

Notes on the Joint Coding of Coefficients Location and Amplitude in Sparse Vectors through Spectral Modeling

Daniele Giacobello

August 27, 2009

Abstract

In this paper, we give an overview of a scheme to jointly code the information about location and amplitude of the nonzero coefficients in a sparse vector, first proposed in [1]. By exploiting the properties and relations between the Linear Prediction and Minimum Variance Distortionless Response spectra, it is able to transform this information into a prediction filter. This scheme allows for systematic trade-offs between bit allocation and accuracy of the sparse representation in a manner hitherto not possible.

1 Introduction

In signal compression, a great deal of attention has been focused lately on sparse representation given that many natural signals (e.g., images or audio) exhibit a clear underlying sparse structure when the proper linear transform is applied. Despite the rich promise of sparse signal representation in coding applications, we believe that their development lacks a clear relation between sparsity and bit rate. In particular, few significant coefficients in a transform domain do not necessarily correspond to a decrease in the number of bits required to describe the signal.

2 Properties of LP and MVDR Spectra

In this section we illustrate the properties of LP and MVDR spectra that are being used in our formulation. Let us consider first consider an input signal consisting on the sum of K real cosine signals

$$u(n) = \sum_{i=1}^K c_i \cos(\omega_i n), \quad (1)$$

and the corresponding correlation sequence

$$r(m) = \sum_{i=1}^K 2S(\omega_i) \cos(\omega_i m), \quad (2)$$

where $S(\omega_i) = |c_i|^2/4$. The input signal exhibits a discrete line spectrum at the positive and negative frequencies $\pm\omega_i$, $i = 1, \dots, K$ with spectral powers $S(\omega_i)$.

Lemma 1. Define $r(m) = \sum_{i=1}^K 2S(\omega_i) \cos(\omega_i m)$, a correlation sequence, a linear prediction filter $A_M(z)$ with order $M = 2K$

$$A_M(z) = 1 + \sum_{i=1}^M a_k z^{-k}, \quad (3)$$

has zeros at the $2K$ positive and negative frequencies $\pm\omega_i$, $i = 1, \dots, K$, i.e. $A_M(e^{\pm\omega_i}) = 0$.

Therefore, an M th order linear prediction filter places its filter zeros at the frequencies of the line spectrum, providing accurate frequency location information, but not amplitude information, since $A_M(e^{\pm\omega_i}) = 0$. It is easy to prove that $A_M(z)$ is a palindromic monic polynomial where all the roots lie on the unit circle [2].

Now we state a property of the MVDR spectrum that we exploit for modeling amplitude information.

Lemma 2. Define $r(m) = \sum_{i=1}^K 2S(\omega_i) \cos(\omega_i m)$, a correlation sequence. The MVDR spectrum $P_{MV}^{(M)}$ of order $M = 2K - 1$ models the powers of the line spectra exactly. $P_{MV}^{(M)}(\omega_i) = S(\omega_i)$.

Details of the MVDR spectral modeling of exponentials are in [3]. With the results on linear prediction modeling of frequency location information, and MVDR modeling of amplitude information, we can state the following.

Theorem 1. Define $r(m) = \sum_{i=1}^K 2S(\omega_i) \cos(\omega_i m)$ with $\omega_i \neq 0, \pi$. Then the prediction error variance $P_e^{(2K-1)}$ and reflection coefficients Γ_m , $m = 1, \dots, 2K-1$ corresponding to a $2K-1$ order linear prediction filter $A_{2K-1}(z)$ based on $r(m)$, are sufficient to recover the line frequency locations ω_i and the spectral powers $S(\omega_i)$ exactly.

Outline of Proof. Note that from the given $r(m)$ and the constraints $\omega_i \neq 0, \pi$, we know that $\Gamma_{2K} = 1$. Consequently the given Γ_m , $m = 1, \dots, 2K-1$, are sufficient to construct the order $2K$ linear prediction filter $A_{2K}(z)$. In particular, given $\Gamma_{2K} = 1$, the relation between the two predictor is [4]:

$$A_{2K}(z) = A_{2K-1}(z) + z^{-2K} A_{2K-1}(z^{-1}). \quad (4)$$

From Lemma 1, $A_{2K}(z)$ has its filter zeros at the frequencies ω_i . The Γ_m , $m = 1, \dots, 2K-1$ and $P_e^{(2K-1)}$ can be used to obtain the order $2K-1$ MVDR spectrum [5]. From Lemma 2, $P_{MV}^{(M)}$ model the spectral powers exactly at the line frequencies, ω_i .

3 Joint coefficients location and amplitude coding

Let us now consider the problem of coding K nonzero coefficients from a sparse vector of length N ($K \ll N$). The K coefficients have positions p_1, \dots, p_K , and are allowed to range from $1 \leq p_i \leq N$. Each coefficient p_i is weighted by $s_i g_i$, where s_i is the sign and g_i is its positive amplitude.

3.1 Encoding

An autocorrelation sequence is constructed as follows:

$$r(m) = \sum_{i=1}^K 2g_i \cos\left(\pi \frac{p_i}{N+1} m\right). \quad (5)$$

This sequence has a corresponding discrete line spectrum in the frequency domain. The “sampling frequency” corresponds to $2(N + 1)$ and the spectral peaks are positioned at the frequencies $\omega_i = \pi/(N + 1)$ which encode the coefficient position information. The line spectra at the frequencies ω_i have corresponding powers g_i which encode the amplitude information. The $2K - 1$ order linear predictor $A_{2K-1}(z)$ and corresponding prediction error variance $P_e^{(2K-1)}$ are computed using the Levinson-Durbin algorithm using $r(m)$. According to Theorem 1, the parameters of the filter are sufficient for recovering all the informations relevant to the perfect reconstruction of the sparse vector.

When the signs are both negative and positive, we can proceed in two ways, one sending the signs separately but we can also easily incorporate this information by modifying the construction of the correlation sequence. By utilizing multiples of $0.5\pi/(N + 1)$, we can incorporate sign information. For negative signs, we follow the convention of adding the fractional frequency value $0.5\pi/(N + 1)$ to the original line frequency. In particular if $s_i = -1$, we add $0.5\pi/(N + 1)$ to the original line frequency $\pi p_i/(N + 1)$ to obtain a new line frequency value $\pi(p_i + 0.5)/(N + 1)$ that is used in the construction of the autocorrelation sequence in (5):

$$r(m) = \sum_{i=1}^K 2g_i \cos\left(\pi \frac{p_i + 0.5}{N + 1} m\right). \quad (6)$$

3.2 Decoding

The order $2K$ linear prediction filter $A_{2K}(z)$ is constructed from $A_{2K-1}(z)$, and $\Gamma_{2K} = 1$ using the relation provided in (4). From Lemma 1, the zeros of $A_{2K}(z)$ are found at the frequencies ω_i which give the information about the locations:

$$p_i = \frac{N + 1}{\pi} \omega_i. \quad (7)$$

The MVDR spectrum is directly computed from the coefficients of $A_{2K-1}(z)$ and the prediction error variance $P_e^{(2K-1)}$. A fast algorithm for computing the MVDR spectrum is given in [5]. From Lemma 2, the $2K - 1$ order MVDR spectrum models the gain informations exactly, i.e. $P_{MV}^{(M)}(\omega_i) = g_i$.

3.3 Example of Joint Coding

Let us provide now a simple example to clarify the concepts illustrated above. Consider a sparse vector $N = 100$, where only $K = 5$ coefficients are nonzero. The position vector is $\mathbf{p} = [12, 34, 69, 74, 83]$ with corresponding gain vector $\mathbf{g} = [1, -10, 32, -52, 7]$. The correlation sequence is computed with the sign modification, the $2K - 1 = 9$ prediction vector $A_9(z)$ and the prediction error P_e^9 are then calculated via the Levinson-Durbin algorithm and transmitted to the decoder (assume for now that no quantization takes place).

The decoder computes the $2K = 10$ order linear prediction filter $A_{10}(z)$ from $A_9(z)$, using $\Gamma_{10} = 1$ (4). The roots of $A_{10}(z)$ are calculated and the information about the location is retrieved using (7) $\hat{\mathbf{p}} = [12, 34, 69, 74, 83]$. The MVDR spectrum of order $2K - 1 = 9$ is constructed using $A_9(z)$ and the prediction error vector P_e^9 and evaluated at the positions of the order $2K = 10$ linear prediction filter's zeros to determine the gain information $\hat{\mathbf{p}} = [1, 10, 32, 52, 7]$. Combining with the sign information we obtain the original gain vector. In Figure 1, we show an example of the order 10 LP spectrum, from which the location information is retrieved, and the order 9 MVDR spectrum from which the gain information is found. These spectra do not need to be computed, they are just shown for clarity of presentation.

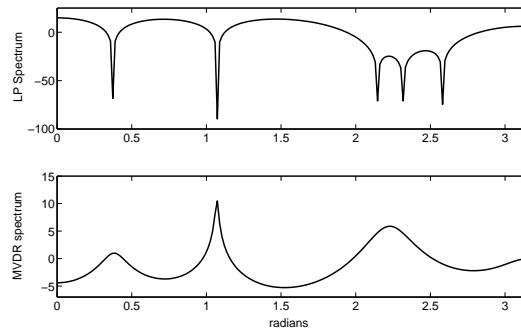


Figure 1: In the top box is shown the LP spectrum. The zeros of the LP filter are used to determine the pulse positions and sign information. The MVDR spectrum below is evaluated at the LP filter zeros used and determines the gain values.

4 Perturbation and Quantization Robustness

In Section 3 we have looked at the encoding-decoding procedure necessary for our method. In synthesis, if we wish to transmit the information regarding K nonzero components (positions and amplitude) in a N length sparse vector, we need an order $2K - 1$ prediction vector $A_{2K-1}(z)$ and the variance of its prediction error $P_e^{(2K-1)}$. This new parameters set carries all the information we need. It is important now to evaluate the robustness of this information to perturbation and quantization noise. In particular, we will require the positions of the K components not to shift. Consider:

$$A_{2K-1}(z) = 1 + \sum_{i=1}^{2K-1} a_i z^{-i} = \prod_{i=1}^{2K-1} (1 - z_i z^{-1}) \quad (8)$$

we know that the roots ω_i of the polynomial $A_{2K}(z)$ constructed from $A_{2K-1}(z)$ through the relation in (4):

$$A_{2K}(z) = 1 + \sum_{i=1}^{2K-1} (a_i + a_{2K-i}) z^{-i} + z^{-2K} \quad (9)$$

are directly related to the positions by (7), therefore we can analyze the sensitivity of this information as the partial derivative $\partial p_m / \partial a_k$, the variation of the m th position (The roots of $A_{2K}(z)$) as a function of the variation of the k th coefficient of A_{2K-1} . To do so, we use the following relations:

$$\begin{aligned} \frac{\partial A_{2K}(z)}{\partial a_k} &= z^{-k} + z^{2K-k}, \\ \frac{\partial A_{2K}(z)}{\partial z_h} &= -z^{-2K} \prod_{i \neq h}^{2K} (z - z_i); \end{aligned} \quad (10)$$

working with this two relations and posing $z = z_h$, we obtain:

$$\left| \frac{\partial z_h}{\partial a_k} \right| = \left| \frac{z_h^{2K-k} + z_h^{4K-k}}{\prod_{n \neq h}^{2K} (z_h - z_n)} \right| \quad (11)$$

since the roots of $A_{2K}(z)$ are complex conjugates and lie on the unit circle $z_i = e^{\pm j\omega_i}$, $i = 1, \dots, K$ the only information relevant to us is the phase ω_i ,

which is related to the position from the formula in 7. Finally, we obtain

$$\left| \frac{\partial p_h}{\partial a_k} \right| = \left| \frac{e^{jp_h \frac{2K-k}{N+1} \pi} + e^{jp_h \frac{4K-k}{N+1} \pi}}{\prod_{n \neq h}^K (e^{j \frac{p_h}{N+1} \pi} - e^{j \frac{p_n}{N+1} \pi})^2} \right|, \quad (12)$$

that relates the variation of the p_h position to the variation of the a_k coefficient. This formula efficiently state what can be intuitively assumed: a clear dependence to the proximity of the locations and the length of the sparse vector N . What we propose now is to find the “weakest link”, i.e. the most sensitive coefficient of a_k of $A_{2K-1}(z)$ related to the shift of the position p_h . In order to do so we propose to construct a *sensitivity matrix* where each value $S_{i,l} = \left| \frac{\partial p_i}{\partial a_l} \right|$. By picking the highest value (i.e., the most sensitive relations between coefficient and position), we found the coefficient a_l most sensitive to quantization. Given this, we find the maximum Δ of quantization that will not move the root location too much (i.e., more than 1/2 of a sample if the sign location is transmitted separately or 1/4 otherwise) by numerical optimization over the single parameter Δ .

Note we might also allow for different delta’s, i.e., different delta’s for each sparse signal but also different delta’s for the individual elements of the sparse vector. This can be done by ranking the sensitivity matrix according to “sensitivity”. The delta’s should then be conveyed using additional bits (side information).

References

- [1] M. N. Murthi and B. D. Rao, “Towards a synergistic multistage speech coder,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp 369-372, 1998.
- [2] C. Magi, T. Bäckström, and P. Alku, “Simple proofs of root locations of two symmetric linear prediction models,” *Signal Processing*, vol. 88, no. 7, pp. 1894–1897, 2008.
- [3] M. N. Murthi and B. D. Rao, “All-pole modeling of speech based on the minimum variance distortionless response spectrum,” *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 221–239, 2000.
- [4] L. R. Rabiner, R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.

- [5] S. Haykin, *Adaptive filter theory*, Prentice-Hall, Englewood Cliffs, NJ, 1991.
- [6] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Speech coding based on sparse linear prediction," *Proc. European Signal Proc. Conf.*, pp. 2524–2528, 2009.
- [7] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 103–106, 2010.
- [8] C. Magi and T. Bäckström, "Properties of line spectrum pair polynomials A review," *Signal Processing*, vol. 86, no. 11, pp. 3286–3298, 2006.
- [9] S. Subasingha, M. N. Murthi, and S.V. Andersen, "Gaussian Mixture Kalman predictive coding of Line Spectral Frequencies," *IEEE Transaction on Audio, Speech and Language Processing*, vol. 17, no. 2, pp. 379–391, 2009.