

MULTIPATH TIME-DELAY ESTIMATION.

Jean-Jacques FUCHS

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

ABSTRACT

A transmitted and known signal is observed at the receiver through more than one path in additive noise. The problem is to estimate the number of paths and for each of them the associated attenuation and delay. It is a frequent problem in sonar, radar and geophysics. We propose an algorithm that is easy to implement, that has a reasonable computational load and seems to be able to solve the problem under more severe conditions (lower SNR) than previous methods.

1. INTRODUCTION

Let the observed signal $y(t)$ be modeled as :

$$y(t) = \sum_{p=1}^P a_p s(t - \tau_p) + n(t) \quad (1)$$

This model describes multipath effects where the emitted signal $s(t)$ is observed at the receiver through more than one path in additive noise $n(t)$. We consider the case where $s(t)$ is a known deterministic signal. a_p denotes the attenuation and τ_p the time delay for path p . The number of paths P is in general unknown. This situation arises in such fields as sonar, radar or geophysics. It amounts to model the effect of the propagation and reflexion as an attenuation and a delay and might well be to simple in many situations. Though (1) is written in continuous time, the processing will deal with discrete time samples and we only consider discrete time signals in the sequel. The delays τ_p are then non-integer multiples of the sampling period taken equal to one and we will assume that the sampled additive noise $n(\cdot)$ is white and gaussian.

2. THE PROBLEM

Observing y_t , the problem is to detect the number of replicas and to estimate $\{a_p, \tau_p\}$ for each of them. Under the gaussian white noise assumption, the maximum likelihood (ML) method leads to :

$$\min \sum_{t=1}^T |y_t - \sum_{p=1}^P a_p s(t - \tau_p)|^2$$

In the case of a single path, the minimum in τ is attained by maximizing $\sum_t y_t s(t - \tau)$. The optimal processing is thus to apply the matched filter i.e. to correlate the observed process $y(t)$ with the known signal $s(t - \tau)$ and to search

for the maximum. When $P > 1$, looking for the P highest peaks in the output of the matched filter is sub-optimal unless the pairwise differences $|\tau_p - \tau_l|$ are large compared to the temporal correlation of the signal $s(t)$. If this restriction on the delays is not satisfied, this technique does not resolve the different paths and is clearly sub-optimal.

We mainly consider this kind of situations and propose an algorithm that allows to detect the number of paths and to resolve them at a reasonable computational cost.

The Maximum Likelihood approach can also handle this situation. But it requires the knowledge of P the number of replicas and will converge to the global minimum only if an excellent initial point is known. For the type of situations we consider, closely spaced replicas, the maximum likelihood function has either not enough or too many local extrema, this approach is essentially unfeasible. As a matter of fact, the ML criterion is used in our procedure to select the best solution among a small number of candidates and to decide how many replicas are needed to explain the observations.

2.1. The model

In the simulations at the end of the paper we will consider the case where the transmitted signal is a windowed linear frequency modulated sinusoid, but the same procedure applies to any situation where the model (1) is valid. As a matter of fact we will not perform any demodulation and work with the raw real data. We are then in a situation where the maximum likelihood has many local extrema and will yield a good estimate only in the case where a precise initial point is already known. The difficulty is then to obtain a good initial point.

The algorithm we propose overcomes these difficulties and performs simultaneously the detection and estimation. It uses as observations the output of the matched filter. Denoting by over-bars the result of this operation and switching to discrete time, relation (1) becomes :

$$\bar{y}_k = \sum_{p=1}^P a_p \bar{s}(k - \tau_p) + \bar{n}_k \quad (2)$$

where $\bar{n}_k$ is no longer white noise. Since the difficulty we are considering is to resolve closely spaced paths and not to detect an isolated extremely weak replica, the localization of a limited zone of interest in this output $\bar{y}_k$ is an easy task

since the SNR's will be reasonable for all the paths. The interesting part of the output of the matched filter will then typically be of length one or two hundred. *Defining precisely how to characterize the interesting part of $\bar{y}_k$ is of course not an easy task. We do not consider this issue here. In the configuration we consider in the simulations the results are quite unsensitive to this choice which becomes delicate only in more difficult scenarios.*

2.2. The scheme of the approach

Let us denote by L , the length of the interesting part of $\bar{y}_k$ and by Y the column vector built upon these samples. This choice also fixes the domain in which the delays are to be sought. If the interesting part of $\bar{y}_k$ extends over L samples, the potential delays will generally belong to an time interval around the middle of Y of length a fraction of L .

Associated with Y , one can build in an obvious way P vectors S_{τ_p} such that (2) can be rewritten as :

$$Y = \sum_{p=1}^P a_p S_{\tau_p} + \text{noise} \quad (3)$$

Observing Y and knowing that it admits such a decomposition, our objective is to reconstruct it as a linear combination of a small number of such vectors, we denote S_m . These vectors of length L are built upon samples of $\bar{s}_k$ and each of it is associated with a given delay. These delays to be chosen among M preassigned values spread over the potential domain of interest, We thus seek a reconstruction of Y of the form :

$$Y \simeq \sum_m \alpha_m S_m + \text{noise} \simeq \underline{S} \underline{\alpha} + \text{noise} \quad (4)$$

with $\underline{S}$ an (L,M) -matrix and $\underline{\alpha}$ an M -dimensional column vector containing the unknown weights α_m . Note that, if the true delays are among the M preassigned values, (3) is precisely of this form and the weighting vector $\underline{\alpha}$ has then exactly P non-zero components. With this linearization procedure, the delay-estimates are hidden in the indices of the non-zero components of $\underline{\alpha}$.

The whole problem is now to define a criterion that allows, observing Y and knowing $\underline{S}$ in (4) to compute the optimal weights $\underline{\alpha}$. Ideally their should be a very small number of "clusters" of non-zero weights, each cluster representing a replica.

3. DELAY ESTIMATION AND INTERPOLATION

3.1. Introduction

One should note at this point that there are two sampling periods involved in this modelisation (4). The first is the sampling period of the data and is taken equal to one, the second is equal to the delay, we denote h , existing between two consecutive columns S_m and S_{m+1} of $\underline{S}$ in (4). Though these two periods need not be linked together we generally choose the second one to be a fraction of the first one. One must realize that, even in the absence of noise, if one of the true delays τ_p in (3) does not belong to the set of

M potential delays we are considering, forcing the identity between S_{τ_p} and $\sum_m \alpha_m S_m$ amounts to solve an interpolation problem. Indeed, in each of the linear equations (rows) of $S_{\tau_p} = \sum_m \alpha_m S_m$ one reconstructs one sample of $\bar{s}(k - \tau_p)$ using equispaced samples $\bar{s}(nh)$ of $\bar{s}(t)$. The weights α_m are thus samples from an interpolating function and our approach which can be seen as a model-fitting or deconvolution approach amounts to estimate the samples of an interpolating function and to deduce the number of replicas and their characteristics from the peaks of the estimated interpolating function. This is exactly what was already proposed in [4] in 1980.

The most well-known interpolating function is the sinus cardinal which works for the reconstruction of functions whose Fourier transform is band-limited provided the sampling period is small enough to satisfy the Nyquist (Shannon) rate. Though in our context, the conditions that allow perfect reconstruction are not all satisfied -the function $\bar{s}(t)$ may not be bandlimited, the number M of available samples is not infinite and there is additive noise present,- it is this feature of the approach that is important to understand.

In general, using weights that are samples of the interpolating function, one reconstructs the value of the signal at a given point from its values at an infinite number of equispaced points. Here, we work the other way round, we know both the interpolating value and the infinite sequence of values and seek estimates of the weights. To make the problem solvable we actually know L interpolating values and seek an unique sequence of weights that allows the reconstruction. In fact, with P the number of replicas, what we will estimate are samples from the superposition of P weighted and shifted interpolating functions.

3.2. Interpolation in case of oversampling

There is one important parameter we are allowed to adjust and this is the sampling period h . Indeed as explained above, the sampling period of the observations y_t is essentially imposed by the application and is only related to the number L of equations in (4). The sampling interval that intervenes in the interpolating function is the time interval h that exists between two contiguous columns of $\underline{S}$ in (4) and this interval h can be chosen. One somehow expects that if one chooses h small enough, the interpolation problem is easier...

Obviously, assuming $\bar{s}(t)$ to be band-limited, h has to be chosen small enough to satisfy the Nyquist rate, but it can indeed be chosen much smaller. If one tries to interpolate an over-sampled signal there are an infinite number of ways to reconstruct it. Indeed, consider for instance an over-sampling ratio of two, one can then reconstruct any point of the function using just the odd samples or the even samples or any convex combination of these two ways. Here since we actually estimate the interpolating function, the precise interpolating function that will come out of the estimation procedure, in case of oversampling, entirely depends upon the criterion we use to minimize $Y - \underline{S} \underline{\alpha}$. Our aim is thus to define a criterion, that will single out an interpolating function that is as close as possible to an impulse. Remember that the α 's we estimate are samples from the superposition of P weighted and shifted interpolating functions and that we want to deduce the number of replicas

and their characteristics from the peaks of the underlying function.

By decreasing h it is always possible to have more unknowns than equations ($M \geq L$) and thus have an infinite number of solutions to $Y = \underline{S} \underline{\alpha}$. The most common way to select one solution is probably to choose the minimum ℓ_2 norm solution. One can show that this leads to weights that are samples from the standard sinus cardinal function scaled by the Nyquist rate. Diminishing h though allowing intuitively for a more localized reconstruction will simply lead to further oversampling the same standard sinus cardinal function. This is thus an extremely bad way of selecting a solution. *What we say is that, under the conditions for which the standard sampling theorem holds, if one oversamples the function to be reconstructed, there are many ways to reconstruct any point between the samples. If one seeks the reconstruction function (or equivalently the weights) with minimal ℓ_2 norm, then one obtains as reconstruction function the sinus cardinal associated with the Nyquist rate. Oversampling the function in order to obtain a more localized reconstruction formula is useless if this criterion is used since it is always the same underlying function that will yield the weights.*

Minimizing the ℓ_1 norm of the interpolating function (weights) happens to be one criterion that leads to pretty localized reconstruction formulas. Indeed one expects that for h small enough, linear interpolation should lead to a close-to-perfect reconstruction and minimizing the ℓ_1 norm of the weights is a criterion that almost leads to this way of reconstructing a point in between two samples. The corresponding interpolating functions are analysed in [5] and are shown in Figure 1 and 2. together with those of minimal ℓ_2 norm.

In Figure 1, we consider an oversampling ratio of two and the resulting sampling period is taken equal to one. One of the curves is the standard sinus cardinal (divided by two) which is the interpolating function with minimal ℓ_2 norm. The *'s on this curve indicate the weights to be assigned to the samples when the midpoint between two samples has to be reconstructed, only the weights to be assigned to the 8 neighboring points on both sides are presented though, of course, an infinite number of samples and weights are needed to achieve a perfect reconstruction. The other curve is the the interpolating function with minimal ℓ_1 norm. The o's on this curve are the weights, with minimal ℓ_1 norm, to be assigned to the samples when the midpoint between two samples has to be reconstructed. Since the function vanishes for $|t| \in [2k-1, 2k]$ with $k > 0$, one notices that except for the two neighboring points in positions $\pm .5$ only one every second sample point is used in the reconstruction so that there are only fours o's on both sides. In Figure 2, the same curves are drawn for an oversampling ratio of three. The resulting sampling period is again taken equal to one. As explained above minimizing the ℓ_2 simply leads to different samples of the same function in the continuous time scale, while minimizing the ℓ_1 norm further improves the localization of the interpolating function. If one increases the oversampling ratio even further, minimizing the ℓ_1 norm of the weights allows to conclude that linear interpolation is close to perfect.

4. THE PROPOSED APPROACH

From the results described above and established in [5], we conclude that minimizing the ℓ_1 norm of the weights should lead to a quite efficient algorithm. The first idea would thus be to solve the following optimization problem :

$$\begin{aligned} \min \quad & \|\underline{\alpha}\|_1 \\ \text{s.t.} \quad & Y - \underline{S} \underline{\alpha} = 0 \end{aligned} \quad \text{LP1}$$

where $\|\underline{\alpha}\|_1$ stands for the ℓ_1 norm of $\underline{\alpha}$. Since this problem can be converted into a linear program [6] it has a unique minimum that is easily and quickly obtained even for large L (the number of linear equality constraints) and M (the number of unknowns) using standard available programs. This is however too simple an approach since it does not take into account the presence of the additive noise (3). Remember that this is -at best- filtered white noise (2) and may have quite a large variance. It is thus unjustified to ask for a perfect match between Y and $\underline{S} \underline{\alpha}$. A first improvement can be obtained by introducing a tolerance in the linear constraints. This leads to :

$$\begin{aligned} \min \quad & \|\underline{\alpha}\|_1 \\ \text{s.t.} \quad & |Y - \underline{S} \underline{\alpha}| \leq \rho \end{aligned} \quad \text{LP2}$$

where ρ is the tolerance, a positive real constant and the inequality constraint has to be taken componentwise and is thus equivalent to $\|Y - \underline{S} \underline{\alpha}\|_\infty \leq \rho$. This problem also can be rewritten as a linear program. The tolerance ρ is an important parameter that has to be tuned. Once this feature is introduced, the inequality $M \geq L$ has no longer to hold since, for ρ large enough, feasible points exist even if it not satisfied. The same holds for the conclusions drawn in section 3.2. As a matter of fact, for the results presented below we introduced a further modification and actually whitened the Y vector and the $\underline{S}$ matrix accordingly. This appears to be beneficial once a limited section (of length L) of the output of the matched filter has been selected but of course uses the assumption that the initial noise (1) is white and gaussian and removes some robustness from the procedure. Similar ideas and further details can be found in [3].

5. A SIMULATION RESULT

Let us describe the scenario for which we present some simulation results below. We take [2] :

$$s_t = w_t \cdot \sin(2\pi(\alpha t^2 + \beta t)) \quad t = 0, 1, \dots, N-1 \quad (5)$$

where $N = 750$, $\alpha = (f_2 - f_1)/2N$, $\beta = f_1$, $f_1 = .1$, $f_2 = .15$ and w_t is a window function equal to :

$$\begin{aligned} w_t &= 0.5 - 0.5 \cos\left(\pi \frac{t}{N_w}\right) & t = 0, 1, \dots, N_w - 1 \\ w_t &= 1. & t = N_w, N - N_w - 1 \\ w_t &= 0.5 - 0.5 \cos\left(\pi \frac{t - N}{N_w}\right) & t = N - N_w, \dots, N \\ w_t &= 0 & \text{otherwise} \end{aligned}$$

A three path received signal is generated as :

$$y_t = s_{t-200} - .8 s_{t-204} + .4 s_{t-220} + n_t, \quad t = 0, \dots, 999 \quad (6)$$

and the gaussian white noise variance v is tuned to yield the desired SNR.

Some simulation results obtained using the above described philosophy are presented below. They are obtained using a more elaborate version than those described above. L is taken equal to 250 is the samples are taken symmetrically around the global maximum of the output of the matched filter. The Y vector is whitened using the inverse of a square root of the covariance matrix of Y computed once and forever with the help of $s(t)$. The potential delays cover a domain placed symmetrically around the maximum of the output of the matched filter of size 60. This means that if the oversampling ratio is taken equal to 5, one has $M = 301$ potential values of the delays. Several values of the tuning parameter are considered for each realization. For each of them, the ML criterion is used to re-estimate the amplitudes and to detect the number of replicas using an Akaike like criterion. The best ML solution is retained among these candidates for each realization.

We performed 400 independent trials of the scenario described in (6) with a noise variance $v = 49.10^{-4}$. This is already a quite difficult configuration. The results are presented in Table 1 for $M = 301$ corresponding to $h = .2$. The procedure correctly (and easily) decided that the number of replicas was three and estimated their characteristics in all the 400 realizations.

replica number	1	2	3
replica SNR in dB	48	46	40
true delay	200	204	220
estimated delay	200.009	203.988	219.993
est. stdt dev. delay	.0245	.0360	.0346
C'R stdt dev. delay	.0220	.0293	.0185
true amplitude	1	- 0.8	0.4
estimated amplitude	1.0034	-.7966	.3994
est. stdt dev. amplit.	.0181	.0190	.0059
C'R stdt dev. amplit.	.0175	.0186	.0059

Table 1. Results over 400 trials for the 3 replicas process.

One should note that the correlation of the s_i (5) is indeed highly oscillatory and adjacent peaks may have nearly equal height making the estimation problem quite difficult. The bound on the variance of the estimates given by the Cramer-Rao inequality, which considers only the curvature of the highest peak is then relevant only for reasonably high SNR's. The method is quite promising because the major difficulty for this type of problem is by far to avoid local minima and the results in the table indicate that this is the case. Obviously the solution given by our method in all the 200 trials is in the domain of attraction of the global optimum.

REFERENCES

- [1] D.H. Johnson and D.E. Dudgeon. *Array Signal Processing. Concept and Techniques*. Signal Processing. Prentice Hall, 1993.
- [2] T.G. Manickam, R.J. Vaccaro and D.W. Tufts. A least squares algorithm for multipath time-delay estimation. *IEEE-T-SP*, 42:3229-3233, nov. 1994.

- [3] J.J. Fuchs. Extension of the Pisarenko method to sparse linear arrays. In *ICASSP*, volume III, pages 2100-2103, may 1995. submitted for publication -in an extended version- to *IEEE-T-SP* in july 1996.
- [4] Y.T. Chan, J.M. Riley and J.B. Plant. A parameter estimation approach to time-delay estimation and signal detection. *IEEE-T-ASSP*, 28:8-16, feb. 1980.
- [5] J.J.Fuchs and B. Delyon. The sampling theorem and time delay estimation. *submitted to SYSID 1997*, sept. 1996.
- [6] D. G. Luenberger. *Introduction to linear and nonlinear programming*. Addison Wesley, 1973.

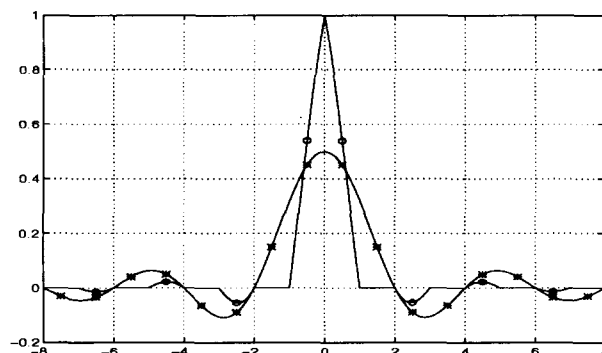


Figure 1. Two examples of the central part of interpolating functions in case of an oversampling ratio of two. The resulting sampling period is taken equal to one. The minimal l1 norm interpolating function decreases rapidly to zero and uses one sample out of two. The minimal l2 norm interpolating function is a sinus cardinal function, decreases slowly towards zero and uses all the samples

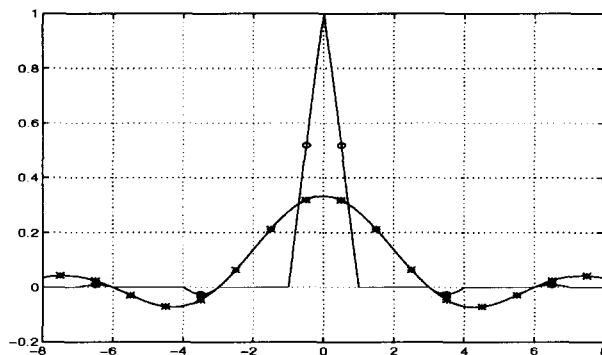


Figure 2. The same as in the figure 1 above for an oversampling ratio of three. The resulting sampling period is again taken equal to one. The minimal l1 norm interpolating function decreases rapidly to zero and uses one sample out of three. The minimal l2 norm interpolating function is the same as in figure 1, decreases slowly towards zero and uses all the samples