# MEAN-SQUARE PERFORMANCE OF THE HYPERSLAB-BASED ADAPTIVE PROJECTED SUBGRADIENT METHOD

*Wemer M. Wee, Masao Yamagishi, and Isao Yamada*

Department of Communications and Computer Engineering, Tokyo Institute of Technology
2-12-1-S3-60 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan
Email: {wemer, myamagi, isao}@sp.ce.titech.ac.jp

## ABSTRACT

This paper is concerned with the mean-square performance of the hyperslab-based adaptive projected subgradient method, a set theoretic estimation tool that has been successfully applied in a wide variety of signal processing tasks. Using energy-conservation arguments, general performance results are derived without restricting the regression data to being Gaussian or white. Numerical simulations are provided to illustrate the theoretical developments.

***Index Terms*—** Adaptive filters, energy conservation, mean-square performance, error nonlinearity, data-reusing algorithms

## 1. INTRODUCTION

The hyperslab-based adaptive projected subgradient method (H-APSM) [1, 2] is a set theoretic estimation tool [3, 4] that has been applied successfully in a variety of signal processing tasks such as adaptive filtering [5], online classification [6–8] and distributed learning [9–15]. The H-APSM has several appealing features that include low complexity, robustness against noise, generality, and design flexibility, which make it well-suited for many applications. In the context of adaptive filtering, its general form serves as a unifying principle for a wide range of schemes that include the normalized least-mean-squares (NLMS) algorithm [16], the set-membership NLMS (SM-NLMS) algorithm [17], the affine projection algorithm (APA) [18, 19], and other projection-based adaptive schemes [5].

Due to its applicability in a wide range of scenarios, there are already several analytical results for the H-APSM [2, 8, 9, 20, 21], although most of them are based on deterministic arguments or are only concerned with the convergence of the estimates of the algorithm. This paper presents yet another result on H-APSM, with a focus on the evaluation of the mean-square performance of the H-APSM as an adaptive filter. More specifically, a main objective of this paper is to characterize the steady-state mean-square error of the aforementioned set theoretic scheme. For this purpose, we rely on well-known energy-conservation arguments [22], which are among the most useful techniques for the analyses of the performance of adaptive algorithms [23–28]. Though the inherent complexity of projection-based algorithms makes the analysis challenging, we attempt to provide a fairly general treatment of the mean-square performance of H-APSM based on the earlier works [22, 25, 26]. We also illustrate that, compared to other data-reusing algorithms such as the APA (see, e.g., [26, 27]), the H-APSM manages to obtain a lower steady-state MSE value even if one reuses more data in its implementation. Moreover, since the H-APSM covers the SM-NLMS algorithm as a special case, the present analysis, in particular the stability result, can also be used to complement the results derived earlier for the said algorithm [28, 29]. Numerical simulations are shown to support our theoretical findings.

This work was supported in part by the Japan Society for the Promotion of Science Grants-in-Aid (24800022, B-21300091). The work of W. M. Wee was supported in part by the Global COE program "Photonics Integration-Core Electronics" and in part by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## 2. PRELIMINARIES

### 2.1. Notation

Throughout this paper, we denote the sets of all nonnegative integers and real numbers by $\mathbb{N}$ and $\mathbb{R}$, respectively. Vectors and matrices are represented by boldface characters. We denote the Euclidean norm by $\|\cdot\|$, the transpose of a vector or matrix by $(\cdot)^{\mathsf{T}}$, the identity matrix of appropriate dimensions by $\boldsymbol{I}$, the trace of a matrix by $\mathrm{Tr}(\cdot)$, and the expectation of a random variable by $\mathbb{E}[\cdot]$.

Let $C$ be a subset of the Euclidean space $\mathbb{R}^M$. If $C$ is nonempty, the distance from $\boldsymbol{h} \in \mathbb{R}^M$ to $C$ is $d(\boldsymbol{h}, C) := \inf \|\boldsymbol{h} - C\|$. If $C$ is also closed and convex, then, for every $\boldsymbol{h} \in \mathbb{R}^M$, there exists a unique point $P_C(\boldsymbol{h}) \in C$, called the projection of $\boldsymbol{h}$ onto $C$, such that $\|\boldsymbol{h} - P_C(\boldsymbol{h})\| = d(\boldsymbol{h}, C)$.

### 2.2. Data Model

Consider noisy measurements $\{d(i)\}_{i\in\mathbb{N}}$ that arise from the linear model

$$d(i) = \boldsymbol{u}_i^{\mathsf{T}} \boldsymbol{h}^{\circ} + v(i), \tag{1}$$

where $\boldsymbol{h}^{\circ}$ is an unknown vector that we wish to estimate, $v(i)$ accounts for modeling errors with variance denoted by $\sigma_v^2$, and $\boldsymbol{u}_i \in \mathbb{R}^M$ denotes the regressor vector with a positive-definite covariance matrix $\boldsymbol{R} := \mathbb{E}[\boldsymbol{u}_i \boldsymbol{u}_i^{\mathsf{T}}]$. For simplicity, the noise $v(i)$ is assumed to be independent of $\boldsymbol{u}_j$ for all $j$ and of $v(j)$ for all $j \neq i$.

### 2.3. The Algorithm

The general form of the adaptive projected subgradient method (APSM) [2, 5, 6, 8–10, 15, 30–38] can be described as follows [2]. Let $\Theta_i : \mathbb{R}^M \to [0, +\infty)$ $(\forall i \in \mathbb{N})$ be a sequence of convex functions and $\mathcal{K} \subset \mathbb{R}^M$ be a nonempty closed convex set. For any arbitrarily given $\boldsymbol{h}_{-1} \in \mathcal{K}$, the APSM generates a sequence of estimates $\{\boldsymbol{h}_i\}_{i\in\mathbb{N}} \subset \mathcal{K}$ using the recursion

$$\boldsymbol{h}_i = \begin{cases} P_{\mathcal{K}}\left(\boldsymbol{h}_{i-1} - \mu \dfrac{\Theta_i(\boldsymbol{h}_{i-1})\Theta_i'(\boldsymbol{h}_{i-1})}{\|\Theta_i'(\boldsymbol{h}_{i-1})\|^2}\right), & \text{if } \Theta_i'(\boldsymbol{h}_{i-1}) \neq 0, \\ \boldsymbol{h}_{i-1}, & \text{if } \Theta_i'(\boldsymbol{h}_{i-1}) = 0, \end{cases} \tag{2}$$

where $\Theta'(\boldsymbol{h}_{i-1}) \in \partial\Theta_i(\boldsymbol{h}_{i-1}) := \{\boldsymbol{r} \in \mathbb{R}^M \mid \Theta_i(\boldsymbol{h}_{i-1}) + \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{h}_{i-1}) \leq \Theta(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^M\} \neq \emptyset$ and $\mu$ is the step-size or relaxation parameter. The APSM satisfies several remarkable properties such as monotone approximation, asymptotic optimality and convergence in norm [2]. In general, different design parameters can lead to several efficient schemes that can be tailored for different types of applications. The commonly used hyperslab-based version H-APSM [2, 5, 8] is obtained by choosing the following parameters. Define $\Theta_i : \mathbb{R}^M \to [0, +\infty)$ by

$$\Theta_i(\boldsymbol{h}) := \sum_{j \in \mathcal{J}_i} \omega_j d^2(\boldsymbol{h}, S_j), \tag{3}$$

where $\{\omega_j\}_{j \in \mathcal{J}_i} \in (0, 1]$, $\sum_{j \in \mathcal{J}_i} \omega_j = 1$ and $\mathcal{J}_i = \{\max(0, i - q + 1), \max(0, i - q + 1) + 1, \ldots, i\}$. Then choosing $\mathcal{K} = \mathbb{R}^M$ in (2) yields the scheme

$$(\forall i \in \mathbb{N}) \quad \boldsymbol{h}_i = \boldsymbol{h}_{i-1} + \mu\Bigg(\sum_{j \in \mathcal{J}_i} \omega_j P_{S_j}(\boldsymbol{h}_{i-1}) - \boldsymbol{h}_{i-1}\Bigg), \tag{4}$$

where $\mu \in [0, \mathcal{M}_i]$ and

$$\mathcal{M}_i := \begin{cases} \dfrac{\sum_{j \in \mathcal{J}_i} \omega_j \|P_{S_j}(\boldsymbol{h}_{i-1}) - \boldsymbol{h}_{i-1}\|^2}{\|\sum_{j \in \mathcal{J}_i} \omega_j P_{S_j}(\boldsymbol{h}_{i-1}) - \boldsymbol{h}_{i-1}\|^2} & \text{if } \boldsymbol{h}_{i-1} \notin \bigcap\limits_{j \in \mathcal{J}_i} S_j \\ 1 & \text{otherwise.} \end{cases}$$

Note that by using a different sequence of convex functions for (3), the step-size range presented above can be doubled [2]. The property set at each iteration in (4) is chosen to take the form of a hyperslab, which is defined as follows:

$$S_j := \{\boldsymbol{h} \in \mathbb{R}^M : |d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h}| \leq \rho\}, \quad j \in \mathbb{N}$$

where $\rho \geq 0$ and $S_j = \mathbb{R}^M$ if $\boldsymbol{u}_j = \boldsymbol{0}$. The projection onto the hyperslab $S_j$ is given by

$$P_{S_i}(\boldsymbol{h}) = \begin{cases} \boldsymbol{h} + \dfrac{d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h} - \rho}{\|\boldsymbol{u}_j\|^2}\boldsymbol{u}_j, & \text{if } d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h} > \rho \\ \boldsymbol{h} + \dfrac{d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h} + \rho}{\|\boldsymbol{u}_j\|^2}\boldsymbol{u}_j, & \text{if } d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h} < -\rho \\ \boldsymbol{h}, & \text{if } \boldsymbol{h} \in S_j \text{ or } \boldsymbol{u}_j = \boldsymbol{0}. \end{cases} \tag{5}$$

## 3. PERFORMANCE ANALYSIS

In order to apply energy-conservation arguments, we illustrate how the H-APSM satisfies a form that is similar to some existing adaptive filtering algorithms. Let $e_j(i) := d(j) - \boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{h}_{i-1}$. Then (5) can be written as

$$P_{S_j}(\boldsymbol{h}_{i-1}) = \boldsymbol{h}_{i-1} + \frac{e_j(i) - f_\rho(e_j(i))}{\|\boldsymbol{u}_j\|^2}\boldsymbol{u}_j,$$

where $f_\rho : \mathbb{R} \to [-\rho, \rho]$ is the continuous piecewise linear function defined by

$$f_\rho(x) := \frac{|x + \rho| - |x - \rho|}{2}.$$

Thus, we can express (4) in the form

$$\boldsymbol{h}_i = \boldsymbol{h}_{i-1} + \mu \sum_{j \in \mathcal{J}_i} \omega_j \frac{e_j(i) - f_\rho(e_j(i))}{\|\boldsymbol{u}_j\|^2 + \varepsilon}\boldsymbol{u}_j,$$

where we added a regularization parameter $\varepsilon > 0$ in the data normalization of the algorithm. Now, introduce the quantities

$$\begin{aligned} \boldsymbol{U}_i &= [\boldsymbol{u}_i \ \boldsymbol{u}_{i-1} \ \ldots \ \boldsymbol{u}_{i-q+1}] \\ \boldsymbol{d}_i &= [d(i) \ d(i-1) \ \ldots \ d(i-q+1)]^{\mathsf{T}} \\ \boldsymbol{W}_i &= \operatorname{diag}\{\omega_i, \omega_{i-1}, \ldots, \omega_{i-q+1}\} \\ \boldsymbol{N}_i &= \operatorname{diag}\{(\|\boldsymbol{u}_i\|^2 + \varepsilon)^{-1}, \ldots, (\|\boldsymbol{u}_{i-q+1}\|^2 + \varepsilon)^{-1}\} \end{aligned}$$

and define the vector-valued estimation error

$$\boldsymbol{e}_i = \boldsymbol{d}_i - \boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{h}_{i-1}.$$

Observe that using the quantities above, the H-APSM is now seen to satisfy the form

$$\boldsymbol{h}_i = \boldsymbol{h}_{i-1} + \mu\boldsymbol{U}_i\boldsymbol{\Pi}_i(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i)). \tag{6}$$

where $\boldsymbol{\Pi}_i = \boldsymbol{W}_i\boldsymbol{N}_i$. It is clear from recursion (6) that the H-APSM satisfies a description similar to other data-reusing adaptive algorithms [27]. The main difference is the presence of the function $f_\rho$ that operates component-wise on the estimation error $\boldsymbol{e}_i$. In relation to the terminologies in [25, 27, 28], one may thus consider the H-APSM as a *data-reusing algorithm with error nonlinearities*.

As the H-APSM is now seen to be similar to other data-reusing algorithms, we may now rely on and build upon the performance analyses developed for those schemes [26, 27]. Define the weight-error vector $\widetilde{\boldsymbol{h}}_i = \boldsymbol{h}^\circ - \boldsymbol{h}_i$. Then (6) can be rewritten in the form

$$\widetilde{\boldsymbol{h}}_i = \widetilde{\boldsymbol{h}}_{i-1} - \mu\boldsymbol{U}_i\boldsymbol{\Pi}_i(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i)). \tag{7}$$

Introduce the a priori and a posteriori weighted estimation errors

$$\boldsymbol{e}_{a,i} = \boldsymbol{U}_i^{\mathsf{T}}\widetilde{\boldsymbol{h}}_{i-1} \quad \text{and} \quad \boldsymbol{e}_{p,i} = \boldsymbol{U}_i^{\mathsf{T}}\widetilde{\boldsymbol{h}}_i.$$

Multiplying both sides of (7) by $\boldsymbol{U}_i^{\mathsf{T}}$ from the left, we find that

$$\boldsymbol{e}_{p,i} = \boldsymbol{e}_{a,i} - \mu\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i\boldsymbol{\Pi}_i(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i)). \tag{8}$$

Combining (7) and (8), we get

$$\widetilde{\boldsymbol{h}}_i + \boldsymbol{U}_i(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{a,i} = \widetilde{\boldsymbol{h}}_{i-1} + \boldsymbol{U}_i(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{p,i}.$$

Equating the weighted norms of both sides of the equation above results in the energy-conservation relation

$$\|\widetilde{\boldsymbol{h}}_i\|^2 + \boldsymbol{e}_{a,i}^{\mathsf{T}}(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{a,i} = \|\widetilde{\boldsymbol{h}}_{i-1}\|^2 + \boldsymbol{e}_{p,i}^{\mathsf{T}}(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{p,i}. \tag{9}$$

An important feature of the relation above is that it has been established without using any approximations. This exact relation describes how the weighted energies of the error quantities evolve in time. One also sees that equation (9) is the energy-conservation relation that is also satisfied by standard data-reusing adaptive algorithms [26, 27].

In the next sections we exploit relation (9) to study the performance of the H-APSM. In order to make the succeeding analysis tractable, we shall sometimes rely on several simplifying assumptions that are usually invoked in the literature, as can be seen in, e.g., [22, 25–27].

### 3.1. Mean Stability

Before we characterize the mean-square performance of the H-APSM, we first determine mean stability conditions by using the weight-error vector recursion (7).

Note that from (1), we have the relation

$$\boldsymbol{e}_i = \boldsymbol{e}_{a,i} + \boldsymbol{v}_i = \boldsymbol{U}_i^{\mathsf{T}}\widetilde{\boldsymbol{h}}_{i-1} + \boldsymbol{v}_i,$$

where $\boldsymbol{v}_i := [v(i)\ v(i-1)\ \ldots\ v(i-q+1)]^{\mathsf{T}}$. We can then express recursion (7) in the form

$$\widetilde{\boldsymbol{h}}_i = \widetilde{\boldsymbol{h}}_{i-1} - \mu\boldsymbol{U}_i\boldsymbol{\Pi}_i\boldsymbol{U}_i^{\mathsf{T}}\widetilde{\boldsymbol{h}}_{i-1} - \mu\boldsymbol{U}_i\boldsymbol{\Pi}_i\boldsymbol{v}_i + \mu\boldsymbol{U}_i\boldsymbol{\Pi}_i f_\rho(\boldsymbol{e}_i).$$

Taking expectations and assuming that

- the noise $v(i)$ is independent and identically distributed (iid) and statistically independent of the regression matrix $\{\boldsymbol{U}_i\}$,

we conclude that the mean of the weight-error vector evolves according to the following dynamics:

$$\mathbb{E}\widetilde{\boldsymbol{h}}_i = (\boldsymbol{I} - \mu\boldsymbol{\mathcal{P}})\mathbb{E}\widetilde{\boldsymbol{h}}_{i-1} + \mu\mathbb{E}[\boldsymbol{U}_i\boldsymbol{\Pi}_i f_\rho(\boldsymbol{e}_i)], \tag{10}$$

where

$$\boldsymbol{\mathcal{P}} := \mathbb{E}[\boldsymbol{U}_i\boldsymbol{\Pi}_i\boldsymbol{U}_i^{\mathsf{T}}].$$

Since $f_\rho(\boldsymbol{e}_i)$ has bounded entries, the algorithm converges in the mean if the matrix $\boldsymbol{I} - \mu\boldsymbol{\mathcal{P}}$ is a stable matrix. This holds if

$$\mu < \frac{2}{\lambda_{\max}(\boldsymbol{\mathcal{P}})},$$

where $\lambda_{\max}$ denotes the maximum eigenvalue of the symmetric matrix $\boldsymbol{\mathcal{P}}$. Now, note that we can simplify $\boldsymbol{\mathcal{P}}$ to

$$\boldsymbol{\mathcal{P}} = \mathbb{E}\left[\frac{\boldsymbol{u}_i\boldsymbol{u}_i^{\mathsf{T}}}{\varepsilon + \|\boldsymbol{u}_i\|^2}\right].$$

Thus, by invoking the analysis on the same matrix $\boldsymbol{\mathcal{P}}$ in [22] that is used for the study of the stability of the NLMS algorithm, we conclude that the H-APSM converges in the mean for any step-size $\mu < 2$. Furthermore, taking the limit of (10) as $i \to +\infty$, we get

$$\lim_{i\to+\infty}\mathbb{E}\boldsymbol{h}_i = \boldsymbol{h}^\circ - \boldsymbol{\mathcal{P}}^{-1}\lim_{i\to+\infty}\mathbb{E}[\boldsymbol{U}_i\boldsymbol{\Pi}_i f_\rho(\boldsymbol{e}_i)]. \tag{11}$$

Hence, we say that the H-APSM as an estimator is asymptotically biased. Note that if we set $\rho = 0$, the second term on the right-hand side of (11) vanishes. That is, if we instead use hyperplanes, i.e., hyperslabs with $\rho = 0$, as property sets, the H-APSM becomes asymptotically unbiased. The following theorem provides a characterization of the stability of the H-APSM in the mean sense.

**Theorem 1** *(Mean stability) Consider data model* (1)*. Then the hyperslab-based APSM* (6) *asymptotically converges in the mean for any initial condition given any step-size $\mu < 2$. Moreover, given any $\rho > 0$, the algorithm is asymptotically biased. Furthermore, the choice of weights $\{\omega_j\}$ does not affect the mean stability and convergence speed of the H-APSM in the mean sense.*

The observations described above regarding the effect of the weights and the asymptotic unbiasedness for $\rho = 0$ coincide with the result in [9] for the hyperplane-based APSM, a special case of the H-APSM used in that work in the context of adaptive networks. Note further that our result is not in conflict with the result established in [21], where a stochastic analysis of the H-APSM has been given, since convergence in the mean is a stronger condition that is not implied by convergence in probability.

### 3.2. Mean-Square Performance

We now derive an approximation to the steady-state performance of the H-APSM. In particular, our goal is to evaluate the steady-state mean-square error (MSE), denoted by $\xi$, which is given by

$$\xi := \lim_{i\to+\infty}\mathbb{E}e^2(i),$$

where

$$e(i) := d(i) - \boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{h}_{i-1}$$

is called the output estimation error at time $i$.

Taking the limit of both sides of the energy-conservation relation (9) and assuming the steady-state condition

$$\mathbb{E}\|\widetilde{\boldsymbol{h}}_i\|^2 = \mathbb{E}\|\widetilde{\boldsymbol{h}}_{i-1}\|^2 \quad \text{as } i \to +\infty,$$

we obtain the variance relation

$$\mathbb{E}[\boldsymbol{e}_{a,i}^{\mathsf{T}}(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{a,i}] = \mathbb{E}[\boldsymbol{e}_{p,i}^{\mathsf{T}}(\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i)^{-1}\boldsymbol{e}_{p,i}].$$

Substituting (8) into the right-hand side of the equation above yields

$$\mu\mathbb{E}[(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i))^{\mathsf{T}}\boldsymbol{A}_i(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i))] = 2\mathbb{E}[\boldsymbol{e}_{a,i}^{\mathsf{T}}\boldsymbol{\Pi}_i(\boldsymbol{e}_i - f_\rho(\boldsymbol{e}_i))], \tag{12}$$

where $\boldsymbol{A}_i := \boldsymbol{\Pi}_i\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{U}_i\boldsymbol{\Pi}_i$. If we neglect the dependency of $\widetilde{\boldsymbol{h}}_{i-1}$ on past noises, we find, after some straightforward manipulations, that

$$\begin{aligned}\mu\{&\mathbb{E}[\boldsymbol{e}_{a,i}^{\mathsf{T}}\boldsymbol{A}_i\boldsymbol{e}_{a,i}] + \mathbb{E}[\boldsymbol{v}_i^{\mathsf{T}}\boldsymbol{A}_i\boldsymbol{v}_i] - 2\mathbb{E}[f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}\boldsymbol{A}_i\boldsymbol{e}_{a,i}]\\ &-2\mathbb{E}[f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}\boldsymbol{A}_i\boldsymbol{v}_i] + \mathbb{E}[f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}\boldsymbol{A}_i f_\rho(\boldsymbol{e}_i)]\}\\ &= 2\mathbb{E}[\boldsymbol{e}_{a,i}^{\mathsf{T}}\boldsymbol{\Pi}_i\boldsymbol{e}_i] - 2\mathbb{E}[\boldsymbol{e}_{a,i}^{\mathsf{T}}\boldsymbol{\Pi}_i f_\rho(\boldsymbol{e}_i)].\end{aligned} \tag{13}$$

In order for us to evaluate the steady-state MSE of the H-APSM, we need to deal with the expectations above. To handle this task we rely on and build upon a simplifying assumption used in the analysis of data-reusing algorithms, in particular, of the APA [26]. That is, we assume at steady-state and for sufficiently small step-sizes that,

- $\boldsymbol{U}_i$ is statistically independent of $\boldsymbol{e}_{a,i}$,
- $\mathbb{E}[\boldsymbol{e}_{a,i}\boldsymbol{e}_{a,i}^{\mathsf{T}}] \approx \mathbb{E}e_a^2(i)\cdot\boldsymbol{I}$, where $e_a(i) := \boldsymbol{u}_i^{\mathsf{T}}\widetilde{\boldsymbol{h}}_{i-1}$,
- $\mathbb{E}[\boldsymbol{v}_i\boldsymbol{v}_i^{\mathsf{T}}] \approx \sigma_v^2\cdot\boldsymbol{I}$,
- $\mathbb{E}[\boldsymbol{e}_{a,i}f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}] \approx \mathbb{E}[e_a(i)f_\rho(e(i))]\cdot\boldsymbol{I}$,
- $\mathbb{E}[\boldsymbol{v}_i f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}] \approx \mathbb{E}[v(i)f_\rho(e(i))]\cdot\boldsymbol{I}$,
- $\mathbb{E}[f_\rho(\boldsymbol{e}_i)f_\rho(\boldsymbol{e}_i)^{\mathsf{T}}] \approx \mathbb{E}[f_\rho^2(e(i))]\cdot\boldsymbol{I}$.

Observe that we considered a simple form of the assumption in [26] to make the resulting expressions simpler. In particular, we note that other approximations for the matrices above can also be motivated as in [26]. The expressions for the expectations involving $f_\rho$ on the other hand can be computed directly as in [29]. Assuming $v(i)$ is white Gaussian, the following expressions can be obtained:

$$\mathbb{E}[v(i)f_\rho(e(i))] = \sigma_v^2\mathrm{erf}\left(\frac{\rho}{\sqrt{2\mathbb{E}e^2(i)}}\right)$$

$$\mathbb{E}[e_a(i)f_\rho(e(i))] = \mathbb{E}e_a^2(i)\cdot\mathrm{erf}\left(\frac{\rho}{\sqrt{2\mathbb{E}e^2(i)}}\right)$$

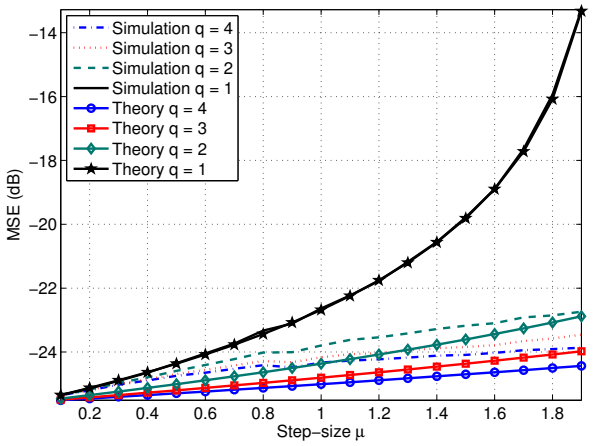


**Fig. 1**: Theoretical and simulated MSE versus step-size curves for the hyperslab-based APSM using correlated Gaussian input with $M = 256$, $\phi = 0.8$, and $\rho = \sigma_v^2$.

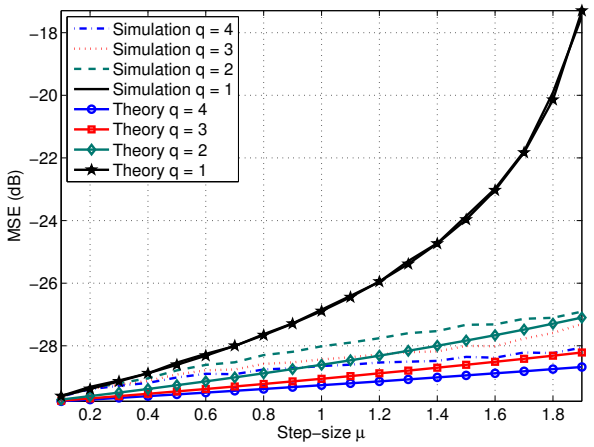


**Fig. 2**: Theoretical and simulated MSE versus step-size curves for the hyperslab-based APSM using correlated Gaussian input with $M = 16$, $\phi = 0.2$, and $\rho = \sigma_v^2$.

$$\mathbb{E}[f_\rho^2(e(i))] = \mathbb{E}e^2(i)\mathrm{erf}\left(\frac{\rho}{\sqrt{2\mathbb{E}e^2(i)}}\right) + \rho^2 - \rho^2\mathrm{erf}\left(\frac{\rho}{\sqrt{2\mathbb{E}e^2(i)}}\right) - \rho\sqrt{\frac{2\mathbb{E}e^2(i)}{\pi}}\exp\left(\frac{-\rho^2}{2\mathbb{E}e^2(i)}\right)$$

Using the property that $\mathrm{Tr}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}) = \mathrm{Tr}(\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{X}})$ for any matrices $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ of compatible dimensions, employing the assumptions regarding the expectations in (13), and noting that $\mathbb{E}e^2(i) = \mathbb{E}e_a^2(i) + \sigma_v^2$, we obtain the following result.

**Theorem 2** *(Steady-state performance) Consider data model* (1)*. Then an approximation to the steady-state MSE $\xi$ of the hyperslab-based APSM* (6) *is the unique solution of the nonlinear equation*

$$\xi = \sigma_v^2 + \Gamma_\mu\left[\sigma_v^2 + \rho^2 - \rho\sqrt{\frac{2\xi}{\pi}}\frac{\exp(-\rho^2/(2\xi))}{1-\mathrm{erf}(\rho/\sqrt{2\xi})}\right], \qquad (14)$$

*where*

$$\Gamma_\mu := \frac{\mu\mathrm{Tr}(\mathbb{E}\boldsymbol{A}_i)}{2\mathrm{Tr}(\mathbb{E}\boldsymbol{\Pi}_i) - \mu\mathrm{Tr}(\mathbb{E}\boldsymbol{A}_i)}.$$

**Remark 1** *Observe that if $q = 1$, i.e., we only consider a single hyperslab,* (14) *reduces to the steady-state MSE equation for the SM-NLMS algorithm [29]. Also, using the analyses presented in [29], one can conclude that the steady-state MSE $\xi$ is a monotonically increasing function of the step-size $\mu$ and is a monotonically decreasing function of $\rho$.*

## 4. SIMULATION RESULTS

We now present simulation results to illustrate the theoretical results presented in this paper. Consider the input vector $\boldsymbol{u}_i$ to have a shift structure with entries generated by passing an iid Gaussian process $\{x(i)\}_{i\in\mathbb{N}}$ with unit variance through the model

$$u(i) = \phi u(i-1) + x(i), \quad i \in \mathbb{N},$$

which is a first-order autoregressive process with a pole at $\phi$. The output $d(i)$ is contaminated using a white Gaussian process $v(i)$ with its variance set so that its signal-to-noise ratio is 30 dB. The original unknown system is generated randomly and all adaptive filter coefficients are initialized to zero. The regularization parameter $\varepsilon$ is fixed to $10^{-3}$. All simulation results are obtained by ensemble averaging over 100 independent trials.

We simulate the steady-state behavior of the H-APSM for $q = 1, 2, 3, 4$ hyperslabs and compare the results with the theory as described in Theorem 2. We consider an unknown system $\boldsymbol{h}^\circ$ with tap lengths $M = 256$ and $M = 16$ using poles $\phi = 0.8$ and $\phi = 0.2$, respectively. The experimental values of the steady-state MSE are calculated by averaging the last 200 samples after $10^5$ iterations. The theoretical steady-state MSE values are obtained numerically from (14). The weighting used for the implementation of the H-APSM is based on simple averaging, i.e., $\omega_j = 1/q$ for every $j$.

We see that, as discussed in Remark 1, the steady-state MSE is a monotonically increasing function of the step-size $\mu$. Also, one notices that given a fixed step-size, a lower steady-state MSE $\xi$ is obtained when one uses a larger number of hyperslabs. In addition to this feature, the low complexity of H-APSM compared to other data-reusing algorithms makes it attractive for many applications. Moreover, observe that although increasing $q$ results in an increase in computational cost, the improvement in the steady-state MSE is significant, especially for the first few increases in the number of hyperslabs. Although the mean-square performance analysis is a challenging task (see, e.g., [22, 26–28, 39]), and despite the use of restrictive simplifying assumptions, colored input or large values of the step-size, we see that we still have a fairly close agreement between theory and practice.

## 5. CONCLUDING REMARKS

We have presented the first mean-square performance evaluation of the hyperslab-based adaptive projected subgradient method. Using energy-conservation arguments, an expression involving the steady-state mean-square error has been derived without restricting the input data to a particular distribution. Simulations supported the theoretical results. Due to space limitations, we have only presented the steady-state performance of the H-APSM. The characterization of its transient behavior will be considered elsewhere.

## 6. REFERENCES


[1] I. Yamada, "Adaptive projected subgradient method: a unified view for projection based adaptive algorithms," *J. IEICE (in Japanese)*, vol. 86, no. 8, pp. 654–658, 2003.

[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] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[4] ——, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 493–506, 1997.

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

[6] K. Slavakis, S. Theodoridis, and I. Yamada, "Adaptive constrained learning in reproducing kernel Hilbert spaces: the robust beamforming case," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4744–4764, 2009.

[7] K. Slavakis, Y. Kopsinis, and S. Theodoridis, "Adaptive algorithm for sparse system identification using projections onto weighted $\ell_1$ balls," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 3742–3745.

[8] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 97–123, 2011.

[9] R. L. G. Cavalcante, I. Yamada, and B. Mulgrew, "An adaptive projected subgradient approach to learning in diffusion networks," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2762–2774, Jul. 2009.

[10] R. L. G. Cavalcante, A. Rogers, N. R. Jennings, and I. Yamada, "Distributed asymptotic minimization of sequences of convex functions by a broadcast adaptive subgradient method," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 739–753, Aug. 2011.

[11] R. L. G. Cavalcante and S. Stańczak, "Robust set-theoretic distributed detection in diffusion networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 3757–3760.

[12] ——, "A distributed subgradient method for dynamic convex optimization problems under noisy information exchange," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 243–256, 2013.

[13] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4692–4707, Oct. 2011.

[14] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, "A sparsity promoting adaptive algorithm for distributed learning," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5412–5425, Oct. 2012.

[15] R. L. G. Cavalcante, S. Stańczak, and I. Yamada, "Cooperative Cognitive Radios with Diffusion Networks," in *Mechanisms and Games for Dynamic Spectrum Allocation*, T. Alpcan, H. Boche, M. Honig, and H. V. Poor, Eds. Cambridge University Press, 2013.

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

[17] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS Algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, 1998.

[18] T. Hinamoto and S. Maekawa, "Extended theory of learning identification," *Trans. Inst. Elect. Eng. Jpn.*, vol. 95, no. 10, pp. 227–234, 1975.

[19] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *IEICE Trans. (in Japanese)*, vol. J67-A, no. 2, pp. 126–132, 1984.

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

[21] S. Chouvardas, K. Slavakis, S. Theodoridis, and I. Yamada, "Stochastic analysis of hyperslab-based adaptive projected subgradient method under bounded noise," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 729–732, 2013.

[22] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[23] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 314–324, 2001.

[24] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 639–652, 2003.

[25] ——, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, 2003.

[26] H.-C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 90–102, 2004.

[27] H.-C. Shin, W.-J. Song, and A. H. Sayed, "Mean-square performance of data-reusing adaptive algorithms," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 851–854, 2005.

[28] W. M. Wee and I. Yamada, "Tracking analysis of adaptive filters with data normalization and error nonlinearities," *IEICE Trans. Fundamentals*, vol. E96-A, no. 11, pp. 2198–2208, Nov. 2013.

[29] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, Sep. 2009.

[30] R. L. G. Cavalcante, I. Yamada, and K. Sakaniwa, "A fast blind multiple access interference reduction in DS/CDMA systems based on adaptive projected subgradient method," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1973–1980, 2004.

[31] R. L. G. Cavalcante and I. Yamada, "Multiaccess interference suppression in orthogonal space-time block coded MIMO systems by adaptive projected subgradient method," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1028–1042, 2008.

[32] K. Slavakis, S. Theodoridis, and I. Yamada, "Online kernel-based classification using adaptive projection algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2781–2796, 2008.

[33] 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, 2005.

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

[35] M. Yukawa, K. Slavakis, and I. Yamada, "Adaptive parallel quadratic-metric projection algorithms," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1665–1680, 2007.

[36] M. Yukawa and I. Yamada, "A unified view of adaptive variable-metric projection algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, 2009.

[37] M. Yukawa, R. C. de Lamare, and I. Yamada, "Robust reduced-rank adaptive algorithm based on parallel subgradient projection and Krylov subspace," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4660–4674, 2009.

[38] M. Yukawa, K. Slavakis, and I. Yamada, "Multi-domain adaptive learning based on feasibility splitting and adaptive projected subgradient method," *IEICE Trans. Fundamentals*, vol. E93-A, no. 2, pp. 456–466, 2010.

[39] S. G. Sankaran and A. A. L. Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1086–1096, 2000.