

SINUSOIDS IN WHITE NOISE: A QUADRATIC PROGRAMMING APPROACH

Nicolas MOAL and Jean-Jacques FUCHS

IRISA/Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex - France
nmoal@irisa.fr - fuchs@irisa.fr

ABSTRACT

We address the problem of the estimation and identification of real sinusoids in white Gaussian noise using a correlation-based method. We estimate a partial covariance sequence from the data and seek a representation of these *new* observations as a superposition of a small number of cosines chosen from a redundant basis and the white noise contribution. We propose to minimize a quadratic program in order to choose a parsimonious decomposition among the many that allow the reconstruction. We develop optimality conditions for the criterion that can be geometrically interpreted and present a dual criterion that has an appealing physical interpretation. Some simulated examples are also presented to show the excellent performance in resolution of the approach.

1. INTRODUCTION

The problem of estimating the frequencies of multiple sinusoids from noisy measurements has received considerable attention for many years. It is a major problem with applications in a number of different fields.

We consider here the very specific case where the process can be modeled as a sum of real sinusoids in additive white noise and propose to estimate the white noise power and the parameters (amplitude and frequency) of the sinusoids as well as their number.

We confine our study to the case where the number of data is rather small (of the order of $T = 100$) and the signal-to-noise ratio (SNR) is quite low, below 10dB.

The major difficulties that one encounters are then :

- the detection of a weak isolated sinusoid,
- the resolution/separation of two closely spaced sinusoids,
- the detection of a weak sinusoid hidden by a stronger one.

In this paper we focus our attention on the resolution problem.

2. THE PROBLEM

We consider the following noise corrupted sinusoidal signal:

$$y(t) = s(t) + e(t) \quad (1)$$

with

$$s(t) = \sum_{i=1}^p A_i \cos(2\pi f_i t + \phi_i), \quad f_i \in]0, 1/2[\quad (2)$$

where $e(t)$ is a zero mean white Gaussian noise with variance σ^2 and ϕ_i are independent random variables uniformly distributed over $[0, 2\pi]$. This insures that the process $s(t)$ is stationary. We address the problem of the estimation of the number of sinusoids p and simultaneously the identification of σ^2 , the amplitudes A_i and the normalized frequencies f_i from a set of data $y(t)$ of length T .

The covariances r_k of $y(t)$ satisfy :

$$r_k = E(y(t)y(t+k)) = \sum_{i=1}^p \frac{A_i^2}{2} \cos(2\pi f_i k) + \sigma_e^2 \delta_k$$

We estimate K of these covariances and denote the so-obtained vector of estimates \hat{r} :

$$\hat{r}_k = \frac{1}{T-k} \sum_{t=1}^{T-k} y(t)y(t+k) \quad (3)$$

\hat{r} has, asymptotically, a Gaussian distribution $\mathcal{N}(r, \Sigma)$. An expression of Σ can be found in [2]:

$$\begin{aligned} \Sigma_{ij} &= \frac{1}{T} \sigma^4 \delta_i \delta_j + \frac{1}{T-i} \sigma^4 \delta_{i,j} \\ &+ \frac{1}{T-i} 2\sigma^2 \gamma_{j-i} + \frac{T-j-i}{(T-i)(T-j)} 2\sigma^2 \gamma_{i+j} \end{aligned}$$

where $j \geq i$ and $\gamma_i = r_i - \sigma^2 \delta_i$.

Given this K -dimensional vector \hat{r} , we want to reconstruct it as a linear combination of similar covariance-vectors associated with sinusoids and the contribution of the white noise. Let A be the $K \times N$ matrix whose last column is the covariance vector associated with white noise and whose other ones are covariance-vectors corresponding to sinusoids at equispaced (angular) frequencies ω_n .

$$A = \begin{pmatrix} 1 & \dots & 1 & 1 \\ \cos \omega_1 & \dots & \cos \omega_{N-1} & 0 \\ \cos 2\omega_1 & \dots & \cos 2\omega_{N-1} & 0 \\ \vdots & & \vdots & \vdots \\ \cos(K-1)\omega_1 & \dots & \cos(K-1)\omega_{N-1} & 0 \end{pmatrix}$$

We discretize uniformly the frequency domain $f \in]0, 1/2[$ with a step $h = .5/N$. We thus have the angular frequencies ranging from $\omega_1 = 2\pi h$ to $\omega_{N-1} = \pi - \omega_1$. We will comment on the choice of the step h later on, but we will always have $N \gg K$, which means that the constrained linear system :

$$\hat{r} = AX, \quad X \geq 0 \quad (4)$$

will, in general, have an infinite number of solutions. The set of solutions is the intersection of a linear variety and the positive orthant. Among them, we seek a solution with $p+1$ or $2p+1$ positive components, an extremely small number when compared to N .

Note that we will also normalize the Euclidean norm of the columns of A to one. It is easy to check that any subset of K columns of A are linearly independent.

3. THE CRITERION

The way we choose to obtain a sparse solution to the constrained linear system (4) is to solve :

$$\min_X \|AX - \hat{r}\|_{\Sigma^{-1}}^2 + \lambda \|X\|_1 \quad X \geq 0 \quad (5)$$

where $\|X\|_Q^2$ stands for $X^T Q X$, $\|X\|_1$ denotes the ℓ_1 -norm of X , which is equal to the sum of the components since $X \geq 0$, and λ is positive parameter which has to be tuned. This is a quadratic program [3]

In the following we will note $\tilde{A} = \Sigma^{-1/2} A$ and $\tilde{r} = \Sigma^{-1/2} \hat{r}$. This transformation whitens the estimation errors in \hat{r} and makes all the K linear equations $\tilde{A}X = \tilde{r}$ equally valid and independent, in a statistical sense.

If $\lambda = 0$, the value of the criterion at the optimum is zero, there are many non-sparse solutions. If λ is very large, we will see below that the solution of (5) is attained at $X^* = 0$. Between these two extreme cases one can expect to find a range of values of λ for which the optimum solution X^* will have only a small number of non-zero components.

Generically, each true frequency in (2) falls between two points of the frequency grid and to reconstruct exactly its contribution in r or \hat{r} a large number of columns of A are needed. By tuning λ , we allow for a more or less approximate reconstruction. We seek a value of λ which is such that each true cosine vector is reconstructed using its two neighbors in A yielding a weighting vector with around $2p$ positive weights plus one weight for the noise contribution. In fact, the additive regularization term $\lambda \|X\|_1$ in the criterion not only allows for these reconstruction errors but it also takes care of the estimation errors $\hat{r} - r$.

For such an optimal choice of λ , the solution X^* has thus $2p + 1$ positive components, with one component for the noise variance and p pairs of neighboring components each associated with one sinusoid. The frequency estimate \hat{f}_i is obtained by linear interpolation of the two columns frequencies and the power \hat{A}_i is the sum of the two weights. In fact, the ℓ_1 norm term creates a bias which has to be corrected, but we will discuss this issue later on. The number of sinusoids is given by the number of significant pairs of non-zero weights.

4. THE OPTIMALITY CONDITION OF THE QUADRATIC PROGRAM

4.1. The optimality condition

Let us rewrite the quadratic program as :

$$\min_X \|\tilde{A}X - \tilde{r}\|_2^2 + \lambda \mathbf{1}^T X, \quad X \geq 0$$

where $\mathbf{1}$ is the vector $[1, \dots, 1]^T$. Its Lagrangian is then :

$$\ell(X, \mu) = \|\tilde{A}X - \tilde{r}\|_2^2 + \lambda \mathbf{1}^T X - \mu^T X, \quad \mu \geq 0$$

First order necessary conditions are (cf [3]):

$$\begin{aligned} \nabla \ell(X, \mu) &= -2\tilde{A}^T(\tilde{r} - \tilde{A}X) + \lambda \mathbf{1}^T - \mu^T = 0 \\ \mu^T X &= 0 \quad X \geq 0, \quad \mu \geq 0 \end{aligned}$$

The second equation $\mu^T X = 0$, which is a sum of positive terms, implies that in each term either X_i or μ_i is equal to zero. Denoting then \bar{X}^* the vector containing the non-zero components of X^* and \bar{A} the associated columns of \tilde{A} , the equations in $\nabla \ell(X, \mu) = 0$ associated with $\bar{\mu} = 0$ yield :

$$\bar{A}^T(\tilde{r} - \bar{A}\bar{X}^*) = \bar{A}^T(\tilde{r} - \bar{A}X^*) = \frac{\lambda}{2} \mathbf{1}$$

Rewriting this relation, we get the following expression for non-zero components of the solution \bar{X}^* of the quadratic program:

$$\bar{X}^* = (\bar{A}^T \bar{A})^{-1} \bar{A}^T \tilde{r} - \lambda/2 (\bar{A}^T \bar{A})^{-1} \mathbf{1} \quad (6)$$

which would allow to compute the optimum if the indices of the non-zero components were known.

The zero components of the optimum are associated with the non-negative components of μ . If $\mu_j \geq 0$, we get then from $\nabla \ell(X, \mu) = 0$

$$\bar{a}_j^T(\tilde{r} - \bar{A}\bar{X}^*) = \bar{a}_j^T(\tilde{r} - \bar{A}X^*) \leq \lambda/2 \quad \forall a_j \notin \bar{A}$$

We can summarize both conditions in

$$\bar{A}^T(\tilde{r} - \bar{A}X^*) \leq \lambda/2 \quad (7)$$

where equality holds for the non-zero components of X .

We present in the appendix some further developments which highlight some other aspects of the criterion and justifies its interest from a physical point of view.

4.2. A geometrical interpretation

Replacing \bar{X}^* in (7) by its expression (6) leads to :

$$\bar{a}_j^T \bar{A} (\bar{A}^T \bar{A})^{-1} \mathbf{1} \leq 1 \quad \text{for} \quad \bar{a}_j \notin \bar{A}$$

With $d = \bar{A} (\bar{A}^T \bar{A})^{-1} \mathbf{1}$, the last relation, and thus (7), becomes :

$$\bar{a}_j^T d \leq 1 \quad (8)$$

This condition has the following geometrical interpretation : the vector d defines a hyperplane that separates the columns present in the scenario (those in \bar{A}) from the others, for the true columns $\bar{a}_j^T d = 1$ while for all the others $\bar{a}_j^T d \leq 1$.

In the simple case where there is no noise, a vector d defining such a separating hyperplane is easily obtained from the theorem by Caratheodory used to obtain the Pisarenko harmonic retrieval method. It is quite easy to build a polynomial which is equal to zero for the few frequencies that one wants to isolate and different from zero for any other frequency. For the case of a single sinusoid at angular frequency ω_1 , this polynomial is : $P(z) = z^2 - 2 \cos \omega_1 z + 1$. Evaluating now the squared modulus of this polynomial and rearranging terms the vector defining such a separating hyperplane is, for instance :

$$d^T = [-2 \cos^2 \omega_1, 4 \cos \omega_1, -1, 0, 0, \dots, 0]$$

In the case of several sinusoids the same method yields a separating hyperplane as long as $2p < K$. This indicates that, for our problem of sinusoids in noise, a *separating hyperplane* always exists. This however does not mean that $d = \bar{A} (\bar{A}^T \bar{A})^{-1} \mathbf{1}$ defines such a hyperplane.

4.3. One sinusoid and white noise

Let us consider the very simple case where :

$$r = a_i + \sigma \epsilon$$

with $\epsilon = [1, 0, \dots, 0]^T$ the column of A associated with the white noise. This is the case of one sinusoid in white noise where more over the frequency of the sinusoid is the i -th point of the grid. Ideally we then seek a value of the parameter λ that yields an optimal solution X^* of our criterion having just two non-zero components,

the components of the columns a_i and ϵ . Defining then the matrix \bar{A} as $\bar{A} = [\bar{a}_i, \epsilon]$, one can easily check that the vector d in (8) becomes :

$$d = ([\bar{a}_i, \epsilon]^T [\bar{a}_i, \epsilon])^{-1} [1, 1]^T = \frac{a_i + \epsilon}{1 + a_i^T \epsilon}$$

so that :

$$a_j^T d = \frac{\bar{a}_j^T \bar{a}_i + \frac{1}{\|a_j\|}}{1 + \frac{1}{\|a_i\|}}$$

It is now easy to check that indeed this vector d is such that $a_i^T d = \epsilon^T d = 1$ and $a_j^T d < 1$ for $j \neq i$. After some further developments, which we cannot include due to limited space, this allows to establish that, at least for this simple case, there always exists a range of values of λ that yield a solution of (5) with exactly two non-zero components, the two expected ones.

One can similarly show that $a_j^T d < 1$ holds for one source between two columns of A , i.e. the same conclusion is valid for an arbitrary sinusoid in white noise.

5. SIMULATION RESULTS

5.1. The complete procedure

We consider the signal composed of sinusoids and white noise (1). We estimate its K first covariances (3) and we solve the quadratic program (5) for one or several values of λ . For a well chosen λ , we then get a sparse X^* with one non-zero component for the white noise and a few pairs of non-zero components, each one corresponding to one sinusoid. We order these pairs by decreasing powers. To choose the order of the model, we first select the first pair, then the first two pairs, ..., and in each case we solve $\min_{\bar{X}} \|\bar{A} \bar{X} - \hat{r}\|^2$ to get the unbiased corresponding solution. From these unbiased values of X we deduce the frequency estimates \hat{f}_i of the sinusoids as the weighted average of the frequencies of the two columns and the corresponding power as the sum of both power. We then apply an Akaike-like test by taking $\|\hat{r} - r(\theta_p^*)\|_{\Sigma^{-1}}^2$ as the maximum likelihood part of the test.

5.2. Tuning the parameters

To completely define our criterion we have two parameters to tune, λ and the discretization step h of the normalized frequencies, that indirectly fixes the number of columns in A .

One can show that, provided the scenario is *separable* in the above mentioned sense, there is a range of λ such that only the *true* columns have non-zero weights. One can show [1] that one has to fix λ of the order of the standard deviation of the estimation errors. Since we whitened these errors by premultiplying both A and \hat{r} by $\hat{\Sigma}^{-1/2}$, this means that λ should be taken of the order of one. In the simulations below, we generally take $\lambda = 2$.

The discretization step h has to be chosen small enough to allow the method to attain the Cramer Rao bounds. On the other hand, one can show (cf. [1]) that the interpolation procedure allows to gain an order of magnitude i.e. that for a step h the accuracy in the estimate is of order h^2 . We thus take h , in the simulations below, reasonably small and adapt it to the accuracy we want to achieve.

5.3. The results

To evaluate the performances of the proposed procedure, we consider the following situation:

$$y(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) + e(t)$$

$$\sigma^2 = E(e(t)^2) = 1$$

$$t = 1, \dots, T$$

$$SNR_i = A_i^2 / 2\sigma^2$$

We systematically perform 50 independent realizations. We give the mean and the variance (w.r.t. to the true parameters) of the estimated parameters, and we indicate the Cramer Rao lower bounds (CRB), which can be found in [5] for example.

We first consider the example treated in [6], which is quite an easy scenario that the periodogram could solve at high SNRs. Therefore we restrict our simulations to the case of lower SNRs 0dB (table 1) and -10dB (table 2). Phases are chosen following [6]: $\phi_1 = 1$ and $\phi_2 = 0$, even if we don't estimate them. We can note that our procedure achieves the Cramer Rao bounds in cases where the SNRs are 20dB worse than in [6].

	Mean	Variance	CRB
$A_1 = 1.41$	1.43	$4.2 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$
$A_2 = 1.41$	1.42	$3.6 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$
$f_1 = .1$.99985	$4.3 \cdot 10^{-9}$	$2.5 \cdot 10^{-9}$
$f_2 = .125$.12502	$4.2 \cdot 10^{-9}$	$2.5 \cdot 10^{-9}$
$\sigma_2 = 1$.99	$1.0 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$

Table 1: SNR=0dB, T=500, $\lambda = 2$, N=500, K=100

	Mean	Variance	CRB
$A_1 = .447$.439	$5.0 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$
$A_2 = .447$.441	$4.1 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$
$f_1 = .1$.09966	$1.56 \cdot 10^{-7}$	$2.43 \cdot 10^{-8}$
$f_2 = .125$.12478	$1.26 \cdot 10^{-7}$	$2.43 \cdot 10^{-8}$
σ_e	1.00	$1.0 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$

Table 2: SNR=-10dB, T=500, $\lambda = 10$, N=2000, K=100

We now consider more difficult scenarios with five times less data ($T = 100$) and closer frequencies. In this case ((2)), the periodogram resolves equipowered sinusoids whose frequency separation is of the order of $\Delta f = \frac{1}{T} = 10^{-2}$ whereas we will consider frequencies separated by $\Delta f/2$ in the examples bellow. In the following we take identical (zero) initial phases for the two sinusoids. Our approach still achieves the CRB.

	Mean	Variance	CRB
$A_1 = 4.47$	4.38	.33	.1
$A_2 = 4.47$	4.35	.21	.1
$f_1 = .1$.0999	$3.2 \cdot 10^{-6}$	$1.7 \cdot 10^{-7}$
$f_2 = .105$.1051	$4.7 \cdot 10^{-7}$	$1.9 \cdot 10^{-7}$
$\sigma_e = 1$	1.00	$2.1 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$

Table 3: SNR=10dB, T=100, $\lambda = 5$, N=5000, K=40

Finally in the table 5 we consider the case where the two sinusoids have not the same power, one at 0dB and the second one at 10dB, while keeping the frequency separation below $\Delta f = 1/T$.

6. CONCLUSION

We have proposed here a new algorithm for the detection and estimation of the parameters of sinusoidal signals in white noise. We

	Mean	Variance	CRB
$A_1 = 1.41$	1.37	.13	.1
$A_2 = 1.41$	1.32	.13	.1
$f_1 = .1$.0991	$6.3 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$
$f_2 = .105$.1056	$3.2 \cdot 10^{-6}$	$1.9 \cdot 10^{-6}$
$\sigma_e = 1$	1.00	$6.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-2}$

Table 4: SNR=0dB, T=100, $\lambda = 5$, N=2000, K=40

	Mean	Variance	CRB
$A_1 = 1.41$	1.32	$6.4 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$
$A_2 = 4.47$	4.07	$2.15 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$
$f_1 = .1$.0986	$7.9 \cdot 10^{-6}$	$1.2 \cdot 10^{-6}$
$f_2 = .107$.10707	$5.9 \cdot 10^{-8}$	$1.3 \cdot 10^{-7}$
$\sigma_2 = 1$	1.00	$6.8 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$

Table 5: $SNR_1=0dB$, $SNR_2=10dB$, T=100, $\lambda = 1$, N=2000, K=100

minimize a quadratic program that has a nice and appealing interpretation in terms of conditions on the matched filter outputs applied to the residues. The criterion is the sum of a quadratic term and a regularization term including an ℓ_1 norm. It is easily minimized using routines present in scientific libraries. The minimization of that criterion yields a parsimonious representation of the observations.

We have derived optimality conditions for the quadratic program that can be geometrically interpreted in terms of the existence of a separating hyperplane between the true components (sinusoids) and the other ones.

Simulation results allow us to assess the performance in resolution of the approach show outstanding performances. We manage to separate two closely spaced sinusoids at very low SNRs. The robustness of the approach to non-white noise is also very good.

7. APPENDIX : THE DUAL CRITERION

Let us rewrite the quadratic program under the form :

$$\min_X \|AX - b\|_2^2 + \lambda \mathbf{1}^T X \quad s.t. \quad X \geq 0 \quad (QP)$$

with A , a $(K \times N)$ full rank matrix ($N \gg K$) and develop some equivalent formulations. We will present a dual problem and establish relationships between the solution to the dual and the original problem [7]. The expression of the dual is interesting because it leads to an appealing physical interpretation. Since the problem (QP) is convex, it is equivalent to :

$$\max_{\mu \geq 0} \{ \min_X \|AX - b\|_2^2 + \lambda \mathbf{1}^T X - \mu^T X \} \quad (P1)$$

For a fixed positive μ , the minimum in X satisfies :

$$2A^T(AX - b) + \lambda \mathbf{1} + \mu = 0$$

since $A^T A$ is not full rank, one cannot draw X as a function of μ from this expression, one can however use it to express μ in terms of X and to eliminate μ in (P1) to obtain the following equivalent formulation :

$$\max_X -X^T A^T A X + b^T b$$

$$s.t. \quad A^T(b - AX) = \frac{\lambda}{2} \mathbf{1} - \frac{1}{2} \mu, \quad \mu \geq 0$$

which, writing $Y = AX$, yields the following dual problem :

$$\min_Y \|Y\|^2 \quad s.t. \quad A^T(b - Y) \leq \frac{\lambda}{2} \mathbf{1} \quad (D)$$

which is a minimum norm problem. To check that the solution Y^* of the dual actually leads to a solution of (QP) (the primal) one verifies that the dual of the dual is nothing but (QP). Indeed applying exactly the same steps as above to (D), one gets :

$$\max_{\mu \geq 0} \{ \min_Y \|Y\|^2 + \mu^T [A^T(b - Y) - \frac{\lambda}{2} \mathbf{1}] \}$$

the minimum in Y now satisfies :

$$2Y + A\mu = 0$$

and replacing Y by its expression, one obtains :

$$\max_{\mu \geq 0} -\frac{1}{2} \{ \|A\mu - b\|^2 + \lambda \mathbf{1}^T \mu \} - \frac{1}{2} b^T b$$

which is equivalent to (QP). The solution X^* of (QP) is thus the Lagrange multiplier vector of the dual and from the optimum Y^* of the dual one can indeed deduce $X^* : x_j^* = 0$ if the j -th constraint in the dual is inactive i.e. if $a_j^T(b - Y^*) < \frac{\lambda}{2}$ and $x_j^* > 0$ if the j -th constraint in the dual is active.

We have thus shown that minimizing (QP) for a given λ amounts also to solve :

$$\min_X \|AX\|^2 \quad s.t. \quad A^T(b - AX) \leq \frac{\lambda}{2} \mathbf{1} \quad (D)$$

i.e. to subtract from the observations in b a quantity AX with least ℓ_2 norm such that when applying the correlator (matched filter) to the residues $(b - AX)$ the output never exceeds $\lambda/2$. This is a quite sensible way to obtain a solution that is certainly much better than in matching pursuit type algorithms [4] that work in an iterative way and subtract successively the best matching component from the current residues until, for instance, all the residues are below a fixed threshold. Here the same is somehow achieved in one single shot.

8. REFERENCES

- [1] J.J.Fuchs, Detection and estimation of superimposed signals, submitted to *Proc. ICASSP*, oct. 1997.
- [2] J.J.Fuchs Multiscale identification of real sinusoids in noise, *Automatica*, Vol.30, No.1, pp. 147-155, 1994.
- [3] D.G.Luenberger, *Linear and non linear programming*, Addison-Wesley, 1984.
- [4] S.G. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE-T.SP*, Vol.41., pp 3397-3415, Dec. 1993.
- [5] B.Porat, *Digital Processing of Random Signals*, Prentice Hall, 1994.
- [6] P.Stoica, R.L.Moses, B.Friedlander, T.Soderstrom, Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements, *IEEE-ASSP*, Vol.37, No.3, March 1989.
- [7] G.A. Watson, A class of programming problems whose objective function contains a norm, *Journal of Approximation Theory*, Vol. 23, pp. 401-411, 1978.