# SELECTING THE BEST AMONG SEVERAL ESTIMATES IN DOA ESTIMATION.

*Jean-Jacques* FUCHS

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

**ABSTRACT**

In this work we present a new direction of arrival (DOA) estimation scheme that applies to uniform linear arrays. We propose to get several estimates by applying the Pisarenko harmonic retrieval (PHR) method to different sets of covariances and to select the best estimate by using the maximum likelihood (ML) criterion. More generally we would like to promote the idea that it is often more efficient to select the best among several estimates rather than to combine them. This requires of course the existence of a decision criterion which indeed conveys additional information.

## 1. INTRODUCTION

In estimation problems, the maximum likelihood (ML) method is generally taken as the reference since it leads to efficient estimators that attain asymptotically the Cramer-Rao bound (CRB). In direction of arrival (DOA) estimation it has only had limited success since the optimization of the likelihood function is computationally complex. For difficult scenarios, it is an ill-conditionned function with numerous local extrema. This means that an excellent initial point has to be known in order to draw the benefits from the ML method and the major problem is actually to find a good initial point.

We consider the standard situation of a linear array with $N$ equispaced sensors receiving $P$ narrow-band uncorrelated sources in spatially white noise. We assume $P$ to be known.

The basic Pisarenko harmonic retrieval (PHR) method uses a minimal set of covariances that has the same number of degrees of freedom as the unknown parameter set. It achieves a change of variables and transforms a set of estimated covariances into an estimate (i.e. a set of parameter estimates). There is no optimization involved in this transformation that is one to one. We apply a similar method to several minimal sets of covariances to obtain several estimates of the unknown parameters. We then evaluate the ML criterion at each of these estimates to select the *best* i.e. the most likely. The resulting estimate is in general quite efficient and its variance close to the Cramer Rao bounds.

While in this DOA estimation context, we propose to get a set of competing estimates by applying the PHR to several minimal sets of covariances other estimates can be proposed and tested. The method of moments can quite generally be used to generate several potential estimates. While optimally combining them is then an alternative to the approach we propose, it is generally more cumbersome to get the optimal weights and the resulting estimate might well be less efficient. The idea is not really new and has been promoted in [1] for the case of real sinusoids in noise. Similar developments have also been proposed in e.g. [2], [3].

This paper is organized as follows. The signal model is described in section 2. In section 3 and 4 we develop the procedure we propose. In section 5 we explain why its statistical properties are hard to assess. Finally simulations are presented in section 6 and in the Appendix.

## 2. SIGNAL MODEL

We consider a linear array with $N$ equispaced sensors with spacing half a wave-length. It receives signals from $P$ far field narrow-band uncorrelated sources in additive white noise. We denote $X_k$ the complex order-$N$ output-vector (snapshot) of the array. These vectors are modeled as independent, identically distributed complex gaussian variables with zero mean and covariance matrix:

$$R = \sum_{p=1}^{P} a_p d\theta(f_p) d\theta(f_p)^* + vI \qquad (1)$$

where $v$ is the power of the additive white noise, $d\theta(f_p)$ is the steering vector at spatial frequency $f_p$ associated with the p-th source whose power is denoted $a_p$. We define the signal to noise ratio (SNR) of source $p$ as $\rho_p = a_p/v$. The spatial frequency $f$ is related to the bearing $\phi$ with respect to broadside by $f = (\sin\phi)/2$. The steering vector at spatial frequency $f$ is then $d\theta(f) = [\exp(2i\pi n f)]$ with $0 \leq n \leq N-1$ and $-\frac{1}{2} \leq f \leq \frac{1}{2}$. The covariance matrix $R$ has thus a Toeplitz structure and is entirely defined by its first column. Since we assume $P$ the number of sources to be known, there are $2P+1$ unknowns: $\theta = \{\ (a_p, f_p)_{p=1,..,P};\ v\}$. An estimate $\hat{R}$ of $R$ is given by:

$$\hat{R} = \frac{1}{T}\sum_{k=1}^{T} X_k X_k^* \qquad (2)$$

where $T$ is the number of snapshots. Under the modeling assumptions made above, $\hat{R}$ in (2) is a sample from a complex Wishart distribution with mean $R$ and $T$ degrees of freedom. Maximizing the likelihood of the observations amounts then to minimize [4] :

$$\ell(\theta) = \log|R(\theta)| + \mathrm{tr}(\hat{R}R(\theta)^{-1}) \qquad (3)$$

with respect to the unknowns in $\theta$, where $|R|$ denotes the determinant of the matrix $R$ and $\mathrm{tr}(R)$ its trace. This is a difficult task since this function is highly non-linear in the unknowns.

## 3. DEVELOPMENT

### 3.1. Getting several estimates

The method of moments [9] consists in replacing the full information, here $\hat{R}$, by a smaller dimensional statistic that still conveys enough information on the unknowns in $\theta$ and to base the estimate on this insufficient statistic. Quite often one uses as statistic a set of moments hence the name of the method. As the number of observations increases, these moments converge in some sense to the true moments that are a function of the unknowns $\theta$. The estimate $\theta$ is then obtained by model-fitting or inversion of this function.

In the DOA estmation case, a first possibility is to consider the set of P+1 covariances denoted $Y_1 = \{r_0,\ r_1, .., r_P\}$ extracted from the first column of the Toeplitz matrix $R$. The set $Y_1$ has 2P+1 real degrees of freedom and its expression as a function of $\theta$ is easily deduced from (1). Caratheodory's theorem [5] establishes that the transformation from $Y_1$ to $\theta$ is generically one to one. The procedure that performs this inversion is known as the basic Pisarenko Harmonic Retrieval (PHR) method [6]. Let $\hat{\theta}_1$ be the estimate of $\theta$ obtained by applying the PHR method to $\hat{Y}_1$ an estimate of $Y_1$. This estimate is consistent asymptotically in $T$ but is known to have poor efficiency.

Many other sets of covariances can be used similarly to $Y_1$ to yield possibly more efficient estimates. We seek those for which the inverse transformation is easy to implement.

Let us consider the sets $Y_k = \{r_0,\ r_k,\ r_{2k}, .., r_{Pk}\}$. For k=1 we get the set considered above and for $k > 1$ it is the set of P+1 covariances obtained by downsampling the covariance sequence by a factor $k$. Though these sets have precisely the right number of degrees of freedom, the transformation from $Y_k$ to $\theta$ is one to many. This is a result of the ambiguity introduced by downsampling. The covariance sequence in $Y_k$ leads to an aliased spectrum and each spatial frequency in the aliased spectrum has generically $k$ potential determinations. More precisely a frequency $F \in ]-\frac{1}{2},\ \frac{1}{2}]$ of the aliased spectrum can be induced by any of the $k$ frequencies $\frac{F}{k} + \frac{\ell}{k} \in ]-\frac{1}{2},\ \frac{1}{2}]$ with $\ell \in \mathbb{Z}$ of the initial spectrum. There are thus generically $k^P$ distinct P-sources scenarios that yield the same $Y_k$ and it remains to find the good determination. We propose to use the ML criterion (3) to select the good determination. This may be quite time consuming and there are several means to drastically reduce this burden. One can use the output of the standard beamformer to locate the frequency domain(s) of interest and only evaluate the ML criterion for determinations falling in these areas. Another mean to solve the ambiguity consists in using simultaneously the estimates given by different $Y_k$'s to localize the frequencies that are stable over several k's. In the two sources scenario described in section 6 the plot of the estimates obtained from 9 different $Y_k$'s is shown in Figure 1. Only the 2 central frequencies around $f_1 = 0$ and $f_2 = .0167$ remain stable while the others move in an orderly fashion

One can also observe in this figure that for a fixed $k$ the distance $1/k$ between 2 determinations is much larger than the resolution which is typically $\frac{1}{N}$. Since the resolution is linked to the size of the domain of attraction of the ML function, this means that for each frequency $F_p$ at most one of the $k$ determinations is within the domain of attraction of the global minimum of the ML function and there is thus no risk of confusion. We will never end up with the wrong determination if we use the ML criterion (3) to select it.

### 3.2. Getting a single 'best' estimate

For an N-sensor array, there are $k_{max} = \text{floor}(\frac{N-1}{P})$ different minimal covariance sets $Y_k$ and thus $k_{max}$ distinct estimates. To come up with an unique one having better efficiency one can either combine them or select the *best* among them.

Finding the optimal way to combine the $\hat{\theta}_k$'s requires the knowledge of the cross-correlation matrix between the estimates which is in general cumbersome to evaluate. This approach seems feasible only in the single source case.

On the other hand, except in the single source case (see the Appendix), the single $\hat{\theta}_k$ that is the most efficient depends upon the scenario and is thus unknown beforehand. Indeed for any fixed value of $k$, it is easy to build a scenario for which 2 or more sources collapse into a single one in $Y_k$ making the corresponding estimate useless.

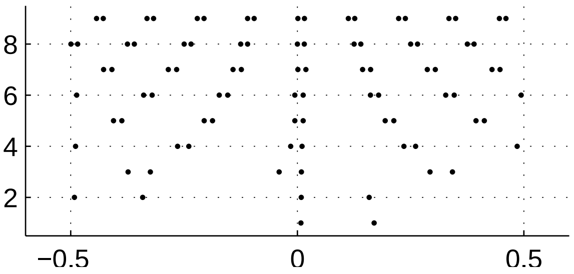


Figure 1: The frequency determinations for downsampling rates k=1 to 9. The 2 central frequencies remain stable while the others move around.

We therefore propose to select the single *best* estimate by using again the ML criterion (3). Once all the estimates $\hat{\theta}_k$ are available, we compare their likelihoods $\ell(\hat{\theta}_k)$ and keep the most likely, we denote $\hat{\theta}_*$. The estimate that is retained in this way is realisation dependent and if one performs a large number of simulations all $\hat{\theta}_k$'s will contribute more or less frequently, see Figure 2. This actually allows $\hat{\theta}_*$ to be more efficient than any of the individual $\hat{\theta}_k$'s.

## 4. DESCRIPTION OF THE ALGORITHM

Given the signal model and the notations defined in section 2 we propose a procedure that allows to estimate the 2P+1 unknowns in $\theta$ using as observations the estimated covariance matrix $\hat{R}$ (2).

We first concentrate the information by toeplerizing $\hat{R}$ to get $\hat{R}_T$. In the analysis and simulations below, we do so by averaging the diagonals. All the information is now conveyed by the $N$ covariances $\{\hat{r}_n\}$ present in the first column of $\hat{R}_T$.

We build $k_{max}$ different subsets of covariances:

$$\hat{Y}_k = (\hat{r}_0,\ \hat{r}_k,\ \hat{r}_{2k}, .., \hat{r}_{Pk}), \quad 1 \le k \le k_{max} = \text{floor}(\tfrac{N-1}{P})$$

from which we deduce $k_{max}$ different estimates $\hat{\theta}_k$. To do so we construct the function $Y_k(\theta)$ that relates $\theta$ to $Y_k$ and apply its inverse to $\hat{Y}_k$ to get $\hat{\theta}_k$.

For k=1, the inverse is unique [5] and the inversion is performed by applying the PHR method [6]. It consists in the following steps:

• built the covariance matrix with Toeplitz structure of order P+1 associated with $\hat{Y}_1$

• $\hat{v}$ the estimate of $v$ is its *smallest* eigenvalue

• compute its *smallest* eigenvector, say $[u_o\ u_1\ ...u_P]$

• the P roots of $u_o + u_1 z + \ ... + u_P z^P = 0$ are then distinct and on the unit circle, rewritte them as $e^{2i\pi \hat{f}_p}$ with $-\frac{1}{2} < \hat{f}_p \le \frac{1}{2}$ the estimates of $f_p$

• obtain the estimates $\hat{a}_p$ by solving a the set of linear equations deduced from $\hat{Y}_k(\theta)$ with $\hat{f}_p$ in place of $f_p$.

For k>1, the inverse transformation is one to many. We first apply the PHR method described above to $\hat{Y}_k$. Let $\hat{F}_p$, $p \in$ (1, P) be the frequency estimates. To each $\hat{F}_p$ one associates $k$ potential determinations:

$$\hat{f}_p(\ell) = \frac{\hat{F}_p}{k} + \frac{\ell}{k}, \qquad \ell \in \mathbb{Z} \ni -.5 < \hat{f}_p(\ell) \le .5 \tag{4}$$

• built $k^P$ distinct estimates $\hat{\theta}$ by supplementing the frequency estimates: one among the $k$ determinations $\hat{f}_p(\ell)$ for $p \in$ (1, P) with the amplitudes estimates.

• evaluate the ML criterion with $R(\hat{\theta})$ in place of $R$ in (3) and retain the most likely $\hat{\theta}_k$ among the $k^P$.

Once the $k_{max}$ estimates $\hat{\theta}_k$ are available retain again the most likely, the one that optimizes (3) to get the estimate $\hat{\theta}_*$.

We gave in section 3.1 some hints that allow to reduce the number of evaluations to be performed from $k^P$ to just a few.

We shall see from the simulations below that $\hat{\theta}_*$ is quite efficient and its variance close to the Cramer Rao bounds. One can nevertheless use $\hat{\theta}_*$ as an initial point of an iterative optimization routine of the likelihood (3) since $\hat{\theta}_*$ is in general consistent, the estimate should converge to the global optimum of the likelihood function.

## 5. STATISTICAL ANALYSIS

A statistical analysis is indeed extremely difficult because the index of the estimate $\hat{\theta}_k$ that is retained is realization-dependent. The fact that the different estimates are correlated further complicates the analysis. This is probably the price to pay for the simplicity of the *fusion* procedure. As in the much more cumbersome optimal weighting schemes the efficiency is better than those of the individual estimates and can be quite close to the CR bounds.

Even in the usually trivial single source case a complete statistical analysis seems out of reach. We indicate in the Appendix how to compute the variances of the $k_{max}$ individual estimates of the unique spatial frequency $f$ taking into account the "toeplerization" of $\hat{R}$ and assuming the ML criterion systematically selects the good determination (7) which is legitimate in this case. We also detail simulation results to highlight how realization dependent our estimate is even in the single source case, see Figure 2. While the variance say $V_k$ of each of the $k_{max}$ individual estimates is quite easy to obtain as well as the correlations between them, it is probably impossible to evaluate the variance of the final estimate $\hat{f}_*$.

As a first step towards an approximate analysis one could probably say that $\hat{f}_*$ is the result of the following model:

• draw a gaussian random variable with mean the true value and variance $V_{cr}$, the CRB

• draw $k_{max}$ gaussian random variables with mean the true value and variance the theoretical variances $V_k$ of the estimates

• keep among these $k_{max}$ values the one that is closest to the CR sample.

Even if it were possible to get a rough estimate of the variance of the so-obtained random variable this analysis takes into account neither the correlation between the $k_{max}$ individual estimates nor their correlations with the first variable, the minimum of the current likelihood function which plays a decisive role.

Since even in the single source case an analysis seems out of reach, let us resort to simulations to assess the potentialities of the approach in more realistic situations.

## 6. SIMULATION RESULTS

The single source case is detailed in the Appendix. To test the resolution power of the approach, we consider two closely spaced equipowered sources with an SNR=0 dB (i.e. $a_p/v = 1$ in (1)), $T = 100$ and $N = 20$ and a spatial frequency separation of $\delta f = 1/(4N)$ i.e. one fourth of the standard resolution. We built $k_{max}$= 8 estimates and the algorithm separates the two sources with a frequency estimate variance slightly higher than twice the CRB. The results are presented in Table 1. for 1000 realizations. Since the standard deviations of the estimates are about a fourth of the source separation the results are meaningful. Few estimation procedures would be able to separate these sources.

| true values | mean | variance | CR bound |
|---|---|---|---|
| $a_1$= 1 | 1.009 | .134 | .092 |
| $a_2$= 1 | .993 | .138 | .092 |
| $f_1$= 0 | -.0005 | 9.00 $10^{-6}$ | 3.72 $10^{-6}$ |
| $f_2$= .0125 | .0131 | 8.88 $10^{-6}$ | 3.72 $10^{-6}$ |

Table 1: The means and variances observed over 1000 independent realizations for the two sources separation case.

We now consider an example taken from [8]. It consists of 3 equipowered sources with an SNR=-5dB which are located at bearings equal to (24, 27, 45) degrees with respect to broadside, $T = 100$ and $N = 15$. The separation of the two close sources is one third of the standard resolution. The results are presented in Table 2. for 1000 independent realizations. The variance on the bearings is again 2 to 3 times the CRB. There are 4 potential $Y_k$'s but the estimates given by $Y_1$ is never retained, those of $Y_2$ quite seldom and those of $Y_3$ and $Y_4$ respectively about one and two thirds of the realization.

In both cases the computational complexity is quite low if one looks at the stability of the different estimates with respect to $k$ (see Figure 2) or starts by locating the areas of interest using the standard beamformer, for instance.

| true values | mean | variance | CR bound |
|---|---|---|---|
| $a_1$= .3162 | .3162 | .0222 | .0120 |
| $a_2$= .3162 | .3238 | .0221 | .0118 |
| $a_3$= .3162 | .3189 | .0018 | .0015 |
| $\varphi_1$= 24 | 23.6144 | 1.192 | .3213 |
| $\varphi_2$= 27 | 27.3014 | 1.163 | .3513 |
| $\varphi_3$= 45 | 45.0206 | .0906 | .0496 |

Table 2: The means and variances observed over 1000 realizations for the three sources case.

## 7. CONCLUSIONS

The general conclusion one can draw from this contribution is that in estimation problems where the ML function is difficult to optimize but simple to evaluate, a quite efficient estimation scheme can be obtained by simply proposing several crude estimates and using the ML criterion to retain the best one. This can for instance be more efficient than (cumbersome) optimal fusion of different sub-optimal estimates working on different data sets.

The additional information that is brought by the likelihood function in the selection procedure plays an important role. We are so far unable to analyse the performance of the procedure. Further investigations are needed.

## 8. REFERENCES

[1] J.J. Fuchs. Multiscale identification of real sinusoids in noise. *Automatica*, vol. 30, pp. 147-155, jan. 1994.

also in Ifac-Ifors *Symposium on Identification and System Parameter Estimation*, Budapest, pp. 1548-1553, july 1991.

[2] P. Stoica and A. Gershman. Maximum Likelihood DOA estimation by data supported grid search. IEEE-Signal Processing Letters, 6, 10, oct. 1999.

[3] B; Volcker and P. Handel. Frequency Estimation from Proper Sets of Correlations IEEE-T-ASSP, 50, pp. 791-802, April 2002.

[4] P. Stoica and A. Nehorai. MUSIC, ML and Cramer Rao bounds: further results and comparisons IEEE-T-ASSP, 37, pp. 720-741, may 1989.

[5] U. Grenander and G. Szego. *Toeplitz forms and their applications.* University of California Press, 1958.

[6] V.F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astron. Soc.*, pages 347–366, 1973.

[7] K.S. Miller. *Complex Stochastic Processes.* Reading, MA, Addison-Wesley, 1974.

[8] P.J. Chung and J; Boehme Comparative convergence analysis of EM and Sage algorithms in DOA estimation. *ICASSP2001*, Salt Lake City, USA, May 2001.

[9] B. Porat. *Digital Processing of Random signals.* Prentice Hall, Englewood Cliffs, NJ, 1994..

## 9. APPENDIX: THE SINGLE SOURCE CASE.

There are 3 reals unknowns $\theta$ = {a, f, v}, we concentrate on the $k_{max}$ distinct estimates of the spatial frequeny $f$ we denote $\hat{f}_k$ in this appendix. One has $\hat{Y}_k = \{\hat{r}_o,\ \hat{r}_k\}$ with $r_o = a + v$ and $r_k = ae^{2ik\pi f} = ae^{ik\omega}$. The basic PHR method applied to an order 2 Toeplitz matrix yields as frequency estimate $\hat{F}_k$ the angle of $\hat{r}_k$. To $\hat{F}_k$ we associate the good determination $\hat{f}_k$ close to the true $f$ (see (4)). The statistical properties of $\hat{f}_k$ thus only depend on those of $\hat{r}_k$ obtained by averaging the elements in the $k$-th diagonal of $\hat{R}$

Under the modeling assumptions made in section 2, $\hat{R}$ in (2) is such that $T\hat{R}$ is a sample of a complex Wishart distribution with parameter matrix $R$ and $T$ degrees of freedom. Defining then $\tilde{R} = \hat{R} - R$, it follows that [7]:

$$T\ E[\mathrm{tr}(\tilde{R}A)\mathrm{tr}(\tilde{R}B) = \mathrm{tr}(RARB) \quad (5)$$

with $A$ and $B$ arbitrary matrices and tr$(A)$ the trace of the matrix $A$. In what follows we assume that $T$, the number of snapshots, is large enough and the modeling errors small enough for a first order approximmation to be valid. We are thus able to evaluate the covariance of the estimates $\hat{F}_k$ from those of the complex random variable $\hat{r}_k$. One has:

$$\hat{r}_k = \frac{1}{N-k}\mathrm{tr}(\hat{R}S^k)$$

where $S$ is the shift matrix with one's on the upper-diagonal Using then (5) with $A = S^k$ and $B$ either $S^k$ or its transpose, one gets introducing $\rho = a/v$:

$$E(\ |\tilde{r}_k|^2) = \frac{1}{T}(\rho^2 + \frac{1+2\rho}{N-k})$$

and $$E(\ \tilde{r}_k^2) = \frac{e^{2ik\omega}}{T}(\rho^2 + \frac{2\max(N-2k,0)}{(N-k)^2}\rho)$$

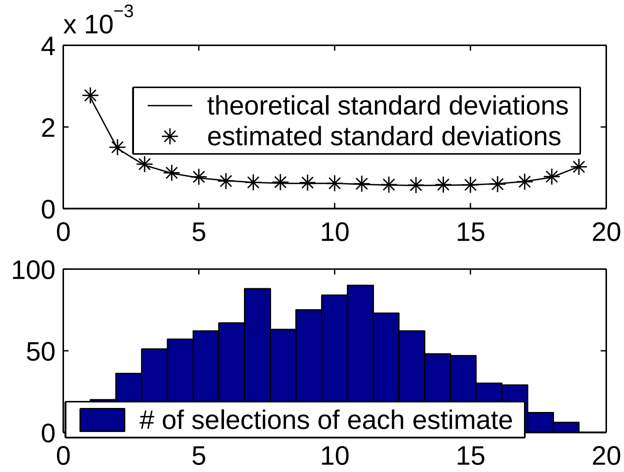


Figure 2: The standard deviations of the different individual frequency estimates and an histogram of their contibutions to the unique estimate over 1000 realizations.

From these two relations, one deduces the covariance of the angle $\hat{\Omega}_k$ of the complex random variable $\hat{r}_k$:

$$E(\ \tilde{\Omega}_k^2) = \frac{1}{2T\rho^2(N-k)}(1 + 2\rho - \frac{2\rho\max(N-2k,0)}{N-k}) \quad (6)$$

the true determination $\hat{\omega}_k$ is obtained from $\hat{\Omega}_k$ by division by $k$ and translation, one has:

$$E(\ \tilde{\omega}_k^2) = \frac{1}{k^2}E(\ \tilde{\Omega}_k^2) \quad (7)$$

This value is always greater than the corresponding Cramer Rao bound [4]:

$$E(\ \tilde{\omega}^2) \geq \frac{6(1+N\rho)}{T(N\rho)^2(N^2-1)} \quad (8)$$

In this single source case, this variance is independent of the direction of arrival and one can seek the value of $k$ that yields the smallest variance. For an SNR of 0dB ($\rho = 1$) it is attained for $k \simeq N/2$ and equals roughly twice the CR bound. It is however not a good idea to only evaluate this single estimate $\hat{\omega}_{N/2}$ because the approach we propose further reduces the variance by almost a factor two to get quite close to the CRB.

To illustrate this observation we present the results we get for a single source at $f = \omega/2\pi = .1$, $a = v = 1$, $T = 100$ snapshots and an array having $N = 20$ sensors. This is of course a trivial example and we certainly would recommend to use the standard beamformer to estimate $f$.

We present in Figure 1. the theoretical (6), (7) and estimated variances (over the 1000 realizations) of each of the 19 individual estimates of the spatial frequency, and the number of times the different estimates have been retained in order to obtain the unique proposed estimate. For all the estimates the bias is negligible.

The standard deviation of the estimate we get with the proposed procedure is .47 × $10^{-3}$ while the minimum of the individual standard deviations (6), (7) is .61 × $10^{-3}$ and the CRB (8) is .45 × $10^{-3}$. From the histogram in Figure 2. one realizes how realization dependent our estimate is and that indeed all 19 estimates contribute with a selection rate that is somehow inversely proportional to their standard deviation.