ADAPTIVE LINEAR FILTERING WITH CONVEX CONSTRAINTS*

P. L. Combettes† *and P. Bondon*‡

†Department of Electrical Engineering, City College and Graduate School, City University of New York, New York, NY 10031, USA.

‡Laboratoire des Signaux et Systèmes - CNRS/Ecole Supérieure d'Electricité Plateau de Moulon, 91192 Gif-sur-Yvette, France.

ABSTRACT

We address the problem of linear mean-square estimation with arbitrary convex constraints for dependent processes. Two algorithms are proposed and their convergence is established. The first algorithm, which is deterministic, covers the case of known correlation structures; the second, which is stochastic and adaptive, covers the case of unknown correlation structures. Since existing algorithms can handle at most one simple constraint this contribution is relevant to signal processing problems in which arbitrary convex inequality constraints are present.

INTRODUCTION

The solution to the unconstrained linear mean-square (LMS) estimation or filtering problem is well known if the joint statistics of the observed and estimated processes are known up to the second order. In many instances these statistics are unknown and an adaptive estimation procedure is employed to approximately optimize performance with incoming data. Adaptive estimation methods based on steepest descent techniques have been applied to a wide range of signal processing problems [1]. One of the most widely used algorithm is the "Widrow LMS" algorithm which employs the gradient method to find the direction of steepest descent and, at each iteration, replaces the true gradient by its instantaneous estimate [11]. The main advantages of the LMS algorithm are its simplicity and the relative low complexity of its implementation. In this unconstrained environment, the convergence of the coefficients of the filter has been analyzed under various statistical assumptions on the underlying processes. Thus, independent processes were considered in [11] and M-dependent processes in [5, 7]; gaussian processes were discussed in [3] and spherically invariant processes in [10].

In practical situations, *a priori* constraints on the optimal filter are often available and they confine it to some feasibility set. The use of the standard LMS algorithm is no longer appropriate and an adaptive constrained optimization procedure is required. In antenna array processing, the LMS algorithm has been studied with a linear [4] and a quadratic [8] equality constraint. In [6], two constraint sets were treated separately: a hypercube and a hypersphere. In the context of image processing, an adaptive algorithm was developed in [12] for the estimation of stack filters with a hypercube constraint. For constant step-sizes, the performance of the stochastic algorithm were studied in [6] in the cases when the input processes are strongly mixing or asymptotically uncorrelated. For decreasing step-sizes, the convergence of the filter coefficients estimates was investigated in [12] under the same statistical assumptions as in [6].

In this paper, we present a constrained LMS algorithm based on the standard gradient projection principle in which each iterate is projected onto the feasibility set to form the update. Assuming that the joint probabilistic attributes of the observed and estimated processes are known up to the second order, we shall obtain a deterministic algorithm that will be shown to converge to the optimal constrained solution. We shall then derive an adaptive constrained LMS algorithm and study its convergence when the input processes are either strongly mixing or asymptotically uncorrelated, and when the step-sizes are either constant or vanishing. In the following developments, the only restriction on the feasibility set is that it be compact and convex. Therefore our results generalize those of the studies mentioned above. The case of multiple convex inequality constraints yielding a bounded feasibility set is also covered by our theoretical analysis.

*The work of the first author was supported by the National Science Foundation under grant MIP-9308609.

NOTATIONS AND ASSUMPTIONS

Let $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ be second-order jointly stationary real random processes on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. We consider the problem of determining the optimal N-tap linear filter $\underline{w}^*$ to estimate the process $(Y_k)_{k\in\mathbb{Z}}$ in terms of the observations of the vector process $(\underline{X}_k)_{k\in\mathbb{Z}} = ([X_k, \cdots, X_{k-N+1}]^t)_{k\in\mathbb{Z}}$ subject to convex constraints on the filter coefficients vector $\underline{w}$. The feasibility set associated with the constraints in $\mathbb{R}^N$ is denoted by S and is supposed to be nonempty, closed, and convex. For $p \in]0, +\infty[$ and $\underline{u} \in \mathbb{R}^N$, we define $\wr\underline{u}\wr_p = [\sum_{n=1}^N |u_n|^p]^{1/p}$ ($\wr\cdot\wr_2$ is the standard euclidean norm and will be denoted by $\|\cdot\|$). We use the notations $R = \mathsf{E}\underline{X}_0\underline{X}_0^t$ and $\underline{r} = \mathsf{E}Y_0\underline{X}_0$; in addition, the eigenvalues of R are $0 < \lambda_1 \leq \cdots \leq \lambda_N$.

DETERMINISTIC ALGORITHM

It is assumed that the correlations R and $\underline{r}$ are known. In the absence of constraints, the optimal filter is given by

$$\underline{u}^* = R^{-1}\underline{r}. \quad (1)$$

Now let

$$\Theta(\underline{u}) = \frac{1}{2}\underline{u}^t R\underline{u} - \underline{u}^t\underline{r}. \quad (2)$$

The constrained mean-square estimation problem can be stated as

$$\min_{\underline{w}\in S} \mathsf{E}|Y_0 - \underline{w}^t\underline{X}_0|^2 = \min_{\underline{w}\in S} \Theta(\underline{w}). \quad (3)$$

Since R is positive definite and S is closed and convex, there exists a unique solution $\underline{w}^*$ to the problem (3). The following proposition characterizes $\underline{w}^*$ and provides an iterative method to compute it. P_S henceforth designates the operator of projection onto S, that is

$$(\forall\underline{w} \in \mathbb{R}^N)\ \ \|P_S(\underline{w}) - \underline{w}\| = \min_{\underline{v}\in S}\|\underline{v} - \underline{w}\|. \quad (4)$$

Proposition 1. [2] Let $0 < \varepsilon < 1$. There is a unique solution to (3) characterized by

$$(\forall\mu \in]0, +\infty[)\ \ \underline{w}^* = P_S((I - \mu R)\underline{w}^* + \mu\underline{r}). \quad (5)$$

In addition, for every $\underline{w}_0 \in \mathbb{R}^N$ and every $(\mu_k)_{k\geq 0} \subset [\varepsilon, 2/\lambda_N - \varepsilon]$, the iterative process

$$(\forall k \in \mathbb{N})\ \ \underline{w}_{k+1} = P_S((I - \mu_k R)\underline{w}_k + \mu_k\underline{r}) \quad (6)$$

converges to $\underline{w}^*$ at a geometric rate. □

In order to obtain an alternative description of the solution to (3), let us equip $\mathbb{R}^N$ with the hilbertian norm

$$\|\underline{u}\|_R = \left[\frac{1}{2}\underline{u}^t R\underline{u}\right]^{1/2}. \quad (7)$$

The following proposition states that the optimal constrained filter is simply the projection of the optimal unconstrained filter onto the feasibility set, relative to the metric of (7).

Proposition 2. [2] $\underline{w}^* = P_S^R(\underline{u}^*)$. □

STOCHASTIC ALGORITHM

When the correlations are unknown, an adaptive procedure is required. To this end, let us define instantaneous cost functions $(\Theta_k)_{k\in\mathbb{Z}}$ by

$$(\forall(k,\underline{u}) \in \mathbb{Z}\times\mathbb{R}^N)\ \ \Theta_k(\underline{u}) = \frac{1}{2}\underline{u}^t\underline{X}_k\underline{X}_k^t\underline{u} - \underline{u}^t Y_k\underline{X}_k. \quad (8)$$

In the unconstrained case, the operation of the adaptive LMS algorithm is based on iterating once the gradient method applied to (8). In the constrained case, we adopt the algorithm ($\underline{w}_0 \in \mathbb{R}^N$, $k \in \mathbb{N}$, $\mu_k \in]0, +\infty[$)

$$\begin{aligned}\underline{w}_{k+1} &= P_S(\underline{w}_k - \mu_k\nabla\Theta_k(\underline{w}_k)) \\ &= P_S((I - \mu_k\underline{X}_k\underline{X}_k^t)\underline{w}_k + \mu_k Y_k\underline{X}_k). \quad (9)\end{aligned}$$

In the following, we analyze the convergence of this algorithm when S is bounded and the input processes are either jointly α-mixing or ρ-mixing. For a constant step-size μ, we give an asymptotic bound on the mean-square deviation of the filter coefficients from their optimal value (Propositions 3) and an asymptotic bound on the additional mean-square signal estimation error (Proposition 4). For vanishing step-sizes $(\mu_k)_{k\geq 0}$, we show that the algorithm (9) converges almost surely and in $\mathcal{L}^p$ (Proposition 5).

Let $\mathcal{F}_l^n$ denote the sub σ-algebra of $\mathcal{F}$ generated by the random variables $(X_k, Y_k)_{l\leq k\leq n}$ and let $\mathcal{L}^2(\mathcal{F}_l^n)$ denote the space of all second-order $\mathcal{F}_l^n$-measurable random variables. We now introduce two notions of asymptotic independence [9].

Definition 1. Let $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ be jointly stationary processes and define for every $i \in \mathbb{N}$

$$\alpha_i = \sup_{(A,B)\in\mathcal{F}_{-\infty}^0\times\mathcal{F}_i^{+\infty}} |\mathsf{P}A\cap B - \mathsf{P}A\mathsf{P}B|. \quad (10)$$

Then $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ are jointly α-mixing if the sequence $(\alpha_i)_{i\geq 0}$ converges to zero.

Definition 2. Let $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ be jointly stationary processes and define for every $i \in \mathbb{N}$

$$\rho_i = \sup_{(U,V)\in\mathcal{L}^2(\mathcal{F}_{-\infty}^0)\times\mathcal{L}^2(\mathcal{F}_i^{+\infty})} \frac{|\mathrm{cov}(U,V)|}{(\mathrm{var}U\,\mathrm{var}V)^{1/2}}. \quad (11)$$

Then $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ are jointly ρ-mixing if the sequence $(\rho_i)_{i\geq 0}$ converges to zero.

The asymptotic bounds on the mean-square deviation and the additional mean-square error, as well as the almost sure and $\mathcal{L}^p$ convergence for jointly α-mixing (respectively ρ-mixing) processes are established under the following Assumption 1 (respectively Assumption 2).

Assumption 1. $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ are jointly α-mixing with mixing sequence $(\alpha_i)_{i\geq 0}$ satisfying:

(i) $\sum_{i\geq 0}\alpha_i^{\delta/(2+\delta)} < +\infty$;

(ii) $\mathsf{E}|X_1Y_k|^{2+\delta} < +\infty,\ 1 \leq |k| \leq N$;

(iii) $\mathsf{E}|X_1X_{k+1}|^{2+\delta} < +\infty,\ 0 \leq |k| \leq N$;

for some $\delta > 0$.

Assumption 2. $(X_k)_{k\in\mathbb{Z}}$ and $(Y_k)_{k\in\mathbb{Z}}$ are jointly ρ-mixing with maximal correlation sequence $(\rho_i)_{i\geq 0}$ satisfying:

(i) $\sum_{i\geq 0}\rho_i < +\infty$;

(ii) $\mathsf{E}|X_1Y_k|^2 < +\infty, 1 \leq |k| \leq N$;

(iii) $\mathsf{E}|X_1X_{k+1}|^2 < +\infty, 0 \leq |k| \leq N$.

Let us remark that, since

$$(\forall i \in \mathbb{N}^*)\ \ \alpha_i \leq \rho_i/4, \quad (12)$$

it is clear that ρ-mixing processes are also α-mixing. Thus the results obtained for α-mixing processes under Assumption 1 hold also for ρ-mixing processes as long as the sequence $(\rho_i^{\delta/(2+\delta)})_{i\geq 0}$ is summable. In Assumption 2, the more stringent condition of summability is imposed on the sequence $(\rho_i)_{i\geq 0}$, but the conditions of existence of moments are weaker than those stated in Assumption 1.

Proposition 3. [2] Take any compact convex set S. Then if either Assumption 1 or Assumption 2 holds, any sequence $(w_k)_{k\geq 0}$ generated by (9) with constant relaxations

$$(\forall k \in \mathbb{N})\ \ \mu_k = \mu \in]0, \min\{2/\lambda_N, 1/(2\lambda_1)\}[\quad (13)$$

satisfies

$$\limsup_{k\to+\infty} \mathsf{E}\|\underline{w}_k - \underline{w}^*\|^2 \leq \mu K_0, \quad (14)$$

where $\underline{w}^*$ is the solution to (3) and K_0 a constant independent from μ. □

Proposition 4. [2] Take any compact convex set S and let

$$\epsilon^* = \mathsf{E}|Y_0 - \underline{w}^{*T}\underline{X}_0|^2 \quad (15)$$

be the minimum attainable mean-square error. Then if either Assumption 1 or Assumption 2 holds, any sequence $(w_k)_{k\geq 0}$ generated by (9) with relaxations as in (13) satisfies

$$(\forall k \in \mathbb{N})\ \ \mathsf{E}|Y_k - \underline{w}_k^T\underline{X}_k|^2 = \epsilon^* + \epsilon_k, \quad (16)$$

where

$$\limsup_{k\to+\infty} \epsilon_k \leq \mu K_1, \quad (17)$$

where K_1 is a constant independent from μ. □

Propositions 3 and 4 generalize the results of [6] in two respects. First of all, S is no longer restricted to be either a hypercube or a hypersphere. Secondly, the relaxation parameter μ is no longer confined to the interval $]0, \min\{2/(\lambda_1+\lambda_N), 1/(2\lambda_1)\}[$. Our next result pertains to convergence in the case of nonconstant relaxations.

Proposition 5. [2] Take any compact convex set S and any sequence $(\mu_k)_{k\geq 0} \subset]0, +\infty[$ such that

(i) $\sum_{k\geq 0}\mu_k = +\infty$;

(ii) $\sum_{k\geq 0}\mu_k^2 < +\infty$.

Then, if either Assumption 1 or Assumption 2 holds, any sequence $(w_k)_{k\geq 0}$ generated by (9) satisfies

$$\lim_{k\to+\infty} \underline{w}_k = \underline{w}^* \ \ \mathsf{P}\text{-almost surely} \quad (18)$$

and

$$(\forall p \in]0, +\infty[)\ \ \lim_{k\to+\infty} \mathsf{E}\wr\underline{w}_k - \underline{w}^*\wr_p^p = 0, \quad (19)$$

where $\underline{w}^*$ is the solution to (3). □

THE CASE OF MULTIPLE CONVEX CONSTRAINTS

Strictly speaking, the above results apply only to problems with one arbitrary convex inequality constraint. However, they can be extended to problems with m compatible convex inequality constraints in a trivial way by simply letting the feasibility set take the form

$$S = \bigcap_{i=1}^{m} S_i, \quad (20)$$

where S_i is the closed and convex set associated with the ith constraint.

This conceptually simple approach may however face serious obstacles in actual applications due to the fact that the operator P_S of projection onto S in (20) is often not known explicitly. In fact, projecting onto an intersection of arbitrary closed and convex sets has until recently remained an open problem in mathematical programming. The key to solving this problem is to proceed via decomposition by observing that, while P_S may be untractable, the operators $(P_i)_{1\leq i\leq m}$ of projection onto the individual sets $(S_i)_{1\leq i\leq m}$ are typically easy to evaluate. Thus, the projection of a vector $\underline{w}$ onto the feasibility set S of (20) can be obtained by the algorithm [2]

$$(\forall n \in \mathbb{N})\ \ \underline{w}^{n+1} = \frac{1}{n+1}\underline{w}^{0} + \frac{n}{m(n+1)}\sum_{i=1}^{m} P_i(\underline{w}^{n}), \tag{21}$$

where $\underline{w}^0 = w$. It is important to note that in this algorithm the m individual projections are averaged at each iteration. They can therefore be computed simultaneously on parallel processors, which makes the cost of the method independent from the number of constraints.

Algorithm (21) can now be combined with the previous ones. Thus, in the case of algorithm (9), we arrive at the following adaptive procedure.

$$
\begin{array}{l}
\text{Fix } \underline{w}_0 \in \mathbb{R}^N \\
\text{for } k = 0, 1, \cdots \\
\quad \left\lfloor \begin{array}{l}
\text{Fix } \mu_k \in]0, +\infty[\\
\underline{w}_k^0 = (I - \mu_k \underline{X}_k \underline{X}_k^t)\underline{w}_k + \mu_k Y_k \underline{X}_k \\
\text{for } n = 0, 1, \cdots \\
\quad \left\lfloor\ \underline{w}_k^{n+1} = \frac{1}{n+1}\underline{w}_k^0 + \frac{n}{m(n+1)}\sum_{i=1}^{m} P_i(\underline{w}_k^n) \right. \\
\underline{w}_{k+1} = \underline{w}_k^{+\infty}
\end{array} \right.
\end{array}
\tag{22}
$$

The choice of the relaxation parameters $(\mu_k)_{k\geq 0}$ can be made as in Propositions 3 and 5. Of course, the theoretically infinite inner iterations on n will be truncated in practice, thereby resulting in an approximate projection onto S. We are currently investigating the effects of these approximate projections on the convergence results presented in this paper. In particular, the case when only a few iterations on n are performed is especially important for on-line applications in which the time allocated for each iteration on k is typically short.

REFERENCES

[1] M. G. Bellanger, *Adaptive Digital Filters and Signal Analysis.* New York: Marcel Dekker, 1987.

[2] P. L. Combettes and P. Bondon, "Constrained Adaptive Filtering for Dependent Processes." Technical report, EE Dept./CUNY and LSS/CNRS, October 1994.

[3] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated gaussian data," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 33, no. 1, pp. 222-230, February 1985.

[4] O. L. Frost, III, "An Algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE,* vol. 60, pp. 926-935, August 1972.

[5] J. K. Kim and L. D. Davisson, "Adaptive linear estimation for stationary M-dependent processes," *IEEE Transactions on Information Theory,* vol. 21, pp. 23-31, January 1975.

[6] A. Krieger and E. Masry, "Constrained adaptive filtering algorithms: Asymptotic convergence properties for dependent data," *IEEE Transactions on Information Theory,* vol. 35, no. 6, pp. 1166-1176, November 1989.

[7] O. Macchi and E. Eweda, "Second-order convergence analysis of stochastic adaptive linear filtering," *IEEE Transactions on Automatic Control,* vol. 28, no. 1, pp. 76-85, January 1983.

[8] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays.* New-York: Wiley, 1980.

[9] M. Rosenblatt, *Stationary Sequences and Random Fields.* Boston, MA: Birkhäuser, 1985.

[10] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Transactions on Signal Processing,* vol. 41, no. 3, pp. 1149-1160, March 1993.

[11] B. Widrow, "Adaptive filters," in *Aspects of Network and Systems,* R. E. Kalman and N. Declaris, Eds. New York: Holt, Rinehart, and Winston, 1971.

[12] L. Yin, J. T. Astola, and Y. A. Neuvo, "Adaptive stack filtering with application to image processing," *IEEE Transactions on Signal Processing,* vol. 41, no. 1, pp. 162-184, January 1993.