A New Approach to System Identification Based on the Generalised Structural Subband Decomposition

Eftychios V. Papoulis and Tania Stathaki
Communications & Signal Processing Group
Imperial College, London

Abstract: The System Identification (SI) problem is addressed from the viewpoint of the Generalised Structural Subband Decomposition (GSSD). We present a System Identification Structure (SIS) for the identification of the Generalised Polyphase Components (GPC's) of the unknown system. Sparsity constraints are imposed on the input for the identification of PC's to be feasible. Significant computational savings, improved tracking capability and a substantial increase in the convergence rate (CR) in coloured input environments, are the main characteristics/features of the proposed approach.

1. Introduction

Adaptive identification of linear time invariant (LTI) systems has been studied extensively and a number of efficient structures and algorithms have been developed for this purpose. Among the various approaches is that of the application of GSSD [3]. This decomposition, although a special case of the generalised sampling theorem and the related Perfect Reconstruction (PR) Filter Banks (FB's), was developed within its own independent context [1]. In [3] it is applied to the input of the unknown system, with the interpolators been chosen a-priori and the sparse subfilters been adapted for the purpose of identification. SSD is a general representation and so the method [3] can be used to identify an arbitrary system S. In that, the adaptation of the coefficients was performed at the initial data rate and the adaptive algorithm was controlled by the composite estimation error; in effect the GPC's of S were identified jointly. The algorithm provided an increased CR for coloured input processes –compared with the Full-band LMS– at the expense of an increase in the computational complexity. Quite recently [4], the Polyphase Decomposition (PD) was incorporated in the classical Subband Adaptive Filtering (SAF) scheme [2], providing an elegant solution that eliminated the need of adaptive cross-terms and the problems associated with them [2]. In this later approach [4], the adaptation was performed at $1/M$ the initial data rate, M being the number of filters used to subband split the input of S and the desired signal. The computational complexity resulted is slightly higher than that of the Full-band LMS. The polyphase components of S were identified jointly.

The present work comes to contribute to a new way that GSSD can be used for the purpose of identification. For S, a particular type of SSD is considered associated with a $\mathbf{T}$ matrix of interpolators. The DCT matrix can be a choice, the Identity would yield to the trivial solution of the PC's and the Hadamard would provide a computationally efficient implementation. The output of S is then analysed using an orthogonal FB corresponding to the particular $\mathbf{T}$ choice. Sparsity constraints are imposed to the input so as the i^{th} FB path to behave like the i^{th} GPC of S. For the identification of each polyphase component an adaptive filter of length N/M is used and is driven by the decimated version of the input.

2. Development of the new System Identification Structure (SIS)

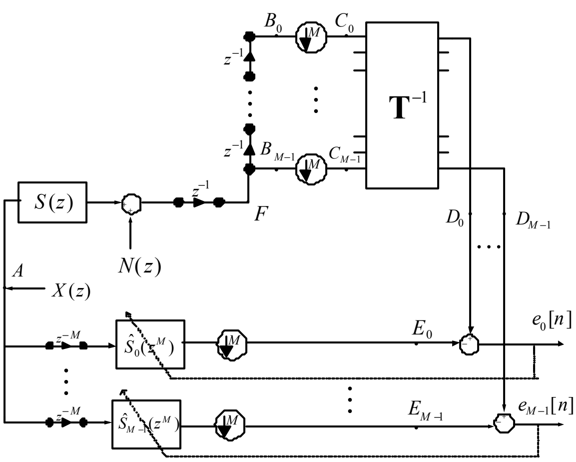

Fig.1. The proposed System Identification Structure

For the structure of Fig.1 let $X(z)$ be the input to the unknown system $S(z)$. Consider the M flow-gram paths, starting at point A and terminating at their corresponding points C_i with $0 \le i < M$. These will be referred as Sub-System Paths (SSP's). The term Identification Paths (IP's) is used to denote the M flow-gram paths between points A and E_i. The significance of this terminology will become apparent shortly. For $d[n]$ as the input to the system, the response of the i^{th} SSP at point B_i is the sequence

$$\left\{ \mathbf{0}_i^T \,\middle|\, s[0], s[1], \ldots, s[M-1], s[M], s[M+1], \ldots \right\} \qquad (1)$$

where $\mathbf{0}_i^T$ denotes the row vector comprised of $M-i$ zeros. The decimation of (1) by M results in $\{0, r_i[n]\}$ which is obtained at C_i; $r_i[n]$ is

$$r_i[n] = \{s[i], s[M+i], s[2M+i], \ldots\} \qquad (2)$$

namely the sequence of the polyphase coefficients that correspond to the i^{th} polyphase component of S. In

Section 3, SIS will be re-established within the context of GSSD; until then the term polyphase when used, will simply denote the sequences r_i of (2) or their Z transforms.

Returning now to the SSP's, we should stress that due to the presence of the decimators, even though linear, they no longer comprise time invariant (TI) networks. By restricting however the input to take non-zero values only at time instants which are integer multiples of M, the TI property is maintained in the sense that a time translation of the input by kM, $k \in Z$ samples, results a corresponding translation of the outputs by k. This suggests that as long as the input $x[n]$ of S is sparse in the sense $\{x[n] = 0 : n \notin kM, k \in Z\}$, the identification of S can be accomplished by the simpler task of identifying independently each one of the M $r_i[n]$ sequences of (2). This is because each one of the SSP's behaves as an LTI network $r_i[n]$ to the class of inputs $X(z^M)$. Since it is so, the effect of the i^{th} SSP can be compensated be the i^{th} IP; in that, the term z^{-M} accounts for the presence of the leading zero vector in (1). Setting all $\hat{s}_i[n]$ to $r_i[n]$ would result in the error vector $\mathbf{e}[n] = (e_1[n], e_1[n], ..., e_M[n])^T$ to be equal to zero for all n. Using the well-known Noble identities [5], we can relocate the decimators in front of the delay elements of the IP's (Fig. 2). In Fig.1, a $M \times M$ transformation matrix $\mathbf{T}^{-1}$ has been included the significance of which will be explained later. For the present we assume that $\mathbf{T}$ is the identity. The part between the points F and D_i of Fig.1 appears in Fig.2 as a M branch filter bank. This equivalent representation is obtained from Fig.1 by moving the decimators at the right of $\mathbf{T}^{-1}$ and replacing what is between the points F and D_i by the orthogonal FB of Fig.2. In that the i^{th} impulse response is the i^{th} row of $\mathbf{T}^{-1}$. The significance of the notation of Fig.2 for the FB filters will become apparent in the next section. The diagram of Fig.2 and the analysis presented so far suggest now the following way to perform the system identification task. This is to use an adaptive algorithm to adjust the weights of $\hat{\mathbf{s}}_i$ based on the corresponding errors e_i. We suggest the use of M FIR adaptive filters, of length L_i each equal to the length of $\mathbf{r}_i$. Each one of them is adapted based on the error e_i and its role is to compensate for the effects of it's corresponding $\mathbf{r}_i$. Due to the architecture of the proposed SIS, the effects of any $\mathbf{r}_k$ for $k \neq i$, are masked at point D_i and so e_i depends only on the coefficient error $\mathbf{e}_i = \mathbf{r}_i - \hat{\mathbf{s}}_i$. Thus a non-zero e_i implies that $\hat{\mathbf{s}}_i$ is not the proper estimate, which will be consequently updated at the next iteration. For our simulations we have chosen the LMS algorithm. The following characteristics make the proposed structure particularly attractive for the purpose of identification:

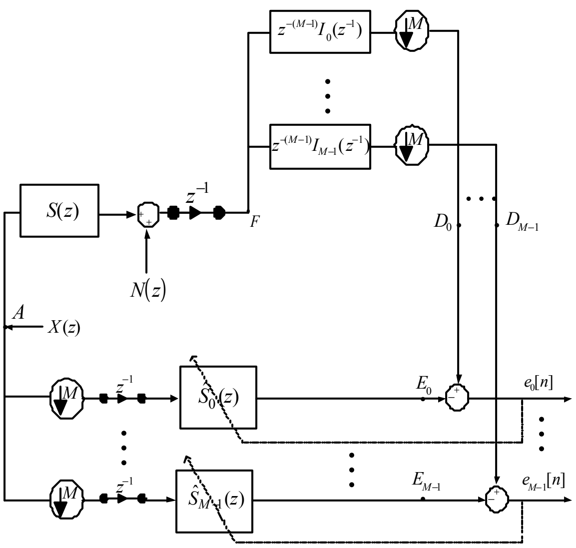

Fig.2. An equivalent representation of the SIS of Fig.1 with the decimators relocated and the matrix replaced by a FB.

C1. The problem of identifying the N coefficients of the system is recast into that of identifying its M polyphase components of length N/M each, on an independent basis. Although the total number of coefficients to estimate remains N, the adaptive filter's length is decreased by a factor of M. By this way the convergence rate (CR) expressed in terms of the number (#) of update cycles increases. The presence of the decimators at the front end of the IP's raises the question whether CR expressed in terms of the # of input samples would increase or decrease. Experimental results for the Full-band LMS, indicate that for large values of N ($N > 256$), CR is inversely proportional to N with $1/2$ being the proportionality factor. In other words, by halving N one would need half the # of update cycles to obtain the same error performance. This suggests that the proposed SIS for large values of N can attain the same error performance this being expressed in terms of the # of input samples.

C2. The adaptation of the coefficients is performed at $1/M$ the input data rate. Since the # of adaptive coefficients remains N, the # of operations per unit time or per input sample is reduced by a factor of M. In the experimental results section we will further discuss the computational complexity of the proposed structure.

C3. The decoupling of the identification of the polyphase components of S, implies that any error in the estimation or any change of one particular of them –due for example to a time variation of the system– will be masked and will not affect the other valid estimates. The proposed structure is thus expected to have improved tracking performance, which is particularly important in situations where S is time varying.

C4. The decimators of the IP's decorrelate the process that appears as input to the adaptive filters –when this is correlated. By this way the convergence rate of the adaptive algorithm is substantially increased.

We now come to the $\mathbf{T}$ matrix of Fig.1. This is an arbitrary non-singular $M \times M$ real coefficient transformation matrix. Its effect is to modify the response of the i^{th} SSP from $\{0, r_i[n]\}$ to $\mathbf{T}^{-1} \cdot \mathbf{R}$, $\mathbf{R}$ being the matrix whose $(i+1)^{th}$ row is the sequence $\{0, r_i[n]\}$. This means that at time $n \geq 0$, instead of obtaining the

$(n+1)^{th}$ column of $\mathbf{R}$ as the response at points D_i, the transformation of this column is obtained instead. This in turn suggests that in order for the zero error condition to hold, $\hat{\mathbf{s}}_i$ should be set not to the actual values of $\mathbf{r}_i$ but to their transformed values under $\mathbf{T}^{-1}$. Thus for the zero error condition we require

$$[\hat{\mathbf{s}}_0^T \mid \hat{\mathbf{s}}_1^T \mid \ldots \mid \hat{\mathbf{s}}_{M-1}^T] = \mathbf{T}^{-1} \cdot [\mathbf{r}_0^T \mid \mathbf{r}_1^T \mid \ldots \mid \mathbf{r}_{M-1}^T] \tag{3}$$

In the above | denotes change of row within the matrices. The elements in between the | signs are row vectors. It is clear that the introduction of $\mathbf{T}_M^{-1} \neq \mathbf{I}_M$ in the structure of Fig.1 still permits the application of the method that we described for the identification of S in exactly the same way. In that, the only effect of $\mathbf{T}^{-1}$ is to change the optimum weights (3) of the adaptive coefficients in the way just described.

3. From the GSSD to the proposed SIS

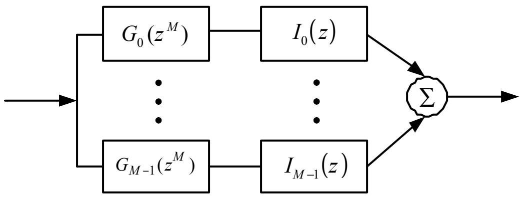

Fig.3. An arbitrary FIR system represented as the parallel interconnection of M systems each comprised by an interpolator and a sparse sub-filter.

We now come to review GSSD and re-derive the proposed SIS within its context. Consider an FIR filter of length N with impulse response $h[n]$ and system function $H(z) = \sum_n h[n] z^{-n}$. Decomposing $h[n]$ into the M sub-sequences $h_r[n] = h[Mn+r]$ with $0 \le r \prec M$, $n \in Z$ and $M \ge 2$, we obtain the following representation of H in algebraic and matrix form

$$H(z) = \sum_{r=0}^{M-1} z^{-r} H_r(z^M) = \begin{bmatrix} 1 & z^{-1} & \ldots & z^{-(M-1)} \end{bmatrix} \cdot \mathbf{I}_M \cdot \begin{bmatrix} H_0(z^M) \\ \vdots \\ H_{M-1}(z^M) \end{bmatrix}$$

$H_r(z)$ are the Z transforms corresponding to $h_r[n]$ and $\mathbf{I}_M$ is the $M \times M$ identity matrix. The vector-matrix representation allows for the following generalisation. Replacing $\mathbf{I}_M$ with an arbitrary non-singular matrix $\mathbf{T}$ and the vector of the Z transforms $\mathbf{H}(z)$ by some $\mathbf{G}(z)$, one can readily solve for the later to obtain $\mathbf{G}(z) = \mathbf{T}^{-1} \cdot \mathbf{H}(z)$, or more explicitly $[G_0(z), \ldots, G_{M-1}(z)]^T = \mathbf{T}^{-1} \cdot [H_0(z), \ldots, H_{M-1}(z)]^T$. This suggests a more general representation of $H(z)$ which now becomes

$$H(z) = \sum_{r=0}^{M-1} I_r(z) G_r(z^M) = \mathbf{Z}^T \cdot \mathbf{T} \cdot \mathbf{G}(z^M) \tag{4}$$

In (4) $I_r(z)$ is the product of $\mathbf{Z}^T$ with the $(r+1)^{th}$ column of $\mathbf{T}$, with $\mathbf{Z}^T$ being the delay row vector $[1, z^{-1}, \ldots, z^{-(M-1)}]$. $G_r(z)$ denotes the $(r+1)^{th}$ element of $\mathbf{G}(z)$ and is known as the r^{th} generalised polyphase component of $H(z)$. An interpretation of (4) as M filters $G_r(z^M) I_r(z)$ connected in parallel is given in the diagram of Fig.3. The r^{th} branch contributes within the frequency support domain of the $I_r(z)$ interpolator, which interpolates the sparse $G_r(z^M)$ polyphase sequence.

We now provide the link between the GSSD and the proposed SIS. Consider $\mathbf{T}$ being Hermitian ($\mathbf{T}^{-1} = \mathbf{T}^H$). For such a choice, the filter coefficients of the FB of Fig. 2 are the columns of $\mathbf{T}$ in reverse order and effectively the time reversed versions of the interpolators associated with any GSSD using $\mathbf{T}$ as its transformation matrix. The notation used in Fig. 2 is now fully justified. For the unknown system S, consider the GSSD under the $\mathbf{T}$ transformation representation (4). For the sparse $X(z^M)$ input, the response at one particular D_i point of Fig.2 is following the (4) representation

$$\left\{ \sum_{r=0}^{M-1} X(z^M) G_r(z^M) I_r(z) z^{-M} I_i(z^{-1}) \right\} \downarrow M \tag{5}$$

with $\downarrow M$ representing the decimation by M operation. Applying the relevant multirate identities [5], (5) equals

$$z^{-1} X(z) \sum_{r=0}^{M-1} G_r(z) \cdot \frac{1}{M} \sum_{k=0}^{M-1} I_r(z^{1/M} W_M^k) I_i(z^{-1/M} W_M^{-k}) \tag{6}$$

The term at the r.h.s. of the multiplicative sign of (6) is the $\downarrow M$ version of the deterministic cross /auto correlation of the interpolators' coefficients $I_{r,i}[n]$ which are simply the $(r+1)^{th}$ and $(i+1)^{th}$ columns of $\mathbf{T}$. This in turn, reduces to $d[r-i]$ due to the orthogonality of $\mathbf{T}$. Effectively, the response to $X(z^M)$ at point D_i becomes $z^{-1} X(z) G_i(z)$, which is exactly our initial argument since it suggests that the i^{th} SSP responses to the sparse $X(z^M)$ as an LTI network with response the i^{th} generalised polyphase component's coefficients $r_i[n]$. The link between the GSSD and the proposed SIS is now established.

4. Experimental results

For the experiments, the input x was obtained by some generating process x_g as $x[nM] = x_g[nM]$, $x[n; n \neq kM] = 0$. In the first set of experiments (Fig.5), x_g was white gaussian noise (WGN). In the second (Fig.4), the output of the AR system $z/(z-0.9)$ driven by WGN. In all cases $N(z)$ was AWGN uncorrelated with x_g. $\mathbf{T}$ was taken to be the Hadamard matrix. All plots were obtained using the optimum for any particular case value of the m LMS parameter; they represent the normalised coefficient error norm (NCEN) in db namely $20 \cdot \log_{10}(\|\hat{\mathbf{w}} - \mathbf{w}_{opt}\|_2 / \|\mathbf{w}_{opt}\|_2)$, versus the # of input samples. Each estimate was obtained by averaging over 50 Monte Carlo simulation results, in each one of which the coefficients of S were chosen at random. Fig.4 demonstrates the decorrelation characteristics of the proposed SIS described before (C4, section 2) and manifests its outstanding performance in coloured input cases. Fig. 5 supports the argument C1 of section 2. In the noisy case ($30 db\ SNR$) we noticed a small deterioration in the performance of SIS compared with the Full-band LMS. This is related to the LMS misadjustment and its dependence on m; increasing M we are allowed to increase m and this in turn increases the misadjustment that becomes more noticeable as SNR decreases.

We now discuss the computational complexity of the proposed algorithm. In each update cycle the # of multiplications needed (to calculate the outputs of the adaptive filters, to update their weights and obtain the FB output), is $M \cdot (N/M + N/M) + M \cdot M$, i.e. $2N + M^2$. This per input sample becomes $(2N + M^2)/M$, N being the length of the S, against $2N$ for the Full-band LMS case. Thus, for a given # of input samples we perform $1/M + M/(2N)$ [1] times the multiplications that would be needed. The # of additions is slightly increased. The proposed SIS is thus computationally very efficient compared with the Full-band LMS. Furthermore, it provides better performance (for non-white inputs) as the results indicated and is in effect an attractive solution to the SI problem.

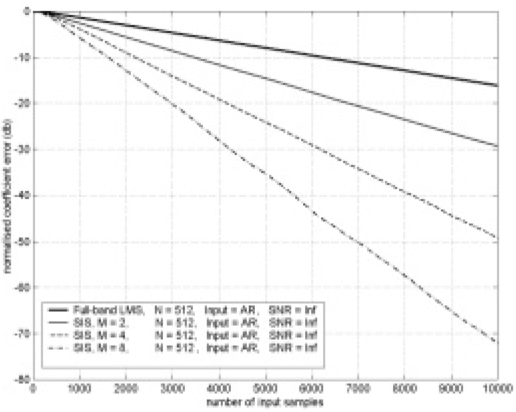

Fig.4 The NCEN for the proposed SIS against that of the Full-Band LMS for the AR input case. $N(z) = 0$

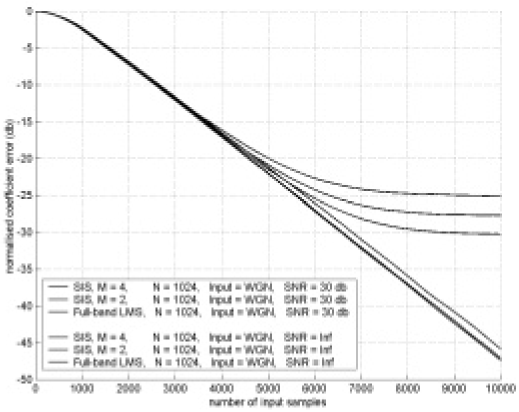

Fig.5 The NCEN for the proposed SIS against that of the Full-Band LMS when the input is a WGN process.

5. Conclusions

A novel and computationally highly efficient algorithm was presented for the purpose of system identification, which significantly outperforms the Full-band LMS when the input process is non-white. The improvement of its performance is traded against the limitation of the range of its applicability, as it requires the input to be a M^{th} band sequence. Still it consists an attractive solution when the input can be properly modified for the purpose of identification.

References

[1] S. Mitra, A. Mahalanobis, T. Saramaki. "A Generalised Structural Subband Decomposition of FIR filers and its Application in Efficient FIR Filter Design and Implementation". IEEE Trans. Circ.&Syst.-II,Vol.40, No.6, June 1993.

[2] A. Gilloire, M. Vetterli. "Adaptive Filtering in Subbands with Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancelation ". IEEE Trans. Sign. Proc. Vol.40, No.8, August 1992.

[3] M. R. Petraglia, S. K. Mitra. "Adaptive FIR Filter Structure Based on the Generalised Subband Decomposition of FIR Filters". IEEE Trans. Circ.&Syst.-II, Vol.40, No.6, June 1993.

[4] S. S. Pradhan, V. U. Reddy. "A New Approach to Subband Adaptive Filtering". IEEE Trans. Sign. Proc. Vol.47, No.3, March 1999.

[5] A. N. Akansu, R. A. Haddad. "Multiresolution Signal Decomposition". 2nd edition, 2001 Academic Press.

[1] This can further be reduced to $1/M$ if $\mathbf{T}$ is the Identity or the Hadamard matrix, as was the case for the experimental results given.