics. From a detection point of view, it is similar to a generalized likelihood ratio approach. The performances are close to the Cramer-Rao bounds for scenarios where few competing methods work. The approach can handle arbitrarily sampled data without further difficulties.

Keywords: detection, estimation, spectral estimation, optimization

1. INTRODUCTION

The problem of estimating the characteristics of real sinusoids from noisy measurement has received attention for many decades and is still a subject of investigation since it has applications in many different fields. There clearly does not exist a universal approach that works optimally for all kind of scenarios. The periodogram [1, 2] has reasonable performances in most cases but has limited resolution capacities.

We consider the specific context where the number of data points is relatively small (below 200, say) and the signal to noise ratio (SNR) below 10 dB. This already covers a wide variety of practical problems. It is also a domain where it is difficult to outperform the basic periodogram (the matched filter) and especially so for the detection problem. Indeed, the main difficulties in this specific context, are the detection of a weak isolated sinusoid, the detection of a weak sinusoid hidden by a stronger one and the resolution of two closely spaced sinusoids as well, of course, as a mixture of these difficulties. Remember that around an SNR of 0 dB, the periodogram separates equi-powered sinusoids whose frequency separation is of order $1/n$ with n the number of data points, the so-called Rayleigh resolution limit. Depending upon the SNR, we improve upon this limit by a factor 2 or 3.

We propose to apply the Global Matched Filter (GMF) [3] to this problem. It would now be called a sparse approximation approach. It consists in representing the observations on a adequately chosen redundant basis. Since the basis is redundant there are many possible representation, it happens that the sparsest, i.e. the one using the fewest basis elements, is close, in some sense, to the exact representation. From a detection point of view, we show that the GMF is similar to a generalized likelihood ratio approach. From an identification point of view, the frequency estimates are excellent but the amplitudes and initial phases estimates are slightly biased. We propose a simple way to unbias them or suggest to use them as initial estimates in a maximum likelihood algorithm.

2. THE MODEL

We consider the following noise corrupted sinusoidal signals

$$y_k = s_k + e_k, \quad \text{with} \quad s_k = \sum_{j=1}^{P} A_j^e \cos(2\pi f_j^e + \varphi_j^e)$$

where $k \in (1, .., n)$. We address the problem of the estimation of the number P of sinusoids and simultaneously the problem of the identification of the characteristics $\{A_j^e,\ f_j^e,\ \varphi_j^e\}$ of each sinusoid which we consider as non-random constant parameters. Note that an estimate of the noise variance σ_e^2 is then easy to obtain a posteriori and that there is no loss in performance by proceeding this way.

For later use, we will describe more compactly the set of n observations $\{y_k\}$ in matrix form

$$\mathbf{y} = \mathbf{s} + \mathbf{e} = \sum_{j=1}^{P} A_j^e \mathbf{c}(f_j^e,\ \varphi_j^e) + \mathbf{e} \tag{1}$$

where $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{e}$ are n-dimensional real vector and $\mathbf{c}(f,\ \varphi)$ an n-dimensional vector function with obvious meaning.

Provided P, the number of sinusoids present in the observations is known, the maximum likelihood (ML) estimates of the $3P$ unknowns are obtained by solving

$$\min_{\{A_j, f_j, \varphi_j\},\ j\in(1,..,P)} \| \mathbf{y} - \sum_{j=1}^{P} A_j \mathbf{c}(f_j,\ \varphi_j)\|^2.$$

But for this approach to work, one needs to initialize the search for the minimum with a quite precise initial point that lies in the domain of a attraction of the global minimum which is known to be quite narrow. The means that the problem is actually to find a good initial point for this algorithm, the minimization itself being then quite straightforward since the problem is then locally convex with Hessian the Fisher information matrix. In summary, the presence of many local extrema for the likelihood function makes the ML approach of no help. Its performance are entirely dependent upon the quality of the point used to initialize it and it should be merely considered as a mean to improve the initial point that has already to be quite excellent for difficult scenarios [4]. Note also that in order to be able to implement this ML approach, the number P of sinusoids has to be known.

Due to limited length, it is difficult to present an overview of the existing methods for such an old standing subject. We refer the reader to the textbooks [1, 2] for a good panorama.

3. THE PROPOSED APPROACH

To solve simultaneously the detection and estimation problem, we propose to implement the GMF based on a redundant complex basis of $m \gg 2n$ cisoids at equispaced frequencies $f_k = k/m, \quad k \in \{0, m-1\}$, or equivalently on the columns vectors of the (n, m)-dimensional matrix, we denote A in the sequel, formed by the n first rows of $\bar{F}$, the conjugate of the m-dimensional square Fourier matrix F.

The Fourier matrix F of order m has component (p, q) equal to $\exp(-2\ i\pi(p-1)(q-1)/m)$. This matrix is symmetric and satisfies $F\bar{F} = mI_m$. The discrete Fourier transform (DFT) of a m-dimensional signal y is then $Y = Fy$ and the inverse discrete Fourier transform (IDFT) of Y is $y = (1/m)\ \bar{F}Y$. Note also that, with a_k the k-th column of A anf for even m which we will assume in the sequel to fix ideas, $a_k = \bar{a}_{m+2-k}$, i.e. the columns in A are pairwise complex conjugate. Similarly, for the reconstructed vector Az to be a real vector, the same hermitian symmetry property must hold for z, i.e. $z_k = \bar{z}_{m+2-k}$ for $k = 2$ to $m/2$.

For $m \gg 2n$ there are an infinite number of ways to reconstruct $\mathbf{y}$ (1) using the columns in A, i.e., there are infinitely many complex m-dimensional vectors z such that $\mathbf{y} = Az$. Now, since we are only interested in $\mathbf{s}$, it is preferable to seek approximate reconstructions of $\mathbf{y}$. Since the noise $\mathbf{e}$ is Gaussian, it is judicious to bound the reconstruction errors in terms of the Euclidean norm, i.e. to introduce the following constraint : $\|Az - \mathbf{y}\|_2^2 \leq \rho$. The larger the value of ρ in this constraint, the larger the set of admissible points $\{z\}$ satisfying it and for an adequately chosen ρ, one can expect the sparsest such z, i.e. the z with the smallest number of non-zero components, to have just $2P$ non zero components and to be close to the true decomposition in (1).

Getting the sparsest vector satisfying $\|Az - \mathbf{y}\|_2^2 \leq \rho$ is unfeasible, it can only be obtained by inspection, by checking all the possibilities [5]. A good and close to optimal choice is obtained by seeking the vector z with minimal ℓ_1-norm [6]. One should thus minimize $\|z\|_1$ subject to $\|Az - \mathbf{y}\|_2^2 \leq \rho$, but we choose to use the following equivalent criterion [3]

$$\min_z \frac{1}{2} \|Az - \mathbf{y}\|_2^2 + h\|z\|_1, \qquad h > 0 \tag{2}$$

with $\|z\|_1 = \sum_k |z_k|$ and $\|z\|_2^2 = \sum_k |z_k|^2$ and h a positive real that has to be tuned, just like ρ would have had to be tuned in the reconstruction error constraint.

We thus propose to minimize (2) to solve both the detection and identification problem we are considering. We will explain below how to tune m and h and how to deduce the estimates from the optimal z. The criterion (2) is the one that is the most used in the sparse representation community usually, though in the real case, i.e. for real A, y and z. We already considered this criterion to solve the same sinusoid-in-noise problem in 1998 [7], long before the birth of the current "sparse representation era", in a more elementary version with no initial phase estimation. We improved this version to allow for initial phases in [8] but still using a real sines basis. Introducing a redundant complex cisoids basis (the A matrix) allows to handle the initial phase without any additional concern and lowers the detection threshold. A similar complex basis case has been considered in [9], where the emphasis is on irregularly sampled data and on the algorithmic side. We put the emphasis on the tuning of the parameters, the unbiasing of the estimates and more importantly elaborate upon the detection properties of this approach which has never been analyzed.

4. IMPLEMENTATION ISSUES

We indicate in this section how to tune the positive real parameter h that appears in the criterion, how to deduce the estimates from the optimal solution z of (2), how to fix the number m of components in the basis, how to detect the number of sinusoids and how to unbias these estimates since the presence of a non zero h introduces a bias.

4.1 Setting the threshold

The value of h will have a major influence upon the detection part of the whole procedure. If the value of h is too large, the procedure may not detect weak sinusoids. On the other hand, if h is too small, there may appear many false alarms, i.e., the procedure will detect and identify sinusoids that do not exist. We will thus set h to achieve optimal detection properties and to do so we introduce the dual of the criterion (2) which is

$$\min_z \|Az\|_2 \qquad \text{under} \qquad \|A^H(Az - \mathbf{y})\|_\infty \leq h \tag{3}$$

where $A^H = \bar{A}^T$ and $\|z\|_\infty = \max_k |z_k|$. This dual which is equivalent to (2), is important because it allows to understand the role played by h [3], as we will realize it below. One can already observe from the constraints in (3) that as h increases, so does the size of the admissible domain. Indeed, for h larger than $\|A^H\mathbf{y}\|_\infty$, $z = 0$ becomes admissible and is then the optimum. Roughly speaking, the larger h, the sparser the optimal z and vice versa.

In the ideal case where the optimal z and associated Az represent exactly the sinusoids $\mathbf{s}$ that are present, the residual vector Az - $\mathbf{y}$ becomes equal to the additive noise $\mathbf{e}$. In the dual constraint, the vector $A^H(\mathbf{y} - Az) = A^H\mathbf{e}$

is then a vector of m scalar complex random (dependent) variables build upon the n independent scalar real Gaussian random variables with mean zero and variance σ_e^2 present in $\mathbf{e}$.

Let us consider one of these complex random variables. It is of the form $a^H\mathbf{e}$ with a an n-dimensional complex vector with norm $\sqrt{n}$ and components of the form $\exp(i\omega)$ that can be decomposed into the sum of its real part (cosines) and imaginary part (sines) so that $a^H\mathbf{e} = v_{\cos}^T\mathbf{e} - iv_{\sin}^T\mathbf{e}$ where the two vectors $v_{\sin}$ and $v_{\cos}$ are quasi-orthogonal and of similar norm.

It follows that $(\sqrt{2/(n\sigma_e^2})a^H\mathbf{e}$ is, approximatively, a complex Gaussian random variable whose real and imaginary part are Gaussian, independent, zero mean and with unit variance and thus that $(2/(n\sigma_e^2))|a^H\mathbf{e}|^2$ is, approximatively, a Chi-square random variable with 2 degrees of freedom.

Now, remember that quite generally, for n independent and identically distributed random variables, say w_i

$$P(\max_{i\in(1,n)} w_i > t) = p \;\Leftrightarrow\; P(\max_{i\in(1,n)} w_i < t) = 1-p$$

$$\Leftrightarrow\; P(w_i < t)^n = 1-p \;\Leftrightarrow P(w_i > t) = 1-(1-p)^{(1/n)} \simeq \frac{p}{n},$$

with $P(w_i < t)$ the probability of the random variable w_i to be smaller than t.

Thus, in order to set the threshold h so that probability that the maximum of the moduli of the m dependent random variables in $A^H\mathbf{e}$ exceeds h is p, we propose to take

$$h \simeq \sigma_e\sqrt{\frac{n}{2}}\,\sqrt{t} \;\text{ with }\; Pr(\chi_2^2 > t) \simeq \frac{p}{n}. \tag{4}$$

We apply the standard result developed above, to a set of n independent random variables w_k that are χ_2^2 since the m dependent variables in $A^H\mathbf{e}$ have only n degrees of freedom. We will see in the simulations below (see Section 5.4) that this quick analysis is quite close to the reality, though leading to a threshold h that is slightly too large.

4.2 Getting the estimates

Having obtained the optimum of (2), denoted z below, we explain how to deduce from z the estimates $\{A_j, f_j, \varphi_j\}$ that characterize the different sinusoids. The optimal z is always such that Az is real, the proof of this property is by contradiction. It is easy to show that if this not the case, one can improve upon z. This means that, for even m, $z_k = \bar{z}_{m+2-k}$ for $k = 2$ to $m/2$.

We first consider the case P=1, where there is a single sinusoid present in the observations. For a well chosen h, the optimal z has then just two non-zero complex conjugate weights. The corresponding weighted couple of conjugate complex basis vectors sums into a real cosine vector with initial phase in $(0, 2\pi)$, positive real weight and frequency on the point of the discretization grid associated with the two vectors. More specifically

$$Az = a_k z_k + a_{m+2-k}z_{m+2-k} = 2\Re(a_k z_k)$$

and with $z_k = \rho_k \exp(i\varphi_k)$, the different estimates are

$$\hat{A} = 2\rho_k, \quad \hat{f} = (k-1)/m \quad \text{ and } \quad \hat{\varphi} = \varphi_k\ .$$

The so-obtained frequency estimate being on the discretization grid may not be the best possible. A better choice of h (slightly smaller than the previous one) will yield a optimal z having four non-zero weights, two neighboring couples of complex conjugate weights that will end up with a frequency estimate that does not lie on the discretization grid and may thus well be closer to the true frequency. More precisely, one has to associate a unique *sinusoid* with the sum of the 4 terms in Az that can already be written as $2\Re(a_k z_k + a_{k+1}z_{k+1})$. We then propose to first obtain the frequency estimate by linear interpolation as follows

$$\hat{f} = \frac{\rho_k f_k + \rho_{k+1} f_{k+1}}{\rho_k + \rho_{k+1}},$$

and to deduce the corresponding unique set of estimates $\hat{A}$ and $\hat{\varphi}$ by solving

$$2\rho_k \cos\varphi_k + 2\rho_{k+1}\cos\varphi_{k+1} = \hat{A}\cos\hat{\varphi}$$

$$2\rho_k \sin\varphi_k + 2\rho_{k+1}\sin\varphi_{k+1} = \hat{A}\sin\hat{\varphi}.$$

In case, there are more than 2 contiguous nonzero components in z (which never happens) one similarly estimates first the frequency by linear interpolation and then proceeds in the same way as above.

If there are P>1 sinusoids present in $\mathbf{y}$, one expects the optimal z to have P disjoints sets of the type described above and the same procedure applies to each set. This, of course, corresponds to the situation where the proposed algorithm does a perfect detection job.

4.3 Detecting the number of sinusoids

An estimate of P, the number of sinusoids present in $\mathbf{y}$, is given by the number of (significant) clusters of non-zero components in the optimal z. A cluster being typically a set of two neighboring couples of complex conjugate weights. For a given scenario, the detection performance of the proposed approach depends mainly upon the choice of h in (2), see Section 4.1 and relation (4) above. While for easy scenarios this choice is quite robust, this is no longer true for difficult scenarios. Globally if h is fixed to get a probability of false alarms equal to p, the procedure will detect an additional (false) sinusoid in p percent of the realizations. This additional sinusoid will have a small amplitude estimate and may thus appear suspicious. More rigorously, if our estimates are used to initialize a ML algorithm as we suggest below to unbias the estimates, it is sensible to implement an information criterion like Minimum Description Length (MDL) [11] or the Bayesian Information Criterion (BIC) [12], together with the ML algorithm. This criterion may well discard some of these false additional sinusoids.

The choice of the number m of columns in A and the associated discretization step in frequency $\delta f = 1/m$ is also important in this context. If m it is too large, it will essentially increase the computation time and lead to a poorly conditioned optimization problem. Remember though that the optimization problem (2) is convex and has generically a unique optimum.

If m is too small, it may happen that the two clusters associated with sinusoids having close frequencies merge which means non-separation of the two sinusoids. This situation is easily avoided by adequately choosing m. Indeed knowing roughly the frequency resolution limit, say Δf, associated with the data set, we propose to fix m so that $\Delta f = 10/m$. As an example, for n equispaced samples and signal to noise ratios around 0 dB, the resolution limit (Rayleigh limit) is known to be about $\Delta f = 1/n$ so that one will typically choose $m = 10n$. Taking account of the symmetry in the weights in the optimal z, this choice allows for about 2 zero weights between two sinusoids that are close in frequencies but nevertheless potentially separable and thus guarantees that two disjoint clusters will be obtained in case the two sinusoids are detected.

4.4 Unbiasing the estimates.

Since even for $\mathbf{e} = 0$, the optimal z will never be such that $Az = \mathbf{s}$, for $h > 0$, it is obvious that the presence of the term $h\|z\|_1$ in (2) induces a systematic bias into the estimates.

In the simplest case where the optimal z, which is in fact a $z(h)$, happens to have only one couple of conjugate nonzero components say $z_k = \rho \exp(i\varphi)$ and $z_{m+2-k} = \bar{z}_k$, their optimal, h dependent, values $[\ z_k\ z_{m+2-k}]^T$ can be shown to satisfy [10]

$$\begin{bmatrix} z_k \\ z_{m+2-k} \end{bmatrix}(h) = \bar{A}^{+}\mathbf{y} - h(\bar{A}^{*}\bar{A})^{-1} \begin{bmatrix} \exp(i\varphi) \\ \exp(-i\varphi) \end{bmatrix}$$

with $\bar{A} = [\ a_k\ \ a_{m+2-k}]$ and $\bar{A}^{+}$ the pseudo-inverse of $\bar{A}$. The second term on the right side above represents the bias due to h. To remove it, even in this simple case, is a difficult task since $\exp(i\varphi) = z_k/|z_k|$ so that z_k appears on both sides of the equation which is thus implicit and admits no explicit form.

A simple way to somehow get rid of this bias, which is most important on the estimates of the amplitudes, and the initial phases is to change only the amplitude estimates. If the current estimates are $\{\ \hat{\rho}_k,\ \hat{f}_k,\ \hat{\varphi}_k\ \}$ for $k = 1$ to P and the current best reconstruction of $\mathbf{y}$ thus of the form $\sum_1^P \hat{\rho}_k \mathbf{c}(\hat{f}_k, \hat{\varphi}_k)$ where $\mathbf{c}(.,.)$ is as in (1), then one can replace the amplitudes estimates $\{\ \hat{\rho}_k\}$ by the components in $C^{+}\mathbf{y}$ where C is the (n,P)-matrix with columns vectors the $\mathbf{c}(.,.)$'s.

There are other, more sophisticated but quite simple, ways to unbias both the amplitudes and the initial phases, but a systematic and efficient way is to use the current estimates to initialize a maximum likelihood procedure. In the case where the current estimates lie in the domain of attraction of the global optimum, this will yield the ML estimates. Though non linear the optimization problem is then convex and the convergence guaranteed for quite simple optimization schemes. We will generally unbias the estimates in this way below in the simulations.

4.5 Summary of the algorithm.

Before presenting some simulation results in the next section, we summarize the algorithm we propose. With the n observed samples we build the column vector $\mathbf{y}$ and the A matrix of dimension (n, m) with m equal to $8n$ or $10n$ and components $a_{p,q} = \exp(-2\ i\pi(p-1)(q-1)/m)$. We then solve

$$\min_z\ \frac{1}{2}\ \|Az - \mathbf{y}\|_2^2 + h\|z\|_1, \qquad h > 0$$

with h as defined in Section 4.1 relation (4), this assumes that one has an approximate value of the additive noise variance σ_e^2, but this is not really an issue. We used the subroutines SeDuMi [13] which is a second order cone programming library to get the optimum. For a quick and transitory test purpose, one can implement the recursion

$$z \to z^{+} = |Z|(A|Z|A^H + hI)^{-1}\mathbf{y},$$

with $|Z|$ =diag($|z|$) and initial condition $z_o = A^{+}\mathbf{y}$ which has been proved to converge to the optimum [14]. We then deduce the estimates from the optimum z as described in Section 4.2 and unbias them. We comment, below in the simulations, on how to decide upon the number of sinusoids that are present.

5. SIMULATION RESULTS.

To evaluate the performance of the proposed procedure, we consider mainly the following situation

$$y_k = A_1 \cos(2\pi f_1 k + \varphi_1) + A_2 \cos(2\pi f_2 k + \varphi_2) + e_k$$

with $k \in (1, 2, ...n = 100)$, e_k a sequence of white Gaussian noise with variance $\sigma_e^2 = 1$ and the signal to noise ratio is defined as

$$\mathrm{SNR}_k = 10 \log_{10} \frac{A_k^2}{2\sigma_e^2}\mathrm{dB}.$$

Our purpose is both to detect the number of sinusoids and to identify their characteristics (A_k, f_k, φ_k). For each scenario, we simulate 1000 independent realizations that differ by their noise sequence and indicate the average mean and standard deviation obtained. We also indicate the corresponding Cramer Rao bounds (CRB), we take $m = 1000$ and fix the value of h at its *nominal* value defined in Section 4.1 relation (4) unless otherwise stated.

Note that, in case a preliminary analysis allows to ascertain that the sinusoids that are present have their frequencies lying in a subset of the frequency domain, one can restrict to the same subset the columns in the A-matrix and diminish the computation time while keeping essentially the same performances. We did not use this possibility in the simulations below.

5.1 Example 1.

The two sinusoids are equipowered with SNR=10dB and the frequency separation is 1/300 i.e. one third of the Rayleigh resolution $1/n$. We take $m = 1000$ and $h = 26.28$ as given by relation (4). Our procedure always separates the two sinusoids and we present, in Table 1, the results obtained both after unbiasing the amplitudes of the estimates by the trivial procedure described in Section 4.4 and after applying a ML algorithm applied to the plain estimates we obtained.

In more than 85% of the realizations, there was no false detection, i.e., only two clusters of non-zero components were present in the optimal z, with in general 2 pairs of non-zero components each. In the remaining realizations,

Table 1: Two equipowered sinusoids in white noise. The frequency separation is 1/3 of the Rayleigh limit. SNR=10 dB, n=100 data points, $h = 26.283$.

	Mean	std. dev.	$\sqrt{mse}$	CRB
$A_1 = 4.4721$	3.084	.4629	1.4630	.9110
$A_1 = 4.4721$	3.055	.4436	1.4845	.9136
$f_1 = 0.2000$	.1991	.0007	.0011	.0007
$f_2 = 0.2033$	.2043	.0007	.0011	.0007
$\varphi_1 = 0$	.6226	.0824	.6280	.3380
$\varphi_1 = .785$	.1461	.0813	.6444	.3364

Estimates of the mean, standard deviation and square root of the mean square error averaged over 1000 independent realizations

	Mean	std. dev.	$\sqrt{mse}$	CRB
$A_1 = 4.4721$	4.5298	.6650	.6675	.9110
$A_1 = 4.4721$	4.5368	.6805	.6836	.9136
$f_1 = 0.2000$	.19999	.0006	.0006	.0007
$f_2 = 0.2033$	.2034	.0006	.0006	.0007
$\varphi_1 = 0$	.0333	.2299	.2323	.3380
$\varphi_1 = .785$	.7517	.2296	.2321	.3364

Estimates of the previous Table improved upon using a few steps of a ML algorithm.

there was an additional cluster of, in general, one pair of non-zero components and their average amplitude was .2486, about one tenth of the amplitude of the clusters corresponding to each true sinusoid.

Unbiasing the estimates obtained by the proposed procedure by using them as initial condition in a ML algorithm does indeed improve the results as expected, and especially so the amplitude and initial phase estimates.

5.2 Example 2.

The two sinusoids are equipowered with SNR=0dB and the frequency separation is 1/200 i.e. one half of the Rayleigh resolution $1/n$. We take $m = 1000$. This is actually a more difficult scenario than in Example 1 (the standard deviation of the CRB on the frequencies is multiplied by 2) and especially so as far as the detection is concerned. To detect systematically the two sinusoids, we had to diminish h from its nominal value, we multiplied it by 3/4. Now this may seem arbitrary but the purpose of fixing h as prescribed in Section 4.1 relation (4) is to limit the number of false alarms, i.e., the number of non-zero components in z not corresponding to existing sinusoids. For smaller h, this number increases, but if the associated amplitude estimate is small enough to be easily declared as false alarms it is not really harmful to take a smaller h. In Table 2, we present the results, unbiased by a ML algorithm and averaged over 1000 realizations.

With this reduced value of h, the 2 major clusters in z always correspond to the true sinusoids but there are additional clusters in 86 percent of the realizations with however (average) maximal amplitude of these clusters equal to .077. This means that these false alarms can be discarded by applying the information criterion (MDL) or (BIC) after the ML algorithm used to unbias the estimates, see Section 4.3.

Table 2: Two equipowered sinusoids in white noise. The frequency separation is 1/2 of the Rayleigh limit. SNR= 0 dB, n=100 data points, $h = 19.712$.

	Mean	std. dev.	$\sqrt{mse}$	CRB
$A_1 = 1.41$	1.534	.385	.403	.315
$A_1 = 1.41$	1.542	.395	.416	.317
$f_1 = 0.200$	.1996	.0059	.0059	.0013
$f_2 = 0.205$	.2051	.0013	.0013	.0013
$\varphi_1 = 0$	.011	.446	.446	.539
$\varphi_1 = 0$	-.008	.443	.443	.543

Estimates of the mean, standard deviation and square root of the mean square error averaged over 1000 independent realizations, unbiased using a ML algorithm.

5.3 Example 3.

We now consider the case where the two sinusoids are not equipowered while keeping the same frequency separation at $1/2n = 1/200$. One of the sinusoids is at 0dB and the other at 10 dB. We take $m = 1000$. In this scenario, the strong sinusoid hides the weak one but at the same time it somehow helps the weak one to emerge. The procedure systematically detects the two sinusoids for the nominal $h = 26.28$, there is an additional detection in 16% of the realizations but the corresponding average amplitude is .0638 which means that it is generally discarded by the MDL-BIC detection test. The results over 1000 independent realizations are presented in Table 3.

Table 3: Two sinusoids in white noise with frequency separation 1/2 of the Rayleigh limit and SNR respectively 0db and 10 dB, n=100 data points, $h = 26.283$.

	Mean	std. dev.	$\sqrt{mse}$	CRB
$A_1 = 1.4142$	1.5787	.4470	.4763	.3153
$A_1 = 4.4721$	4.5080	.4111	.4127	.3179
$f_1 = 0.2000$	.1999	.0013	.0013	.0013
$f_2 = 0.2050$	.2050	.0004	.0004	.0004
$\varphi_1 = 0$	.0216	.4614	.4619	.5399
$\varphi_1 = 0$	.0100	.1638	.1641	.1719

Estimates of the mean, standard deviation and square root of the mean square error averaged over 1000 independent realizations, unbiased using a ML algorithm.

5.4 Detection of a single sinusoid

The detection of a sinusoid with known frequency in additive white Gaussian noise can be performed by comparing the output of the matched filter to a threshold [15, 16]. This test procedure is moreover optimal (corresponds to the likelihood ratio test) in the sense that it maximizes the detection probability for a given probability of false alarm. In the case where the frequency of the sinusoid is not known, a suboptimal strategy (generalized likelihood ratio test) is to project the observation on the corresponding manifold and to retain the model leading to the projection with largest modulus. In the present case, this can be achieved by looking at the modulus of the Discrete Fourier Transform and searching for the frequency that leads to the maximum.

To set the threshold, one then has to evaluate the probability density function of these outputs (modulus of the Discrete Fourier Transform) and one fixes the threshold to achieve a given probability of false alarm under H_0. But this is precisely equivalent to the approach we have taken in Section 4.1 since $\|A^H \mathbf{y}\|_\infty$ represents the modulus of the maximum of the Fourier Transform evaluated at $m/2$ frequencies. The proposed procedure is thus exactly an implementation of this strategy.

We simulated 1000 independent realizations, of noise only observations $\mathbf{y} = \mathbf{e}$ and for $h = 26.283$, the value proposed in Section 4.1, relation (4), the procedure detected a sinusoid (had a optimal z not identically zero) for 134 among the 1000 realizations. In all these case it detected a single sinusoid. We unbiased the estimate using the ML algorithm and applied the MDL-BIC criterion to the output. This last step discarded 39 only out of the 134 false alarms. This means that our approximate evaluation of the statistical properties of the moduli of the outputs of the Fourier Transform under H_0 gives a threshold that is slightly too small since it led to a probability of false alarm of 13% while it was tuned to get 10%.

We also simulated a single sinusoid with $A = \sqrt{2}$, i.e. SNR=0dB. For the same value of h, the procedure detects systematically the sinusoid and sometimes detects a second one sinusoid. This is not a surprise since the value of h has been fixed such the probability of false alarms is 10 % and since the presence of a true sinusoid does not preclude the existence of false alarms. We do not present the results obtained over 1000 independent realizations, the proposed procedure attains the Cramer-Rao bounds. Among the 1000 realizations, a second sinusoid is detected in 127 realizations (about 13 percent of the realizations) and the MDL-BIC criterion rejected it in 41 of these 127 realizations.

6. CONCLUDING REMARKS

We have presented an algorithm that allows to simultaneously detect the number of sinusoids present in the observations and to identify their characteristics. Its performance are excellent for SNR's around 0dB and a few hundreds of observations and to our knowledge very few methods having similar performances exist. Especially so if one takes into account the fact that the proposed procedure also applies in case of arbitrary sampling schemes, i.e. for non equispaced data samples. We have shown that for the detection of a single sinusoid, it can be seen as achieving a generalized likelihood ratio test. Since, in the present context, the likelihood ratio test corresponds to the matched filter, it is indeed fair to call a procedure that corresponds to the generalized likelihood ratio test, the Global (or Generalized) Matched Filter (GMF), a term that we introduced in [3] for this type of algorithms.

REFERENCES

[1] B. Porat, "Digital Processing of Random Signals. Theory and Methods." *Prentice Hall*, New Jersey, 1994.

[2] P. Stoica and R. Moses, "Introduction to Spectral Analysis." *Prentice Hall*, New Jersey, 1997.

[3] J.J. Fuchs, "On the Application of the Global Matched Filter to DOA Estimation with Uniform Circular Arrays", *IEEE-T-SP*, vol. 49, p. 702–709, April 2001.

[4] P. Stoica et al, "Maximum Likelihood Estimation of Multiple Sinusoids from Noisy Measurements." *IEEE-T-ASSP* 17, 178–392, Mar. 1989.

[5] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, 24, 21, 227-234, April 1995.

[6] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. on I.T.*, 49, 12, 3320-3325, Dec. 2003.

[7] N. Moal and J.J. Fuchs, "Sinusoids in White Noise: a Quadratic Programming Approach." *Proc. IEEE ICASSP*, vol. III, pp. 2221-2224, Seattle, 1998.

[8] J.J. Fuchs, "On the use of sparse representations in the identification of line spectra" *Proc. IFAC World Congress*,10225-10229, Seoul, 2008.

[9] S. Bourguignon, H. Carfantan and T. Bohm, "SparSpec : a new method for fitting multiple sinusoids with irregularly sampled data", *Astronomy and Astrophysics*, vol. 462, pp. 379-387, 2007.

[10] J.J. Fuchs. "More on sparse representations in arbitrary bases". *IEEE Trans. on I.T.* 50, 6, 1341-1344, June 2004, also in *13th IFAC SYSID*, 1357-1362, Rotterdam, 2003.

[11] J. Rissanen, "Minimum Description Length Denoising". *IEEE Trans. Information Theory*, 46(7), pp. 2537-2543

[12] G. Schwarz, "Estimating the dimension of a model". *Annals of Statistics*, 6(2):461-464, 1978.

[13] SeDuMi, http://sedumi.mcmaster.ca

[14] J.J. Fuchs, "Convergence of a Sparse Representations Algorithm Applicable to Real or Complex Data", *IEEE Journ. of STSP* 1, 598-605, Dec. 2007.

[15] H.V. Poor. "An Introduction to Signal Estimation and Detection". *Springer Texts in E.E.* 1994.

[16] K.W. Helstrom. "Elements of Signal Detection and Estimation". *Prentice Hall* 1995.