SPARSE VECTOR LINEAR PREDICTION WITH OPTIMAL STRUCTURES

Davorka Petrinović, Davor Petrinović
Department of Electronic Systems and Information Processing
Faculty of Electrical Engineering and Computing, Zagreb, Croatia
davorka.petrinovic@fer.hr , davor.petrinovic@fer.hr

Abstract: *A modification of a classical Vector Linear Prediction (VLP) technique is proposed in this paper, enabling significant reduction in complexity. The proposed sparse VLP technique (sVLP) is based on predictors with reduced number of nonzero elements. For a given input vector process, a design procedure for obtaining optimal sparse predictor structures and matrix elements is described. The effectiveness of the sVLP is evaluated on interframe predictive coding of Line Spectrum Frequencies (LSF) and compared to the classical VLP based on Switched-Adaptive Interframe Prediction scheme. The loss of the prediction gain due to sparse structures is calculated using various design parameters. Simulation results prove that a 6-fold reduction in complexity of prediction can be achieved causing only insignificant loss of the prediction gain and coder performance.*

Keywords: *Vector Linear Prediction, SIVP, Line Spectrum Frequencies, low complexity*

1 INTRODUCTION [#]

There are numerous examples of vector processes in geophysics, electrical engineering, biomedical engineering and many other areas. For a particular class of vector processes, each component of the current vector can be estimated based on linear combination of all components of certain number of preceding vectors. This technique, known as Vector Linear Prediction (VLP), is frequently used for signal coding by quantization of the prediction residual, i.e. the difference between the original and the predicted vector, instead of the original vector. The gain that can be achieved by this technique depends on the degree of correlation between consecutive vectors. The correlation is modeled by the predictor represented by one or more matrices, depending on the prediction order. The optimal predictors minimizing the energy of prediction residual within the analysis frame (block of input vectors) can be found by solving so called normal equations as in [1].

Although VLP can maximally exploit interframe as well as intraframe correlation by predicting each component of the current vector from all components of previous vectors, it is a reasonable assumption that for a certain class of processes not all components of preceding vectors contribute equally to the prediction. Elements of the optimal predictor matrices for such processes have different magnitudes. Some are greater and therefore contribute more to prediction, while the others are much smaller and less significant. If the smaller elements are made equal to zero, this should not have a great impact on the prediction gain, but can at the same time reduce the amount of computation.

VLP is commonly used for predictive vector quantization of LSF parameters of the speech signal. Vector predictors for this process exhibit the sparse property discussed above, as also mentioned in [2]. Some authors even employ diagonal (scalar) predictors [3] and [4], but none of these papers offer detailed discussion of the issues concerning sparse predictor structure or comparison to the full predictor case.

In our previous work [5], the use of sparse multidiagonal predictor matrices was proposed as a compromise between the diagonal and the full predictor case. The number of diagonals was chosen as a design parameter, determining the total number of nonzero elements (complexity). Comparative analysis was performed for all structures (from diagonal to the full case). The design procedure for obtaining matrix nonzero elements that are optimal for the given multidiagonal structure was also proposed.

This paper proposes more general sparse predictor structures that are not predefined as in multidiagonal case but rather depend on the actual correlation of the input vector process. In another words, sparse predictor matrix will have nonzero elements on locations that contribute mostly to the prediction gain, while all the others will be set to zero. The number of nonzero elements is chosen as a design parameter. Several

[#] This work was supported by the Ministry of Science and Technology of Croatia under project No. 036024.

techniques for obtaining optimal sparse predictor structures are investigated.

2 OPTIMAL SPARSE PREDICTORS

The procedure for sparse predictor matrix design will be illustrated on the example of the first-order VLP, having only one predictor matrix $\mathbf{A}_1=\{a_{i,j}\}$, but can be easily extended to any prediction order. The prediction error vector [#] $\mathbf{e}(n)$ is defined using the following expression:

$$\mathbf{e}(n) = \mathbf{x}(n) - \mathbf{A}_1\mathbf{x}(n-1) \ , \qquad (1)$$

where $\mathbf{x}(n\text{-}1)$ and $\mathbf{x}(n)$ are consecutive samples of the k-dimensional input vector process. The procedure starts with the conventional VLP predictor design [1] resulting with a full $k \times k$ predictor matrix having all nonzero elements. It is obtained by minimizing the energy E of the prediction residual $\mathbf{e}(n)$.

$$E = \sum_{n=1}^{P} \| \mathbf{e}(n) \|^2 = \sum_{i=1}^{k} E_i \quad , \quad E_i = \sum_{n=1}^{P} e_i^2(n) \qquad (2)$$

where $e_i(n)$ denotes the prediction error sequence of the i^{th} component. Since E can be expressed as a sum of the components' residual energies, E_i, it is minimal when all of the E_i are also minimal. This leads to an optimal predictor determined by covariance matrices $\mathrm{C}_{11}=\{c_{i,j}\}$ and $\mathrm{C}_{10}=\{b_{i,j}\}$ as follows:

$$\mathbf{C}_{11} \cdot \mathbf{A}_1^{\mathsf{T}} = \mathbf{C}_{10} \qquad (3)$$

$$\mathbf{C}_{1r} = \sum_{n=1}^{P} \mathbf{x}(n-1)\, \mathbf{x}^{\mathsf{T}}(n-r) \quad , \quad r = 0, 1. \qquad (4)$$

Covariance matrices are calculated from the input process comprised of $P+1$ vectors $\mathbf{x}(0)$ to $\mathbf{x}(P)$.

The basic idea presented in this paper is the reduction of the VLP complexity by setting most of the predictor matrix elements to zero. These elements should be chosen in a way that the increase of the energy of the prediction residual ΔE is minimal. It is easily shown that setting any $a_{i,j}$ to zero results with the increase of:

$$\Delta E = 2a_{i,j} \sum_{n=1}^{P} e_i(n) x_j(n-1) + a_{i,j}^2 \sum_{n=1}^{P} x_j^2(n-1) \qquad (5)$$

where $x_j(n\text{-}1)$ corresponds to the j^{th} component of $\mathbf{x}(n\text{-}1)$. The i^{th} row of the optimal predictor $\mathbf{A}_1$ was obtained provided that vectors $[e_i(1), e_i(2),.., e_i(P)]^{\mathsf{T}}$ and $[x_j(0), x_j(1),.., x_j(P\text{-}1)]^{\mathsf{T}}$ are orthogonal for all input observables, j=1,...,k. Therefore, the first term in (5) vanishes for all combinations of $i, j \in \{1,...,k\}$ and ΔE becomes:

$$\Delta E = a_{i,j}^2 \cdot c_{j,j} = a_{i,j}^2 \cdot (P-1)\sigma_j^2 \qquad (6)$$

where σ_j^2 is the variance of the j^{th} component, for a zero-mean input process.

Nonzero elements of the new sparse predictor $\breve{\mathbf{A}}_1 = \{\, \breve{a}_{i,j} \,\}$ will be located on positions where the products $\sigma_j^2 a_{i,j}^2$ are above the chosen threshold Λ, while the remaining elements are set to zero. The simplest procedure for obtaining the values of nonzero elements is to retain their values from the original full VLP matrix i.e.

$$\breve{a}_{i,j} = \begin{cases} a_{i,j} \ , \ \sigma_j^2 a_{i,j}^2 > \Lambda \\ 0 \ , \ \text{otherwise} \end{cases} . \qquad (7)$$

However, sparse predictor designed this way does not minimize the energy of the prediction residual for this structure. In another words, there is another predictor with same positions but different values of nonzero elements that results with minimal E. Therefore, the nonzero elements must be recalculated to compensate for the loss induced by zeroing. This procedure is explained next.

Number and positions of nonzero elements in each row of the sparse predictor can be different, so the rows of the predictor have to be treated separately. Let $J^{(i)}$ denote the set of indices that correspond to positions (columns) of nonzero elements in the i^{th} row of matrix $\breve{\mathbf{A}}_1$

$$J^{(i)} = \left\{ j \mid \breve{a}_{i,j} \neq 0 \right\} \quad , \ i = 1, 2, ..., k \ , \qquad (8)$$

while the reduced row vectors $\breve{\mathbf{a}}^{(i)\mathsf{T}}$ contain only these nonzero elements

$$\breve{\mathbf{a}}^{(i)\mathsf{T}} = \{\, [\, \breve{a}_{i,j} \,] \mid j \in J^{(i)} \,\} \ , \ i = 1, 2, ..., k. \qquad (9)$$

Analogously, let $\breve{\mathbf{x}}^{(i)}(n-1)$ represent the reduced input vector obtained by extracting the rows defined by indices in $J^{(i)}$ from the input vector $\mathbf{x}(n\text{-}1)$. Vectors $\breve{\mathbf{x}}^{(i)}(n-1)$ and $\breve{\mathbf{a}}^{(i)\mathsf{T}}$ are used for prediction of $x_i(n)$, resulting with a new prediction error sequence given by:

$$e_i'(n) = x_i(n) - \breve{\mathbf{a}}^{(i)\mathsf{T}} \breve{\mathbf{x}}^{(i)}(n-1). \qquad (10)$$

Optimal nonzero elements in the i^{th} row of the sparse predictor are found by minimizing the energy E_i' of the residual $e_i'(n)$. This leads to an independent set of linear equations for each row i of the sparse predictor:

$$\breve{\mathbf{C}}_{11}{}^{(i)} \cdot \breve{\mathbf{a}}^{(i)} = \breve{\mathbf{c}}_{10}{}^{(i)} \qquad (11)$$

Matrix $\breve{\mathbf{C}}_{11}{}^{(i)} = \{\, \breve{c}_{l,m}^{(i)} \,\}$ and vector $\breve{\mathbf{c}}_{10}{}^{(i)} = \{\, \breve{b}_l^{(i)} \,\}$ can be found from the input process $\mathbf{x}(n)$ as in:

[#] Throughout the paper, vectors without transposition are assumed to be column vectors.

$$\breve{\mathbf{C}}_{11}{}^{(i)} = \sum_{n=1}^{P} \breve{\mathbf{x}}^{(i)}(n-1)\,\breve{\mathbf{x}}^{(i)\mathsf{T}}(n-1) \qquad (12)$$

$$\breve{\mathbf{c}}_{10}{}^{(i)} = \sum_{n=1}^{P} \breve{\mathbf{x}}^{(i)}(n-1)\,x_i(n) \qquad (13)$$

Dimensions of $\breve{\mathbf{C}}_{11}{}^{(i)}$, $\breve{\mathbf{c}}_{10}{}^{(i)}$ and $\breve{\mathbf{a}}^{(i)}$ depend on i and are determined by the number of elements in $J^{(i)}$ (from 0 to k). Actually, there is no need to re-compute $\breve{\mathbf{C}}_{11}{}^{(i)}$ and $\breve{\mathbf{c}}_{10}{}^{(i)}$ in (12) and (13), since their elements $\breve{c}_{l,m}^{(i)}$ and $\breve{b}_{l}^{(i)}$ are already contained in matrices $\mathrm{C}_{11}=\{c_{i,j}\}$ and $\mathrm{C}_{10}=\{b_{i,j}\}$, and can simply be extracted from them:

$$\left.\begin{aligned} \breve{c}_{l,m}^{(i)} &= c_{J^{(i)}(l),\,J^{(i)}(m)} \\ \breve{b}_{l}^{(i)} &= b_{J^{(i)}(l),\,i} \end{aligned}\right\} \; i = 1,2,\ldots,k \qquad (14)$$

where $J^{(i)}(l)$ denotes the l^{th} element of $J^{(i)}$. After solving (11), the elements of $\breve{\mathbf{a}}^{(i)\mathsf{T}}$ are placed back on their original positions in the i^{th} row of sparse matrix $\breve{\mathbf{A}}_1$ according to the indices in $J^{(i)}$. By calculating nonzero elements of the sparse predictor matrix as described, these elements attain the optimal value for the obtained sparse structure.

The whole process can be repeated by comparing the elements of this sparse predictor to the threshold again. The threshold can be the same as in the first iteration since some of the new elements may now fall below the threshold, but can also be higher if a structure with less nonzero elements is required. Whenever the structure is changed, the design of the optimal sparse predictor should be repeated to compensate for the removed elements.

2.1 Criteria for sparse structure determination

In each design iteration, a certain number of predictor elements are made equal to zero, i.e. the ones with $\sigma_j^2 a_{i,j}{}^2 < \Lambda$. Several criteria for threshold determination were investigated and compared in this paper:

a) Λ is chosen such that a fixed number q of predictor elements with the smallest $\sigma_j^2 a_{i,j}{}^2$ are set to zero in each iteration.

b) The total weighted energy E_{A1} of the predictor is calculated using the following expression

$$E_{\mathrm{A}_1} = \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_j^2 a_{i,j}^2 \qquad (15)$$

and all elements with $\sigma_j^2 a_{i,j}{}^2 < \Lambda$, are set to zero, with threshold $\Lambda = p \cdot E_{\mathrm{A1}}$ determined as the percentage p of E_{A1}.

c) A maximum possible number of predictor elements with the smallest $\sigma_j^2 a_{i,j}{}^2$ are set to zero in each iteration but with restriction that the total weighted energy of the remaining components is at most p percent below E_{A1} before removal.

3 APPLICATION OF THE SPARSE VLP IN SPEECH CODING

The proposed sparse VLP (sVLP) method was applied to interframe predictive coding of LPC parameters of the speech signal. The Line Spectrum Frequencies (LSF) representation of the LPC speech model was chosen as the input vector process, since consecutive LSF vectors are highly correlated. Prediction orders higher than the first result with insignificant prediction gain improvement [2], so the first order was chosen.

The correlation properties of LSF vectors are non-stationary and can change very rapidly from vector to vector. Therefore, vector predictor must be adapted to this change by using a vector-based switched-adaptive approach [2]. This technique is known as Switched-Adaptive Interframe Vector Prediction (SIVP) and is based on prediction using one predictor from the pre-designed set. Selection of the predictor is independent for each LSF vector.

Interframe coding of LSF vectors based on SIVP scheme results in lower overall complexity as well as lower bit-rate, compared to memoryless speech spectrum quantization. The lowest bit-rate is achieved if SIVP scheme is combined with vector quantization (VQ) of the prediction residual [6]. To reduce the complexity and storage requirements of the VQ, it is favorable to use as many switched matrices in SIVP as possible. On the other hand, to achieve the highest possible prediction gain, the selection of the switched predictor should be performed by exhaustive search through all of the N predictors. However, using a large number of matrices automatically means greater computational load, so application of sVLP technique for such a scheme seems as a reasonable choice.

The procedure for design of the SIVP predictor set with sparse predictor matrices is explained next. It starts by designing the set of full predictors by an iterative procedure in which each predictor is optimized to achieve maximal prediction gain for the class of input vectors that are predicted with that particular predictor [6], [7]. Integration of the proposed sparse VLP technique into the SIVP scheme imposes some modifications of the original design procedure.

The goal of the procedure is to determine a set of N sparse predictors with total number of nonzero elements equal to S, and with optimized structure and matrix elements for each of the switched predictors. This can be obtained by two nested loops where the inner loop represents the modification of the standard full predictor iterative design procedure. The difference is in the way the predictors are calculated from the covariance matrices as explained in the second section. Each of the switched predictors is optimized to achieve the maximal prediction gain for a given structure. The structure of all switched predictors is changed in the outer loop according to one of the proposed criteria. The structures obtained in one iteration of the outer loop are fixed for the whole iterative SIVP design performed in the inner loop. The procedure terminates upon reaching the desired number of nonzero elements S.

Since the overall complexity of the sVLP prediction with exhaustive search is proportional to S, the complexity reduction factor $\gamma = S / Nk^2$ was chosen as a design parameter, where Nk^2 is the total number of elements of switched predictors for the classical (full) VLP.

4 SIMULATION RESULTS

The speech database used for design contains 20 minutes of speech, spoken by 2 male and 2 female speakers. It was analyzed using robust LPC analysis [8] with pre-emphases and bandwidth expansion of 10 Hz. The total number of 60000 LSF vectors of dimension k=10 were obtained from speech segments of 25 ms with 8kHz sampling rate, and with the LPC frame rate of 50 frames/s.

SIVP design was performed for several number of switched predictors: N = 2, 4, 8, 16 and 32. To evaluate the performance of the sVLP compared to the classical VLP, two groups of simulations were carried out. Firstly, the analysis was performed for the case without the quantizer in the loop (open-loop prediction). The loss in prediction gain L due to sparse structure was calculated as the difference between the open loop prediction gain of the full predictor G_{pF}, and of the sparse predictor, G_{pS}, i.e. $L = G_{pF} - G_{pS}$. For both cases the open loop prediction gain (in dB) was calculated using the conventional expression:

$$G_p = 10\log_{10}\left(\mathrm{E}(\| \mathbf{x}(n) \|^2) / \mathrm{E}(\| \mathbf{e}(n) \|^2)\right), \quad (16)$$

where $\mathrm{E}(\,)$ denotes statistical expectation. The sparse predictors were designed with complexity reduction factors γ ranging from 100% (baseline full VLP) to 10% (10-fold reduction in complexity).

The results of these simulations are illustrated in Fig. 1 for N=32. It can be observed that all proposed criteria, a), b) and c) result with similar, relatively small loss, with c) being the best of them. However, criterion a) is simpler for actual implementation since it results with the same number of nonzero elements in each of the switched matrices. Furthermore, it is obvious that the results obtained by all criteria proposed in this paper are much better then those obtained with multidiagonal predictors [5] also shown in Fig. 1. This is due to fact that the structures of the switched predictors were not fixed as for the multidiagonal case, but were optimized in the design procedure and thus adjusted to the correlation properties of the input process.

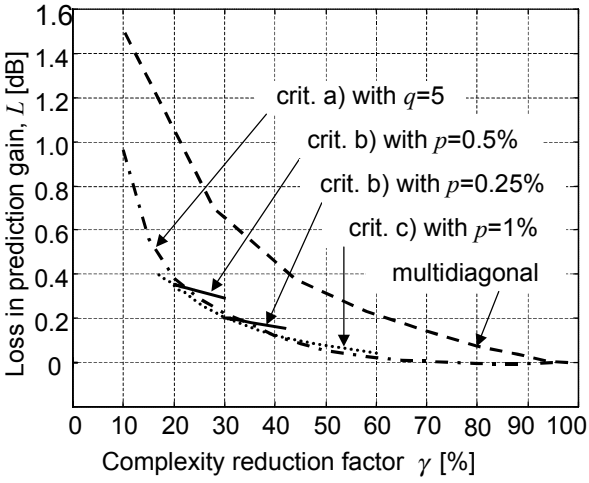

Figure 1. Loss in prediction gain vs. complexity reduction factor γ for several sparse predictor design criteria, with N=32

The influence of the number of switched predictors on the performance of sparse VLP is demonstrated in Fig. 2. together with the corresponding results of the baseline predictor.

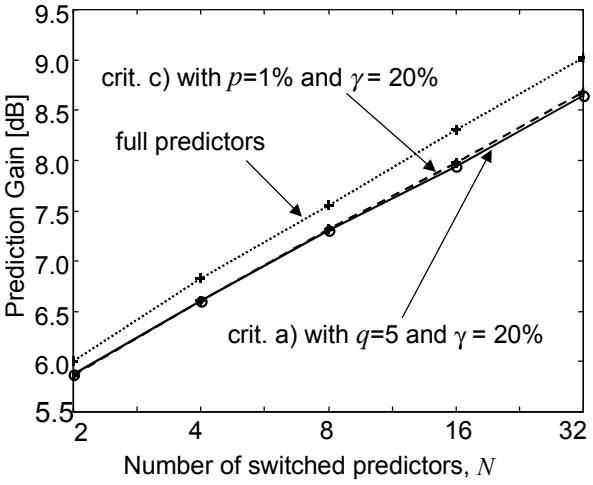

Figure 2. Open loop prediction gain vs. number of predictors, obtained with full predictors and sparse predictors with γ=20%

The prediction gain of sVLP for γ=20% (5-fold reduction) obtained with two sparse design criteria, a) and c), was calculated for different number of switched predictors. As can be seen, the loss of the prediction gain due to sparse structure increases with N, varying from 0.1 to 0.4 dB. However, this loss can be easily justified by the significant reduction of complexity and storage requirements.

In the second part of the assessment, both baseline and sparse predictors were integrated into the actual SIVP LSF coder with VQ. Coders were evaluated using 10 utterances, each 1 minute long, spoken by 10 different speakers, and prepared in the same way as the training base. The employed sparse switched predictors had an average of 16.4 nonzero elements (approx. 6-fold reduction) and were designed with criterion c). Quantization was performed using full search split VQ with two subvectors having 4 and 6 components, respectively. The average spectral distortion for different VQ resolutions is shown in Fig. 3., where the total number of bits required for coding of one LSF vector (frame) is shown on the x axis. In all cases 5 bits are used to specify the switched predictor (N=32). The difference in $\overline{SD}$ between the two group of coders is relatively small and varies from 0.029dB @ 19bit / fr. to only 0.020 dB @ 23bit / fr. This suboptimality is equivalent to approx. 0.3 bit at all VQ resolutions.

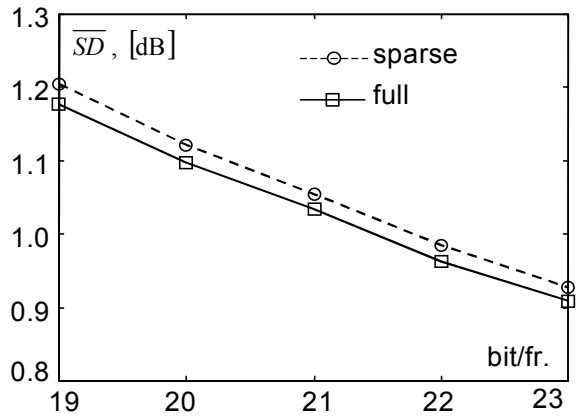

Figure 3. Average spectral distortion of the baseline and sVLP (γ=16%) coder vs. bit / fr.

5 CONCLUSION

The results presented in this paper suggest that for certain vector processes (such as LSF in speech coding) sparse VLP can be effectively employed. Predictors with structures optimized to a given vector process offer prediction gain that is only slightly lower than the one with full predictors. Favorable results obtained with the open loop evaluation of the sparse VLP were further confirmed by closed loop evaluation in the actual coder. It was shown that a 6-fold complexity reduction can be achieved by sacrificing only 0.3 bits per frame. The proposed method is especially suitable for coders with large number of switched predictors.

REFERENCES

[1] Chen J.H., Gersho A., Covariance and autocorrelation methods for vector linear prediction, in *Proc. ICASSP*, 1987, pp.1545-1548

[2] Yong M., Davidson G., Gersho A., Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction, in *Proc. ICASSP*,1988, vol.1, pp. 402-405

[3] Skoglund J., Linden J., Predictive VQ for noisy channel spectrum coding: AR or MA?, in *Proc. ICASSP*, 1997, vol.2, pp. 1351-1354

[4] Eriksson T., Linden J., Skoglund J., Interframe LSF quantization for noisy channels, in *IEEE Trans. on Speech and Audio Processing*, vol. SAP-7, no. 5, pp. 495-509, 1999

[5] Petrinović D., Petrinović D., Sparse vector linear predictor matrices with multidiagonal structure, in *Proc. of the 6th European Conference on Speech Comm. and Technology (EUROSPEECH '99)*, Budapest, Hungary, 1999, vol. 3. , pp. 1483-1486

[6] Wang S., Paksoy E., Gersho A., Product code vector quantization of LPC parameters, in *Speech and Audio Coding for Wireless and Network Applications*, editors: B. S. Atal, V. Cuperman, A. Gersho, Kluwer Academic Publishers, 1993, pp. 251-258

[7] Gersho A., Gray R. M., *Vector quantization and signal compression*, Kluwer Academic Publishers, 1992

[8] Lee, C.H., On robust linear prediction of speech, in *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-36, no. 5, pp. 642-650, 1988