COMPARISON OF TWO EIGENSTRUCTURE ALGORITHMS FOR LOSSLESS MULTIRATE FILTER OPTIMIZATION

Dong-yan Huang[1] *Phillip A. Regalia*[1] *Maurice Bellanger*[2]

[1]SIM/Institut National des Télécommunications, 9 rue Charles Fourier, 91011 Evry cedex France
huang@int-evry.fr and regalia@int-evry.fr
[2]CNAM / Electronique, 292 rue Saint-Martin, 75634 Paris cedex 03 France
bellang@cnam.fr

ABSTRACT

This paper compares the eigenstructure and modulation algorithms, which are used for two-channel lossless FIR filter optimization. We study the effects of eigenvalue separation of the input covariance matrix and the step size on their convergence behavior. First, we show that the convergence rate of two algorithms increases as the separation of eigenvalues of the covariance matrix increases. The modulation algorithm (MA) converges more rapidly than the eigenstructure one (EA) because of its better eigenvalue separation. Second, the necessary condition for which the two algorithms converge is derived. Simulations are presented which support th analysis.

1. INTRODUCTION

Adaptive algorithms for the two-channel lossless FIR filters (Fig. 1) optimization have recently been developed in various real-time applications such as wavelet analysis and data compression [1, 2, 3, 4]. These update algorithms seek recursively the rotation angles $\{\theta_k\}_{k=0}^{M-1}$ such that the variance of the output $E[y_2^2(n)]$ is "small" or minimized with respect to some predefined criterion. Comparatively few results are available, however, characterizing the convergence properties, which are characterized by two factors: convergence behavior and the steady-state mean-square error (MSE).

In [3] two eigenstructure algorithms—eigenstructure and modulation—are proposed which offer many advantages over stochastic gradient algorithms [1, 2], including computational simplicity and freedom from local minima. However, the nonlinearity and non-gradient nature of these two algorithms render their performance analysis a difficult task. Ordinary differential equation (ODE) analysis of an adaptive algorithm is by now a well established subject and has already been used for assessing the performance analysis of various adaptive algorithms[5, 6, 7]. In this paper, we shall also use this tool to analyze the performance of the above mentioned algorithms.

Specifically, we shall compare adaptive lossless FIR filter banks governed by the two algorithms proposed in [3] in terms of convergence behavior. Theorectical analysis and computer simulations are conducted, leading to useful guidelines for selecting the best algorithm. Performance analyses of the two algorithms in the case of stationary inputs lead to the following conclusion. With appropriate

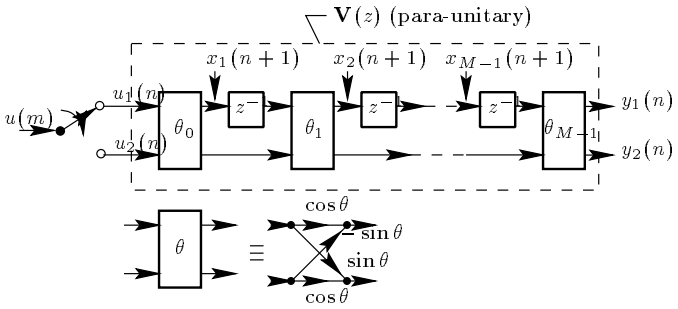

Figure 1. Two-channel lossless filter bank

choice of the adaptation parameter, the convergence rate of both algorithms increases as the spread of eigenvalues of the input covariance matrix increases. The modulation algorithm converges more rapidly than the eigenstructure one due to its better eigenvalue separation. A necessary convergence condition for both algorithms is derived. The paper is organized as follows. Section 2 is concerned with the problem formulation and the assumptions used throughout the paper. Section 3 studies the influence of the eigenvalue spread of the covariance matrix on the algorithms and determines the stepsize μ for which the algorithms converge. Section 4 presents simulation results for validating our theorectical analysis. Conclusion are drawn in Section 5.

2. PROBLEM FORMULATION

Fig. 1 is a block diagram of the two-channel lossless FIR filter bank, which may be described as

$$\begin{bmatrix} \mathbf{x}(n+1) \\ \mathbf{y}(n) \end{bmatrix} = Q_1(\Theta) \left\{ \underbrace{q_2(\Theta) \left\{ \begin{bmatrix} A(\Theta) & B(\Theta) \\ c_2(\Theta) & d_2(\Theta) \end{bmatrix} \right.}_{Q(\Theta)} \begin{bmatrix} c_2(\Theta) & d_2(\Theta) \end{bmatrix} \right. \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{u}(n) \end{bmatrix} \tag{1}$$

where $\mathbf{x}(\cdot) = [x_1(\cdot), \cdots, x_M(\cdot)]^T$ is the state vector, $\mathbf{u}(\cdot) = [u_1(n), u_2(n)]^T$ is the input vector, and $\mathbf{y}(\cdot) = [y_1(\cdot), y_2(\cdot)]^T$ is the ouput vector.

At each iteration n, the eigenstructure algorithm updates $\Theta(\cdot) = [\theta_0(\cdot), \cdots, \theta_{M-1}(\cdot)]^T$, the vector of rotation angles, according to

$$\Theta(n+1) = \Theta(n) - \mu y_2(n) \Gamma(n) \begin{bmatrix} \mathbf{x}(n+1) \\ y_1(n) \end{bmatrix} \tag{2}$$

where μ is a constant step-size; and $\Gamma = \text{diag}\,[\gamma_1, \cdots, \gamma_M]$ with $\gamma_M = 1, \gamma_k(n) = \gamma_{k+1}(n)\cos\theta_k(n)$.

The eigenstructure algorithm does not converge if the odd-indexed terms of the input autocorrelation are negligible compared to the even-indexed terms. For this reason, we suggested an alternative: the modulation algorithm. It is driven exclusively by the odd-indexed terms of the input autocorrelation:

$$\begin{aligned}\Theta(n+1) &= \Theta(n) - \mu\Gamma(n) \\ &\times \left\{ y_2(n) \begin{bmatrix} \mathbf{x}(n+1) \\ y_1(n) \end{bmatrix} - \hat{y}_2(n) \begin{bmatrix} \hat{\mathbf{x}}(n+1) \\ \hat{y}_1(n) \end{bmatrix} \right\} \end{aligned} \tag{3}$$

where $\hat{\mathbf{x}}(n+1), \hat{y}_2(n), \hat{y}_1(n)$ are obtained from a duplicate structure to Fig. 1, but driven by a modulated version of the input:

$$\hat{\mathbf{u}}(m) = \begin{bmatrix} u(2n) & -u(2n-1) & u(2(n-1)) & \cdots \end{bmatrix}$$

The vector $\Theta(n)$ is assumed to be randomly time-varying. Suppose Θ_* is the the solution given by the algorithm (2) or (3); the angle error between $\Theta(n)$ and Θ_* is denoted by

$$\varepsilon(n) \triangleq \Theta(n) - \Theta_* \tag{4}$$

The following assumptions are used throughout the paper:

1 The sequences $\{u(n)\}$ and $\{\varepsilon(n)\}$ are mutually independent;

2 $\{\varepsilon(n)\}$ is a stationary sequence of independent zero mean vectors;

3 The sequence $\{u(n)\}$ is stationary, zero mean, and the covariance matrix $R \triangleq E\left[\mathbf{u}(n)\mathbf{u}^T(n)\right]$ is positive definite.

3. CONVERGENCE BEHAVIOR

The transient phenomena during convergence can be studied using the ODE method as a by-product, as mentioned in the Introduction. Generally, one associates an adaptive algorithm with a corresponding ODE by checking a number of conditions based on which averaging can take place. Then the behavior of the adaptive algorithm can be predicted, in an average sense, by the solutions of the associated ODE. In doing so, one converts a stochastic discrete time process into its averaged version, a deterministic continuous process, which is easier to analyze.

For slow adaptation, the convergence properties of the engienstructure algorithm (2) can be described by the solution $\Theta(t)$ of an associated ODE of the form

$$\begin{aligned}\frac{d\Theta(t)}{dt} &= \underbrace{-\Gamma(\Theta)E\left\{ \begin{bmatrix} \mathbf{x}(n) \\ y_1(n) \end{bmatrix} y_2(n) \mid_\Theta \right\}}_{H(\Theta(n),\mathbf{x}(n))} \\ &= -\Gamma(\Theta)Q_1(\Theta)K(\Theta)q_2^T(\Theta) \end{aligned} \tag{5}$$

where

$$K(\Theta) \triangleq E\left\{ \left(\begin{bmatrix} \mathbf{x}(n) \\ \mathbf{u}(n) \end{bmatrix} \begin{bmatrix} \mathbf{x}^T(n) & \mathbf{u}(n) \end{bmatrix} \mid_\Theta \right\} \tag{6}$$

For a sufficiently small but constant step size μ, $\Theta(n)$ converges in probability to a convergent point Θ_* of (2) if and only if this same value of Θ_* is an attractive stationary point of the differential equations (5).

The modulation algorithm (3) can likewise be studied by using its ODE, as given by

$$\begin{aligned}\frac{d\Theta(t)}{dt} &= \underbrace{-\Gamma(\Theta)E\left\{ y_2(n) \begin{bmatrix} \mathbf{x}(n) \\ y_1(n) \end{bmatrix} - \hat{y}_2(n) \begin{bmatrix} \hat{\mathbf{x}}(n) \\ \hat{y}_1(n) \end{bmatrix} \right\}}_{H(\Theta(n),\mathbf{x}(n))} \\ &= -\Gamma(\Theta)\left\{ Q_1(\Theta) \underbrace{\left[K(\Theta) - \widehat{K}(\Theta)\right]}_{K_0(\Theta)} q_2^T(\theta) \right\} \\ &= -\Gamma(\Theta)Q_1(\Theta)K_0(\Theta)q_2^T(\Theta) \end{aligned} \tag{7}$$

Using $K_g(\Theta)$ to denote $K(\Theta)$ in (6) or $K_0(\Theta)$ in (7), we observe that (5) and (7) assume the common form

$$\frac{d\Theta(t)}{dt} = -\Gamma(\Theta)Q_1(\Theta)K_g(\Theta)q_2^T(\Theta)$$

A stationary point Θ_* of the algorithms, in which the right hand side of the above equation vanishes, is attained if and only is the column vector $K_g(\Theta_*)q_2^T(\Theta_*)$ is orthogonal to the M rows of $Q_1(\Theta_*)$,i.e.,

$$K_g(\Theta_*)q_2^T(\Theta_*) = \lambda(K_g(\Theta_*))q_2^T(\Theta_*) \tag{8}$$

Simulations indicate that $\lambda = \lambda_{min}$ is always obtained. This is why we call (2) and (3) eigenstructure algorithms. Unfortunately, about the convergence behavior, no general result seems to be avaible for either algorithm because of their nonlinearity in the rotation angles. The convergence behavior of these algorithms is determined by two factors: (1) the step-size parameter μ, and (2) the covariance matrix of the input. We shall study the influence of these factors by way of simulation examples.

3.1. Influence of the Eigenvalues Spread of the Covariance Matrix

First, we examine the effect of the eigenvalue separation of the covariance matrix $K(\Theta)$ in (6) and $K_0(\Theta)$ in (7) on the convergence speed of the algorithms (2) and (3), respectively, for a fixed step size μ. As has been shown above, the eigenstructure algorithms (2) and (3) consist in seeking the vector Θ_* such that it satisfies an eigenstructure equation (8). In fact, the decomposition (8) cannot be calculated exactly because of round off error in computing, therefore we shall investigate how the eigenvectors are affected by perturbation. We have the following property:

Property 1 *Assume that the covariance matrix $K_g(\Theta_*) \in C^{(M+1)\times(M+1)}$ has distinct eigenvalues $\lambda_1 > \cdots > \lambda_{M+1}$ and $F \in C^{(M+1)\times(M+1)}$ satisfies $\| F \|_2 = 1$. If the matrix $K_g(\Theta_*)$ is perturbed by a small amount δF, the sensitivity $q_2^T(\Theta_*)$ is in the form:*

$$q_{2,\Theta_*}^T(\delta) \approx q_{2,\Theta_*}^T(0) + \sum_{i=1}^{M} \frac{y_i^H F q_2^T(\Theta_*)}{(\lambda_{M+1}(K_g(\Theta_*)) - \lambda_i)y_i^H x_i} x_i + o(\delta^2) \tag{9}$$

where $q^T_{2,\Theta_*}(0) = q^T_2(\Theta_*)$, $\| x_i(\delta) \|_2 = 1, \| y_i(\delta) \|_2 = 1, i = 1:M$.

Thus the sensitivity of $q^T_{2,\Theta_*}(\delta)$ depends upon eigenvalue sensitivity and the separation of $\lambda_{M+1}(K_g(\Theta_*))$ from the other eigenvalues. For a derivation of this property, see Golub and Van Loan [8, pp344-345]. Suppose that λ is a simple eigenvalue of $K_g(\Theta_*)$ and that x and y satisfy $K_g x = \lambda x$ and $y^H K_g = y^H \lambda$ with $\| x \|_2 = \| y \|_2 = 1$. Using classical results from function theory, it can be shown that in a neighborhood of the origin there exist differentiable $x(\delta)$ and $\lambda(\delta)$ such that

$$(K_g + \delta F)x(\delta) = \lambda(\delta)x(\delta) \quad \| F \|_2 = 1 \quad \| x(\delta) \|_2 = 1$$

where $\lambda(0) = \lambda = \lambda(K_g(\Theta_*))$ and $x(0) = x(\Theta_*)$. As $K_g(\Theta_*)$ has distinct eigenvalues, a continuity argument ensures us that for all δ in some neighborhood of the origin we have

$$\begin{array}{c} (K_g(\Theta_*) + \delta F)x_i(\delta) = \lambda_i(\delta)x_i(\delta) \qquad \| x_i(\delta) \|_2 = 1 \\ y_i(\delta)^H (K_g(\Theta_*) + \delta F) = \lambda_i(\delta) y_i(\delta)^H \qquad \| y_i(\delta) \|_2 = 1 \\ i = 1:(M+1) \end{array}$$

where each $x_i(\delta)$, $y_i(\delta)$ and $\lambda_i(\delta)$ is differentiable. Finally, we have the following result:

$$x_k(\delta) \approx x_k(0) + \sum_{i=1(i\neq k)} \frac{y_i^H F x_k}{(\lambda_k - \lambda_i) y_i^H x_i} x_i + o(\delta^2)$$

As the two algorithms give us $q^T_2(\Theta_*) = x_{M+1}(0)$, we obtain naturally the property above. This property shows that the convergence speed of both eigenstructure algorithms is proportional to the separation of eigenvalues of a covariance matrix. The modulation algorithm in general has better eigenvalue separation than the eigenstructure one. This is confirmed by more rapid convergence in our simulations compared to the latter.

3.2. Step Size Selection

Now we study the effect of the step-size μ on the convergence behavior of the algorithms. For the algorithms to be stable, we must choose μ such that two types of convergence are satisfied:

1 Convergence in the mean, which means that the expectation of the vector $\Theta(n)$ approaches the stationary point Θ_* as the number of iterations n tends to infinity;

2 Convergence in the mean square, which means that the asymtotic values of $E\left[\varepsilon(n)\varepsilon(n)^T\right]$ of the mean-squared error is finite.

To derive the first two moments of the estimate of the vector $\Theta(n)$, it is more convenient to work with the angle error (4). Knowledge of the first two moments of the angle error $\varepsilon(n)$ is important for establishing conditions that ensure convergence of the algorithms. Obtaining analytic expression for $E\left[\varepsilon(n)\varepsilon(n)^T\right]$ is difficult, though. Therefore, we examine only the first moment of $\varepsilon(n)$, i.e., the expectation $E[\varepsilon(n)]$, to determine the necessary condition of convergence to apply with the assumptions of Section 2. The method we shall use is based on local linearization about a stationary point Θ_*.

To obtain simpler expressions, we summarize the algorithms (2) and (3) in the general form

$$\Theta(n+1) = \Theta(n) - \mu H(\Theta(n), \mathbf{x}(n)) \tag{10}$$

Subtracting the stationary point vector Θ_* from both sides of the above equation and using the definition of equation (4), we can rewrite the algorithm (10) in terms of the angle error as follows:

$$\varepsilon(n+1) = \varepsilon(n) - \mu H(\Theta_* + \varepsilon(n), \mathbf{x}(n)) \tag{11}$$

Suppose that $\varepsilon(n)$ is sufficiently small such that the locally linearized model

$$H(\Theta_* + \varepsilon(n), \mathbf{x}(n)) = H(\Theta_*, \mathbf{x}(n)) + \sum_{i=0}^{M-1} \Delta\theta_i \frac{\partial H(\Theta)}{\partial \theta_i} \mid_{\Theta_*}$$

applies. We can then express equation (11) as

$$\varepsilon(n+1) = \varepsilon(n) - \mu \sum_{i=0}^{M-1} \Delta\theta_i \frac{\partial H(\Theta)}{\partial \theta_i} \mid_{\Theta_*} - \mu H(\Theta_*, \mathbf{x}(n))$$

This can be rewritten as

$$\varepsilon(n+1) = (I - \mu \frac{\partial H(\Theta, \mathbf{x}(n))}{\partial \Theta} \mid_{\Theta_*}) \varepsilon(n) - \mu H(\Theta_*, \mathbf{x}(n)) \tag{12}$$

As a consequence of our assumptions, the angle error vector $\varepsilon(n)$ is independent of $\{u(n)\}$. Hence taking the mathematical expectation of both sides of equation (12), we get

$$E(\varepsilon(n+1)) = E(I - \mu \frac{\partial H(\Theta_*, \mathbf{x}(n))}{\partial \Theta}) E(\varepsilon(n)) - \mu E(H(\Theta_*, \mathbf{x}(n))) \tag{13}$$

If we denote

$$A = E\left[\frac{\partial H(\Theta, \mathbf{x}(n))}{\partial \Theta}\right] \mid_{\Theta = \Theta_*}$$

which depends only on the input $\{u(n)\}$ at a stationary point Θ_* and

$$E[H(\Theta_*, \mathbf{x}(n))] = \mathbf{0},$$

we may then simplify Eq. (13) as follows:

$$E[\varepsilon(n+1)] = (I - \mu A) E[\varepsilon(n)]$$

We therefore deduce the necessary condition that the two algorithms converge:

Property 2 *The transient component of* Θ *does not exhibit oscillations, if the step-size parameter* μ *satisfies*

$$0 < \mu < \frac{2}{\sum_{i=0}^{M-1} \lambda_i(A)}$$

where the $\lambda_i, i = 0, \cdots, M-1$ *are the eigenvalues of the matrix* A *and* M *is the number of rotation angles.*

In the following, we shall present simulations for verifying these propreties.

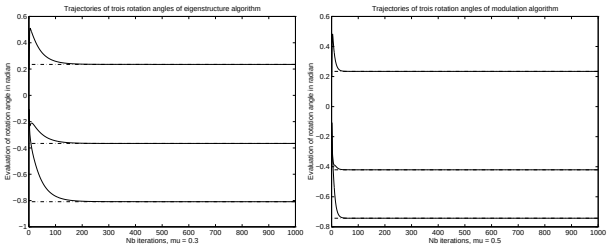

Figure 2. Trajectories of trois rotation angles of EA (left) and MA (right), $\chi(R) = 3$

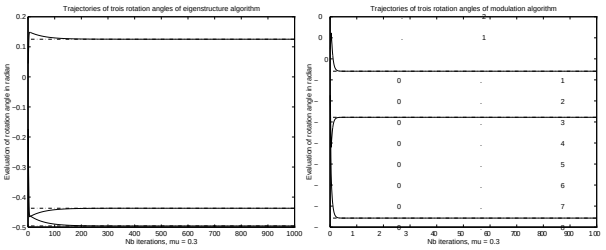

Figure 3. Trajectories of trois rotation angles, $\chi(R) = 10$

4. SIMULATION RESULTS

We present the simulation results for examining the transient behavior of these two algorithms applied to a predictor that on a real-valued autoregressive (AR) process which is described by the second-order difference equation

$$u(n) + a_1\, u(n-1) + a_2 u(n-2) = v(n)$$

where the sample $v(n)$ is drawn from a white-noise process of zero mean and variance σ_v. The transient behavior of two algorithms is evaluated under two scenarios:

- Varying the eigenvalue separation of input $\{u(n)\}$ for a fixed step size μ
- Varying the step-size parameter μ for a fixed eigenvalue separation of input

In the first experiment, the step size μ is fixed at 0.3, and the evaluations are made for two sets of AR parameters following $(a_1 = -0.9750, a_2 = 0.95, \chi(R) = 3), (a_1 = -1.5955, a_2 = 0.95, \chi(R) = 10)$. Figures (2) and (3) show the mean rotation angle evolutions for the two choices of inputs. We observe that the simulations confirm that well separated eigenvalues leads to rapid convergence, while poorly separated eigenvalues lead to slow convergence, and also that the modulation algorithm converges more rapidly than the eigenstructure one.

In the second experiment, the step size μ is varied for the input $\chi(R) = 10$. In particular, we examine the transient behavior of EA for $\mu_{max} = 2/\sum_{i=1}^{M+1}(A) = 0.3906$ and MA for $\mu_{max} = 0.5$. The corresponding results are shown in the left of Fig. 4 for EA and in the right of Fig. 4 for MA. We observe that when μ approaches the value μ_{max}, the transient behavior of two algorithms begins to exhibit oscillations. These simulations support the theorectical analyses.

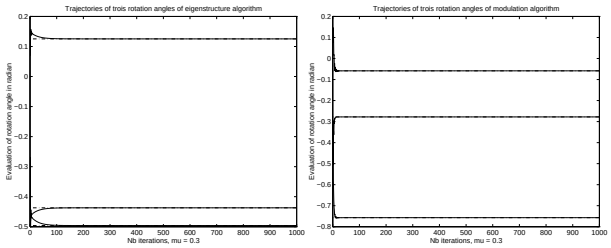

Figure 4. EA (left) for $\mu = 0.4$, MA (right) for $\mu = 0.5$

5. CONCLUSION

In this paper, we have analyzed and compared two algorithms in terms of the convergence behavior. Based on both analysis and simulation we can see that the convergence rate of the two algorithms increases as the separation of eigenvalues of the input covariance matrix increases. The modulation algorithm has a more rapid convergence than the eigenstructure one due to its better eigenvalue separation. The necessary condition for the algorithms to converge is derived. Further study of the asymptotic covariance properties of the filter coefficients is presently underway.

REFERENCES

[1] P. Delsarte, B. Macq, and D. T. M. Slock. Signal adapted multiresolution transform for imag coding. *IEEE Trans. Information Theory*, 38(2):897–904, March 1992.

[2] B. Macq and J. Y. Mertes. Optimization of linear multiresolution transforms for scene adaptive coding. *IEEE Trans. Signal Processing*, 41(12):3568–3572, Dec. 1993.

[3] P.A. Regalia and D. Y. Huang. Attainable error bounds in multirate adaptive lossless fir filters. In *Proc. IEEE int. Conf. Acoust., Speech, Signal Processing*, volume V, pages 1460–1463, Detroit. Michigan, May 1995. IEEE.

[4] D.-Y. Huang and P.A. Regalia. An adaptive projection algorithm for multirate filter bank optimization. *EUSIPCO*, september 1996.

[5] A. Benveniste, M. Metivier, and P. Priouret. *Algorithmes Adaptatifs et Approximation Stochastiques.* Masson, Paris, 1987.

[6] H. Fan and X.Q. Liu. GAL and LSL revisited: New convergence results. *IEEE Trans. Signal Processing*, 41(1):55–66, Jan. 1993.

[7] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 45:59–83, 1995.

[8] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, London, 1984.