V. Conclusion

In this paper, an analytical characterization of DS/CDMA global noise has been presented, which is based on applying HOS-based concepts to a non-Gaussian parametric distribution model. Results have been presented for the worst case of deviation from Gaussianity, which demonstrate the different behavior of the chosen HOS parameter, namely, normalized kurtosis, for different choices of PN sequences. Moreover, the performed analysis allowed us to show that the Gaussian approximation, which is usually considered in literature, underestimates the BER in few-user systems when certain PN sequences (e.g., Gold sequences) are employed for spreading, whereas it can be effectively used in high safety applications when other sequences are adopted (e.g., EOE-Gold).

Appendix A

The computation of the second- and fourth-order statistical moments of the r.v. Z for complex PN sequences is here presented. The value of $E(Z_{k,1}^2)$ for complex PN sequences is given by

$$E\left(Z_{k,1}^2\right) = \frac{1}{2T}\sum_{l=0}^{N-1}\int_{lT_c}^{(l+1)T_c}\left[|R_{k,1}(\tau)|^2 + |\hat{R}_{k,1}(\tau)|^2\right]d\tau \quad \text{(A1.1)}$$

where $R_{k,1}(\tau) \hat{=} \int_0^{\tau} a_k(t-\tau)a_1^*(\tau)\,d\tau$ and $\hat{R}_{k,1}(\tau) \hat{=} \int_{\tau}^{T} a_k(t-\tau)a_1^*(\tau)\,d\tau$. These quantities are linked directly to the aperiodic cross-correlation functions among complex PN sequences, as seen in [4] and [6]. The expression reported in (2.4) is obtained after mathematical manipulation. The mentioned *polynomial function* $f_v(\cdot)$ is analytically defined as

$$f_v(x,y,z,w) \hat{=} x^2 + y^2 + z^2 + w^2 + xy + zw. \quad \text{(A1.2)}$$

Similar considerations can be made for the computation of the fourth-order statistical moment. The value of $E(Z_{k,1}^4)$, in the case of complex PN sequences, is given by

$$\begin{aligned} E\left(Z_{k,1}^4\right) = \frac{3}{8T}\int_0^T & \left[\left(A_{k,1}^4 + \hat{A}_{k,1}^4\right) + 6\left(A_{k,1}^2\hat{A}_{k,1}^2\right)\right] \\ &+ \left[\left(B_{k,1}^4 + \hat{B}_{k,1}^4\right) + 6\left(B_{k,1}^2\hat{B}_{k,1}^2\right)\right] \\ &+ 2\left(A_{k,1}^2 + B_{k,1}^2 + \hat{A}_{k,1}^2\hat{B}_{k,1}^2 + A_{k,1}^2\hat{B}_{k,1}^2 + \hat{A}_{k,1}^2 B_{k,1}^2\right. \\ &\left. + 4A_{k,1}B_{k,1}\hat{A}_{k,1}\hat{B}_{k,1}\right)d\tau \end{aligned} \quad \text{(A1.3)}$$

where $A_{k,1} \hat{=} \mathrm{Re}\{R_{k,1}(\tau)\}$, $\hat{A}_{k,1} \hat{=} \mathrm{Re}\{\hat{R}_{k,1}(\tau)\}$ $B_{k,1} \hat{=} \mathrm{Im}\{R_{k,1}(\tau)\}$, and $\hat{B}_{k,1} \hat{=} \mathrm{Im}\{\hat{R}_{k,1}(\tau)\}$. As $A_{k,1} = a_{k,l}T_c + b_{k,l}(\tau - lT_c)$, $\hat{A}_{k,1} = \hat{a}_{k,l}T_c + \hat{b}_{k,l}(\tau - lT_c)$, $B_{k,1} = c_{k,l}T_c + d_{k,l}(\tau - lT_c)$, $B_{k,1} = \hat{c}_{k,l}T_c + \hat{d}_{k,l}(\tau - lT_c)$ for some integer l, for which $lT_c \le \tau \le (l+1)T_c$ [1], the expression (2.5.1) can be obtained after some purely mathematical computations. The *polynomials functions* $f_q(\cdot)$, $g_q(\cdot)$, $h_q(\cdot)$ of (2.5.1) are defined as

$$\begin{aligned} f_q(x,y,z,w) \hat{=} &(x^4 + z^4) + 2(x^3y + z^3w) \\ &+ 2(x^2y^2 + z^2w^2) + (xy^3 + zw^3) \\ &+ \frac{1}{5}(y^4 + w^4) + 6(x^2z^2 + xyz^2 + x^2zw) \\ &+ 2(x^2w^2 + y^2z^2 + 4xyzw) \\ &+ 3(xyw^2 + y^2zw) + \frac{6}{5}y^2w^2 \end{aligned} \quad \text{(A1.4)}$$

$$\begin{aligned} g_q(x,y,z,w) \hat{=} &\frac{2}{5}y^2w^2 + (xyw^2 + y^2zw) \\ &+ \frac{2}{3}(x^2w^2 + 4xyzw + y^2z^2) \\ &+ 2(x^2zw + xyz^2) + 2x^2z^2 \end{aligned} \quad \text{(A1.5)}$$

and

$$\begin{aligned} &h_q(x,y,z,w,m,n,o,p) \\ &\quad \hat{=} 8xzmo + 4[xz(mp + no) + mo(xw + yz)] \\ &\qquad + \frac{8}{3}[xznp + (xw + yz)(mp + no) + ywmo] \\ &\qquad + 2[np(xw + yz) + yw(mp + no)] + \frac{8}{5}ywnp. \end{aligned} \quad \text{(A1.6)}$$

References

[1] K. Yao, "Error probability of asynchronous spread spectrum multiple access communication systems," *IEEE Trans. Commun.*, vol. COMM-25, pp. 803–809, Aug. 1977.

[2] D. Laforgia, A. Luvison, and V. Zingarelli, "Bit error rate evaluation for spread spectrum multiple access systems," *IEEE Trans. Commun.*, vol. COMM-32, pp. 660–669, June 1984.

[3] J. S. Lehnert and M. B. Pursley, "Error probabilities for binary direct-sequence spread spectrum communications with random signature sequences," *IEEE Trans. Commun.*, vol. COMM-35, pp. 87–98, Jan. 1987.

[4] M. B. Pursley, "Performance evaluation for phase-coded spread spectrum multiple-access communication—Part I: System analysis," *IEEE Trans. Commun.*, vol. COMM-25, pp. 795–799, Aug. 1977.

[5] A. Lam and F. Ozluturk, "Performance bounds of DS/SSMA communications with complex signature sequences," *IEEE Trans. Commun.*, vol. 40, pp. 1607–1614, Oct. 1992.

[6] D. Sarwate and M. B. Pursley, "Correlation properties of pseudorandom and related sequences," *Proc. IEEE*, vol. 68, pp. 593–619, May 1980.

[7] H. Fukumasa, R. Kohno, and H. Imai, "Design of pseudonoise sequences with good odd and even correlation properties for DS/CDMA," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 828–835, June 1994.

[8] A. Tesei and C. S. Regazzoni, "HOS-based generalized noise pdf models for signal detection optimization," *Signal Process.*, vol. 65, no. 2, pp. 267–281, Mar. 1998.

[9] M. Kendall, A. Stuart, and J. K. Ord, *Advanced Theory of Statistics: Distribution Theory*, 5th ed. London, U.K.: C. Griffin, 1987, vol. 1.

Multipath Time-Delay Detection and Estimation

Jean-Jacques Fuchs

***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. We propose a deconvolution approach with an additive regularization term built around an ℓ_1 norm. The underlying optimization problem is transformed into a quadratic program and is, thus, easily and quickly solved by standard programs. The procedure is able to handle more severe conditions than previous methods.**

I. Introduction

Let the observed signal $z(t)$ be modeled as

$$z(t) = \sum_{p=1}^{P} a_p h(t - \tau_p) + e(t). \quad (1)$$

This model describes multipath effects where the emitted signal $h(t)$ is observed at the receiver through more than one path in additive

Manuscript received April 1, 1997; revised May 27, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ali H. Sayed.

The author is with IRISA/Université de Rennes I, Rennes, France (e-mail: fuchs@irisa.fr).

Publisher Item Identifier S 1053-587X(99)00166-X.

noise $e(t)$. We consider the case where $h(t)$ 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 modeling the effect of the propagation and reflexion as an attenuation and a delay. This model might well be too simple in many situations since we consider that there are no scattering effects and that the attenuations are real numbers. Although (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 noninteger multiples of the sampling period taken equal to one.

The classical method for estimating the arrival times is matched filtering, which consists of correlating the received with the transmitted waveform. The resolution of this approach is limited by the width of the main lobe of the autocorrelation of the transmitted signal. To improve the performance, a number of methods have been proposed. Some of them are maximum likelihood techniques; they often suffer from high computational cost and need a precise initial point. An interesting alternative, combining matched filtering and deconvolution, is the iterative method of projection onto convex sets (POCS). The method of complex to real least squares (CRALS) time delay estimation focuses on the resolution of overlapping signals with noninteger delays.

In this correspondence, we propose a method that is based on a deconvolution approach with an ℓ_1 regularization term. The importance of this term is established in an original reconstruction theorem [6] for oversampled signals that shows that minimizing the ℓ_1 norm of the reconstruction function yields a function with much faster decay than the usual sinus cardinal function. It is this feature that allows for the resolution of closely spaced paths in more difficult scenarios. The optimization of the convex criterion is achieved using a standard quadratic programming routine.

In the next section, we formalize the problem and introduce the philosophy of the approach. In Section III, we comment on the connection between delay estimation and interpolation and state the interpolation theorem for oversampled bandlimited signals. The ℓ_1 regularized deconvolution criterion is presented in Section IV, where the whole procedure is described. Some simulation results are presented in the last section that allow for comparisons with existing schemes. The proposed method outperforms both POCS and CRALS by about 20 dB.

II. The Problem

Observing $z(t)$ in (1), 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 and for P known, the maximum likelihood (ML) method leads to

$$\min_{\{a_p,\tau_p\}} \sum_{t=1}^{T} \left| z(t) - \sum_{p=1}^{P} a_p h(t-\tau_p) \right|^2. \tag{2}$$

For $P = 1$, the minimum in τ is then obtained by maximizing $\sum_t z(t)h(t-\tau)$. This amounts to correlation of the observed process $z(t)$ with the known signal $h(t-\tau)$ and to a search for the maximum, i.e., to apply the matched filter. When $P > 1$, looking for the P highest peaks in the output of the matched filter is suboptimal unless the pairwise differences $|\tau_p - \tau_l|$ are large compared with the temporal correlation of the signal $h(t)$. If this restriction on the delays is not satisfied, this technique does not resolve the different paths and is clearly suboptimal.

To handle these more complex scenarios, the maximum likelihood approach (2) 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, i.e., closely spaced replicas, the maximum likelihood function has many local extrema, and (2) 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 candidate solutions and to decide how many replicas are needed to explain the observations.

A. The Model

The algorithm we propose uses as observations the output of the matched filter or the signal $z(t)$ itself if the matched filter cannot be used. When applying the matched filter to $z(t)$, (1) becomes

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

where $y(t)$, $s(t)$, and $n(t)$ are the outputs of the matched filter when applied, respectively, to $z(t)$, $h(t)$, and $e(t)$. The signal $s(t)$ is thus the autocorrelation of $h(t)$, and $n(t)$ is no longer white noise. Switching to discrete time, we rewrite the previous relation as

$$y_k = \sum_{p=1}^{P} a_p s(k-\tau_p) + n_k. \tag{3}$$

Since the difficulty we are considering is to resolve closely spaced paths and not to detect an isolated extremely weak replica, the localization of one or several limited zones of interest in this output y_k is an easy task since the SNR's will be reasonable for all the paths. In the case where there are several well-separated zones, each of them can be processed separately. We only consider the processing of one of these zones. Its length will then be reasonable and not exceed a few hundred samples.

B. The Philosophy of the Approach

Let us denote by L the length of the interesting part of y_k and by Y the column vector built on these samples. This choice also fixes the domain in which the delays are to be sought. The potential delays will generally belong to a time interval around the middle of Y whose length is a fraction of L.

Associated with Y, there is a noise vector N built from samples of n_k and P vectors S_{τ_p} such that (3) can be rewritten as

$$Y = \sum_{p=1}^{P} a_p S_{\tau_p} + N. \tag{4}$$

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 column-vectors built on samples of s_k. We denote S_m these vectors of length L that are built similarly to S_{τ_p}. Each of it is associated with a given delay. These delays to be chosen among M preassigned values cover the potential domain of interest. We thus seek a reconstruction of Y of the form

$$Y = \mathbf{S}\mathbf{w} + E \tag{5}$$

where $\mathbf{S}$ is an (L, M) matrix, and $\mathbf{w}$ is an M-dimensional column-vector containing the unknown weights w_m. E denotes a vector modeling the reconstruction error that has yet to be specified. If the true delays τ_p are among the M preassigned values, (4) is precisely of this form. Taking $E = N$, a possible weighting vector $\mathbf{w}$ then has exactly P nonzero components. For $M > L$, other weighting vectors do exist in general, even in this ideal case.

For $M > L$, there are many solutions $\mathbf{w}$ to (5), and the problem is then to find a criterion that allows retrieval of the *best* one, i.e., one with few nonzero components at the true locations. The delay estimates are then hidden in the indices of the nonzero components of $\mathbf{w}$ and the number of replicas in the number of clusters of nonzero weights.

III. Delay Estimation and Interpolation

A. Introduction

There are two sampling periods involved in the modelization (5). The first is the sampling period of the data that is taken equal to one, and the second is equal to the delay, which we denote h, as existing between two consecutive columns S_m and S_{m+1} of $\mathbf{S}$ in (5). To simplify the exposition, let us consider the simplest case where $Y = S_\tau$. A typical equation (row) in (5) then has the form

$$s(t-\tau) = \sum_m s(t-mh)w_m. \tag{6}$$

This clearly indicates that the weights $\{w_m\}$ we are seeking are samples from an interpolation or reconstruction function. The most well-known interpolating function is the sinus cardinal function, which works for the reconstruction of functions whose Fourier transform is bandlimited, provided the sampling period is small enough to satisfy the Nyquist (Shannon) rate. Indeed, if $s(t)$ is a signal whose Fourier transform vanishes for $|f| \geq f_{\max}$ and if the sampling period h satisfies $h \leq \frac{1}{2f_{\max}}$, the standard reconstruction or interpolation theorem yields

$$s(t-\tau) = \sum_{m=-\infty}^{+\infty} s(t-mh)\,\mathrm{sinc}\left(\frac{1}{h}(mh-\tau)\right) \tag{7}$$

where $\mathrm{sinc}(\cdot)$ denotes the (unscaled) sinus cardinal given by $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

Comparing (6) and (7) confirms this interpretation of the weights in $\mathbf{w}$. Our approach thus amounts to estimating the samples of an interpolating function and deducing the number of replicas and their characteristics from the peaks of the estimated interpolating function. This is exactly what was proposed for a different problem in [5] and solved using a least squares approach.

In general, using weights that are samples of the interpolating function, we reconstruct the value of the signal at a given point from its values at an infinite number of equispaced points. In the present context, we work the other way around. We know both the interpolating value $s(t-\tau)$ and the sequence of values at equispaced samples $s(t-mh)$ and seek the samples $\mathbf{w}$ from an interpolating function. To make the problem solvable, we actually know a large number L of interpolating values and seek the common sequence of weights $\mathbf{w}$ that simultaneously permits the L reconstructions.

Another major difference is that we are allowed to adjust the sampling period h that intervenes in the interpolating function (7, 6). We expect that the smaller h is, the easier the interpolation problem will be. For h small enough, linear interpolation should correspond to a close-to-perfect reconstruction.

B. Interpolation in Case of Oversampling

Assuming $s(t)$ is bandlimited, h has to be chosen small enough to satisfy the Nyquist rate, but it can indeed be chosen much smaller. There are an infinite number of ways to reconstruct an oversampled signal. Indeed, consider, for instance, an oversampling ratio of two. We can then reconstruct the function at any point using just the odd samples, or the even samples, or any convex combination of these two ways. Here, since we estimate the interpolating function, the precise function that will come out of the estimation procedure, in case of oversampling, entirely depends on the criterion we use to solve $Y = \mathbf{S}\mathbf{w}$. Our objective is thus to define a criterion that

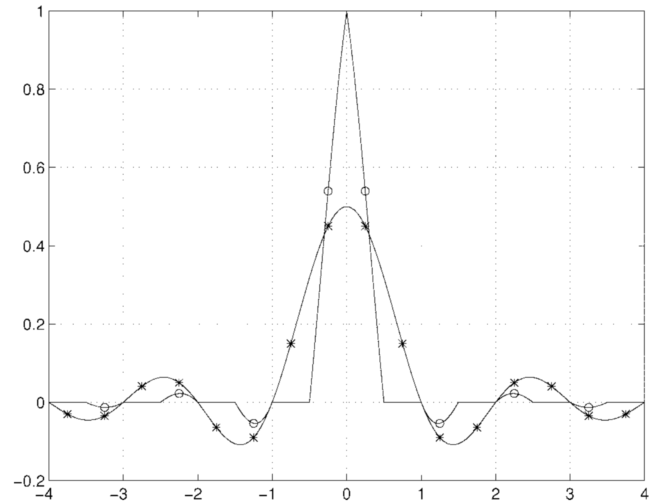

Fig. 1. Central parts of two interpolating functions in case of an oversampling ratio of two. The sampling period is equal to one half. The minimal $l2$ norm interpolating function is a sinus cardinal function, decreases slowly toward zero, and uses all the samples. The minimal $l1$ norm interpolating function decreases rapidly to zero and uses one sample out of two.

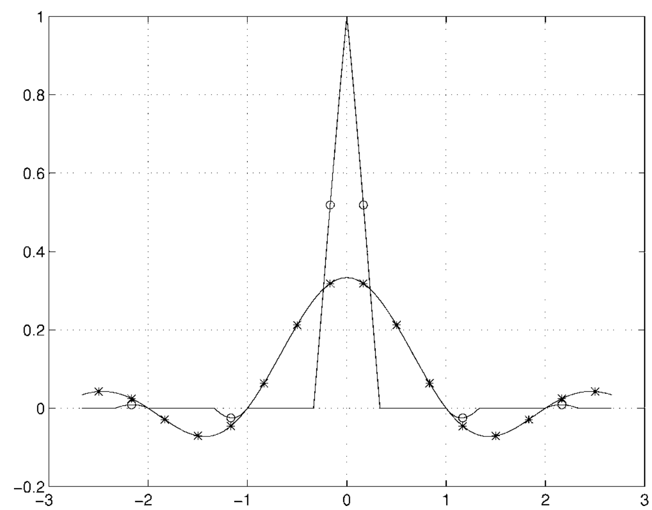

Fig. 2. Same as in Fig. 1 for an oversampling ratio of three. The resulting sampling period is equal to one third. The minimal $l2$ norm interpolating function is the same as in Fig. 1, decreases slowly toward zero, and uses all the samples. The minimal $l1$ norm interpolating function decreases rapidly to zero and uses one sample out of three.

will single out an interpolating function that takes advantage of the oversampling and comes as close as possible to an impulse. For such a function, it is then easy to deduce the number of replicas and the delays from the weights.

1) Minimum ℓ_2 Norm: When a problem has an infinite number of solutions, a common way to select a specific solution is to choose a minimum norm solution. The usual choice is to consider the ℓ_2 norm because it is easy to compute. Unfortunately, we can prove that in our case, this leads to weights that are samples from the standard sinus cardinal function scaled by the Nyquist rate, regardless of the true oversampling rate. Diminishing h though allowing intuitively for a more localized reconstruction will simply lead to further oversampling the *same* sinus cardinal function. In our context, this is, thus, probably the worst way to select a solution.

The proof is not presented. An illustration is given in Figs. 1 and 2.

2) Minimum ℓ_1 Norm: Another norm for which the solution is not too difficult to compute is the ℓ_1 norm. It happens that minimizing the ℓ_1 norm of the interpolating function (weights) is one criterion that leads to fairly localized reconstruction formulas, i.e., to an interpolating function that *narrows* and tends toward an *impulse* as h decreases. *The details of this result and its proof are omitted.* More information can be found in [6], where the functional optimization problem is further detailed. The interpolating functions $\psi(t)$ are shown in Figs. 1 and 2, together with those of minimal ℓ_2 norm for oversampling factors l equal to 2 and 3, respectively.

In Fig. 1, $l = 2$, and the resulting sampling period is, thus, one half. 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 eight neighboring points on both sides are presented, however. Of course, an infinite number of samples and weights are needed to achieve a perfect reconstruction. The other curve is the interpolating function $\Psi(t)$ with minimal ℓ_1 norm. The o's on this curve are the weights to be assigned to the samples when the midpoint between two samples has to be reconstructed. The function vanishes for $|t| \in [k-1/2, k]$ with $k > 0$, and we can notice that except for the two neighboring points with abscissae ± 0.25, only one in every second sample point is used in the reconstruction so that there are only four o's on both sides. In Fig. 2, the same curves are drawn for an oversampling ratio $l = 3$. The resulting sampling period is now one third. As explained above, miminizing the ℓ_2 simply leads to different samples of the same function in the continuous time scale, whereas minimizing the ℓ_1 norm further improves the localization of the interpolating function. We can see on Fig. 2 that except for the two neighboring samples at $\pm 1/6$, the weights attributed to the other samples are already quite small. If one further increases the oversampling ratio, minimizing the ℓ_1 norm of the weights almost leads to linear interpolation.

C. Comments

The ℓ_1 norm appears to achieve what we want pretty well. It chooses among all the possible solutions to $Y = \mathbf{S}\mathbf{w}$ (a linear variety) an extremely parsimonious one. It is known [8] that minimizing the ℓ_1 norm leads to sparse solutions. Here, in an interpolation context, we have shown that it further leads to solutions that are localized. In the presence of several replicas, we will, of course, estimate a linear combination of such sampled reconstruction functions, and the *localization* property is thus important in order to evaluate the number of replicas and the associated delays.

IV. Development

From the results described above, we conclude that minimizing the ℓ_1 norm of the weights should lead to quite an efficient algorithm. The first idea is thus to solve the optimization problem

$$\begin{aligned} &\min_{\mathbf{w}} \ \|\mathbf{w}\|_{\mathbf{1}} \\ &\text{s.t.} \ \ Y - \mathbf{S}\mathbf{w} = 0 \end{aligned} \tag{LP1}$$

where $\|\mathbf{w}\|_{\mathbf{1}}$ stands for the ℓ_1 norm of $\mathbf{w}$. This problem can be converted into a linear program [7]. It has a unique mininum that is easily and quickly obtained, even for large L and M using standard programs such as the Simplex algorithm, which is available in any scientific program library. This is, however, somehow too simple an approach since it does not take into account the presence of the additive noise N in (4). It is thus unjustified to ask for a perfect match between Y and $\mathbf{S}\mathbf{w}$.

Remember that this noise is either white noise or filtered white noise (3) and may have quite a large variance. This latter case happens when the white noise corrupting the data $\{z_t\}$ in (1) passes through the matched filter to become $\{n_k\}$ in (3). We should indeed take into account this information to *whiten* the observation vector Y. Let $\sigma_e^2\Sigma$ denote the covariance matrix of the noise N in (4). Σ is a known symmetric Toeplitz matrix built from the signal $\{h_t\}$ and models the effect of the matched filter. We propose to take the symmetric square root of this order-L matrix and to premultiply both Y and the (L, M) matrix $\mathbf{S}$ by $\Sigma^{-1/2}$ to obtain $\tilde{Y}$ and the matrix $\tilde{\mathbf{S}}$. From a statistical point of view, it is then natural to replace the constraint $Y - \mathbf{S}\mathbf{w} = 0$ with a constraint on the sum of the squared residues $\|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2 \le B$. A similar constraint is actually proposed in [2]. The ℓ_2 norm, especially for Gaussian random variables, has an appealing interpretation since for parametric models, it often corresponds to a maximum likelihood type approach.

The new optimization problem then reads

$$\begin{aligned} &\min_{\mathbf{w}} \ \|\mathbf{w}\|_{\mathbf{1}} \\ &\text{s.t.} \ \|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2 \le B \end{aligned} \tag{LS1}$$

where B stands for a bound that has yet to be fixed. This optimization problem is again convex and straightforward to solve. It is equivalent to the deconvolution criterion

$$\min_{\mathbf{w}} \|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1. \tag{D}$$

The equivalence between (D) and (LS1) can be established in the following way. Let λ be fixed, and denote $\mathbf{w}^*$ as the optimum of (D). If we now take B in (LS1) equal to $\|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}^*\|_2^2$, then (LS1) has this same optimum $\mathbf{w}^*$. Inversely, the λ in (D) that corresponds to a given B in (LS1) is nothing but the inverse of the Lagrange multiplier of (LS1) [7] at the optimum. Fixing B in (LS1) or λ in (D) is thus equivalent, although, of course, the relation existing between both variables is implicit.

The criterion (D) is the one we retain in the sequel. It is a deconvolution criterion with a penalization or regularization term using the ℓ_1 norm.

V. Deconvolution with an ℓ_1 Regularization Term

Let us adapt the model in order to include the whitening step of the observations. We can now consider that we observe a L-dimensional vector $\tilde{Y}$ that admits a model [compare with (4)]

$$\tilde{Y} = \sum_{p=1}^{P} a_p \tilde{S}_{\tau_p} + E \tag{8}$$

where the expression of $\tilde{S}_\tau$ as a function of τ is known, and E is a zero mean random vector with covariance matrix $\sigma_e^2 I$. The problem is to identify both P and the $\{a_p, \tau_p\}$. To do so, we propose to solve

$$\min_{\mathbf{w}} \|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1 \tag{D}$$

where the positive scalar parameter λ has yet to be fixed. If we introduce new variables $w_i^+ = \max(w_i, 0)$, $w_i^- = \max(-w_i, 0)$ and replace w_i by $w_i^+ - w_i^-$ and $|w_i|$ by $w_i^+ + w_i^-$, this unconstrained nonsmooth optimization problem is converted into a quadratic program where these new variables w_i^+ and w_i^- are constrained to be greater or equal to zero [7]. Its unique solution is easily and quickly obtained, even for large number of unknowns, using standard programs available in any scientific program library.

A. Tuning the Parameter λ

For λ too small or too large, the optimal solution $\mathbf{w}^*$ of (D) is useless. Indeed, if λ is taken equal to zero, we are left with $\min_{\mathbf{w}} \|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2$, and since there are, generally, more unknowns than equations $(M > L)$, the minimum is zero and is attained for all points in a linear variety with, generically, between L and M nonzero components. No useful conclusion can be drawn from them. If λ is taken too large (larger than $2\|\tilde{\mathbf{S}}^{\mathbf{T}}\tilde{Y}\|_\infty$), all the components of the solution $\mathbf{w}^*$ will be zero. Somewhere in between these two values, there may be a set of values of λ for which the solution to (D) easily leads to the *exact* values of $P, \{a_p, \tau_p\}$. Such a solution $\mathbf{w}^*$ has its nonzero weights concentrated in P clusters around the true delays $\{\tau_p\}$. A cluster will consist typically of two nonzero weights from which the delay is deduced by linear interpolation. This ideal is, however, difficult to expect for difficult scenarios so that we rather look for a λ that slightly overestimates the number of replicas that yields one or two false replicas. This strategy does not allow us to miss a true replica. The false replicas are easily eliminated by simple thresholding or a better statistical test since their amplitudes are generally quite small.

The simplest idea is to solve (D) for several values of λ and to try to detect the optimal value of λ and true solution $\mathbf{w}^*$ by other means such as statistical tests. Except for the computational cost, this is an extremely efficient way to solve the estimation and identification problem whose performance constitutes a upper limit to those we can expect. A possible way to achieve the selection is described below in Section V-B.

Let us present an approach that allows us to tune λ. Using the same reasoning as in Section IV, we can verify that λ in (D) is the Lagrange multiplier associated with the unique constraint in the *equivalent* optimization problem

$$\begin{aligned} &\min_{\mathbf{w}} \|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}\|_2^2 \\ &\text{s.t. } \|\mathbf{w}\|_1 \le B' \end{aligned}. \tag{LS2}$$

It is with this interpretation of λ that we use to tune it. For fixed (B'), (LS2) has a unique optimum attained at, say, $\mathbf{w}^*(B')$. The value of the associated Lagrange multiplier $\lambda^*(B')$ gives the sensitivity of the optimal value of the criterion $\|\tilde{Y} - \tilde{\mathbf{S}}\mathbf{w}^*(B')\|_2^2$ to a variation of B' [7]. The maximum likelihood (ML) approach applied to our model (8), for a fixed number Q of replicas, amounts to a solution of

$$\min_{\{a_q, \tau_q\}} \left\| \tilde{Y} - \sum_{q=1}^{Q} a_q \tilde{S}_{\tau_q} \right\|_2^2.$$

If this problem is solved for $Q > P$, false replicas modeling the additive noise E will be present in the solution. Such a false replica that models the noise solves $\min_{\{a,\tau\}} \|E - a\tilde{S}_\tau\|_2^2$. The optimum attained for a delay, which is denoted τ^*, has amplitude $E^T\tilde{S}_{\tau^*}/\tilde{S}_{\tau^*}^T\tilde{S}_{\tau^*}$ and the decrease it induces in the criterion is $(E^T\tilde{S}_{\tau^*})^2/\tilde{S}_{\tau^*}^T\tilde{S}_{\tau^*}$. In order to allow for a few false replicas and using the sensitivity interpretation of the Lagrange multilpler, we thus propose to take λ equal to the order of magnitude of $E^T\tilde{S}_{\tau^*}$, which is the inverse of the ratio of these two values:

$$\lambda = O(\sigma_e \|\tilde{S}_\tau\|_2). \tag{9}$$

B. Summary of the Procedure

For data $\{z_n\}$ following a model like (1), we propose to first process them using the matched filter if this remark applies. This lets us to improve the SNR and generally leads to a smaller set of observations $\{y_k\}$ that verify a similar model (3). In matrix form, we now have (4), where Y, the observation vector, is of dimension L. The additive noise N is then no longer white, and we propose to *rewhiten* it to obtain the model in (8), where the noise E is white again. The model in (8) indeed has the same form as the model we started with (1). We also build a (L, M)-matrix $\tilde{\mathbf{S}}$ with accordingly filtered column-vectors $\tilde{S}_m$. To each column, with index m, is associated a delay so that M different preassigned delays are proposed to allow for the reconstruction of the observations $\tilde{Y}$.

We then solve (D) for one (or several) value of λ of the order given in (9). The optimum $\mathbf{w}^*$ has, in general, a small number of nonzero weights (typically between $2P+4$ and $2P+10$). The nonzero weights are either isolated or appear in pairs, and to each "cluster," we associate an unique replica with an amplitude $\hat{a}_p$ equal to the sum of the weights and a delay $\hat{\tau}_p$ obtained by linear interpolation. We order these replicas by decreasing amplitudes. Due to the additive regularization term in (D), the amplitudes are biased, and we re-evaluate them from the data for an increasing number Q of potential replicas by solving the linear least squares fit

$$C(Q) = \min_{\{a_q\}} \left\| \tilde{Y} - \sum_{1}^{Q} a_q \tilde{S}(\hat{\tau}_q) \right\|_2^2. \tag{10}$$

To decide on the number of replicas needed to explain the observations, we use a minimum description length type test (MDL) [9]. For an increasing number Q of replicas, let $C^*(Q)$ denote the value of the minimum of (10). MDL applied to this situation amounts to taking $\hat{P}$, which is the estimate of the P number of replicas in the observations as

$$\hat{P} = \arg\min_{Q} C^*(Q) + 2Q \operatorname{Log} L\hat{\sigma}_e^2. \tag{11}$$

If (D) is solved for several values of λ, the same scheme is applied for each λ, and we have to select the best representation among these competing ones. The corresponding values $C^*(Q)$ are again used to achieve this selection. In this last case, we can say that our approach yields a small number of potential representations/solutions and that the maximum likelihood criterion, together with an MDL like detection test, is used to choose the one to be retained.

VI. Simulation Results

We now present some simulation results to allow for comparisons with two other approaches [1], [2]. Our method may be computationally more demanding than the others but appears to work at much lower SNR's. For the scenario in [1], where the threshold SNR is clearly indicated, we gain about 20 dB, i.e., the SNR below which outliers start to appear is lowered by about 20 dB. The first approach is known as the method of complex to real least squares time delay estimation (CRALS). The second approach is based on the method of projection onto convex sets (POCS) [4] that is applied to the present context in [2]. In both cases, the transmitted signal $h(t)$ is a linear FM signal, and the received signal $z(t)$ is generated with three replicas having integer delays and real amplitudes (attenuations). $z(t) = \sum_1^3 a_p h(t - \tau_p) + e(t)$.

We define the SNR for the pth path to be

$$\text{SNR}_p = 10 \log \frac{a_p^2 \sum h_t^2}{\sigma_e^2} \tag{12}$$

This definition (10 log [energy of the signal/variance of the noise]) is also the one used in [1]. It is generally considered when it comes to detection of a known deterministic signal in white noise and is actually the SNR at the output of the matched filter. In [2], another definition is considered: $10\log$ (variance of the signal/variance of the noise) with variance of the signal defined as its energy divided by the number of samples. The difference with (12) is quite important and is a function of the number of points in the signal. For a signal length of 450 samples as in [2], the second definition yields a value that is $10\log 450 \simeq 26$ dB below the one given by (12).

A. Example 1

This example is taken from [1]

$$h_t = u_t \cdot \sin(2\pi(\alpha t^2 + \beta t)) \quad t = 0, 1, \dots N-1 \tag{13}$$

where $N = 750$, $\alpha = (f_2 - f_1)/2N$, $\beta = f_1 = 0.1$, $f_2 = .15$, and w_t is a window function equal to

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

A three path received signal is generated as

$$z_t = h_{t-200} - .8h_{t-204} + .4h_{t-220} + e_t, \quad t = 0, \dots, 999 \tag{14}$$

and the Gaussian white noise variance σ_e^2 is tuned to yield the desired SNR's according to (12).

Some simulation results obtained using the above described algorithm are presented below. The data $\{z_t\}$ are processed by the matched filter. The number of samples that are used to build the Y vector is $L = 250$ samples. They are taken symmetrically around the global maximum of the output of the matched filter. This vector is whitened using the inverse of the symmetric square root of the covariance matrix Σ of Y computed once and forever using $\{h_k\}$. 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, we have $M = 301$ potential values of the delays. Three different values of λ are considered for each realization. They correspond to 80, 100, and 120% of $(\hat{\sigma}_e \|\tilde{S}_\tau\|_2)/10$ [see (9)], where $\hat{\sigma}_e$ is an estimate evaluated on the data. For each of them, the ML criterion (10) is used to re-estimate the amplitudes and to detect the number of replicas (11). The best ML solution is retained among these candidates for each realization.

We performed 500 independent trials of the scenario (14) with a noise variance $\sigma_e^2 = 64.10^{-4}$. The SNR's (12) are then 47, 45, and 39 dB. There are typically about ten nonzero weights among the $2M$ components of $\mathbf{w}$. This is a difficult configuration since it is 20 dB below the threshold SNR observed in [1]. The results are presented in Table I for $M = 301$ corresponding to $h = 0.2$. The test (11) correctly decided that the number of replicas was three for all the 500 realizations.

Results closer to the Cramér–Rao (CR) bound can easily be obtained, if desired, by performing a local search using our results as initial estimates since these are always in the domain of attraction of the true optimum. If we further decrease the SNR's some outliers appear: Either the detection test decides wrongly, or a false path is retained. For some realizations, due to the highly oscillatory nature of $s(t)$, the global optimum is then no longer around the *true* value. The CR bounds, which consider only the curvature of the highest peak, are then no longer relevant, and more elaborate bounds should be considered.

B. Example 2

This example is taken from [2] and has also been considered in [3]. The linear frequency-modulated signal has time radial-bandwidth product equal to 450 and a frequency bandwith equal to $B_f = 1/(2\pi)$. The carrier frequency is taken equal to $1/\pi$. A three-path channel with delays $\{10, 12, 50\}$ (in samples) and amplitudes $\{1, 0.9, 0.5\}$ in white Gaussian noise is simulated. Refer to [2] and [3] for details on the scenario. The replicas with delay 10 and 12 are too close to be resolved in the baseband matched filter output since the standard resolution is $1/B_f \simeq 6$ samples. Our approach uses, as the Y vector, the real and imaginary part of the matched filter applied to the signal around the carrier frequency. For comparison purposes, as in [2] and [3], we take $L = 60$ and $M = 60$, i.e., only integer delays are considered. As indicated in [3], noninteger delays are an important issue, and proposing only integer delays in the estimation part may introduce some prior information when the true delays are integers.

TABLE I
RESULTS OVER 500 TRIALS FOR THE THREE REPLICAS PROCESS OF EXAMPLE 1

replica number	1	2	3
replica SNR in dB	47	45	39
true delay	200	204	220
estimated delay	200.009	203.987	219.996
est. stdt dev. delay	.0449	.0653	.0569
CR stdt dev. delay	.0251	.0334	.0212
true amplitude	1	- 0.8	0.4
estimated amplitude	1.0057	-.7954	.3985
est. stdt dev. amplit.	.0234	.0252	.0077
CR stdt dev. amplit.	.0200	.0213	.0067

In Fig. 3, we present the (modulus of the) output of the matched filter at SNR $= 21$ dB (-5 dB for the SNR definition used in [2]), and the corresponding M-dimensional weight vector $\mathbf{w}$ that we obtain is presented in Fig. 4. In this case, 17 out of the 60 weights are nonzero. This quite high number of nonzero weights is due to the fact that no oversampling ($l = 1$) is performed. For this SNR, that is, 20 dB below the one considered in [2], our approach solves the three paths, except for about one realization in 500 for which the detection test makes a wrong decision.

C. Robustness Issues

The choice of the location and the size of the observations that are used to build Y is one of these issues. Our approach is quite robust in this respect. The results presented for Example 1 are obtained by taking a quite large set placed symmetrically around the global maximum of the output of the matched filter. This choice has been made to highlight the robustness of our approach and to be sure to introduce no prejudice. Since the true delays all lie on one side of this maximum, better performances could be obtained by focusing the set more closely around the true locations. A second issue is the selection of the size and the location of the potential delays that have to be proposed. This is the main issue in standard deconvolution approaches and is known as the indicator set selection problem. It is the subject of much effort. It directly influences the conditioning of the $\mathbf{S}$ matrix and, thus, the results of most approaches. Due to the presence of the additive ℓ_1-norm regularization term in (D), our method is again quite unsensitive in this respect. This selection is somehow automatically done by the ℓ_1-norm that induces parsimony, as shown in Section

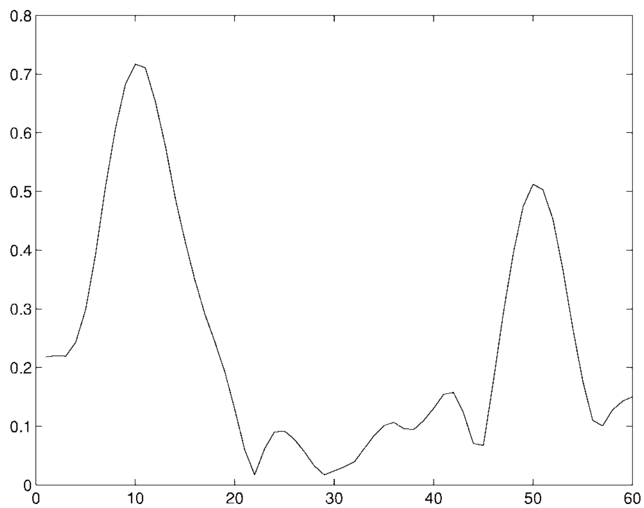

Fig. 3. Modulus of the output of the matched filter for Example 2 as a function of the delay in samples.

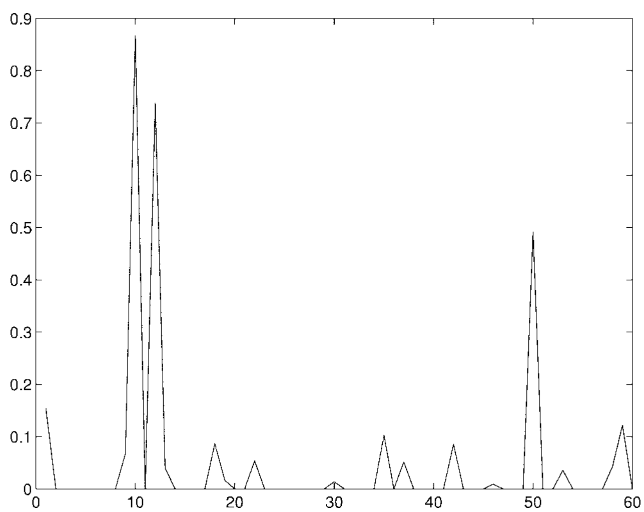

Fig. 4. Typical weight-vector **w** by the proposed method for Example 2.

III-B2. Again, in the simulations of Example 1, the potential delays simply cover a domain placed symmetrically around the maximum of the output of the matched filter. Choosing a better placed, smaller set would certainly help.

Two further parameters that are specific to our approach and need to be tuned are the oversampling rate l and the hyperparameter λ. The choice of l is not crucial; it has to be taken to allow for performances comparable to the CR bounds and is thus easily fixed. The choice of the parameter λ is more delicate, and we systematically solve the problem for several values of λ of the order given in (9) and then select the solution using the ML criterion and the MDL test. The order of magnitude given in (9) is obtained using many angles of attack, but a more precise result would be helpful. Further investigations are under progress. This is indeed more a computational load issue than a robustness issue.

D. Computational Complexity

Although the conditioning of the matrix **S** has little influence on the result of the algorithm, its influence on the computational complexity might be important, and the number of iterations needed to converge to the unique optimum depends on it. Little is said in the litterature about the computational cost of a quadratic programming algorithm (we use the NAG E04NCF program). It seems that the number of operations required to perform one iteration of the quadratic program is, in our case, proportional to the square of $\min(L, M)$ and that the number of iterations is, in general, considered to be linear in this same number. The computational load would thus mainly be cubic in $\min(L, M)$. It is difficult to compare these figures with those of the other techniques. Let us note that according to [10], to solve a standard linear least squares problem with n unknowns and m equations $(m > n)$, the number of flops is equal to $(n^2/2)(m + n/3)$, where cubic terms are also present. The difference in computational load between the proposed method and others may thus be quite small.

VII. Conclusion

We have considered signals that can be represented as the sum of an unknown number of amplitude-scaled and time-shifted replicas of a known pulse shape. For such situations, we present an algorithm that determines both the number of such replicas and the amplitude and arrival time of the individual paths.

Our method has several advantages over most existing ones. It does not rely on an initialization procedure and does not require an initial point. The indicator set selection problem that determines the performance of most deconvolution methods has a minor influence in our case. Diminishing its size in order to improve the conditioning of the **S** matrix and the variance of the estimates is not an issue in our approach. The proper selection is somehow done by the additive ℓ_1-norm regularization term, which ensures that only a small number of the available weights will be nonzero at the optimum. Moreover, since the proposed method includes a detection scheme that discards weak spurious paths, it requires no prior knowledge about the number of paths.

The computational complexity is reasonable. If only degraded performance is needed, we can diminish the resolution of the method, and this will decrease the computational load. The method is, in fact, extremely versatile in this respect.

Simulation results on classical examples taken from the litterature indicate that the performance of the method are excellent. As for most other similar methods, no theoretical analysis of the statistical performance is available at the moment. Investigations are underway.

References

[1] T. G. Manickam, R. J. Vaccaro, and D. W. Tufts, "A least squares algorithm for multipath time-delay estimation," *IEEE Trans. Signal Processing*, vol. 42, pp. 3229–3233, Nov. 1994.

[2] Z. Kostic, M. I. Sezan, and E. L. Titlebaum, "Estimation of the parameters of a multipath channel using set theoretic deconvolution," *IEEE Trans. Commun.*, vol. 40, pp. 1006–1011, June 1992.

[3] T. G. Manickam and R. J. Vaccaro, "A noniterative deconvolution method for estimating multipath channel responses," in *Proc. ICASSP*, May 1993, pp. I-333–I-336.

[4] D. Youla and H. Webb, "Image restauration by the method of convex projections part 1—Theory," *IEEE Trans. Med. Imag.*, vol. 1, pp. 81–94, Oct. 1992.

[5] Y. T. Chan, J. M. Riley, and J. B. Plant, "A parameter estimation approach to time-delay estimation and signal detection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 8–16, Feb. 1980.

[6] J. J. Fuchs and B. Delyon, "The sampling theorem and time delay estimation," in *Proc. 11th IFAC Symp. Syst. Ident. SYSID*, Fukuoka, Japan, July 1997, vol. 2, pp. 505–510.

[7] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison Wesley, 1973.

[8] R. M. Leahy and B. D. Jeffs, "On the design of maximally sparse beamforming arrays," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1178–1187, Aug. 1991.

[9] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[10] G. G. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1983.