ON ADAPTIVE NOISE CANCELLING BASED ON INDEPENDENT COMPONENT ANALYSIS

Hyung-Min Park, Sang-Hoon Oh, and Soo-Young Lee

Brain Science Research Center
and also Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
Taejon 305-701, Korea
mhp@eeinfo.kaist.ac.kr

ABSTRACT

We present a method to deal with adaptive noise cancelling based on independent component analysis (ICA). Although popular least-mean-squares (LMS) algorithm removes noise components based on second-order correlation, the proposed ICA-based algorithm can utilize higher-order statistics. Additionally, extending to transform-domain adaptive filtering (TDAF) methods, normalized ICA-based algorithm is derived to improve convergence rates. Experimental results show that the proposed ICA-based algorithm provides much better performances than conventional LMS approach in real-world problems.

1. INTRODUCTION

Adaptive noise cancelling is an approach to reduce noise based on reference noise signals [1]. Fig. 1 shows the typical adaptive noise cancelling system. In conventional adaptive noise cancelling systems, a signal s is transmitted over a channel to a sensor and a noise n_0 is added in the sensor from the noise source so that the combined signal and noise $s + n_0$ form the primary input signal. Another sensor receives a noise signal n_1 through another channel, and this sensor provides the reference input signal. The goal is to get a system output u which is the best least squares estimate of the signal s.

The most popular algorithm for noise cancellation is LMS algorithm [1][2], which adapts as

$$\Delta w(k) \propto u(t) n_1(t-k), \tag{1}$$

where the output u is

$$u(t) = s(t) + n_0(t) - \sum_{k=1}^{K} w(k) n_1(t-k). \tag{2}$$

This work was supported by the Brain Science & Engineering Research Program sponsored by Korean Ministry of Science and Technology.

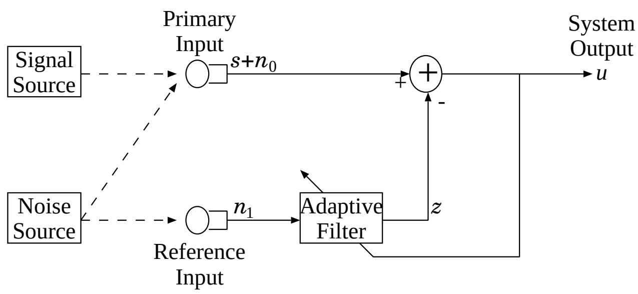


Fig. 1. Adaptive noise cancellation scheme

It decorrelates system output signal from the reference noise signal and removes noise components of the primary input signal based on second-order statistics only. However, there may exist many other components in the primary input signal which depend on the noise reference signal through higher-order statistics.

In this paper, we present a method that improves performances of adaptive noise cancelling systems by using ICA. This method can remove noise components of the primary input signal based on statistical independence, which incorporates both second-order and higher-order statistics. In addition, it is extended to TDAF methods to improve convergence rates, and normalized version of the adaptation rule is derived. Experimental results show that the performances of the ICA-based algorithm are much better than those of the popular LMS algorithm.

2. LEARNING RULE

ICA was proposed to recover independent sources from given sensor signals in which the sources have been mixed through unknown channels [3][4]. Bell and Sejnowski proposed to learn the unmixing matrix $\mathbf{W}$ by minimizing the mutual information among components of $\mathbf{y} = g(\mathbf{u})$, where g is a nonlinear function approximating the cumulative density function (cdf) of the sources and $\mathbf{u}$ denotes recovered sources [3]. They showed that for super-Gaussian signals minimizing the mutual information between components of $\mathbf{y}$ is equal to maximizing the entropy of $\mathbf{y}$. Lee *et al.* and Torrkola have addressed blind separation of convolved sources [5][6].

Learning rules of adaptive filter coefficients in the noise cancellation system can be derived by maximizing entropy. To derive the learning rules conveniently, we set dummy output $v = n_1$. It doesn't make any difference during derivation using entropy maximization because noise n_1 is independent of signal s and it doesn't introduce any other parameters. In this system, the Jacobian can be expressed as

$$J = \frac{\partial y_1}{\partial x}\frac{\partial y_2}{\partial n_1} - \frac{\partial y_1}{\partial n_1}\frac{\partial y_2}{\partial x} = \frac{\partial y_1}{\partial u}\frac{\partial y_2}{\partial v}, \tag{3}$$

where $x = s + n_0$, and y_1 and y_2 are outputs of nonlinear functions approximating the cdfs of the signal and the noise, respectively. By maximizing $log|J|$, the learning rule of each coefficients can be obtained as follows:

$$\begin{aligned} \Delta w(k) &\propto \frac{\partial}{\partial w(k)} log|J| = \left(\frac{\partial y_1}{\partial u}\right)^{-1} \frac{\partial}{\partial w(k)}\left(\frac{\partial y_1}{\partial u}\right) \\ &= \varphi(u(t))n_1(t-k), \end{aligned} \tag{4}$$

where the score function $\varphi(u)$ is

$$\varphi(u) = -\left(\frac{\partial y_1}{\partial u}\right)^{-1} \frac{\partial^2 y_1}{\partial u^2}. \tag{5}$$

The difference between the LMS algorithm and the ICA-based approach comes from existence of the score function. Introducing nonlinearity to the LMS algorithm has been studied by many researchers to improve the properties and the performances [7][8][9]. Especially, Douglas and Meng generalized the LMS algorithm by using a nonlinearity acting on the error for system identification problem [7]. They provided methods for optimizing the nonlinearity to minimize the misadjustment for a given convergence rate. Under i.i.d. signal assumption, the same nonlinearity was derived using linearized approximation near convergence, but in general derivation, nonlinearity was different assuming zero-mean white Gaussian (noise) data. However, the ICA-based learning rule can be derived only with independency between the signal and the noise sources. And, Douglas and Meng analyzed convergence behavior near convergence with i.i.d. signal [7]. The coefficient-error vector $\mathbf{v}(t) = \mathbf{w}(t) - \mathbf{w}^*$ can be expressed as

$$E[\mathbf{v}(t+1)] = (\mathbf{I} - \mu E[\varphi'(s(t))]\mathbf{R})E[\mathbf{v}(t)], \tag{6}$$

where $\mathbf{R}$ is the correlation matrix of the reference input vector. It exactly matches the corresponding form derived for the LMS algorithm, and an approximate bound on the step size μ can be determined to guarantee convergence in the mean with eqn. (6).

The LMS algorithm decorrelates output signal from the reference input to remove corrupting noise component which is correlated to the reference input. However, the ICA-based approach makes output signal independent of the secondary input. The independence involves higher-order statistics including the second-order statistics, i.e. correlation. When noise signals are obtained from a noise source through different channels, there may exist many components which depend on each other without correlation. When the ICA-based approach is used for adaptive noise cancelling, these noise components can be cancelled, being different from the LMS algorithm. Therefore, the ICA-based approach provides better performances than the conventional LMS algorithm.

3. EXTENSION TO TRANSFORM-DOMAIN ADAPTIVE FILTERING (TDAF) METHODS

The LMS algorithm is the most widely used real time adaptive filtering algorithm due to its low computational complexity and memory requirement. The convergence speed of the LMS algorithm is governed by the spread of eigenvalues of the autocorrelation matrix of the input data, and by reducing the eigenvalue spread, the convergence speed can be enhanced [10][11]. The recursive least squares algorithm is known to achieve near optimum convergence rates by forming an estimate of the inverse of the input autocorrelation matrix and automatically whitening the input data. Unfortuately, the computational complexity is large and is not easily carried out in real time within the resource limitations of practical applications. TDAF methods have been devised to pre-whiten the input data using unitary transforms. The best transform for this purpose is the Karhunen-Loéve transform (KLT). However, the KLT is signal-dependent, and usually cannot be computed in real time. Therefore, it is replaced by simpler transforms. Marshall *et al.* compared several computationally appealing unitary transforms for convergence enhancement [10].

Fig. 2 shows the TDAF structure. In the TDAF structure, a vector of K past input samples is processed by a unitary transform $\mathbf{T}$ to produce the transform output vector $\mathbf{r}(t)$ which is given by

$$\mathbf{r}(t) = [r_1(t) \;\; r_2(t) \;\; \cdots \;\; r_K(t)]^T = \mathbf{T}\mathbf{n}_1(t), \tag{7}$$

where $\mathbf{n_1}(t) = [n_1(t-1) \;\; n_1(t-2) \;\; \cdots \;\; n_1(t-K)]^T$. The filter output can be expressed as

$$z(t) = \mathbf{w}^T\mathbf{r}(t), \tag{8}$$

where $\mathbf{w} = [w(1) \;\; w(2) \;\; \cdots \;\; w(K)]^T$ is an adaptive filter coefficient vector.

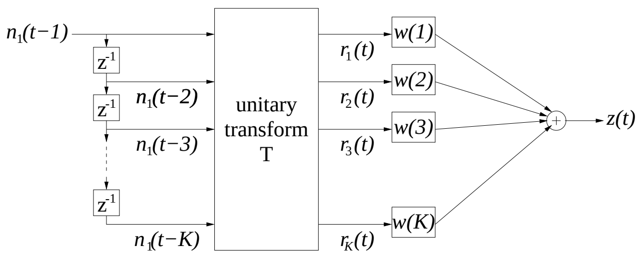


Fig. 2. The TDAF structure

The LMS algorithm for TDAF must have power normalization for improved convergence. In this structure, normalized LMS algorithm is [11]

$$\Delta w(k) \propto \frac{1}{\hat{\sigma}_k^2(t)} u(t) r_k(t), \tag{9}$$

where $\hat{\sigma}_k^2(t)$ is estimates of the input power at kth branch given by

$$\hat{\sigma}_k^2(t) = \alpha \hat{\sigma}_k^2(t-1) + (1-\alpha) r_k^2(t), \quad 0 < \alpha < 1. \tag{10}$$

If the power in a branch is too small, the effective step size becomes too large resulting in convergence failure. To prevent such situations, lower bound of the effective step size was set.

In the same procedure, normalized ICA-based learning rules for TDAF were derived with the same notation for the normalized LMS algortihm as follows:

$$\Delta w(k) \propto \frac{1}{|\hat{\sigma}_k(t)|^q} \varphi(u(t)) r_k(t), \tag{11}$$

where q depends on the form of the score function $\varphi(\cdot)$ [12]. $|\hat{\sigma}_k(t)|^q$ is estimates of the "qth input power" at kth branch expressed as

$$|\hat{\sigma}_k(t)|^q = \alpha |\hat{\sigma}_k(t-1)|^q + (1-\alpha) |r_k(t)|^q, \quad 0 < \alpha < 1. \tag{12}$$

In the same way with the normalized LMS algorithm, lower bound of the effective step size was set.

In addition, Mahalanobis *et al.* provided a method which decomposed an adaptive filter [11]. The overall transform matrix $\mathbf{T}_M$ by decomposing an adaptive filter is discussed in Appendix. Examining $\mathbf{T}_M$, it can be divided into Walsh-Hadamard transform (WHT) matrices, each of which is submatrix with smaller number of rows and columns, by rearranging rows. The WHT sub-matrix can be replaced with any unitary transform matrix since it also produces unitary matrix $\mathbf{T}_M$ which can be applied to the TDAF methods. Using the sub-matrix transform, efficient TDAF methods can be implemented for the adaptive filter with long length. In this paper, this method is applied with an 8×8 unitary submatrix transform.

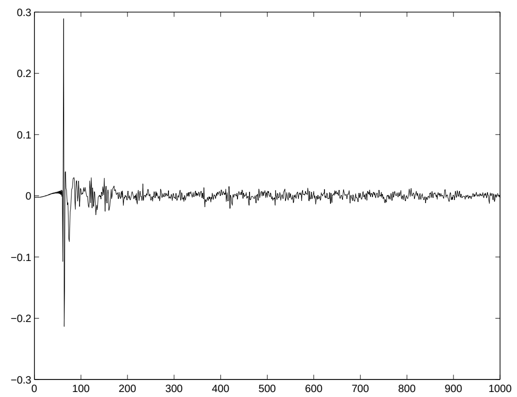


Fig. 3. The measured transfer function in a normal office room

4. EXPERIMENTAL RESULTS

We have compared the performances of the ICA-based approach with those of the LMS algorithm. Artificially generated i.i.d. signals and recorded sources were mixed using both a simple simulation filter and a measured filter in a normal office room. The follwing results are compared in terms of signal-to-noise ratio (SNR), which we define for the output u in the typical adaptive noise cancelling system as the total power of the components caused by the signal source versus that caused by the noise source,

$$SNR = \frac{\langle (s(t))^2 \rangle}{\langle (n_0(t) - \sum_{k=1}^{K} w(k) n_1(t-k))^2 \rangle}. \tag{13}$$

The transfer functions from the signal source to the primary input and from the noise source to the reference input are simple linear scales. To obtain desired initial SNRs, the proper scale values were used. For the simple simulation filter, the transfer function h_{12} from the noise source to the primary input was [6]

$$h_{12}(z) = 0.4z^{-20} - 0.2z^{-28} + 0.1z^{-36}. \tag{14}$$

Fig. 3 shows the transfer function h_{12} which was measured in a normal office room. For the two mixing transfer functions, the numbers of taps of adaptive filter coefficients were 128 and 1024, respectively. Assuming that the primary and the secondary inputs pick up signals with appropriate powers, we have normalized mixture powers properly (generally to 1) and it prevents severe mismatching between recovered signal levels and the nonlinear function. (We have not normalized to exactly match recovered signal levels with the nonlinear function.) All experiments were conducted with several step sizes, and the best performance is shown.

Table 1. SNRs of output signals for artificially generated i.i.d. signals after convergence for the simple simulation mixing filter (dB)

Signal and Noise	Initial SNRs	SNRs after convergence		
		LMS algorithm	ICA-based approach	
			$\varphi = sign$	$\varphi = tanh$
Laplacian	-3.0	30.9	38.0	31.3
	10.0	30.9	38.3	31.7
Gaussian	-3.0	30.6	28.7	30.3
	10.0	30.6	28.7	30.0

4.1. Experiments for artificially generated i.i.d. signals

Table 1 displays the SNRs of output signals for artificially generated i.i.d. signals. The simple simulation filter was used as the transfer function h_{12}. Each signal was composed of 160000 samples. For the ICA-based approach, two different score functions were used. $sign(\cdot)$ and $tanh(\cdot)$ can be used as the score functions by assuming that the probability density functions (pdfs) of the output signals u approximate Laplacian and Gaussian distributions, respectively. Although the ICA-based approach introduces the score function, additional computation is negligible because output of the score function is commonly used for the learning rules of all adaptive filter coefficients. Especially, computational requirements are reduced with $sign(\cdot)$ as the score function because one multiplication can be replaced with just sign change.

For the Laplacian source signals, the performances of the ICA-based approach were better than those of the LMS algorithm. From these results, it can be reasoned that there may be many components in the primary input which depend on the reference signal in higher-order statistics and these noise components can be cancelled by the ICA-based learning rule. For the Gaussian source signals, however, the ICA-based approach provided almost the same SNRs as or a little worse than the LMS algorithm. Gaussian signals can be described by only the first and second-order statistics without higher-order statistics. Therefore, the ICA-based approach which can utilize higher-order statistics does not have any advantage over the LMS algorithm. If one use the score function which is not adequate to the original signal (for example, $sign(\cdot)$ for Gaussian signals), the performances can be degraded because the nonlinear function mismatches the cdf of the signal.

4.2. Experiments for real recorded signals

To perform experiments for real recorded signals, we used speech, car noise, or music as the noises. Another speech was used as the signal s. Korean sentences were recorded for the speech, and the car noise and the music were obtained in NOISEX-92 CD-ROMs and a Korean popular song, respectively. Each signal had 10 second length with $16kHz$ sampling rate. It is known that speech signal approximately follows Laplacian distribution and $sign(\cdot)$ was used as the score function in these experiments.

Table 2. SNRs of ICA-based approach and LMS algorithm for three different noises after convergence for the simple simulation filter (dB)

Signal	Noise	Initial SNRs	SNRs after convergence	
			LMS algorithm	ICA-based approach
Speech	Car	-3.0	27.0	28.4
Speech	Speech	-3.0	25.0	43.2
Speech	Music	-3.0	33.9	55.7

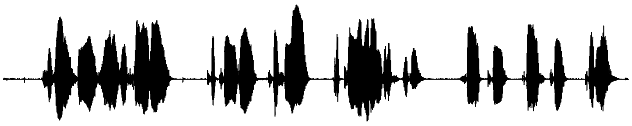
Table 3. SNRs of ICA-based approach and LMS algorithm for three different noises after convergence for the measured filter in a normal office room (dB)

Signal	Noise	Initial SNRs	SNRs after convergence	
			LMS algorithm	ICA-based approach
Speech	Car	-3.0	21.0	26.8
Speech	Speech	-3.0	21.5	38.7
Speech	Music	-3.0	21.7	41.8

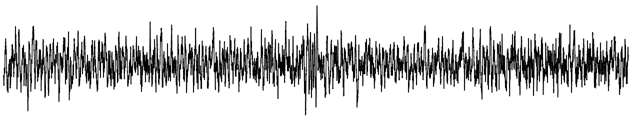
Table 2 and 3 display the SNRs of the two algorithms for the three different noises after convergence for the simple simulation filter and the measured filter, respectively. The SNRs of the ICA-based approach were superior to those of the LMS algorithm. These results show that the ICA-based approach can remove dependent components through higher-order statistics for real recorded signals as well. When the adaptive filter coefficients were learned by the ICA-based approach, Fig. 4 displays the signal and noise sources, the primary input signal, and the system output signal after convergence for the car noise and the simple simulation mixing filter.

4.3. Experiments for TDAF methods

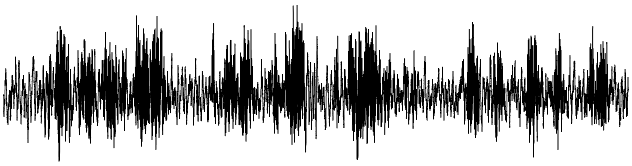
In experiments for TDAF methods, the signal and the noise sources were, respectively, the same speech and car noise signals used in the previous experiments which showed slow convergence rates. When the normalized ICA-based approach was used as the adaptation rule, q in eqn. (11) was set to 1 because $sign(\cdot)$ was used as the score function. As



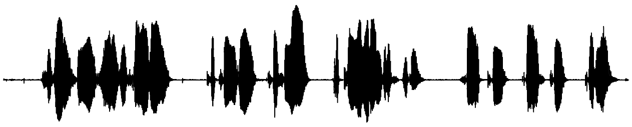
(a) Signal source



(b) Noise source



(c) Primary input signal

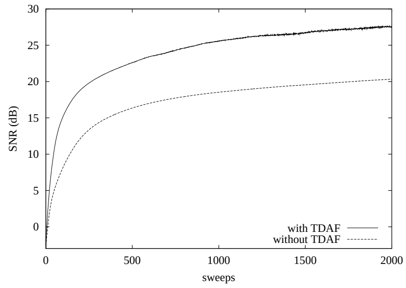


(d) System output signal

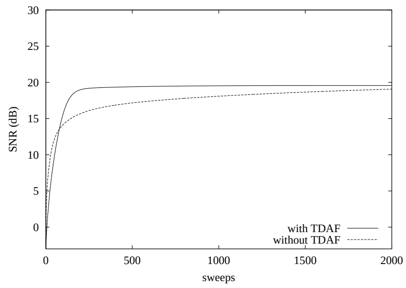
Fig. 4. Signal waveforms after convergence of ICA-based approach to adaptive noise cancelling for the car noise and the simple simulation mixing filter

the unitary transform, we chose the discrete cosine transform (DCT) which is popular in speech processing and had given fast convergence rates for several classes of input signals [10].

Fig. 5 shows learning curves with and without TDAF for each algorithm for the measured filter in a normal office room in the beginning. For the ICA-based algorithm, convergence speed was significantly improved by TDAF with the almost same SNR after convergence. For the LMS algorithm, however, there was no obvious difference with TDAF because relatively large step sizes gave higher SNRs after convergence and had fast convergence speed in the beginning in case that the TDAF method was not used. It is worthy of note that the ICA-based algorithm with TDAF is comparable with the LMS algorithm in the beginning, not to mention SNR after convergence.



(a) ICA-based approach



(b) LMS algorithm

Fig. 5. Comparison of learning curves with and without TDAF methods

5. CONCLUSION

In this paper, a method to adaptive noise cancelling based on ICA was proposed and the ICA-based learning rule was derived. The method was compared with the LMS algorithm through the experiments for several noise signals and mixing filters. By including higher-order statistics, the proposed ICA-based approach gave better performances than the conventional LMS algorithm. Additionally, the TDAF method was applied to derive the normalized ICA-based learning rule and it improved convergence rates.

The ICA-based approach makes the output independent of the reference input as much as possible, whereas the LMS algorithm makes the output uncorrelated to the reference input. Correlation is just the second-order statistics. But, they can have dependency through higher-order statistics, and the ICA-based approach can remove the dependency which involves statistics of all orders. Therefore, the performances of the ICA-based approach were better than those of the LMS algorithm.

6. REFERENCES

[1] Bernard Widrow, Jr. John R. Glover, John M. McCool, John Kaunitz, Charles S. Williams, Robert H. Hearn, James R. Zeidler, Jr. Eugene Dong, and Robert C. Goodlin, "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, December 1975.

[2] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall, Eaglewood Cliffs, 3rd edition, 1996.

[3] Anthony J. Bell and Terrence J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[4] Te-Won Lee, *Independent Component Analysis*, Kluwer Academic Publishers, Boston, 1998.

[5] Te-Won Lee, Anthony J. Bell, and Russell H. Lambert, "Blind separation of delayed and convolved sources," in *Advances in Neural Information Processing Systems 9*, Cambridge, MA, 1997, pp. 758–764, MIT Press.

[6] Kari Torkkola, "Blind separation of convolved sources based on information maximization," *IEEE workshop on Neural Networks for Signal Processing*, 1996.

[7] Scott C. Douglas and Teresa H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," *IEEE Transaction on Signal Processing*, vol. 42, pp. 1335–1351, June 1994.

[8] Scott C. Douglas and Teresa H.-Y. Meng, "Normalized data nonlinearities for lms adaptation," *IEEE Transaction on Signal Processing*, vol. 42, no. 6, pp. 1352–1365, June 1994.

[9] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transaction on Automatic Control*, vol. 12, no. 3, pp. 282–287, June 1967.

[10] Daniel F. Marshall, W. Kenneth Jenkins, and J. J. Murphy, "The use of orthogonal transforms for improving performance of adaptive filters," *IEEE Transaction on Circuits and Systems*, vol. 36, no. 4, pp. 474–484, April 1989.

[11] A. Mahalanobis, Sewoong Song, Sanjit K. Mitra, and Mariane R. Petraglia, "Adaptive fir filters based on structural subband decomposition for system identification problems," *IEEE Transaction on Circuits and Systems-II*, vol. 40, no. 6, pp. 375–381, June 1993.

[12] Shun-ichi Amari, Scott C. Douglas, Andrzej Cichocki, and Howard Hua Yang, "Novel on-line algorithms for blind deconvolution using natural gradient approach," in *Proceedings of the 11th IFAC Symposium on System Identification*, Kitakyushu, Japan, July 1997, pp. 1057–1062.

A. APPENDIX

The transfer function of an adaptive filter with even length K can be expressed as

$$\begin{aligned} H(z) &= \sum_{k=0}^{K-1} h(k) z^{-k} \\ &= (1+z^{-1})G_0(z^2) + (1-z^{-1})G_1(z^2), \end{aligned} \quad \text{(A.1)}$$

where

$$G_0(z) = \frac{1}{2}[\sum_{k=0}^{K/2-1} h(2k)z^{-k} + \sum_{k=0}^{K/2-1} h(2k+1)z^{-k}], \quad \text{(A.2)}$$

$$G_1(z) = \frac{1}{2}[\sum_{k=0}^{K/2-1} h(2k)z^{-k} - \sum_{k=0}^{K/2-1} h(2k+1)z^{-k}]. \quad \text{(A.3)}$$

$G_0(z^2)$ and $G_1(z^2)$ have $K/2$ nonzero coefficients, and the sections $1+z^{-1}$ and $1-z^{-1}$ are called ”interpolators”. If $K/2$ is even, we can further apply this decomposition to the subfilters $G_0(z^2)$ and $G_1(z^2)$ as follows:

$$H(z) = \sum_{m=0}^{M-1} I_m(z) G_m(z^M), \quad \text{(A.4)}$$

where filter length $K = M \cdot P = 2^L \cdot P$. $I_m(z)$ is the interpolator and $G_m(z^M) = \sum_{n=0}^{P-1} g_m(n) z^{-nM}$ is the subfilter of the mth band.

In matrix notation, a general M-band decomposition can be written as

$$\mathbf{h}_K = \sum_{m=0}^{M-1} \mathbf{A}_m \mathbf{g}_m = \mathbf{T}_M^T [\mathbf{g}_0^T \ \mathbf{g}_1^T \ \cdots \ \mathbf{g}_{M-1}^T]^T = \mathbf{T}_M^T \mathbf{w}, \quad \text{(A.5)}$$

where $\mathbf{h}_K$ is the filter vector of length K, and the vector $\mathbf{g}_m = [g_m(0) \ \ g_m(1) \ \ g_m(2) \ \ \cdots \ \ g_m(P-1)]^T$. The $K \times P$ subband transform matrices $\mathbf{A}_m$ are determined by the interpolator structures. "Unnormalized" unitary matrix $\mathbf{T}_M = [\mathbf{A}_0 \ \ \mathbf{A}_1 \ \ \cdots \ \ \mathbf{A}_{M-1}]^T$ is the overall transform matrix applied to the input vector $\mathbf{n}_1(t)$ since

$$\mathbf{h}_K^T \mathbf{n}_1(t) = \mathbf{w}^T \mathbf{T}_M \mathbf{n}_1(t). \quad \text{(A.6)}$$