- signal-to-reconstruction-noise ratio of the filter bank $10\log_{10}\sum x(n)^2 - 10\log_{10}\sum\{x(n) - \hat{x}(n)\}^2$, where $\hat{x}(n)$ is the reconstructed signal given white noise input.

Our algorithm obtained the solution for design example 1 in 8, 6, 5, 4, 3, 3, 3, and 2 iterations of the inner loop procedure, and it obtained the solution for design example 2 in 9, 5, 5, 3, 3, 2, 2, 2, and 2 iterations of the inner loop procedure. The frequency responses of the two design examples are plotted in Figs. 1 and 2, respectively.

It is well known that the stopband ripple profiles of the prototype filter will affect the overall reconstruction error of the modulated filter bank. Using our proposed algorithm, we are able to control the ripple profiles in the prototype filter's stopband to minimize the aliasing error as well as the overall reconstruction error. From Table I, it can be seen that both examples achieved SNR in excess of 100 dB. In comparison with the modulated filter bank used in MPEG-1 audio codec layers 1 and 2 (which achieves an SNR of about 86 dB using the same test sequence), our design example 2 performs significantly better with an SNR of 108 dB and equiripple stopband attenuation of -124 dB while maintaining the same computational complexity as the modulated filter bank used in MPEG-1.

V. Conclusions

In this correspondence, we presented an efficient algorithm for the design of M-channel cosine-modulated near PR filter banks. The filter bank design is formulated as an unconstrained quadratic programming problem with respect to the prototype linear-phase lowpass FIR filter. Typically, only a few iterations are needed to obtain a solution optimal in the weighted minimax sense. The proposed algorithm provides flexible control of the ripples in the prototype filter's stopband, the overall filter bank transfer function, and the aliasing components. Good M-channel cosine-modulated filter banks with stopband attenuation and signal-to-reconstruction-noise ratio exceeding 100 dB can be easily designed using our algorithm.

References

[1] Y. C. Lim, R. H. Yang, and S. N. Koh, "The design of weighted minmax quadrature mirror filters," *IEEE Trans. Signal Processing*, vol. 41, pp. 1780–1788, May 1993.

[2] D. Estaban and C. Galand, "Application of quadrature mirror filters to split band voice coding scheme," in *Proc. IEEE ICASSP*, 1977, pp. 191–195.

[3] A. N. Akansu and M. S. Kadur, "Subband coding of video with adaptive vector quantization," in *Proc. IEEE ICASSP*, Apr. 1990, pp. 2109–2112.

[4] P. L. Chu, "Quadrature mirror filter design for an arbitrary number of equal bandwidth channels," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 203–218, Feb. 1985.

[5] R. V. Cox, "The design of uniformly and nonuniformly spaced pseudo-quadrature mirror filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1090–1096, Oct. 1986.

[6] R. D. Koilpillai and P. P. Vaidyanathan, "Cosine-modulated FIR filter banks satisfying perfect reconstruction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 40, pp. 770–783, Apr. 1992.

[7] T. Q. Nguyen, "Near-perfect-reconstruction pseudo-QMF banks," *IEEE Trans. Signal Processing*, vol. 42, pp. 65–75, Jan. 1994.

[8] T. Q. Nguyen and R. D. Koilpillai, "The theory and design of arbitrary-length cosine-modulated filter banks and wavelets, satisfying perfect reconstruction," *IEEE Trans. Signal Processing*, vol. 44, pp. 473–483, Mar. 1996.

[9] G. D. T. Schuller and M. J. T. Smith, "New framework for modulated perfect reconstruction filter banks," *IEEE Trans. Signal Processing*, vol. 44, pp. 1941–1954, Aug. 1996.

[10] H. Xu, W. S. Lu, and A. Antoniou, "Efficient iterative design method for cosine-modulated QMF banks," *IEEE Trans. Signal Processing*, vol. 44, pp. 1657–1667, July 1996.

[11] C. D. Creusere and S. K. Mitra, "A simple method for designing high-quality prototype filters for M-band pseudo-QMF banks," *IEEE Trans. Signal Processing*, vol. 43, pp. 1005–1007, Apr. 1995.

[12] S. C. Chan and C. W. Kwok, "Perfect reconstruction modulated filter banks without cosine constraints," in *Proc. IEEE ICASSP*, 1993, pp. 189–192.

Analysis of Momentum Adaptive Filtering Algorithms

Rajesh Sharma, William A. Sethares, and James A. Bucklew

***Abstract*— This correspondence analyzes the momentum LMS algorithm and other momentum algorithms using asymptotic techniques that provide information regarding the almost sure behavior of the parameter estimates and their asymptotic distribution. The analysis does not make any assumptions on the autocorrelation function of the input process.**

I. Introduction

The least mean squares (LMS) algorithm [1], [2] has become one of the most popular adaptive filtering algorithms due to its inherent simplicity and robustness. However, LMS often converges slowly. To remedy this, several modifications of LMS have been proposed over the years. One such modification is the momentum LMS (MLMS) adaptive algorithm first proposed by Proakis [3]. Roy and Shynk [4] demonstrated that MLMS can be viewed as an approximation to the conjugate gradient algorithm. The MLMS is useful in applications where error bursting is a problem. The MLMS recursion is

$$W_{k+1} = W_k + \mu(D_k - W_k^T X_k)X_k + \alpha(W_k - W_{k-1}) \quad (1)$$

where $W_k = [w_k^1, w_k^2, \cdots, w_k^d]^T \in \mathbb{R}^d$ is the parameter estimate at the kth iteration, D_k is a real valued desired response, $X_k = [x_k^1, x_k^2, \cdots, x_k^d]^T \in \mathbb{R}^d$ is the input process, $\alpha \in (-1, 1)$ is the momentum factor, and $\mu > 0$ is the step size.

An analysis of MLMS is carried out in [4], in which convergence of $E[W_k]$ and $E[W_k W_k^T]$ are studied under assumptions that $\{X_k\}$ is a stationary Gaussian process, and W_k is independent of X_k. The theoretical analysis is valid for small α, i.e., $|\alpha| \ll 1$. It is shown via theoretical analysis and simulations that for $\alpha > 0$, MLMS offers an improvement in convergence rate relative to LMS. However, this increase is offset by a corresponding increase in the misadjustment. Interestingly, when $\alpha < 0$, MLMS has a slower convergence rate and lower misadjustment than LMS. It is shown that MLMS becomes unstable when $|\alpha| \to 1$. In addition, the misadjustment is quantified in terms of α and the statistics of the inputs.

The current correspondence extends these results based on the theoretical framework of [5]. The analysis is valid for all $\alpha \in (-1, 1)$, and we show that these values of α result in stable MLMS algorithms. Although the analysis in [4] was for Gaussian processes, our analysis does not make any assumptions on the distribution or correlation of

Manuscript received August 30, 1996; revised October 28, 1997. The associate editor coordinating the review of this paper and approving it for publication was Dr. Akihiko Sugiyama.

R. Sharma is with the Signal and Image Processing Department, ERIM International, Ann Arbor, MI 48113 USA.

W. A. Sethares and A. Bucklew are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA.

Publisher Item Identifier S 1053-587X(98)03261-9.

the input, and no independence assumptions between W_k and X_k are made. In fact, in most applications, W_k and X_k are not independent. Our results hold for a large class of input and desired response processes. Furthermore, the theoretical approach gives almost sure behavior of the parameter estimates $\{W_k\}$ rather than mean squared behavior or mean behavior. From a practical point of view, this is pleasing since the results hold with probability one (w.p. 1), i.e., for essentially any realization of the input process $\{X_k\}$ and the desired response process $\{D_k\}$. Our results confirm the observations made in [4] regarding convergence rates and misadjustment and their relation with α. Furthermore, the asymptotic distribution of W_k is shown to be Gaussian with mean $W^* = E[X_k X_k^T]^{-1} E[X_k D_k]$ and covariance matrix $(\mu/1-\alpha)\Sigma$, where Σ depends on the correlation of the input process and the desired response process.

The main results used in our analysis are stated and proved in [6]. These are extensions of the results found in [5]. Although these results are technical, their usefulness is demonstrated in the later sections where strong statements are made regarding the behavior of the parameter estimates of MLMS. In Section II, the results are applied to MLMS and are extended to adaptive algorithms with multiple momentum terms in Section III. In Section IV, examples and simulations are presented to illustrate the applicability of the results.

II. Momentum LMS

For our analysis of (1), define $W_\mu(t) = W_{[t/\mu]}$ for $t \in [0, \infty)$, where $[t/\mu]$ is the largest integer less than t/μ. The μ dependence of $W_\mu(t)$ is noted by the subscript, and on occasion, it will be necessary to note the α dependence and write $W_{\mu,\alpha}(t)$. The parameter estimates $\{W_k\}$ also depend on μ and α, that is, W_k is sometimes written as $W_k^{\mu,\alpha}$ or as W_k^μ to emphasize the dependence. The Euclidean norm of a vector $x \in \mathbb{R}^d$ is denoted by $|x|$.

To study the convergent behavior of W_k, fix α and assume the following:

H1) $\{(X_k, D_k)\}_{k=0}^\infty$ is a zero mean stationary ergodic random process. Assume $E[|X_0|^2]$ and $E[|D_0|^2]$ are finite.

H2) $R = E[X_0 X_0^T]$ is positive definite.

H3) Assume, for simplicity, the algorithm is initialized with $W_0 = 0$.

If $P = E[X_k D_k]$, then the optimum solution is given by $W^* = R^{-1}P$. R will be positive definite for almost all processes encountered in applications. Let $\{\lambda_m\}_{m=1}^d$ denote the eigenvalues of R. Since R is symmetric, $R = Q\Lambda Q^T$, where Λ is a $d \times d$ diagonal matrix containing the eigenvalues of R, and $Q = [q_1, \cdots, q_d]$ is an orthogonal matrix, i.e., $Q^T Q = I_{d\times d}$.

Using [6, Th. 1], we obtain the following result: w.p. 1., for any $T > 0$ and for fixed $\alpha \in (-1, 1)$

$$\lim_{\mu\to 0} \sup_{t\le T} |W_{\mu,\alpha}(t) - \mathcal{W}_\alpha(t)| = 0 \tag{2}$$

where

$$\mathcal{W}_\alpha(t) = \beta \int_0^t e^{-\beta R(t-\tau)} P \, d\tau = W^* - \delta_\alpha(t) \tag{3}$$

and

$$\delta_\alpha(t) = \sum_{i=1}^d \frac{e^{-\beta\lambda_i t}}{\lambda_i} q_i q_i^T p \tag{4}$$

where

$$\beta = \frac{1}{1-\alpha}. \tag{5}$$

Note that (2) implies that for almost all realizations (w.p. 1.) of the input process and desired response, the parameter estimates $\{W_k\}$ approximately follow the evolution of $\mathcal{W}_\alpha(t)$ over finite time intervals for small stepsize μ. That is, for any $T > 0$

$$\lim_{\mu\to 0} \max_{0\le k\le [T/\mu]} |W_k^{\mu,\alpha} - \mathcal{W}_\alpha(k\mu)| = 0. \tag{6}$$

Furthermore, (6) implies that for small μ, the number of iterations of (1) required for W_k to enter some ball about W^* is proportional to $1/\mu$. It is important to note that although (6) does imply that the parameter estimates will enter a ball about W^*, it does not imply that the parameter estimates will remain in the ball about W^* forever. This is a consequence of the asymptotic distribution of W_k, which is shown to be Gaussian later on in this section.

Let $\overline{h}$: $\mathbb{R}^d \to \mathbb{R}^d$ be defined as $\overline{h}(w) = (P - Rw)$. For small μ, the behavior of MLMS can be determined by studying the deterministic ordinary differential equation (ODE)

$$\dot{w} = \beta \overline{h}(w) \tag{7}$$

for which $\mathcal{W}_\alpha(t)$ is the unique solution corresponding to the initial condition $\mathcal{W}_\alpha(0) = 0$. For instance, since

$$\frac{\partial \beta \overline{h}(W^*)}{\partial w} = -\beta R$$

is negative definite, W^* is locally stable [5], [7]. Furthermore, it follows from (3) that $\lim_{t\to\infty} \mathcal{W}_\alpha(t) = W^*$, and the rate at which $\mathcal{W}_\alpha(t)$ converges to W^* depends on the rate at which $\delta_\alpha(t)$ tends to zero as $t \to \infty$. Therefore, W^* is a globally stable equilibrium point of the ODE (7) since given any initial condition, the solution to the ODE decays to W^*.

Recall that (2) holds for fixed α. It follows from (2) that given $\epsilon > 0$ and $T > 0$, there exist a μ_0 possibly depending on α such that for all $\mu \le \mu_0$

$$\max_{0\le k\le [T/\mu]} |W_k^{\mu,\alpha} - \mathcal{W}_\alpha(k\mu)| < \epsilon. \tag{8}$$

Typically, one is interested in the behavior of (1) for different momentum factors α with the stepsize μ fixed (see Section IV). Applying [6, Corollary 1], it follows that μ_0 does not depend on α if α is restricted to lie in $[-\alpha^*, \alpha^*]$, where $\alpha^* \in (0, 1)$ That is, α^* can be chosen arbitrarily close (but not equal) to 1. The precise statement is w.p. 1, for any $T > 0$

$$\lim_{\mu\to 0} \sup_{\alpha\in[-\alpha^*,\alpha^*]} \sup_{t\le T} |W_{\mu,\alpha}(t) - \mathcal{W}_\alpha(t)| = 0. \tag{9}$$

Therefore, given $\epsilon > 0$ and $T > 0$, there exists a $\mu_0 > 0$ such that for any $\alpha \in [-\alpha^*, \alpha^*]$ and all $\mu \le \mu_0$, (8) holds. Hence, for a small fixed stepsize μ, the deviation of $W_k^{\mu,\alpha}$ from $\mathcal{W}_\alpha(k\mu)$ over finite time intervals will be small for any $\alpha \in [-\alpha^*, \alpha^*]$. Thus, for small μ, the rate at which W_k converges to a ball about W^* for various α depends on the rate at which $\delta_\alpha(t)$ decays to zero. Observe that

$$|\delta_\alpha(t)|^2 = \sum_{i=1}^d \left(\frac{q_i^T p}{\lambda_i^2}\right)^2 e^{(-2\beta\lambda_i t)}.$$

Hence, by increasing $\alpha > 0$, it follows that $\delta_\alpha(t)$ approaches zero more rapidly, whereas if $\alpha < 0$ is decreased, $\delta_\alpha(t)$ approaches zero more slowly. Therefore, α can be thought of as a convergence rate accelerator relative to the convergence rate of LMS. Moreover, if $\alpha \in (0, 1)$ is increased, the parameter estimates of MLMS will enter some ball about W^* earlier than those for LMS for small μ. This fact has been well established by numerous simulations [4]. Likewise, if $\alpha \in (-1, 0)$ is decreased, the parameter estimates in MLMS will enter some ball about W^* later than the parameter estimates in LMS [4].

As shown in [4], this faster convergence for $\alpha \in (0, 1)$ is offset by an increase in the misadjustment of W_k about the optimum weight vector W^*. The misadjustment can be investigated by studying the distributional behavior of $\{W_k\}$. The central limit result (see [6, Th. 2]) shows that the parameter estimates are asymptotically distributed like a Gaussian random vector with mean W^* and covariance matrix that depends on the statistics of $\{X_k\}$ and $\{D_k\}$.

To obtain the central limit result, we need to make an additional assumption (H4 or H4') regarding the process $\{(D_k, X_k)\}$.

H4) Let $Z_k = (D_k - X_k^T W^*)X_k$. Assume

$$\sqrt{\mu} \sum_{k=0}^{[t/\mu]-1} Z_k \Rightarrow B(t)$$

where B is $\mathbb{R}^d$-dimensional Brownian motion on $[0, \infty)$ with mean zero and has covariance matrix

$$R_L = E[Z_0 Z_0^T] + \sum_{n=1}^{\infty} E[Z_0 Z_n^T + Z_n Z_0^T].$$

H4') $D_k = X_k^T W^* + U_{k+1}$, where $\{U_k\}_{k=1}^{\infty}$ is a sequence of real valued i.i.d. random variables independent of $\{X_k\}_{k=0}^{\infty}$.

In [8, ch. 4], it is shown that H4 holds for a large collection of random processes. Essentially, it is required that the sequence $\{Z_k\}$ be weakly dependent (ϕ-mixing with $\Sigma_n \quad \phi^{1/2}(n) < \infty$). Furthermore, if $\{Z_k\}$ is a function of a ϕ-mixing process, then under certain conditions on the function, H4 will hold [8, p. 182]. Assumption H4' is common in the analysis of LMS type algorithms (see Section IV). Under assumption H4', with $\hat{W}_k = W^* - W_k$, (1) can be rewritten as

$$\hat{W}_{k+1} = \hat{W}_k - \mu(\hat{W}_k^T X_k + U_{k+1}) + \alpha(\hat{W}_k - \hat{W}_{k-1}). \qquad (10)$$

We can repeat the previous convergence analysis and obtain results regarding the convergent behavior of the parameter estimate errors $\{\hat{W}_k\}$, which can then be translated into the behavior of the parameter estimates $\{W_k\}$.

Assuming either H4 or H4', the central limit results follow. We carry out the details, assuming H4 holds in order to stay consistent with the previous analysis, where the behavior of the parameter estimates was studied. Note that R_L has the following decomposition: $R_L = \overline{Q\Lambda Q}^T$, where $\overline{Q} = [\overline{q}_1, \overline{q}_2, \cdots, \overline{q}_d]$ is an orthogonal matrix, and $\overline{\Lambda}$ is a diagonal matrix of eigenvalues $\{\overline{\lambda}_i\}_{i=1}^d$ of R_L. Applying [6, Th. 2] and using the properties of Ornstein–Uhlenbeck processes [9], it can be shown that asymptotically, W_k is mean zero and has covariance $\beta\Sigma$, where

$$\Sigma = \sum_{k=1}^{d}\sum_{l=1}^{d}\sum_{m=1}^{d} q_k q_k^T \overline{q}_l \overline{q}_l^T q_m q_m^T \frac{\overline{\lambda}_l}{\lambda_k + \lambda_m}. \qquad (11)$$

That is, for small μ [5], it follows that W_k is asymptotically Gaussian with mean W^* and covariance $\mu\beta\Sigma$. Note that as $\alpha \to 1$, the covariance of W_k becomes unbounded since $\beta \to \infty$. This divergence of the asymptotic variance of W_k reinforces the conclusion in [4] regarding the instability of MLMS when $\alpha \to 1$.

If H4' holds, then for small μ, W_k is asymptotically distributed like a Gaussian random vector with mean zero and covariance

$$\frac{\mu\beta E[U_1^2]}{2} I_{d\times d}. \qquad (12)$$

Example 1: Assume that H1–H3 and H4' hold. Let μ_{LMS} be a small stepsize, and let $\mu_{MLMS} = (1-\alpha)\mu_{\text{LMS}}$, where $\alpha \in (-1, 1)$. From (12), it follows that $W_k^{\mu_{\text{LMS}},0}$ and $W_k^{\mu_{\text{MLMS}},\alpha}$ will have the same covariance matrix. Furthermore, it is easy to see using (3) that their convergence rate to W^* will be identical.

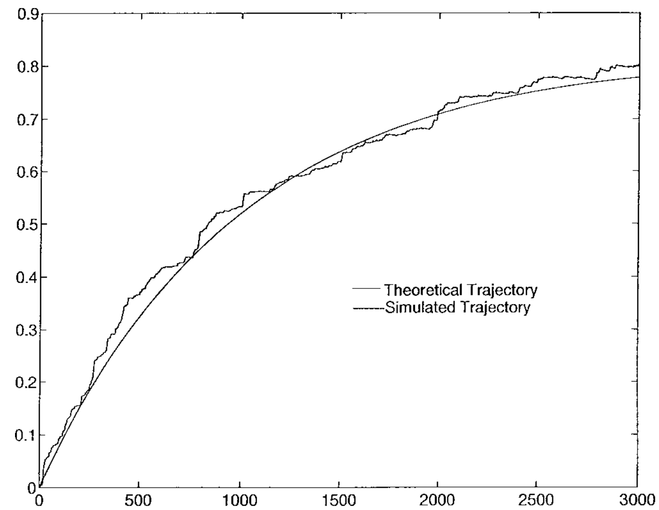

Fig. 1. Theoretical and simulated trajectories for $\alpha = 0$.

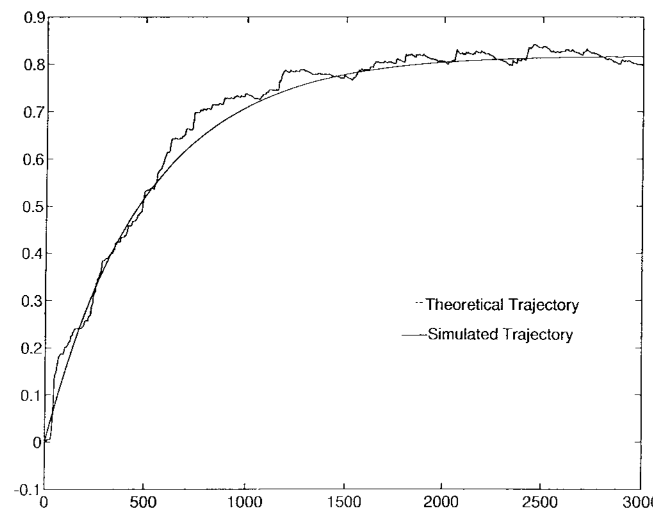

Fig. 2. Theoretical and simulated trajectories for $\alpha = 0.5$.

III. Other Adaptive Algorithms with Momentum Factors

The results have an obvious extension to other algorithms of the form

$$W_{k+1} = \alpha_0 W_k + +\mu H(W_k, Y_k, U_{k+1}) \qquad (13)$$

since the particular form of H was not used in the previous analysis (see [6]). As an example, consider the sgn-sgn adaptive algorithm (see [5, sec. III-A]). With momentum factor α, the update for the parameter estimate error becomes

$$\hat{W}_{k+1} = \hat{W}_k - \mu\text{sgn}(X_k)\text{sgn}(X_k^T \hat{W}_k + U_k) + \alpha(\hat{W}_k - \hat{W}_{k-1}) \qquad (14)$$

where sgn(a) is the signum function. Assume the parameter estimates are initialized to zero, which implies that the parameter estimate error $\hat{W}_0$ is W^*. Assume that $\{U_k\}$ is a zero mean i.i.d symmetric sequence with probability distribution function $\eta(\cdot)$ and bounded, continuous density $f_u(\cdot)$ with $f_u(0) > 0$ and $\alpha \in (-1, 1)$. Assume $\{X_k\}$ is a stationary ergodic sequence (with distribution function F) independent of $\{U_k\}$. Let $W_{\mu,\alpha}(t) = \hat{W}_{[t/\mu]}$, and let $\beta = (1/1-\alpha)$. Referring to the computations in [5], it follows using [6, Th. 1] that

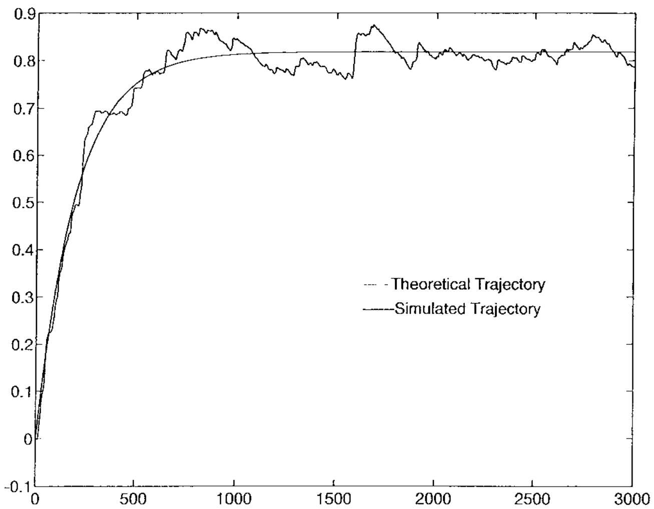

Fig. 3. Theoretical and simulated trajectories for $\alpha = 0.8$.

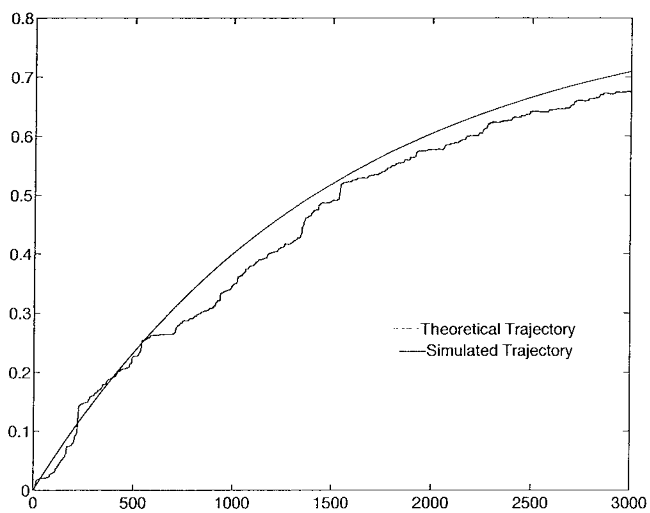

Fig. 4. Theoretical and simulated trajectories for $\alpha = -0.5$.

for any $\epsilon > 0, \alpha^* \in (0, 1)$ and $T > 0$

$$\lim_{\mu \to 0} \mathcal{P}(\sup_{\alpha \in [-\alpha^*, \alpha^*]} \sup_{t \le T} |W_\mu(t) - \mathcal{W}_\alpha(t)| > \epsilon) = 0$$

where $\mathcal{W}_\alpha$ is the unique solution of $\dot{w} = -\beta \hat{h}(w)$ with initial condition $w(0) = W^*$, and

$$\hat{h}(w) = \int \text{sgn}(x)[1 - 2\eta(-x^T w)]\, dF(x).$$

Observe that

$$\frac{\partial \hat{h}}{\partial w}(0) = -2\beta f_u(0) E[\text{sgn}(X_0) X_0^T]. \tag{15}$$

If the eigenvalues of $E[\text{sgn}(X_0)X_0^T]$ have strictly positive real parts, then the sgn-sgn adaptive algorithm with momentum updating will be locally stable for any $\alpha \in (-1, 1)$. Next, to simplify the computations needed to calculate the asymptotic distribution of W_k, assume that the input vector $X_k = [x_k^1, x_k^2, \cdots, x_k^d]^T$ consists of i.i.d. symmetric components such that $E[X_0 X_0^T] = \sigma_x^2 I_{d \times d}$, where $\sigma_x^2 > 0$. Then, as in [5], at equilibrium, $\hat{W}_k$ is asymptotically distributed like a Gaussian random vector with mean 0 and covariance

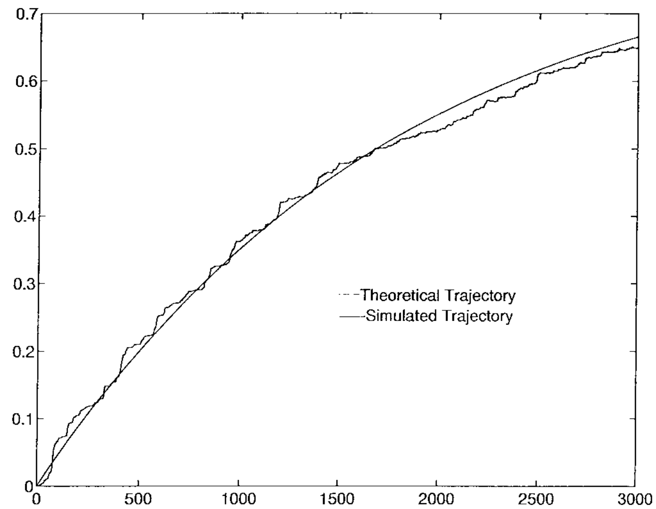

Fig. 5. Theoretical and simulated trajectories for $\alpha = -0.8$.

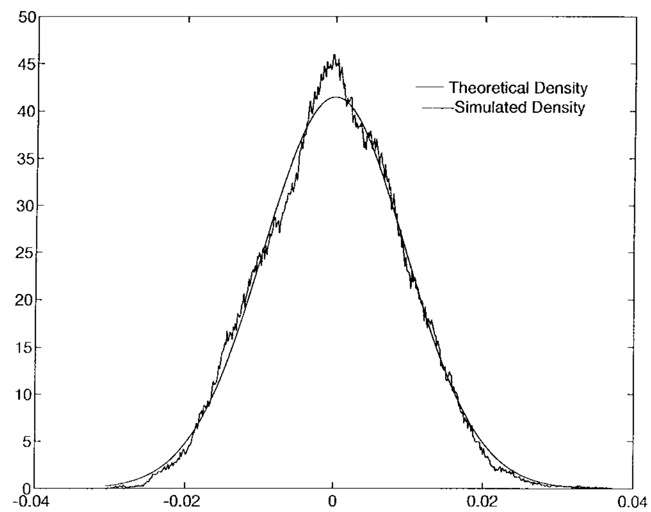

Fig. 6. Theoretical and simulated densities for $\alpha = -0.8$.

$$\frac{\beta \mu}{4 f_u(0) \sigma} I_{d \times d}$$

where $\sigma > 0$, and $E[X_0 \text{sgn}(X_0^T)] = \sigma I_{d \times d}$.

IV. Application to First-Order Linear Prediction

Let $\{X_k\}_{k=-\infty}^{\infty}$ be a first-order autoregressive process

$$X_{k+1} = r X_k + e_{k+1}$$

where $|r| < 1$, $\{e_k\}_{k=-\infty}^{\infty}$ is a sequence of i.i.d. random variables with $E[e_n] = 0$, and $E[e_n^2] = \sigma^2$. Then, note that $\{X_k\}_{k=-\infty}^{\infty}$ is stationary ergodic [10, ch. 6], and for $k \ge 0$

$$E[X_0 X_n] = \frac{\sigma^2 r^n}{(1 - r^2)}.$$

The first-order linear prediction problem is to estimate X_{k+1} as a scalar multiple of X_k such that the mean squared error is minimized.

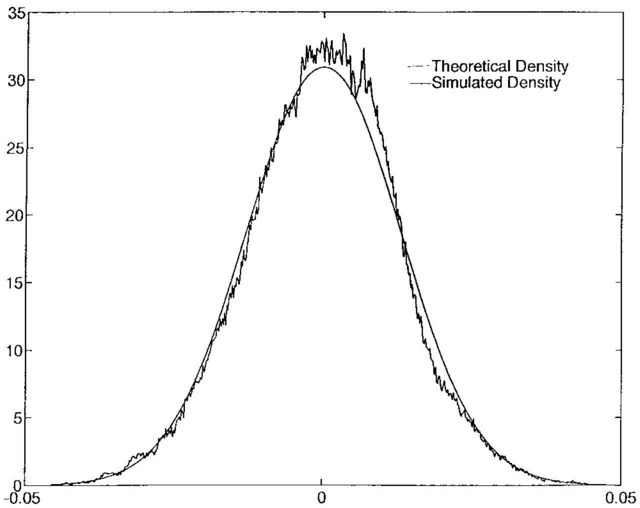

Fig. 7. Theoretical and simulated densities for $\alpha = 0$.

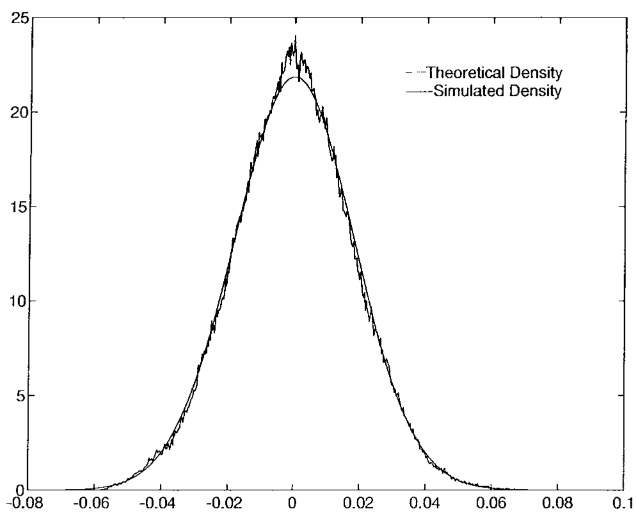

Fig. 8. Theoretical and simulated densities for $\alpha = 0.5$.

The MLMS algorithm for solving this problem is

$$W_{k+1} = W_k + \mu(X_{k+1} - W_k X_k)X_k + \alpha(W_k - W_{k-1}).$$

Note that $R = E[X_0^2] = (\sigma^2/1 - r^2)$, and $p = (\sigma^2 r/1 - r^2)$. Assumptions H1–H3 and H4' are clearly satisfied. Therefore, since $W_k \approx \mathcal{W}_\alpha(k\mu)$, (3) and (12) imply that

$$W_k \approx r - \frac{e^{-\beta R \mu k}}{R} p$$

and W_k is distributed like a Gaussian random variable with mean r and variance $\beta\mu\sigma^2/2$.

Experiment: Figs. 1–5 summarize how well the simulated trajectory matches the theoretical trajectory predicted by (3) for $r = 0.8182$ and $\mu = 0.001$. To determine the steady-state distribution of the parameter estimate errors $(\hat{W}_k = r - W_k)$, 1 million iterations of the algorithm were run, and the last 500 000 were used to compute the simulated error densities. The plots show both the simulated and theoretical densities. In Fig. 6, $\alpha = -0.8$, and in Fig. 7, $\alpha = 0$, whereas $\alpha = 0.5$ in Fig. 8.

V. Conclusions

Asymptotic results were derived for the MLMS and other momentum algorithms. The analysis was based on the results of [5]. The effect of the momentum factor on convergence was studied, and expressions for the asymptotic distribution of the parameter estimates were derived.

References

[1] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[3] J. G. Proakis, "Channel identification for high speed digital communications," *IEEE Trans. Automat. Contr.,* vol. AC-19, pp. 916–922, Dec. 1974.

[4] S. Roy and J. J. Shynk, "Analysis of the momentum LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 38, Dec. 1990.

[5] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Trans. Inform. Theory,* vol. 39, pp. 966–978, May 1993.

[6] R. Sharma, W. A. Sethares, and J. A. Bucklew, "Analysis of momentum adaptive filtering algorithms," Tech Rep. ECE-95-6, Dept. Elect, Comput. Eng., Univ. Wisconsin, Madison, 1995.

[7] J. A. Bucklew and W. A. Sethares, "The covering problem and μ-dependent adaptive algorithms," *IEEE Trans. Signal Processing,* vol. 42, pp. 2616–2627, Oct. 1994.

[8] P. Billingsley, *Convergence of Probability Measures.* New York: Wiley, 1968.

[9] L. Arnold, *Stochastic Differential Equations: Theory and Applications.* New York: Wiley, 1974.

[10] R. Durrett, *Probability: Theory and Examples*. San Francisco, CA: Wadsworth and Brooks/Cole, 1991.

Occam Filters for Stochastic Sources with Application to Digital Images

Balas Natarajan, Konstantinos Konstantinides, and Cormac Herley

***Abstract*—An Occam filter employs lossy data compression to separate signal from noise. Previously, it was shown that Occam filters can filter random noise from deterministic signals. Here, we show that Occam filters can also separate two stochastic sources, depending on their relative compressibility. We also compare the performance of Occam filters and wavelet-based denoising on digital images.**

I. Introduction

A practical problem in real signal processing systems is the treatment of noise-corrupted signals. A commonly used noise-removal approach is the Wiener filter, which is a linear filter that weights the spectrum of the signal by an amount that depends on the noise strength at a given frequency. Recently, there have been a number of alternative nonlinear approaches to noise removal, representative of which is the soft-thresholding approach introduced by Carlson *et al.* [3] and furthered by Donoho and others [5], [6]. In this approach,

Manuscript received February 18, 1997; revised August 7, 1997. The associate editor coordinating the review of this paper and approving it for publication was Dr. Frans M. Coetzee.

The authors are with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail: balas@hpl.hp.com).

Publisher Item Identifier S 1053-587X(98)02010-8.