ACCELERATION OF ADAPTIVE PARALLEL PROJECTION ALGORITHMS BY PAIRWISE OPTIMAL WEIGHT REALIZATION

Masahiro YUKAWA & Isao YAMADA

Dept. of Communications & Integrated Systems, Tokyo Institute of Technology,
Meguro-ku, Tokyo, 152-8552, Japan.
E-Mail: {masahiro,isao}@comm.ss.titech.ac.jp

ABSTRACT

The adaptive Parallel Subgradient Projection (PSP) technique improves the convergence speed, in noisy environment, of linear-projection-based algorithms (e.g., NLMS and APA), with low computational complexity. The technique utilizes weighted average of the metric projections onto a series of closed half-spaces which contain, with high probability, unknown system to be identified. So far, mainly for simplicity, uniform weighting has been used. However, it is of great interest to develop more strategic weighting for further improvements of convergence, where the weight design should also be with low computational complexity.

This paper presents a novel weighting technique named *Pairwise Optimal WEight Realization PSP (POWER-PSP)*. For each pair of half-spaces, the proposed technique realizes the exact metric projection onto their intersection. Even for q (≥ 3) half-spaces, the technique can approximate, in computationally efficient way, the exact projection onto their intersection by applying the same idea to certain hierarchical structure of half-spaces. Simulation results exemplify that the proposed technique yields drastic improvements of convergence speed and robustness against noise, while keeping linear computational complexity.

1. INTRODUCTION

A simple linear-projection-based adaptive algorithm so-called NLMS (Normalized Least Mean Squares) [1, 2] has been playing a central role, in the presence of possibly non-stationary inputs (e.g., the acoustic echo canceling problem [3]), due to its simplicity and robustness against noise; e.g., [4, 5]. To raise the convergence speed of NLMS, particularly for highly-colored excited inputs like speech signals, APA (Affine Projection Algorithm) was proposed [6, 7], which is based on projection onto certain linear variety (cf. the footnote 1 in the next page). Unfortunately, however, it is observed that APA is seriously unstable under noisy condition (Recently certain inherent sensitivity of APA to noise was investigated through a simple statistical analysis on the set-membership to the projected linear varieties [8]). Moreover, even in low noise conditions, fast convergence can be just achieved by the use of projection onto low dimensional linear variety, which increases computational costs. On the other hand, the RLS (Recursive Least Squares) algorithm is well known to exhibit fast convergence for stationary inputs at the expense of a large increase in computational complexity [2]. For computational efficiency, Fast RLS (FRLS) has been also proposed (see [9] and references therein). However, for non-stationary inputs including speech signals, it is observed that the (N)LMS algorithm exhibits better tracking behaviour than (F)RLS [9, 10].

To alleviate the drawbacks in the aforementioned algorithms, an efficient convex-projection-based scheme named *adaptive Parallel Subgradient Projection (PSP) algorithm* has been proposed [8] (As other convex-projection-based schemes, the so-called *set-membership approaches* (e.g., SM-NLMS[11] / F-SM-NLMS[12]) were also developed. In [13], a unified view for projection-based algorithms, surely including NLMS, APA, set-membership approaches and the adaptive PSP algorithm, has been presented). In [8], stochastic property sets [which are closed convex; see Sec. 2 (1)] are introduced, which is designed to contain, with high probability, the unknown system to be identified, based on the statistics of noise process (see Example 2 in Appendix), which yields robustness against noise. Then, since computing the projection onto such a convex set requires, in general, huge computational complexity, the algorithm alternatively uses projection (well-approximating it and being just computed with linear computational complexity) onto certain closed half-space, which exactly contains the stochastic property set [see Sec. 2 (2)]. Note that the intersection of several number of half-spaces which are constructed as above is expected to be a sufficiently small set containing the unknown system with high probability. Therefore, it would be a natural and meaningful strategy to find a point in the intersection. Based on this strategy, the adaptive PSP algorithm utilizes so-called *parallel projection*; i.e., weighted average of projections onto the half-spaces. So far, as a simple realization of the weighted average, uniform weights have been commonly used. However, in such a case, the update is not necessarily in the best direction toward the intersection. The use of appropriate weights, instead of uniform weights, is expected to improve the convergence speed. Therefore, it is of great interest to develop strategic adaptive weighting techniques, which should also be with simple computation.

In this paper, we propose an efficient adaptive weighting technique named *Pairwise Optimal WEight Realization PSP (POWER-PSP)* to improve convergence speed. Computation of the weights giving the exact direction toward the intersection of a large number of half-spaces, unfortunately, requires extra computational complexity comparable to APA. To avoid it, we introduce certain hierarchical structure of half-spaces. The proposed technique firstly computes, for each pair of half-spaces, the exact metric projection onto their intersection. Even for q (≥ 3) half-spaces, by applying the same idea to the hierarchical structure, the technique can approximate, with moderate increase of computations [see Remark 1 (a)], the exact projection onto their intersection, which leads to a reasonable direction of update. Numerical examples demonstrate that the proposed technique exhibits much faster convergence compared with regularized RLS, regularized APA, and NLMS even in highly noisy condition. Moreover, the increase of q in POWER-PSP accelerates the convergence without any degradation of miss-identification level in steady state.

2. PRELIMINARIES

Throughout the paper, we use the following notations. Let $\mathbb{N}$ and $\mathbb{R}$ denote the sets of all non-negative integers and real numbers, respectively. Define also $\mathbb{N}^* := \mathbb{N}\backslash\{0\}$. Let $k \in \mathbb{N}$ denote the time index. Given $N \in \mathbb{N}^*$, let $\mathscr{H} := \mathbb{R}^N$ be a real Hilbert space equipped with the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^t \boldsymbol{y}$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathscr{H}$, and its induced norm $\|\boldsymbol{x}\| := \left(\boldsymbol{x}^t \boldsymbol{x}\right)^{1/2}$, $\forall \boldsymbol{x} \in \mathscr{H}$, where the superscript t stands for transposition. For any nonempty closed convex set $C \subset \mathscr{H}$, the *projection operator* $P_C : \mathscr{H} \to C$ is defined by $\|\boldsymbol{x} - P_C(\boldsymbol{x})\| = \min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|, \forall \boldsymbol{x} \in \mathscr{H}$. The notation $|S|$ stands for the cardinality of a set S.

In this paper, we address the following adaptive filtering estimation problem, as is also depicted in Fig. 1. Let $(u_k)_{k \in \mathbb{N}} \subset \mathbb{R}$ be the input sequence, and define the sequence of input vectors $(\boldsymbol{u}_k)_{k \in \mathbb{N}} \subset \mathscr{H}$ as $\boldsymbol{u}_k := [u_k, u_{k-1}, \cdots, u_{k-N+1}]^t \in \mathscr{H}$, $\forall k \in \mathbb{N}$. For $r \in \mathbb{N}^*$, let $\boldsymbol{U}_k := [\boldsymbol{u}_k, \boldsymbol{u}_{k-1}, \cdots, \boldsymbol{u}_{k-r+1}] \in \mathbb{R}^{N \times r}$, $\forall k \in \mathbb{N}$. From $(\boldsymbol{U}_k)_{k \in \mathbb{N}}$, the observable data process $(\boldsymbol{d}_k)_{k \in \mathbb{N}} \subset \mathbb{R}^r$ is produced as $\boldsymbol{d}_k := \boldsymbol{U}_k^t \boldsymbol{h}^* + \boldsymbol{n}_k$, $\forall k \in \mathbb{N}$, where $\boldsymbol{h}^* \in \mathscr{H}$ is the (linearly modeled) unknown system to be estimated and $\boldsymbol{n}_k := [n_k, n_{k-1}, \cdots, n_{k-r+1}]^t \in \mathbb{R}^r$, $\forall k \in \mathbb{N}$, is the noise vector. The problem is to estimate $\boldsymbol{h}^*$ by the adaptive filter $\boldsymbol{h} \in \mathscr{H}$ based on the observable data $(\boldsymbol{U}_l)_{l \leq k}$ and $(\boldsymbol{d}_l)_{l \leq k}$. We define the following stochastic property set:

This work was supported in part by JSPS grants-in-Aid (C15500129).

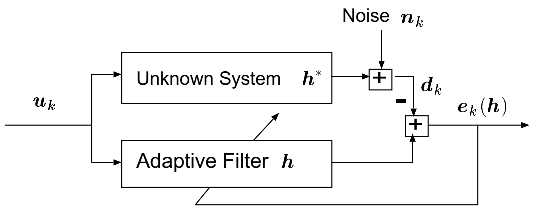

Figure 1: Adaptive Filtering Scheme

$$C_k(\rho) := \left\{ \boldsymbol{h} \in \mathscr{H} : g_k(\boldsymbol{h}) := \|\boldsymbol{e}_k(\boldsymbol{h})\|^2 \leq \rho \right\}, \forall k \in \mathbb{N}, \quad (1)$$

where $\boldsymbol{e}_k : \mathscr{H} \to \mathbb{R}^r$ is the estimation residual function defined by $\boldsymbol{e}_k(\boldsymbol{h}) := \boldsymbol{U}_k^t \boldsymbol{h} - \boldsymbol{d}_k$, $\forall k \in \mathbb{N}$, and $\rho \geq 0$ (NOTE: $C_k(\rho)$ is obviously closed convex). Noticing that $\boldsymbol{e}_k(\boldsymbol{h}^*) = \boldsymbol{n}_k$, the membership probability Prob$[\boldsymbol{h}^* \in C_k(\rho)]$ is equivalent to Prob$[\|\boldsymbol{n}_k\|^2 \leq \rho]$. Therefore, by setting ρ to an appropriate value (cf. Example 2), the membership $\boldsymbol{h}^* \in C_k(\rho)$ is guaranteed with high probability, which leads to stable behavior (strictly saying, monotone approach to the set $C_k(\rho)$ at each time $k \in \mathbb{N}$) of the algorithm [see Remark 1 (c)]. While computation of the direct projection onto the closed convex set $C_k(\rho)$ requires, in general, huge computational complexity, the projection onto the closed half-space[1] $H_k^-(\boldsymbol{h}) := \{\boldsymbol{x} \in \mathscr{H} : \langle \boldsymbol{x} - \boldsymbol{h}, \nabla g_k(\boldsymbol{h})\rangle + g_k(\boldsymbol{h}) \leq 0\} (\supset C_k(\rho))$ has a simple closed-form expression as follows:

$$P_{H_k^-(\boldsymbol{h})}(\boldsymbol{h}) = \begin{cases} \boldsymbol{h} + \frac{-g_k(\boldsymbol{h})}{\|\nabla g_k(\boldsymbol{h})\|^2} \nabla g_k(\boldsymbol{h}), & \text{if } \boldsymbol{h} \notin H_k^-(\boldsymbol{h}), \\ \boldsymbol{h}, & \text{otherwise.} \end{cases} \quad (2)$$

Note that $P_{H_k^-(\boldsymbol{h})}(\boldsymbol{h}) \cong P_{C_k(\rho)}(\boldsymbol{h})$ [8]. Given $q \in \mathbb{N}^*$, define the set of time indices $I_k := \{\iota_k^{(1)}, \iota_k^{(2)}, \cdots, \iota_k^{(q)}\} \subset \mathbb{N}$, $\forall k \in \mathbb{N}$, and the weights $w_\iota^{(k)} > 0$, $\forall \iota \in I_k$ $(\subset \mathbb{N})$, $\forall k \in \mathbb{N}$, to satisfy $\sum_{\iota \in I_k} w_\iota^{(k)} = 1$, $\forall k \in \mathbb{N}$. Then the adaptive PSP algorithm is given as follows.

Algorithm 1 [8] **(Adaptive PSP (Parallel Subgradient Projection) Algorithm)** *Suppose that a sequence of closed convex sets $(C_\iota(\rho))_{\iota \in I_k} \subset \mathscr{H}$ is defined as in (1). Let $\boldsymbol{h}_0 \in \mathscr{H}$ and define a sequence $(\boldsymbol{h}_k)_{k \in \mathbb{N}} \subset \mathscr{H}$ by*

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k + \mu_k \left(\sum_{\iota \in I_k} w_\iota^{(k)} P_{H_\iota^-(\boldsymbol{h}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right), \forall k \in \mathbb{N}, \quad (3)$$

where $\mu_k \in [0, 2\mathscr{M}_k]$ with

$$\mathscr{M}_k := \begin{cases} \dfrac{\sum_{\iota \in I_k} w_\iota^{(k)} \left\| P_{H_\iota^-(\boldsymbol{h}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|^2}{\left\| \sum_{\iota \in I_k} w_\iota^{(k)} P_{H_\iota^-(\boldsymbol{h}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|^2}, & \text{if } \boldsymbol{h}_k \notin \bigcap_{\iota \in I_k} H_\iota^-(\boldsymbol{h}_k), \\ 1, & \text{otherwise.} \end{cases}$$

In the following section, we focus on the design of weights $w_\iota^{(k)}$.

3. MAIN RESULTS

In this section, we firstly present a simple closed form expression to give the projection onto intersection of a pair of closed half-spaces. Based on this fact, we propose the POWER-PSP algorithm, a computationally efficient scheme to achieve a reasonable direction of update for more than two hyperplanes.

3.1 Projection onto Intersection of Pair of Closed Half-Spaces

For a given pair of hyperplanes Π_i $(\subset \mathscr{H})$, $i = 1, 2$, let Π_i^- $(\subset \mathscr{H})$, $i = 1, 2$, be the closed half-spaces whose boundary hyperplanes are Π_i, $i = 1, 2$, respectively. Then, the following proposition holds.

[1]Given $\boldsymbol{v} \in \mathscr{H}$ and a closed subspace $M \subset \mathscr{H}$, the translation of M by $\boldsymbol{v}$ defines the *linear variety* $V := \boldsymbol{v} + M := \{\boldsymbol{v} + \boldsymbol{m} : \boldsymbol{m} \in M\}$. If $\dim(M^\perp) = 1$, V is called *hyperplane*, which can be expressed as $(V =) \Pi = \{\boldsymbol{x} \in \mathscr{H} : \langle \boldsymbol{a}, \boldsymbol{x} \rangle = c$, for some $(0 \neq) \boldsymbol{a} \in \mathscr{H}$ and $c \in \mathbb{R}\}$. With the hyperplane, we can define a *closed half-space* $\Pi^- := \{\boldsymbol{x} \in \mathscr{H} : \langle \boldsymbol{a}, \boldsymbol{x} \rangle \leq c\}$.

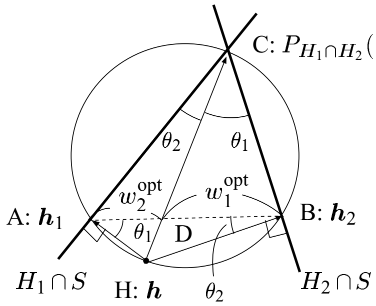

Figure 2: Geometric interpretation of Theorem 1 in the cases of $\cos\theta_1$, $\cos\theta_2 > 0$. $S := \{\boldsymbol{s} = \boldsymbol{h} + \alpha(\boldsymbol{h}_1 - \boldsymbol{h}) + \beta(\boldsymbol{h}_2 - \boldsymbol{h}) \in \mathscr{H} : \alpha, \beta \in \mathbb{R}\}$.We can easily see that $P_{H_1 \cap H_2}(\boldsymbol{h}) \in S$.

Proposition 1 *For any $\boldsymbol{x} \in \mathscr{H}$,*

$$P_{\Pi_1^- \cap \Pi_2^-}(\boldsymbol{x}) = \begin{cases} P_{\Pi_1^-}(\boldsymbol{x}), & \textit{if } P_{\Pi_1^-}(\boldsymbol{x}) \in \Pi_2^- \\ \textit{(the roles of } 1 \textit{ and } 2 \textit{ are interchangeable)}, & \\ P_{\Pi_1 \cap \Pi_2}(\boldsymbol{x}), & \textit{if } P_{\Pi_2^-}(\boldsymbol{x}) \notin \Pi_1^-, P_{\Pi_1^-}(\boldsymbol{x}) \notin \Pi_2^-. \end{cases} \quad (4)$$

The next theorem gives a useful way to compute $P_{\Pi_1 \cap \Pi_2}(\boldsymbol{x})$ from $\boldsymbol{x}$, $P_{\Pi_1}(\boldsymbol{x})$ and $P_{\Pi_2}(\boldsymbol{x})$.

Theorem 1 *For given different vectors $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathscr{H}$ and $\boldsymbol{h} \in \mathscr{H}$ (i.e., $\boldsymbol{h} \notin \{\boldsymbol{h}_1, \boldsymbol{h}_2\}$, $\boldsymbol{h}_1 \neq \boldsymbol{h}_2$), suppose that a pair of hyperplanes, $\forall i = 1, 2$, (see Fig.2)*

$$H_i := H_i(\boldsymbol{h}, \boldsymbol{h}_i) := \{\boldsymbol{x} \in \mathscr{H} : \langle \boldsymbol{h}_i - \boldsymbol{h}, \boldsymbol{x} - \boldsymbol{h}_i \rangle = 0\} \subset \mathscr{H}, (5)$$

satisfy $H_1 \cap H_2 \neq \emptyset$. Then, we have

$$P_{H_1 \cap H_2}(\boldsymbol{h}) = \boldsymbol{h} + \mathscr{M} \left\{ w_1^{\text{opt}} \boldsymbol{h}_1 + w_2^{\text{opt}} \boldsymbol{h}_2 - \boldsymbol{h} \right\}, \quad (6)$$

where

$$\mathscr{M} := \frac{w_1^{\text{opt}} \|\boldsymbol{h}_1 - \boldsymbol{h}\|^2 + w_2^{\text{opt}} \|\boldsymbol{h}_2 - \boldsymbol{h}\|^2}{\left\| w_1^{\text{opt}} \boldsymbol{h}_1 + w_2^{\text{opt}} \boldsymbol{h}_2 - \boldsymbol{h} \right\|^2}, \quad (7)$$

$$w_i^{\text{opt}} := \frac{\cos\theta_i \sin\theta_i}{\cos\theta_1 \sin\theta_1 + \cos\theta_2 \sin\theta_2}, \quad \forall i = 1, 2, \quad (8)$$

$$\cos\theta_1 := \frac{\langle \boldsymbol{h}_1 - \boldsymbol{h}, \boldsymbol{h}_1 - \boldsymbol{h}_2 \rangle}{\|\boldsymbol{h}_1 - \boldsymbol{h}\| \|\boldsymbol{h}_1 - \boldsymbol{h}_2\|}, \quad \cos\theta_2 := \frac{\langle \boldsymbol{h}_2 - \boldsymbol{h}, \boldsymbol{h}_2 - \boldsymbol{h}_1 \rangle}{\|\boldsymbol{h}_2 - \boldsymbol{h}\| \|\boldsymbol{h}_2 - \boldsymbol{h}_1\|}.$$

NOTE: The denominators in (7) and (8) are non-zero due to $H_1 \cap H_2 \neq \emptyset$. By (5), apparently $P_{H_i}(\boldsymbol{h}) = \boldsymbol{h}_i$ $(i = 1, 2)$.

(Sketch of Proof):

(I) If $\cos\theta_i > 0, \forall i = 1, 2$: First of all, by $\angle HAC = \angle HBC = \pi/2$ [rad] and the fact of elementary geometry on inscribed angle, we get $\angle HCB = \angle HAB$ and $\angle HCA = \angle HBA$ (see Fig. 2). Then, noting the areas of $\triangle$HDB and $\triangle$HDA in Fig. 2,

$$\begin{aligned} w_1^{\text{opt}} : w_2^{\text{opt}} &= HB\cos\theta_1 : HA\cos\theta_2 \\ &= \cos\theta_1 \sin\theta_1 : \cos\theta_2 \sin\theta_2, \end{aligned} \quad (9)$$

which verifies that (8) gives the weights that realize the direction toward the intersection $H_1 \cap H_2$.

(II) If $\cos\theta_1 < 0$ or $\cos\theta_2 < 0$: We can prove in a similar way to (I). □

By (4) and (6), $P_{H_1^- \cap H_2^-}(\boldsymbol{h})$ can be easily computed, where H_i^- $(i = 1, 2)$ is a closed half-space with the boundary hyperplane H_i.

3.2 Pairwise Optimal WEight Realization

Followed by the discussion in Secs. 1 and 2, Prob$[\boldsymbol{h}^* \in \bigcap H_\iota^-(\boldsymbol{h})]$ is sufficiently high, hence we can expect that $P_{\bigcap H_\iota^-(\boldsymbol{h})}(\boldsymbol{h})$ is a reasonable candidate as an approximation of $\boldsymbol{h}^*$. Unfortunately, extension of Proposition 1 (Theorem 1) to the case of more than 2 half-spaces (hyperplanes) is not a straightforward task, and $P_{\bigcap H_\iota^-(\boldsymbol{h})}(\boldsymbol{h})$ requires comptational load like APA. To approximate $P_{\bigcap H_\iota^-(\boldsymbol{h})}(\boldsymbol{h})$ as much as possible while saving computational costs, we present a weighting technique below, which applies Proposition 1 and Theorem 1 pairwise to a series of closed half-spaces.

First of all, for given $q \in \mathbb{N}^*$, define the control sequence $I_k^{(0)} := \{\iota_k^{(1)}, \iota_k^{(2)}, \cdots, \iota_k^{(q)}\} \subset \mathbb{N}$, $\forall k \in \mathbb{N}$, where $\iota_k^{(i)}$'s $(i = 1, 2, \cdots, q)$ represent the time indices used in the 0-th stage at time k in the POWER-PSP (see below). Then, $\forall k \in \mathbb{N}$, $\forall m = 1, 2, \cdots, M$ $(M \in \mathbb{N}^*)$, define inductively $I_k^{(m)} \subset \{(\iota_1, \iota_2) : \iota_1, \iota_2 \in I_k^{(m-1)}, \iota_1 \neq \iota_2\}$ such as $1 = |I_k^{(M)}| \leq |I_k^{(M-1)}| \leq \cdots \leq |I_k^{(1)}| \leq |I_k^{(0)}| = q$. The POWER-PSP algorithm is given as follows.

Algorithm 2 (POWER-PSP Algorithm) *Suppose that a sequence of closed convex sets $(C_k(\rho))_{k \in \mathbb{N}} \subset \mathscr{H}$ is defined as in* (1). *Let $\boldsymbol{h}_0 \in \mathscr{H}$ be an arbitrary chosen initial vector and $(\lambda_k^{(m)})_{k \in \mathbb{N}} \subset [0, 2]$ ($\forall m = 0, 1, \cdots, M$) the sequence of step size satisfying $\prod_{m=0}^{M} \lambda_k^{(m)} \leq 2$. Then, define a sequence $(\boldsymbol{h}_k)_{k \in \mathbb{N}} \subset \mathscr{H}$ by the following three steps:*

Step 1: *(The* 0*-th Stage)*

$$\boldsymbol{h}_{k,\iota}^{(0)} := \boldsymbol{h}_k + \lambda_k^{(0)} \left(P_{H_\iota^-(\boldsymbol{h}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right), \quad \forall \iota \in I_k^{(0)}, \tag{10}$$

where $P_{H_\iota^-(\boldsymbol{h}_k)}(\boldsymbol{h}_k)$, $\forall \iota \in I_k^{(0)}$, is computed by (2).

Step 2: *(The* 1*-st to M-th Stage)*

$\boldsymbol{h}_{k,\iota}^{(m)} \in \mathscr{H}$ ($\forall k \in \mathbb{N}, \forall \iota \in I_k^{(m)}, \forall m = 1, 2, \cdots, M$) is produced according to the hierarchical procedure as below:

for $m := 1$ **to** M **do**

$$\boldsymbol{h}_{k,\iota}^{(m)} := \boldsymbol{h}_k + \lambda_k^{(m)} \mathscr{M}_{k,\iota}^{(m)} \left\{ w_{k,\iota,1}^{(m)} \boldsymbol{h}_{k,\iota_1}^{(m-1)} + w_{k,\iota,2}^{(m)} \boldsymbol{h}_{k,\iota_2}^{(m-1)} - \boldsymbol{h}_k \right\},$$

$$\forall k \in \mathbb{N}, \quad \forall \iota = (\iota_1, \iota_2) \in I_k^{(m)}, \tag{11}$$

where the weights are defined, $\forall k \in \mathbb{N}$, $\forall \iota = (\iota_1, \iota_2) \in I_k^{(m)}$, $\forall (i, j) \in \{(1, 2), (2, 1)\}$, as

$$w_{k,\iota,i}^{(m)} := \begin{cases} 0, \ \text{if } cs_{k,\iota,1}^{(m)} + cs_{k,\iota,2}^{(m)} = 0 \text{ or } c_{k,\iota,i}^{(m)} < 0, \\ 1, \quad \text{if } c_{k,\iota,j}^{(m)} < 0, \\ cs_{k,\iota,i}^{(m)} / (cs_{k,\iota,1}^{(m)} + cs_{k,\iota,2}^{(m)}), \quad \text{otherwise}, \end{cases} \tag{12}$$

$$cs_{k,\iota,i}^{(m)} := c_{k,\iota,i}^{(m)} \sqrt{1 - c^2{}_{k,\iota,i}^{(m)}},$$

$$c_{k,\iota,i}^{(m)} := \begin{cases} 0, \quad \text{if } \boldsymbol{h}_{k,\iota_i}^{(m-1)} - \boldsymbol{h}_k = 0, \\ 1/2, \quad \text{if } \boldsymbol{h}_{k,\iota_1}^{(m-1)} - \boldsymbol{h}_{k,\iota_2}^{(m-1)} = 0, \\ \dfrac{\left\langle \boldsymbol{h}_{k,\iota_i}^{(m-1)} - \boldsymbol{h}_k, \boldsymbol{h}_{k,\iota_i}^{(m-1)} - \boldsymbol{h}_{k,\iota_j}^{(m-1)} \right\rangle}{\left\| \boldsymbol{h}_{k,\iota_i}^{(m-1)} - \boldsymbol{h}_k \right\| \left\| \boldsymbol{h}_{k,\iota_i}^{(m-1)} - \boldsymbol{h}_{k,\iota_j}^{(m-1)} \right\|}, \quad \text{otherwise}, \end{cases}$$

and $\mathscr{M}_{k,\iota}^{(m)}$ is defined, $\forall k \in \mathbb{N}$, $\forall \iota = (\iota_1, \iota_2) \in I_k^{(m)}$, as

$$\mathscr{M}_{k,\iota}^{(m)} := \begin{cases} 1, \quad \text{if } w_{k,\iota,1}^{(m)} \boldsymbol{h}_{k,\iota_1}^{(m-1)} + w_{k,\iota,2}^{(m)} \boldsymbol{h}_{k,\iota_2}^{(m-1)} - \boldsymbol{h}_k = 0, \\ \dfrac{w_{k,\iota,1}^{(m)} \left\| \boldsymbol{h}_{k,\iota_1}^{(m-1)} - \boldsymbol{h}_k \right\|^2 + w_{k,\iota,2}^{(m)} \left\| \boldsymbol{h}_{k,\iota_2}^{(m-1)} - \boldsymbol{h}_k \right\|^2}{\left\| w_{k,\iota,1}^{(m)} \boldsymbol{h}_{k,\iota_1}^{(m-1)} + w_{k,\iota,2}^{(m)} \boldsymbol{h}_{k,\iota_2}^{(m-1)} - \boldsymbol{h}_k \right\|^2}, \\ \text{otherwise}. \end{cases}$$

end;

Step 3: *(Update)*

$$\boldsymbol{h}_{k+1} := \boldsymbol{h}_{k,\iota}^{(M)}, \quad \iota \in I_k^{(M)}. \tag{13}$$

Note that, in Step 3, $|I_k^{(M)}| = 1$. In addition, in Step 2, in the m-th stage, $\boldsymbol{h}_{k,\iota}^{(m)}$ $(\iota = (\iota_1, \iota_2) \in I_k^{(m)})$ is obtained just from $\boldsymbol{h}_k$, $\boldsymbol{h}_{k,\iota_1}^{(m-1)}$ and $\boldsymbol{h}_{k,\iota_2}^{(m-1)}$ $(\iota_1, \iota_2 \in I_k^{(m-1)})$ in the same way as $P_{H_1 \cap H_2}(\boldsymbol{h})$ is obtained just from $\boldsymbol{h}$, $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ in Theorem 1. The following proposition shows that Algorithm 2 is a special example of Algorithm 1, the adaptive PSP algorithm.

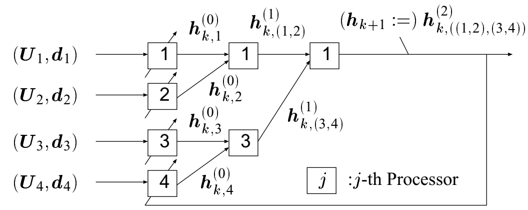

Figure 3: An efficient example of implementation of hierarchical structure in Algorithm 2 $(q = 4)$. For brevity, the subscripts $\iota_k^{(l)}$ $(l = 1, 2, 3, 4)$ are abbreviated to l.

Proposition 2 (POWER-PSP) *The vectors $\boldsymbol{h}_{k+1} \in \mathscr{H}$, $\forall k \in \mathbb{N}$, [in* (13)*] generated by Algorithm 2 can be expressed in the form of* (3) *by rearranging appropriately. In this case, the corresponding weight $w_\iota^{(k)}$ is determined automatically by $w_{k,\iota,i}^{(m)}$'s and $\mathscr{M}_{k,\iota}^{(m)}$'s shown in step 2, and satisfy $w_\iota^{(k)} > 0$ and $\sum_{\iota \in I_k} w_\iota^{(k)} = 1$. Moreover, the corresponding step size μ_k [in* (3)*] is given by $\mu_k = \mathscr{M}_k \prod_{m=0}^{M} \lambda_k^{(m)}$, $\forall k \in \mathbb{N}$.*

Next, a systematic design of hierarchical structure (design of $I_k^{(m)}$, $\forall k \in \mathbb{N}$, $\forall m = 1, \cdots, M$) is given below.

Example 1 (Binary-Tree Like Implementation) *For given $I_k^{(0)}$ ($\forall k \in \mathbb{N}$), we suggest a systematic design of $I_k^{(m)}$'s ($m = 1, \cdots, M$), $\forall k \in \mathbb{N}$, as shown in Fig. 3, which we call binary-tree like implementation of Algorithm 2. That is, for example, for $I_k^{(0)} := \{\iota_k^{(1)}, \iota_k^{(2)}, \iota_k^{(3)}, \iota_k^{(4)}\}$, we define $I_k^{(1)} := \{(\iota_k^{(1)}, \iota_k^{(2)}), (\iota_k^{(3)}, \iota_k^{(4)})\}$ and $I_k^{(2)} := \{((\iota_k^{(1)}, \iota_k^{(2)}), (\iota_k^{(3)}, \iota_k^{(4)}))\}$.*

Note that in Example 1, q $(= |I_k^{(0)}|)$ is not necessarily in the form of 2^M (If $q \neq 2^M$, we can dispense with appropriate nodes). We give some important remarks below.

Remark 1

(a) *A simple inspection of* (10) *and* (11) *implies that the proposed technique is (i) free from the computational load of solving a system of linear equations and (ii) well-suited for q concurrent processors (since $\left|I_k^{(m)}\right| \leq \left|I_k^{(0)}\right| = q$, $\forall m = 1, \cdots, M$). With such concurrent processors, the number of multiplications imposed on each processor, at each iteration, is $(M + 1) \times O(N)$.*

(b) *The number of stages in binary-tree like implementation is no more than $M + 1 = \log_2 q + 1$ (see Example 1).*

(c) *Proposition 2 guarantees that Algorithm 2 has the property of monotone approach; i.e., $\|\boldsymbol{h}_{k+1} - \boldsymbol{h}^*\| \leq \|\boldsymbol{h}_k - \boldsymbol{h}^*\|$ if $\boldsymbol{h}^* \in \bigcap_{\iota \in I_k^{(0)}} H_\iota^-(\boldsymbol{h}_k)$ (For detail, see* [8, Proposition 1]*).*

4. NUMERICAL EXAMPLES

To verify the efficacy of the proposed algorithm (POWER-PSP), we compare POWER-PSP with NLMS, APA and RLS for estimating $\boldsymbol{h}^* \in \mathscr{H} := \mathbb{R}^{256}$. In these simulation tests, we apply the algorithms to the acoustic echo canceling problem[2] (e.g., [3]) by using the USASI[3] signal, which is known as speech-like wide sense stationary process, as the input $(u_k)_{k \in \mathbb{N}}$ by following the way in [8]. We use white noise, as $(n_k)_{k \in \mathbb{N}}$, with SNR$:= 10 \log_{10} \left(E\left\{z_k^2\right\} / E\left\{n_k^2\right\} \right) = 10$dB $(z_k := \langle \boldsymbol{u}_k, \boldsymbol{h}^* \rangle)$, which is a usual condition in acoustic echo canceling problem. We evaluate the normalized system mismatch defined as System Mismatch$(k) := 10 \log_{10} \|\boldsymbol{h}^* - \boldsymbol{h}_k\|^2 / \|\boldsymbol{h}^*\|^2$ [dB], $\forall k \in \mathbb{N}$.

[2]Multi-channel acoustic echo cancellation (e.g., [14]) is also an interesting application. However, to examine purely the effect of the proposed weighting technique, we focus on the simplest single channel case.

[3]The USASI (USA Standards Institute) generation routine is characterized as an ARMA model, which can be found in http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/txt/usasi.txt.

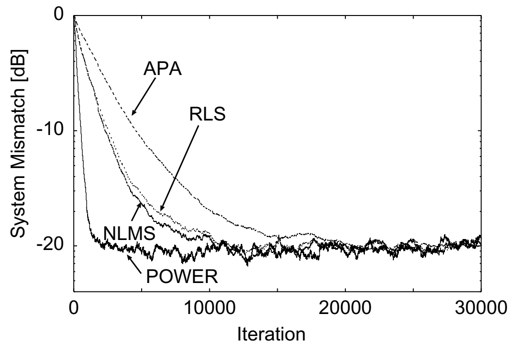

Figure 4: POWER-PSP (proposed) with $q = 64$ versus NLMS, APA and RLS under SNR=10dB. Both APA and RLS are regularized.

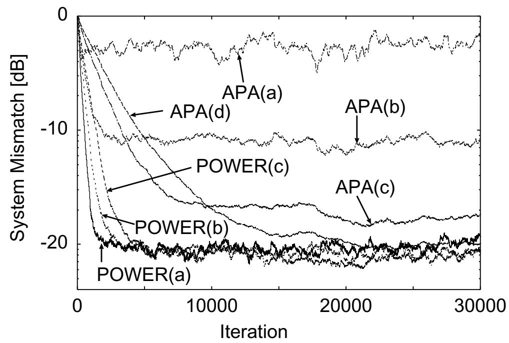

Figure 5: POWER-PSP with (a) $q = 64$, (b) $q = 32$, (c) $q = 16$ versus regularized APA with (a) $r = 64$, (b) $r = 10$, (c) $r = 3$, (d) $r = 2$ under SNR=10dB.

In all algorithms used in the first experiment shown in Fig. 4, we employ the parameters, with the best performance among our experiments, that achieve the level of around -20dB in steady state. For NLMS, the step size is set to 0.15. APA (RLS) has a version called FAP (FRLS), which is computationally efficient but is similar or somewhat inferior to the original exact algorithm in convergence behavior (see [9] and references therein). Since we concentrate on the performance of the algorithms, we do not consider such fast implementations. On the contrary, to stabilize the convergence behavior of APA and RLS (because of their sensitivity to background noise), we use the regularized versions of them (see [9] and references therein); i.e., $\delta_{APA}I$ and $\delta_{RLS}I$ are added beforehand to the factors whose inverses are necessary, respectively. Both δ_{APA} and δ_{RLS} take the value of 20 times the power of the input signal. For APA, $r = 2$ (r: dimension of affine projection) with step size 0.04 is used. For RLS, we set the forgetting factor λ to 0.98 and let the initial matrix of the sample covariance matrix be $\delta^0_{RLS}\mathrm{diag}\left\{\lambda^{N-1}, \lambda^{N-2}, \cdots, 1\right\}$ ($\delta^0_{RLS} = 0.01$) by following a recommendation given in [2]. For the proposed algorithm, we design the stochastic property set by $r = 1$ and $\rho = \rho_3 = 0$ (see (1) and Example 2), and the step sizes are set to $\lambda^{(m)}_k = 1$ (for $m = 0, 1, \cdots, M-1$), $\lambda^{(M)}_k = 0.04$. Moreover, for the present numerical examples, we focus on the binary-tree like implementation (see Example 1) with $I^{(0)}_k := \{k, k-1, \cdots, k-q+1\}$, $q = 64 = 2^6 (M = 6)$, $\forall k \in \mathbb{N}$. We observe that POWER-PSP outperforms the other existing methods. Furthermore, it is also observed that, due to high level noise, the simple NLMS exhibits better performance than regularized APA and RLS.

In the second experiment shown in Fig. 5, we examine the effects of increases of q (in POWER-PSP) and r (in APA). We employ (a) $q = 64$, (b) $q = 32$, (c) $q = 16$ for POWER-PSP, and (a) $r = 64$, (b) $r = 10$, (c) $r = 3$, (d) $r = 2$ for APA. The other conditions are the same as the first experiment. We observe that the speed of convergence is accelerated by increasing q, while keeping the low level of system mismatch in the steady state and the linear computational complexity [see Remark 1 (a)]. The observed instability of APA with large r is due to high level noise (cf. the statistical analysis on set-membership in [8]).

Acknowledgement

The authors would like to express their deep gratitude to Prof. K. Sakaniwa of Tokyo Institute of Technology for fruitful discussions.

REFERENCES

[1] J. Nagumo and J. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, 1967.

[2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd edition, 1996.

[3] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control — an application of very-high-order adaptive filters," *Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, July 1999.

[4] 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, 1993.

[5] J. A. Apolinário Jr., S. Werner, P. S. R. Diniz, and T. I. Laakso, "Constrained normalized adaptive filters for CDMA mobile communications," in *Proc. of EUSIPCO '98*, 1998, vol. IV, pp. 2053–2056.

[6] T. Hinamoto and S. Maekawa, "Extended theory of learning identification," *Trans. IEE Japan*, vol. 95, no. 10, pp. 227–234, 1975, in Japanese.

[7] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *IEICE Trans.*, vol. 67-A, no. 5, pp. 126–132, 1984, in Japanese.

[8] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1091–1101, 2002.

[9] S. L. Gay, "The fast affine projection algorithm," in *Acoustic signal processing for telecommunication*, S. L. Gay and J. Benesty, Eds. Kluwer, 2000.

[10] B. Hassibi, A. H. Sayed, and T. Kailath, "H^∞ optimality of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 2, pp. 267–280, 1996.

[11] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. H. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Trans. Signal Processing Letters*, vol. 5, no. 5, pp. 111–114, 1998.

[12] L. Guo, A. Ekpenyong, and Y. H. Huang, "Frequency-domain adaptive filtering — A set-membership approach," in *Proc. of the 37th Asilomar Conference on Signals, Systems and Computers*, Nov. 2003, to appear.

[13] I. Yamada and N. Ogura, "Adaptive projected subgradient method and its applications to set theoretic adaptive filtering," in *Proc. of the 37th Asilomar Conference on Signals, Systems and Computers*, Nov. 2003, to appear.

[14] H. Buchner, J. Benesty, and W. Kellerman, *Adaptive Signal Processing*, chapter Multi-Channel Frequency-Domain Adaptive Filtering with Application to Multi-Channel Acoustic Echo Cancellation, pp. 95–128, Springer, 2003.

Appendix

Under the standard assumption of the noise process of zero mean i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$, a systematic design of the stochastic property set $C_k(\rho)$ was proposed based on the following simple formulae for ρ that rely only on r and on the variance σ^2 of the corrupting noise process $(n_k)_{k \in \mathbb{N}}$.

Example 2 [8] **(Design of Stochastic Property Sets)** *$\rho_1 := (r + \sqrt{2r})\sigma^2 \geq \rho_2 := r\sigma^2 \geq \rho_3 := \max\{(r-2)\sigma^2, 0\}$, where ρ_3: peak of pdf, ρ_2: mean value, ρ_1: (mean value) + (standard deviation), of random variable $\|\boldsymbol{e}_k(\boldsymbol{h}^*)\|^2$ (= $\|\boldsymbol{n}_k\|^2$), which obeys χ^2 distribution [see (1)].*