ADAPTIVE PARALLEL VARIABLE-METRIC PROJECTION ALGORITHM —AN APPLICATION TO ACOUSTIC ECHO CANCELLATION

Masahiro Yukawa and *Isao Yamada*

Dept. of Communications and Integrated Systems, Tokyo Institute of Technology, JAPAN.
E-mails: {masahiro,isao}@comm.ss.titech.ac.jp

ABSTRACT

In this paper, we propose a novel adaptive filtering algorithm named *adaptive parallel variable-metric projection (APVP) algorithm*, which includes the proportionate normalized least mean square (PNLMS) algorithm as its special example. The proposed algorithm is based on parallel projection (onto multiple closed convex sets) with *time-varying metrics*. A convergence analysis of the proposed algorithm is presented with the aid of the *adaptive projected subgradient method*. Numerical examples demonstrate that the proposed algorithm realizes echo cancellation superior to the conventional algorithms.

Index Terms— adaptive filtering algorithm, Adaptive Projected Subgradient Method, proportionate NLMS, variable metric

1. INTRODUCTION

The adaptive projected subgradient method (APSM) [1–3] has given a flavor of *the fixed point theory (of quasi-nonexpansive mapping)* [4] to the adaptive filtering problem [5]. Precisely speaking, APSM produces an algorithmic solution to the following *time-varying* optimization problem: minimize asymptotically a sequence of non-negative convex functions (time-varying objective functions) over a closed convex set in a real Hilbert space. Moreover, APSM has been proven to be quite effective in real-world applications [3]; e.g., stereophonic acoustic echo cancellation [6], blind multiple access interference suppression in DS/CDMA systems [7, 8], and adaptive beamforming [9]. On the other hand, the proportionate normalized least mean square (PNLMS) algorithm [10, 11] (or its extended version: proportionate affine projection algorithm (PAPA) [12, 13]) has been reported to perform even better than NLMS (or APA) in case of sparse *estimandum*, a system to be estimated, such as in acoustic/network echo cancellation.

This paper throws a bridge between APSM and the proportionate-type algorithms by extending APSM from a *constant* metric to a *variable* one. Thanks to the benefits from proportionate-type algorithms, we may improve the performance of APSM. Indeed, the variable-metric version of APSM produces an efficient algorithm, named *adaptive parallel variable-metric projection (APVP) algorithm*, improving convergence behavior even in noisy environments. By employing a simple metric, the computational complexity (excluding the complexity to design the time-varying metric) of the APVP algorithm is kept linear w.r.t. filter length (see Remark 1). A convergence analysis of APVP is presented in Sec. 3 (all the proofs are omitted due to lack of space). In Sec. 4, APVP is applied to the acoustic echo cancellation problem, where the proportionate-type algorithms are introduced as special cases of APVP. Numerical examples demonstrate that APVP significantly improves echo cancellation ability compared with the PAPA [12] and exponentially weighted stepsize projection (ESP) [14] algorithms.

2. PROPOSED ADAPTIVE ALGORITHM

Let us start with brief mathematical preliminaries. Throughout the paper, $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and nonnegative integers, respectively. Given a positive definite matrix ($\mathbb{R}^{N\times N} \ni$) $\boldsymbol{Q}_k \succ \boldsymbol{O}$ ($N \in \mathbb{N}^* := \mathbb{N}\setminus\{0\}$, $k \in \mathbb{N}$), define the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{Q}_k} := \boldsymbol{x}^T\boldsymbol{Q}_k\boldsymbol{y}$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$. Then, its induced (quadratic) norm is given as $\|\boldsymbol{x}\|_{\boldsymbol{Q}_k} := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle_{\boldsymbol{Q}_k}}$, $\forall \boldsymbol{x} \in \mathbb{R}^N$, and its associated *quadratic-metric* is defined as $d_{\boldsymbol{Q}_k}(\boldsymbol{a}, \boldsymbol{b}) := \|\boldsymbol{a}-\boldsymbol{b}\|_{\boldsymbol{Q}_k}$, $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$. Given a continuous convex function $\phi : \mathbb{R}^N \to \mathbb{R}$, the set of all its *subgradients* at any $\boldsymbol{y} \in \mathbb{R}^N$ is called the *subdifferential* of ϕ at $\boldsymbol{y}$, defined with the (variable) inner product $\langle\cdot,\cdot\rangle_{\boldsymbol{Q}_k}$ as $\partial_{(\boldsymbol{Q}_k)}\phi(\boldsymbol{y}) := \{\boldsymbol{a} \in \mathbb{R}^N : \langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{a}\rangle_{\boldsymbol{Q}_k} + \phi(\boldsymbol{y}) \le \phi(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^N\} \neq \emptyset$. Suppose that there exists an $\boldsymbol{x} \in \mathbb{R}^N$ s.t. (such that) $\phi(\boldsymbol{x}) \le 0$. Then, selecting a subgradient $\phi' : \mathbb{R}^N \to \mathbb{R}^N$ (i.e., $\phi'(\boldsymbol{x}) \in \partial_{(\boldsymbol{Q}_k)}\phi(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathbb{R}^N$), define a mapping $T^{(\boldsymbol{Q}_k)}_{\mathrm{sp}(\phi)} : \mathbb{R}^N \to \mathbb{R}^N$ by

$$T^{(\boldsymbol{Q}_k)}_{\mathrm{sp}(\phi)}(\boldsymbol{x}) := \begin{cases} \boldsymbol{x} - \dfrac{\phi(\boldsymbol{x})}{\|\phi'(\boldsymbol{x})\|^2_{\boldsymbol{Q}_k}}\phi'(\boldsymbol{x}) & \text{if } \phi(\boldsymbol{x}) > 0, \\ \boldsymbol{x} & \text{otherwise.} \end{cases} \quad (1)$$

The operator $T^{(\boldsymbol{Q}_k)}_{\mathrm{sp}(\phi)}$ is called *the subgradient projection* relative to ϕ. Note in (1) that ϕ' depends on $\boldsymbol{Q}_k$. From (1), we assume, without loss of generality, $\|\boldsymbol{Q}_k\|_{\mathcal{F}} \le 1$ in the following, where $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm.

The problem is formulated as follows [1, 2]: minimize asymptotically a sequence of nonnegative convex objective functions $\Theta_k : \mathbb{R}^N \to [0, \infty)$, $k \in \mathbb{N}$. APSM has mathematically been proven to achieve this goal [1, 2]. For the purpose of improving the convergence behavior, we present its variable-metric version below.

Scheme 1 (Variable-Metric Version of the Adaptive Projected Subgradient Method) *For an arbitrarily given $\boldsymbol{h}_0 \in \mathbb{R}^N$, generate a sequence $(\boldsymbol{h}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^N$ by*

$$\boldsymbol{h}_{k+1} := \begin{cases} \left[(1-\lambda_k)I + \lambda_k T^{(\boldsymbol{Q}_k)}_{\mathrm{sp}(\Theta_k)}\right](\boldsymbol{h}_k) & \text{if } \Theta'_k(\boldsymbol{h}_k) \neq \boldsymbol{0}, \\ \boldsymbol{h}_k & \text{otherwise,} \end{cases}$$

where $\lambda_k \in [0, 2]$, $\forall k \in \mathbb{N}$ (I and $\boldsymbol{0}$ denote the identity mapping and the zero vector, respectively).

We remark that it is possible to use a convex-projection operator $P^{(\boldsymbol{Q}_k)}_C$ (or a *strongly attracting nonexpansive mapping* T_k) as in the original APSM [1, 2] (or the extended APSM [15]), which is useful, e.g., in multiple access interference suppression in CDMA wireless communication systems [7]. Here, $P^{(\boldsymbol{Q})}_C : \boldsymbol{h} \mapsto \arg\min_{\boldsymbol{a}\in C} d_{\boldsymbol{Q}}(\boldsymbol{a}, \boldsymbol{h})$, for any given $\boldsymbol{Q} \succ \boldsymbol{O}$, stands for the projection operator associated with a nonempty closed convex set C w.r.t. a metric $d_{\boldsymbol{Q}}$.

Let $(\boldsymbol{h}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^N$ be a sequence of adaptive filtering vectors. To compute $\boldsymbol{h}_{k+1}$ from $\boldsymbol{h}_k$ at each $k \in \mathbb{N}$, we construct q ($\in \mathbb{N}^*$) closed convex sets $C^{(k)}_i$, $i = 1, 2, \cdots, q$, that are defined by means of observable data so as to contain the estimandum $\boldsymbol{h}^*$ with high reliability (see Sec. 4.2). Define the weights to those data-dependent

This work was done while the first author was with Dept. of Communications and Integrated Systems, Tokyo Institute of Technology. From April 2007, he will work for RIKEN in Japan as a Special Postdoctoral Researcher.

closed convex sets as $w_i^{(k)} \in (0,1]$, $i = 1, 2, \cdots, q$, $k \in \mathbb{N}$, satisfying $\sum_{i=1}^q w_i^{(k)} = 1$. Application of Scheme 1 to[1]

$$\Theta_k(\boldsymbol{h}) := \begin{cases} \dfrac{1}{L_k} \displaystyle\sum_{i=1}^q w_i^{(k)} d_{\boldsymbol{Q}_k}[\boldsymbol{h}_k, C_i^{(k)}] d_{\boldsymbol{Q}_k}[\boldsymbol{h}, C_i^{(k)}] \\ \text{if } L_k := \sum_{i=1}^q w_i^{(k)} d_{\boldsymbol{Q}_k}[\boldsymbol{h}_k, C_i^{(k)}] \neq 0, \\ 0 \quad \text{otherwise,} \end{cases}$$

yields the proposed algorithm, as shown below (Here, $d_{\boldsymbol{Q}_k}(\boldsymbol{a}, C) := \min_{\boldsymbol{b} \in C} \|\boldsymbol{a} - \boldsymbol{b}\|_{\boldsymbol{Q}_k}$ stands for the distance from an arbitrary point $\boldsymbol{a} \in \mathbb{R}^N$ to a closed convex set C).

Algorithm 1 (Adaptive Parallel Variable-Metric Projection (APVP) Algorithm) *For an arbitrarily chosen initial vector $\boldsymbol{h}_0 \in \mathbb{R}^N$, generate a sequence of adaptive filtering vectors $(\boldsymbol{h}_k)_{k \in \mathbb{N}} \subset \mathbb{R}^N$ as*

$$\boldsymbol{h}_{k+1} := \boldsymbol{h}_k + \lambda_k \mathcal{M}_k \left[\sum_{i=1}^q w_i^{(k)} P_{C_i^{(k)}}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right], \forall k \in \mathbb{N}, \quad (2)$$

where $\lambda_k \in [0,2]$ is the step size and

$$\mathcal{M}_k := \begin{cases} \dfrac{\displaystyle\sum_{i=1}^q w_i^{(k)} \left\| P_{C_i^{(k)}}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|_{\boldsymbol{Q}_k}^2}{\left\| \displaystyle\sum_{i=1}^q w_i^{(k)} P_{C_i^{(k)}}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|_{\boldsymbol{Q}_k}^2} & \text{if } \boldsymbol{h}_k \notin \displaystyle\bigcap_{i=1}^q C_i^{(k)}, \\ 1 & \text{otherwise.} \end{cases}$$

An efficient design of the weights $w_i^{(k)}$ has been addressed in [16]. A remark on Algorithm 1 is given below.

Remark 1 (Computational Complexity of Algorithm 1)

(a) *(Inherent parallelism) Algorithm* 1 *has the inherently parallel structure* [17]*, because each projection in* (2) *can be computed independently (thus in parallel). In fact, in addition that the algorithm is relevant to parallel implementation, it has a fault tolerance nature (see* [16, Sec. V]*).*

(b) *(Complexity) Note firstly that the complexity shown below excludes the computation to design $\boldsymbol{Q}_k$ (e.g., in* [11]*, the 'strobe down' technique is introduced in PNLMS to reduce the complexity for designing $\boldsymbol{Q}_k$; see Example 1). By employing an 'efficient' metric (i.e., the matrix $\boldsymbol{Q}_k$ has a special structure such as diagonal), the overall computational complexity (the number of multiplications/divisions) is kept $O(N)$ [If $\boldsymbol{Q}_k$ has no special structure, then the algorithm requires matrix-vector multiplications]. In addition, by employing q concurrent processors, the computational complexity imposed on each processor at each iteration is approximately $(2r+4)N$. For more precise discussion on computational complexity, see* [18]*. The proposed algorithm can significantly raise, by increasing q, convergence speed while keeping low time consumption, which is a great advantage especially in real-time applications including the acoustic echo cancellation.*

3. A CONVERGENCE ANALYSIS

In this section, we present conditions for Scheme 1 to have the remarkable properties of APSM, (a) asymptotic optimality and (b) convergence. First of all, the following proposition is straightforwardly verified by [2, Theorem 2(a)].

[1] The factor $d_{\boldsymbol{Q}_k}[\boldsymbol{h}_k, C_i^{(k)}]$, $i = 1, 2, \cdots, q$, is constant in terms of $\boldsymbol{h}$. Indeed, $d_{\boldsymbol{Q}_k}[\boldsymbol{h}_k, C_i^{(k)}]$ is an automatically-determined weighting factor and gives a large weight to a set 'far' from $\boldsymbol{h}_k$ in the sense of the metric $d_{\boldsymbol{Q}_k}$.

Proposition 1 (Monotone Approximation) *Suppose, for $(\boldsymbol{h}_k)_{k \in \mathbb{N}}$ generated by Scheme 1, $\boldsymbol{h}_k \notin \Omega_k := \{\boldsymbol{h} \in \mathbb{R}^N : \Theta_k(\boldsymbol{h}) = \Theta_k^* := \inf_{\boldsymbol{x} \in \mathbb{R}^N} \Theta_k(\boldsymbol{x})\} \neq \emptyset$, $k \in \mathbb{N}$. Then, with $\lambda_k \in \left(0, 2\left(1 - \frac{\Theta_k^*}{\Theta_k(\boldsymbol{h}_k)}\right)\right)$, we have $\left\|\boldsymbol{h}_{k+1} - \boldsymbol{h}^{*(k)}\right\|_{\boldsymbol{Q}_k} < \left\|\boldsymbol{h}_k - \boldsymbol{h}^{*(k)}\right\|_{\boldsymbol{Q}_k}$, $\forall \boldsymbol{h}^{*(k)} \in \Omega_k$.* □

The essential difference of Scheme 1 from APSM is the time-varying metric $d_{\boldsymbol{Q}_k}$. However, taking a careful look at Proposition 1, the monotonicity holds *w.r.t. different metric* at each iteration. This causes difficulty in a convergence analysis, because the key to prove those properties is monotonicity of the sequence $(\|\boldsymbol{h}_k - \boldsymbol{h}^*\|)_{k \geq \kappa_0}$, $\forall \boldsymbol{h}^* \in \Omega := \bigcap_{k \geq \kappa_0} \Omega_k$, for some $\kappa_0 > 0$ and a *fixed* norm $\|\cdot\|$. For an analytical reason, we consider the following assumptions.

Assumption 1 *(used also in* [2]*)*

(a) *There exists $k_0 \in \mathbb{N}$ s.t. (i) $\Theta_k^* \big(:= \inf_{\boldsymbol{x} \in \mathbb{R}^N} \Theta_k(\boldsymbol{x})\big) = 0$, $\forall k \geq k_0$ and (ii) $\Omega := \bigcap_{k \geq k_0} \Omega_k \neq \emptyset$.*

(b) *$(\Theta_k'(\boldsymbol{h}_k))_{k \in \mathbb{N}}$ is bounded.*

(c) *There exists a hyperplane $\Pi \subset \mathbb{R}^N$ s.t. $\mathrm{ri}_\Pi(\Omega) \neq \emptyset$, where $\mathrm{ri}_\Pi(\Omega) := \{\boldsymbol{h} \in \Omega : \exists \epsilon > 0 \text{ s.t. } B(\boldsymbol{h}, \epsilon) \cap \Pi \subset \Omega\}$ is the relative interior of Ω w.r.t. Π. Here, $B(\boldsymbol{h}, \epsilon) := \{\boldsymbol{x} \in \mathbb{R}^N : \|\boldsymbol{x} - \boldsymbol{h}\| < \epsilon\}$ is an open ball; the norm $\|\cdot\|$ can be arbitrary due to the norm equivalency for finite-dimensional vector spaces.*

Assumption 2

(a) *There exist $k_1 \in \mathbb{N}$ and a well-defined norm $\|\cdot\|_{c1}$ s.t.*

$$\|\boldsymbol{h}_k - \boldsymbol{h}^*\|_{c1}^2 - \|\boldsymbol{h}_{k+1} - \boldsymbol{h}^*\|_{c1}^2 \geq \nu_1 \frac{\Theta_k^2(\boldsymbol{h}_k)}{\|\Theta_k'(\boldsymbol{h}_k)\|_{c1}^2},$$
$$\exists \nu_1 > 0,\ \forall \boldsymbol{h}^* \in \Omega,\ \forall k \geq k_1,\ \text{for } \Theta_k'(\boldsymbol{h}_k) \neq \boldsymbol{0}.$$

(b) *There exist $k_2 \in \mathbb{N}$ and a well-defined norm $\|\cdot\|_{c2}$ s.t.*

$$\|\boldsymbol{h}_k - \boldsymbol{h}^*\|_{c2}^2 - \|\boldsymbol{h}_{k+1} - \boldsymbol{h}^*\|_{c2}^2 \geq \nu_2 \|\boldsymbol{h}_k - \boldsymbol{h}_{k+1}\|_{c2}^2,$$
$$\exists \nu_2 > 0,\ \forall \boldsymbol{h}^* \in \Omega,\ \forall k \geq k_2.$$

We are now ready to show the remarkable properties below.

Proposition 2 *(Properties of $(\boldsymbol{h}_k)_{k \in \mathbb{N}}$ generated by Scheme* 1*)*

(a) *(Boundedness, Asymptotic optimality) Under Assumptions* 1(a) *and* 2(a)*, $(\boldsymbol{h}_k)_{k \in \mathbb{N}}$ is bounded. In addition, under Assumption* 1(b)*, $\lim_{k \to \infty} \Theta_k(\boldsymbol{h}_k) = 0$.*

(b) *(Convergence) Under Assumptions* 1(a)*,* 1(c) *and* 2(b)*, $(\boldsymbol{h}_k)_{k \in \mathbb{N}}$ converges to a point $\widehat{\boldsymbol{h}} \in \mathbb{R}^N$. In addition, under Assumptions* 1(b)*, and* 2(a)*, $\lim_{k \to \infty} \Theta_k(\widehat{\boldsymbol{h}}) = 0$ provided that there exists bounded $(\Theta_k'(\widehat{\boldsymbol{h}}))_{k \in \mathbb{N}}$ where $\Theta_k'(\widehat{\boldsymbol{h}}) \in \partial_{(\boldsymbol{Q}_k)} \Theta_k(\widehat{\boldsymbol{h}})$, $\forall k \in \mathbb{N}$* □

Remark 2 *In static environments, employing, e.g., the metric reproducing PNLMS/PAPA (see Example 1), the variation of $\boldsymbol{Q}_k$ tends to taper in steady state (i.e., the metric $d_{\boldsymbol{Q}_k}$ in steady state is nearly constant), thus Assumption 2 would be natural. (NOTE: Assumption 2 automatically holds in case of a constant metric.) If the variation of $\boldsymbol{Q}_k$ does not taper, then one could stop adjusting $\boldsymbol{Q}_k$.*

In dynamic environments (more specifically, when the estimandum is highly time-variant), it is not obvious that Assumption 2 holds. In such environments, however, the adaptive filter should keep tracking the time-varying estimandum all the time, thus the monotone approximation at each iteration (see Proposition 1) is a more desirable property for stability than convergence.

4. APPLICATION TO ACOUSTIC ECHO CANCELER

4.1. Acoustic Echo Cancellation Problem

The acoustic echo cancellation (AEC) problem [19] is formulated as follows. Let $k \in \mathbb{N}$ denote the time index and $N \in \mathbb{N}^* := \mathbb{N}\backslash\{0\}$ the length of echo canceler $\boldsymbol{h}_k$. For notational simplicity, we let the estimandum (i.e., echo impulse response) $\boldsymbol{h}^* \in \mathbb{R}^N$. With a sequence of input signals $(u_k)_{k\in\mathbb{N}} \subset \mathbb{R}$, let $(\boldsymbol{u}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^N$ be a sequence of input vectors defined as $\boldsymbol{u}_k := [u_k, u_{k-1}, \cdots, u_{k-N+1}]^T$. For $r \in \mathbb{N}^*$, define $\boldsymbol{U}_k := [\boldsymbol{u}_k, \boldsymbol{u}_{k-1}, \cdots, \boldsymbol{u}_{k-r+1}] \in \mathbb{R}^{N\times r}$ (usually $r \ll N$). Also define the noise vector as $\boldsymbol{n}_k := [n_k, n_{k-1}, \cdots, n_{k-r+1}]^T \in \mathbb{R}^r$, $\forall k \in \mathbb{N}$, where $(n_k)_{k\in\mathbb{N}}$ is a sequence of additive noise process. We introduce the linear model for the data process $(\boldsymbol{d}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^r$: $\boldsymbol{d}_k := \boldsymbol{U}_k^T\boldsymbol{h}^* + \boldsymbol{n}_k$. The goal of the echo cancellation is to remove (or cancel) the echo part $\boldsymbol{U}_k^T\boldsymbol{h}^*$ in $\boldsymbol{d}_k$ by subtracting the output of adaptive (linear) filter $\boldsymbol{h}_k \in \mathbb{R}^N$, $k \in \mathbb{N}$, as $\boldsymbol{d}_k - \boldsymbol{U}_k^T\boldsymbol{h}_k$. Since $\boldsymbol{h}_k \approx \boldsymbol{h}^*$ implies successful echo cancellation, the problem can be interpreted as *system identification* (i.e., identify an estimandum $\boldsymbol{h}^*$ by means of input-output relations), which is also known as adaptive filtering.

A special example of Algorithm 1 for AEC is given below.

Example 1 *Let, in Algorithm 1,* $C_1^{(k)} := V_k := \{\boldsymbol{h} \in \mathbb{R}^N : \boldsymbol{U}_k^T\boldsymbol{h} = \boldsymbol{d}_k\}$*,* $\forall k \in \mathbb{N}$*, with* $q = 1$*. Then we get the following algorithm:* $\boldsymbol{h}_{k+1} := \boldsymbol{h}_k + \lambda_k\left[P_{V_k}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) - \boldsymbol{h}_k\right]$*. In particular*[2]*,* $\boldsymbol{Q}_k := \boldsymbol{G}_k^{-1} / \left\|\boldsymbol{G}_k^{-1}\right\|_{\mathcal{F}}$ *yields the proportionate APA (PAPA) algorithm* [12, 13] *(or the proportionate NLMS (PNLMS) algorithm* [10, 11] *if* $r = 1$*). Here,* $\boldsymbol{G}_k := diag(\alpha_1^{(k)}, \alpha_2^{(k)}, \cdots, \alpha_N^{(k)}) \succ \boldsymbol{O}$ *with* $\alpha_i^{(k)} := \gamma_i^{(k)}/(\sum_{j=1}^N \gamma_j^{(k)})$, $\forall i = 1, 2, \cdots, N$, $\gamma_i^{(k)} := \max\left\{\sigma L_{\max}^{(k)}, |h_i^{(k)}|\right\}$*, and* $L_{\max}^{(k)} := \max\{\delta, |h_1^{(k)}|, |h_2^{(k)}|, \cdots, |h_N^{(k)}|\}$*.* σ *and* δ *are small positive constants;* $h_i^{(k)}$ *the ith component of* $\boldsymbol{h}_k$*; and* $diag(\cdots)$ *stands for a diagonal matrix (the arguments are scalars or square matrices). Modified designs of the matrix* $\boldsymbol{G}_k$ *are also proposed, e.g., in* [13, 20–22].

4.2. Efficient Echo Canceler by Algorithm 1

We present a more efficient realization of Algorithm 1 for AEC; we show (i) how to design the data-dependent closed convex sets $(C_i^{(k)})_{i=1}^q$, $k \in \mathbb{N}$, containing the estimandum $\boldsymbol{h}^*$ with high reliability, and (ii) how to compute the projection onto those sets.

As a first step, we define the stochastic property set[3] [23]: $C_k(\rho) := \left\{\boldsymbol{h} \in \mathbb{R}^N : g_k(\boldsymbol{h}) := \|\boldsymbol{e}_k(\boldsymbol{h})\|_{\boldsymbol{I}}^2 - \rho \le 0\right\}$, $\forall k \in \mathbb{N}$, where $\boldsymbol{e}_k : \mathbb{R}^N \to \mathbb{R}^r$, $\boldsymbol{h} \mapsto \boldsymbol{U}_k^T\boldsymbol{h} - \boldsymbol{d}_k$, is the error (or residual) function and $\rho \ge 0$ a parameter governing the membership probability that $\boldsymbol{h}^* \in C_k(\rho)$; noise statistics are involved in the design of ρ [23, Ex. 1]. Since the computational cost for the direct projection onto $C_i(\rho)$ is prohibitive in general, we introduce the following outer approximating half-space [6, 16, 23] (see [24] for other outer approximation): $(C_i^{(k)} :=) H_i^-(\boldsymbol{h}_k) := \{\boldsymbol{x} \in \mathbb{R}^N : \langle \boldsymbol{x} - \boldsymbol{h}_k, \nabla_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k)\rangle_{\boldsymbol{Q}_k} + g_i(\boldsymbol{h}_k) \le 0\} \supset C_i(\rho)$, where $\nabla_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k) = 2\boldsymbol{Q}_k^{-1}\boldsymbol{U}_i\boldsymbol{e}_i(\boldsymbol{h}_k) \in \partial_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k) := \{\boldsymbol{a} \in \mathbb{R}^N : \langle \boldsymbol{x} - \boldsymbol{h}_k, \boldsymbol{a}\rangle_{\boldsymbol{Q}_k} + g_i(\boldsymbol{h}_k) \le g_i(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^N\}$. Note that $\partial_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k) = \{\nabla_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k)\}$ in this (differentiable) case.

The projection onto $H_i^-(\boldsymbol{h}_k)$ has the following simple closed-form expression: $P_{H_i^-(\boldsymbol{h}_k)}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) = \boldsymbol{h}_k - \dfrac{g_i(\boldsymbol{h}_k)}{\left\|\nabla_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k)\right\|_{\boldsymbol{Q}_k}^2}$ $\nabla_{(\boldsymbol{Q}_k)} g_i(\boldsymbol{h}_k)$, if $\boldsymbol{h}_k \notin H_i^-(\boldsymbol{h}_k)$, $P_{H_i^-(\boldsymbol{h}_k)}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) = \boldsymbol{h}_k$, otherwise. NOTE [23]: $P_{H_i^-(\boldsymbol{h}_k)}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k) \cong P_{C_i(\rho)}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k)$; and $P_{H_i^-(\boldsymbol{h}_k)}^{(\boldsymbol{Q}_k)}(\boldsymbol{h}_k)$ requires only $O(N)$ complexity.

[2] Although $\boldsymbol{Q}_k := \boldsymbol{G}_k^{-1}$ is used in [10, 11], we use its normalized version. The normalization makes the algorithm more stable when we use *regularization*, while nothing is changed without regularization.

[3] For unified notation, $\|\cdot\|_{\boldsymbol{I}}$ will be used for the Euclidean norm.

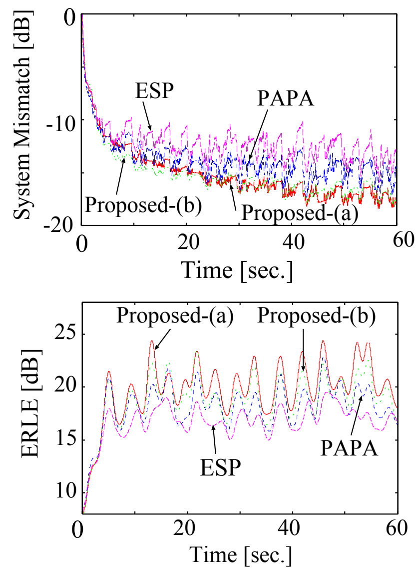

Fig. 1. A comparison among the proposed, PAPA, and ESP algorithms in system mismatch and ERLE.

Table 1. Steady-state performance of the proposed, ESP and PAPA algorithms.

Algorithm	Prop(a)	Prop(b)	PAPA	ESP
System Mismatch [dB]	-17.3	-16.8	-15.0	-13.3
ERLE [dB]	20.5	19.3	18.6	17.1

5. NUMERICAL EXAMPLES

We compare the performance of the proposed algorithm (Algorithm 1) with ESP [14] and PAPA [12] in the AEC problem under the following conditions. The input signal $\boldsymbol{u}_k$ is English-native-male's speech recorded at sampling rate 8 kHz. The noise $\boldsymbol{n}_k$ is white with SNR = 10 dB. A real impulse response[4] $\boldsymbol{h}^* \in \mathbb{R}^N$ recorded in a small room is used for $N = 1024$.

We adopt two measures: (i) system mismatch defined as $10\log_{10}(\|\boldsymbol{h}^* - \boldsymbol{h}_k\|_{\boldsymbol{I}}^2/\|\boldsymbol{h}^*\|_{\boldsymbol{I}}^2)$ at kth iteration, and (ii) the Echo Return Loss Enhancement (ERLE) [19]. To obtain smooth ERLE curves, after calculating instantaneous ERLE at kth iteration, we pass the instantaneous one through a smoothing filter three times[5].

For the metric $d_{\boldsymbol{Q}_k}$ in the proposed algorithm, we use the following: (a) $d_{\boldsymbol{G}_k^{-1}}$ designed in the same way as PNLMS [10, 11] (see Example 1); and (b) $d_{\boldsymbol{A}^{-1}}$ designed in the same way as ESP (see [14, 25]; we employ the exponential factor $\gamma = 0.99569$ for better performance). Proposed-(b), a constant version of APVP, has been presented in [26]. For proposed-(a) and proposed-(b), we set $r = 1$, $q = 8$, $\lambda_k = 0.4$, $\forall k \in \mathbb{N}$, $w_i^{(k)} = 1/q$, $\forall i = 1, 2, \cdots, q$, $\forall k \in \mathbb{N}$, and $\rho = \rho_3 (= 0)$; $f_r(\rho_3)$ gives the peak value of the probability

[4] We employed the impulse response available at http://www.echochamber.ch/responses/960/rooms.zip; the name of the file is "1960small_room.wav". As the frequency of the downloaded sample is in fact 44.1 kHz, we convert it by the matlab command 'resample' into 8 kHz.

[5] Precisely, $\mathrm{ERLE}(k) := \mathrm{ERLE}_{\mathrm{tmp}}^{(3)}(k)$ with $\mathrm{ERLE}_{\mathrm{tmp}}^{(i+1)}(k) := \left(\sum_{j=\psi(k-\ell)}^{k+\ell} \mathrm{ERLE}_{\mathrm{tmp}}^{(i)}(j)\right) / [k + \ell - \psi(k-\ell) + 1]$, for $i = 0, 1, 2$, where $\ell = 5000$ and $\psi(n) := \max\{n, 0\}$ for any integer n.

density function f_r of the random variable $\xi := \|\boldsymbol{n}_k\|_{\boldsymbol{I}}^2$ [23]. For PAPA, we set $r = 8$ and $\lambda_k = 0.05$, $\forall k \in \mathbb{N}$, with the same metric as proposed-(a). For ESP, we set $r = 8$ and $\lambda_k = 0.02$, $\forall k \in \mathbb{N}$, with the same metric as proposed-(b). The step size for each algorithm was tuned so that all the algorithms attain almost the same initial convergence speed.

Fig. 1 illustrates the simulation results, where we see that the proposed algorithm significantly outperforms the conventional ones. More precisely, Table 1 gives the steady-state performance; system mismatch and ERLE in Table 1 express the values averaged uniformly over the last 1.0×10^5 and 3.2×10^5 samples (12.5 [sec.] and 40 [sec.]), respectively. We observe that, compared with ESP, proposed-(a) gains approximately 4 dB and 3 dB in system mismatch and ERLE, respectively, while compared with PAPA approximately 2 dB both in system mismatch and ERLE.

Finally, we remark that proposed-(a) achieves more than 1 dB better ERLE than proposed-(b) although we select a best metric in proposed-(b) among the constant metrics in the form of $d_{\boldsymbol{A}^{-1}}$. This implies that, in highly time-varying situations such as mobile telecommunications (where it is difficult to attain always a good constant metric), the difference between the use of the variable metric [proposed-(a)] and the use of the constant one [proposed-(b)] is expected to be more apparent (a comparison in such a situation between the standard and generalized-proportionate APA algorithms has been presented in [21]).

6. CONCLUSION

This paper has proposed the adaptive parallel variable-metric projection (APVP) algorithm and has presented its convergence analysis. By employing a reasonable metric such as the one used in PNLMS, the proposed algorithm converges, under natural conditions, to a point optimal in an asymptotic sense. The simulation results have shown that the proposed algorithm achieves echo cancellation better than the ESP and PAPA algorithms with linear computational complexity. Those remarkable improvements are realized by bringing an idea of PNLMS to APSM with the use of variable-metric.

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

7. REFERENCES

[1] I. Yamada, "Adaptive projected subgradient method: A unified view for projection based adaptive algorithms," *The Journal of IEICE*, vol. 86, no. 8, pp. 654–658, Aug. 2003, in Japanese.

[2] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numer. Funct. Anal. Optim.*, vol. 25, no. 7&8, pp. 593–617, 2004.

[3] I. Yamada, K. Slavakis, M. Yukawa, and R. L. G. Cavalcante, "Adaptive Projected Subgradient Method and its applications to robust signal processing," in *Proc. IEEE ISCAS*, 2006, pp. 269–272, (invited paper).

[4] D. P. Mandic and I. Yamada, "Machine learning and signal processing applications of fixed point theory," *Tutorial proposal for IEEE ICASSP2007*, accepted.

[5] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 4th edition, 2002.

[6] M. Yukawa, N. Murakoshi, and I. Yamada, "Efficient fast stereo acoustic echo cancellation based on pairwise optimal weight realization technique," *EURASIP J. Appl. Signal Processing*, vol. 2006, Article ID 84797, 15 pages, 2006.

[7] R. L. G. Cavalcante, I. Yamada, and K. Sakaniwa, "A fast blind MAI reduction based on adaptive projected subgradient method," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1973–1980, Aug. 2004.

[8] M. Yukawa, R. L. G. Cavalcante, and I. Yamada, "Efficient blind MAI suppression in DS/CDMA systems by embedded constraint parallel projection techniques," *IEICE Trans. Fundamentals*, vol. E88-A, no. 8, pp. 2062–2071, Aug. 2005.

[9] K. Slavakis, M. Yukawa, and I. Yamada, "Robust Capon beamforming by the Adaptive Projected Subgradient Method," in *Proc. IEEE ICASSP*, 2006, pp. 1005–1008.

[10] S. L. Gay, "An efficient fast converging adaptive filter for network echo cancellation," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 1998, pp. 394–398.

[11] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

[12] T. Gänsler, J. Benesty, S. L. Gay, and M. M. Sondhi, "A robust proportionate affine projection algorithm for network echo cancellation," in *Proc. IEEE ICASSP*, 2000, pp. 793–796.

[13] S. Werner, J. A. Apolinário Jr., P. S. R. Diniz, and T. I. Laakso, "A set-membership approach to normalized proportionate adaptation algorithms," in *Proc. EUSIPCO, Special session on partial-update adaptive filters and sparse systems*, 2005.

[14] S. Makino and Y. Kaneda, "Exponentially weighted stepsize projection algorithm for acoustic echo cancellers," *IEICE Trans. Fundamentals*, vol. E75-A, no. 11, pp. 1500–1508, Nov. 1992.

[15] K. Slavakis, I. Yamada, and N. Ogura, "Adaptive projected subgradient method over the fixed point set of strongly attracting non-expansive mappings," *Numer. Funct. Anal. Optim.*, vol. 27, no. 7&8, pp. 905–930, 2006.

[16] M. Yukawa and I. Yamada, "Pairwise optimal weight realization —Acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4557–4571, Dec. 2006.

[17] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithm, and Optimization*, Oxford University Press, 1997.

[18] M. Yukawa, K. Slavakis, and I. Yamada, "Adaptive parallel quadratic-metric projection algorithms," *IEEE Trans. Audio, Speech and Language Processing*, accepted.

[19] 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," *IEEE Signal Processing Mag.*, vol. 16, no. 4, pp. 42–69, July 1999.

[20] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, 2002, pp. 1881–1884.

[21] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc IEEE ICASSP*, 2004, pp. 161–164.

[22] H. Deng and M. Doroslovački, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Processing Lett.*, vol. 12, no. 3, pp. 181–184, Mar. 2005.

[23] 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, May 2002.

[24] N. Ogura and I. Yamada, "A deep outer approximating half space of the level set of certain quadratic functions," *J. Nonlinear Convex Anal.*, vol. 6, no. 1, pp. 187–201, 2005.

[25] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 101–108, Jan. 1993.

[26] M. Yukawa and I. Yamada, "Adaptive quadratic-metric parallel subgradient projection algorithm and its application to acoustic echo cancellation," in *Proc. EUSIPCO*, Sept. 2006.