An Iterative Quadratic Method for High Resolution Delay Estimation with Known Waveform

Umberto Spagnolini

Abstract—The iterative delay estimation method is based on its analogy with parameter estimation of sinusoids in correlated noise. It assumes a-priori knowledge of echo waveform, their number, and the structure of the covariance matrix of noise. The advantages of this method are improved resolution as compared to the conventional matched filter approach and convergence, in a few iterations, to maximum likelihood estimate.

I. INTRODUCTION

Delay estimation in a multipath environment with a known pulse waveform is, in remote sensing applications, a recurrent problem that has received much attention in the past. The most common methods for many applications are still based on a matched filter approach (or a modification) that is known to have resolution limited by pulse width (Tc). When the fine structure of the investigated media is of any interest, the backscattered echoes have delays that are closely spaced when compared to Tc. Multiple delays can be estimated by minimizing the weighted residuals between modeled and measured data according to the Maximum Likelihood (ML) criterion. Algorithms for multipath delay estimation differ by the optimization strategy used to avoid the local minima of this multimodal objective function.

Optimization by exhaustive search is known to be impractical and the global minimum needs appropriate methods for the search. Alternating projection methods [10], estimate methods, and maximize methods [2] (or any combination) all rely on direct optimization of the likelihood function using iterative methods. Manickam et al. [5] suggested exploiting the properties of the envelope of the analytic signals to reduce the local minima in delay estimation. Since the delay estimation problem after Fourier transformation is similar to frequency estimation of exponential signals (or the direction of arrivals in array processing), the ML estimator can be reparameterized in terms of prediction polynomials [9].

Optimization of the likelihood function can be obtained by the iterative quadratic ML (IQML) method [3], [1]. The IQML is known to provide an equivalent representation of the ML estimator that converges better (when compared to other iterative methods that deal directly with the likelihood function) to the global minimum even when relative delays are closer than Tc (i.e., the matched filter resolution). Since convergence of the IQML is sensitive to the initialization (mainly for low SNR’s), the proposal of this correspondence is to estimate the delays first by exploiting the advantage of the envelope of analytic representations of signals [5] and then, using this estimate as initialization, by refining the delay estimate with real valued signals (two-step IQML). Since the envelope is known to have less local minima but a greater delay error compared to real signals, the second IQML step becomes necessary.

1) Notation: Lowercase (uppercase) indicates the sampled (continuous) in time (frequency) domain, lowercase (uppercase) bold denotes vectors (matrices), (·)T is the matrix transpose, (·)* is the complex conjugate, (·)H is the Hermitian transposition (i.e., complex conjugate transpose (·)* = (·)H), ||x||2 = xHx is the norm, and ||x||A2 = xHAx is the weighted norm. PA = A(AHA)−1A is the projection matrix onto the column space of the matrix A, and P−1 A = I − PA is its complementary projection, the nested matrix operators are compactly written in standard matrix notation ((A−1)−1/2 becomes A−1/2).

II. THE PREDICTION POLYNOMIAL REPRESENTATION OF MLE

The sampled measured signal in any specific experiment can be modeled as a superposition of L echoes:

\[ s(iT) = \sum_{l=1}^{L} a_l w(iT - \tau_l) + n(iT), \quad 0 \leq i \leq M - 1 \]  (1)

\[ w(t) \] is the known real valued pulse waveform with time resolution Tc, \( a_l \) and \( \tau_l \) are respectively the amplitude and delay of the echo due to the \( l \)th target, \( n(iT) \) is the zero mean Gaussian noise and \( T \) is the sampling interval. Provided that \( M \) samples of the observation window fully contain the delayed waveforms, the received signal (1) can be rewritten in terms of the discrete Fourier transform (DFT)

\[ S(k) = W(k) \sum_{l=1}^{L} a_l e^{j2\pi k \tau_l} + N(k), \quad k = 0, 1, \ldots, M - 1 \]  (2)

where the delay of the \( l \)th echo \( \phi_l = -2\pi \tau_l / (MT) \) has been normalized with respect to the observation interval \( M \cdot T \). The relationship (2) can be rewritten in matrix notation

\[ \mathbf{S} = \mathbf{W} \mathbf{D} \mathbf{\phi} \mathbf{a} + \mathbf{N} \]  (3)

where \( \mathbf{S} = [S(0), S(1), \ldots, S(M-1)]^T \), \( \mathbf{W} = \text{diag}[W(0), W(1), \ldots, W(M-1)] \), \( \mathbf{a} = [a_1, a_2, \ldots, a_L]^T \), and \( \mathbf{D} \mathbf{\phi} = [d(\phi_1), d(\phi_2), \ldots, d(\phi_L)] \). Each vector \( d(\phi_l) = [1, e^{j2\pi \phi_l}, \ldots, e^{j2\pi (L-1)\phi_l}]^T \) depends on the delay \( \phi_l \). The ML estimation of the amplitudes \( \mathbf{a} \) and delays \( \mathbf{\phi} = [\phi_1, \phi_2, \ldots, \phi_L] \) can be obtained from the minimization of the weighted error

\[ \mathcal{E}(\phi, \mathbf{a}) = ||\mathbf{S} - \mathbf{W} \mathbf{D} \mathbf{\phi} \mathbf{a}||_Q^{-1} \]  (4)

Here, \( Q = E[\mathbf{N} \mathbf{N}^H] \) denotes the known covariance matrix of the DFT of noise process. For notational convenience, dependence on the delays \( \mathbf{\phi} \) is understood (e.g., \( \mathbf{D} = \mathbf{D}(\phi) \)). Minimization of (4) belongs to the class of separable nonlinear optimization problems that can be carried out in two steps, first with respect to \( \mathbf{a} \) and then with respect to \( \mathbf{\phi} \). This latter optimization

\[ \hat{\phi} = \arg\min_{\phi} \{ \mathbf{S}^H Q^{-1/2} \mathbf{P}^{-1}_{\hat{\phi}} Q^{-1/2} \mathbf{S} \} \]  (5)

is nonlinear and needs to be carried out as an expensive multidimensional search for the set \( \phi \) (\( \mathbf{D} = Q^{-1/2} \mathbf{W}, \) and \( Q^{-1/2} \) denotes the Cholesky factor of \( Q^{-1} \)).
The method discussed here is based on the analogy between the delay model after DFT (2) and the model for the estimation of the frequency of sinusoids or the direction of arrivals in uniform linear array, the only difference is the presence of the waveform spectrum $W(k)$, which can be compensated as described later. The number of echoes $L$ is assumed to be known or to be correctly estimated (e.g., by a hypothesis testing method). The basic idea for delay estimation relies on the finite order linear prediction polynomial of the signal according to the model (2). Let

$$b(z) = b_0 \sum_{l=0}^{L} (z^{-l} - e^{j\omega l}) = \sum_{l=0}^{L} b_l z^{-l}, \quad |b_0| \neq 0 \quad (6)$$

be the $L$-th-order polynomial characterized by the coefficients $b^T = [b_0, b_1, \ldots, b_L]$, the roots of the polynomial are the delays $\phi$ as $b(e^{j\omega l}) = 0, \forall \ell = 1, \ldots, L$.

According to the definition above, vector $b$ is orthogonal to each subset of length $L+1$ extracted from the vector $d(\phi_l)$. The polynomial coefficients can be arranged in an $(M-L) \times M$ Sylvester matrix, in which each column contains the shifted versions of the vector $b$ as

$$B^H = \begin{bmatrix} b_L \ldots b_0 & 0 \ldots 0 \\ 0 & b_L \ldots b_0 \\ \vdots & \vdots & \vdots \\ 0 \ldots 0 & b_L \ldots b_0 \end{bmatrix} \quad (7)$$

it can be easily verified that $B^H D = 0$ or equivalently $B$ is orthogonal to the column space of $D(\phi)$. Since the matrix $B = Q^{1/2} W^{1/2} B^{H}$ spans the null space of matrix $D$ as $B^H D = B^H D = 0$ or $P_{\phi} = P_B$, the delay estimate (5) can be rewritten in terms of the estimate $\hat{b}$ of the coefficients of the polynomial $b(z)$. From the coefficients $b$, the delays can be found from the phases of the roots of the polynomial $\sum_{l=0}^{L} b_l z^{-l}$. According to the above properties, the objective function (4) needs to be minimized with respect to vector $b$ as

$$\mathcal{E}(b) = \tilde{S}^H B (B^H Q B)^{-1} B^H \tilde{S} \quad (8)$$

where the noise covariance matrix $\tilde{S} = W^{-1} Q W^{1/2}$ and the spectrum of the signal $S = W^{-1} S = [S(0), S(\lambda_1), \ldots, S(M-1)]^T$ are both deconvolved according to the DFT of the waveform $W$. For convenience, related to analytical signal, $S$ is sorted for increasing frequency bins ($S(k)$ and $W(k)$ are periodic with a period of $M$) $S(k) = S(k-M/2) W(k-M/2)$ for $k = 0, 1, \ldots, M-1$. It can be shown that optimization of the objective function (8) represents an attractive alternative for the delay estimation, as it is an exact expression of the ML criterion; 2) it has a computational advantage compared to a multidimensional search, as it converges into a few iterations of a quadratic function; and 3) it shows an improved resolution compared to conventional matched filter methods. Discussion on topics 1 and 2 can be found in the literature for the parameters estimation of exponential signals [3], [1]. Here, we consider only the main aspects of the optimization algorithm for the delay estimation. A few remarks are in order to differentiate the IQML for delay estimation.

**Remark 1:** Spectra compensation $\tilde{S} = W^{-1} S$ can lead to instability problems for small (or null) values of $\{W(k)\}_{k=0}^{M-1}$. The frequency interval can be limited to any continuous subset where the signal is sufficiently strong [4]. Alternatively, some (small) term $\varepsilon$ can be added to stabilize the compensation $\tilde{S} = S(k-M/2) W(k-M/2)$. Here we suggest $\varepsilon \approx (1/100) \max_{k} \{W(k)\}$.

**Remark 2:** The structure of the covariance matrix depends on the correlation of the Gaussian noise. The analysis can be restricted to white noise, $Q = \sigma_n^2 I$, or correlated noise with the same power density spectra as signal (e.g., stationary clutter), $Q = \sigma_n^2 W W^H$. The number of echoes can be found from the phases of the roots of the polynomial $b(z)$.

The structure of the covariance matrix depends on the correlation of the Gaussian noise. The analysis can be restricted to white noise, $Q = \sigma_n^2 I$, or correlated noise with the same power density spectra as signal (e.g., stationary clutter), $Q = \sigma_n^2 W W^H$. The number of echoes can be found from the phases of the roots of the polynomial $b(z)$.
where $C^{(b)} = (B^{(b)H}QB^{(b)})^{-1/2}$ has size $(M - L) \times (L + 1)$ and is obtained from the Cholesky factor of $B^{(b)H}QB^{(b)}$, the signal $\{ S(k) \}_{k=0}^{M-L-1}$ is arranged into an $(M - L) \times (L + 1)$ matrix

$$
S = \begin{bmatrix}
\hat{S}(L) & \hat{S}(L-1) & \cdots & \hat{S}(0) \\
\hat{S}(L+1) & \hat{S}(L) & \cdots & \hat{S}(1) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{S}(M-1) & \hat{S}(M-2) & \cdots & \hat{S}(M-L-1)
\end{bmatrix}.
$$

(11)

The polynomial coefficients need to be constrained to have conjugate symmetry $b_{L-k} = b_k^*$, as this choice lets the polynomial have the roots on the unit circle [1]. This constraint can be expressed by partitioning the vector $b$ and the matrix $C^{(b)}$ into a set of independent coefficients (i.e., the real and imaginary parts of polynomial coefficients $b$) that need to be separated for even and odd $L$ (see [8] and [9] for a detailed discussion of the numerical implementation).

Minimization of (10) needs to be constrained to avoid the trivial solution $b = 0$. Here, we have chosen $\text{Re}\{b_0\} = 1$ or $\text{Im}\{b_0\} = 1$ according to [8], even if other constraints are equally possible (see [1] for a detailed discussion). The constrained minimization can be rewritten as an unconstrained problem, unknowns $b$ can be obtained as a solution of normal equations, or it can be obtained by Q-R decomposition according to the method used for direction of arrivals estimation [8].

A. Two-Step IQML

Practical use shows that the IQML converges in 2–4 iterations and, for high SNR, is not highly sensitive to the initial value for polynomial coefficients $b^{(0)}$. The initial estimate $b^{(0)}$ can be chosen according to the specific problem. A simple choice could be $B^{(b)H}B^{(b)} = I$ or $b^{(0)} = -b^{(0)} = 1$ (i.e., all delays are equally spaced within the observation window $\phi^{(0)} = 2\pi t/L$ or the direct estimate of the delays by using the ESPRIT algorithm [9]. In this correspondence, we propose to iterate the IQML algorithm first on the envelope of analytic representation of the signal $\{ s(t) \}_{t=0}^{T-1}$ by initializing the search with

$$
\phi^{(0)} = 2\pi t/L.
$$

Next, the IQML is exploited for real signals (two-step IQML). The advantage arises from the objective function in the optimization (5), as it shows less local minima when carried out with the envelope. However, since the delay estimate carried out on the envelope is known to be less accurate [7], the second step is needed to improve the accuracy of the delay estimates.

Analytic signals are complex-valued with a real and imaginary part that are their respective Hilbert transforms. The DFT of the analytic signal is simply related to $S(k)$ by multiplying $S(k)$ for $(1 + \text{sign}_M(k))$ with $\text{sign}_M(k) = 1$ for $1 \leq k < M/2$, $\text{sign}_M(k) = 0$ for $k = 0, M/2$, and $\text{sign}_M(k) = -1$ for $M/2 + 1 < k \leq M - 1$. Therefore, the prediction polynomial can be rewritten for analytic signals. The only meaningful differences are the use of $(M^2 + 1)$ elements vector $S = [S(0), 2S(1), \ldots, 2S(M/2 - 1), S(M/2)]^T$, and matrix $W = \text{diag}\{W(0), 2W(1), \ldots, 2W(M/2 - 1), W(M/2)\}$ in (9). The main advantage is that the likelihood function is less affected by local minima when $\mathcal{L}(\phi, a)$ is rewritten in terms of signal envelope, it should be noticed that in this case the estimates $\{a_\ell\}_{\ell=1}^M$ are complex valued. The IQML iterations are not reproposed as they are similar to those described up to now, what is needed is redefinition according to the vectors $S$ and $W$ described above. Let $b^{(1)} = \hat{b}^{(1)}_0$ be the prediction polynomial at convergence for signal envelope, the initialization of the prediction polynomial $b^{(0)}$ for real signals is based on these estimates: $b^{(0)} = b^{(1)}_0$.

IV. EXAMPLES

Examples demonstrate the ability of the IQML method to estimate the delays with closely spaced echoes (or below the resolution of matched filter method). All the tests have been carried out with the second-order derivative of a Gaussian pulse $w(t) = g(t) = (1 - (t/T_w)^2)e^{-2\pi^2} / T_w^2$ as it is representative of some remote sensing waveforms [7]. Fig. 1 shows a simulation of a model with two echoes whose delays $\{\tau_1, \tau_2\}$ are reducing, while the SNR is increasing (for $Q = \sigma_n^2 I$, it is $\text{SNR} = \sum_k w^2(kT) / \sigma_n^2$). Visual inspection indicates that the delays estimated by two-step IQML...
(i.e., white dots are superimposed on the simulated signals), are consistent with the model even when the relative delay $\Delta \tau_{ij} \equiv |\tau_{ij} - \tau_j|$ is close to (or below) $T_w$.

Quantitative analysis in terms of the RMSE of the delay estimates and the probability of resolution ($P_{res}$) for varying SNR and delays $\Delta \tau_{ij}$ is shown in Fig. 2. Two delays are resolved when $|\tau_{ij} - \tau_j| < \Delta \tau_{ij}/2$ for $\ell = 1$ and $2$ simultaneously for both echoes. In all these examples, the pulse waveform is assumed to preserve the bandwidth properties, but it has a known phase distortion $\theta$ (i.e., $w(t) = g_\theta(t) = \text{Re}\{[g(t) + j\tau[g(t)]\exp[5\theta]], \tau[\cdot]\}$ is the Hilbert transform). To stress the convergence properties of IQML in all the tests, the initialization is randomly chosen within the observation window. For the parameters in Fig. 2, the value of $P_{res}$ when the estimated delays are randomly distributed within $M \cdot T$ is $P_{res} \approx (\Delta \tau_{ij} / MT)^2 \leq 0.02$. The RMSE for a uniform distribution of delays within $\Delta \tau_{ij}$ (i.e., the delays that passed the resolution test) is $\Delta \tau_{ij} / 2\sqrt{3}$. Both values represent the bound for low SNR and/or $\Delta \tau_{ij} \ll T_w$. When the SNR is high and $\Delta \tau_{ij} > 2T_w$, the RMSE attains the Cramér–Rao Bound (CRB) for the delay estimation (see e.g., [7]). In any case, $P_{res} > 0.5$ for $\Delta \tau_{ij} > T_w/2$ and SNR > 40 dB. For $\Delta \tau_{ij} < T_w$ and low SNR, the delay estimates are dominated by the local minima of the objective function. In this case, the IQML for signal envelope is enough to guarantee good performances when evaluated in terms of $P_{res}$ and RMSE. A simple explanation of this contradicting result is that even if the signal envelope has lower effective bandwidth (or lower resolution), the IQML method can efficiently maximize the likelihood function that for a signal envelope has a lower number of local extrema compared to the direct use of $s(t)$ [5]. Asymptotically, for large SNR and $\Delta \tau_{ij} \gg T_w$, the RMSE of two-step IQML is, as expected, lower than for IQML for signal envelope. Fig. 2 shows the case of SNR = 60 dB.

An example from monostatic GPR experiments for pavement profiling is shown in Fig. 3. Some details on the specific application as well as on the measurement setup to obtain $w(t)$ are discussed in [6]. From core samples, shallow reflections are known to contain some closely spaced reflections where the delays need to be estimated to infer the fine structure of the asphalt layers. In this case, the resolution of the matched filter approach is known to be insufficient [6]. Two-step IQML has been used with $L = 9$, delays at the first step (initialization for the envelope) are equally spaced within the observation window consisting of $M = 256$ samples. In this case, some of the false alarms are due to the overparameterization of the IQML (i.e., $L$ is too large), and post-processing can be exploited for their reduction.

V. CONCLUSIONS

The ML method can be used to estimate the delays of closely spaced echoes when the pulse waveform is known. This correspondence has shown that the multidimensional search of ML algorithms can be avoided by a suitable parameterization of the nullspace that leads to the optimization of a quadratic cost function. The IQML algorithm is sensitive to initialization but, for practical purposes, it converges in a few steps. The two-step IQML has been proposed to cope with local minima of likelihood function. The approach is effective as it increases the probability of resolution for closely spaced echoes. Examples confirm that the use of the signal envelope could be recommended to estimate echo delay or any related properties.

REFERENCES