STEADY-STATE PERFORMANCE OF HYPERSLAB PROJECTION ALGORITHM

Noriyuki Takahashi and *Isao Yamada*

Department of Communications and Integrated Systems, Tokyo Institute of Technology
2-12-1-S3-60 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
Email: {takahashi,isao}@comm.ss.titech.ac.jp

ABSTRACT

This paper presents an analysis of the steady-state mean-square error of an adaptive filtering algorithm using the metric projection onto a closed hyperslab, which we refer to as the *hyperslab projection algorithm (HSPA)*. HSPA is not only a generalization of both the *normalized least mean square (NLMS)* algorithm and the *set-membership NLMS (SM-NLMS)* algorithm but also a special case of the *adaptive parallel subgradient projection (PSP)* method. It is known that HSPA possesses both fast convergence and robustness against noise. The approach of this paper is to employ the *energy conservation relation*, which enables us to avoid the transient analysis of HSPA. Under different assumptions, we obtain two results, which are generalizations of well-known results of the steady-state performance of NLMS. Extensive simulations show the good match between the theories and experiments.

Index Terms— Adaptive filters, steady-state performance, projection, hyperslab, energy conservation relation

1. INTRODUCTION

Adaptive filters based on orthogonal projections, such as the *normalized least mean square (NLMS)* algorithm and the *affine projection algorithm (APA)*, are widely used as low computational cost algorithms, and there are many works on performance analyses of these algorithms; see e.g. [1–3] and references therein. It is also known that sensitivities to noise of these algorithms are overcame in the framework of the *set-theoretic (set-membership) adaptive filtering* [4–8], where orthogonal projections are replaced by *convex projections*. However, in general, statistical performance analyses of set-theoretic algorithms are challenging problems due to nonlinearities of convex projections.

In this paper, we tackle a steady-state performance analysis of an adaptive algorithm using the projection onto a closed *hyperslab*, which we refer to as the *hyperslab projection algorithm (HSPA)*. HSPA might be the simplest set-theoretic adaptive algorithm as well as a slight generalization of the *set-membership NLMS (SM-NLMS)* algorithm [4, 6], which is a generalization of NLMS. In addition, HSPA is also a special case of the *adaptive parallel subgradient projection (PSP)* algorithm [5], which uses multiple hyperslabs simultaneously. Although it has been experimentally shown that SM-NLMS and PSP possess good steady-state performances and robustness against noise [4–6], there are few statistical analysis on the steady-state performances of these algorithms (an available analysis can be found in [7]). This motivates us to analyze the steady-state performance of HSPA statistically.

The approach of this paper is to employ the *energy conservation relation* [1–3], which is one of powerful techniques for analyses of adaptive algorithms. This technique enables us to avoid a transient analysis of HSPA. Since HSPA is a generalization of SM-NLMS and a special example of PSP, our result covers not only the steady-state performance of SM-NLMS but also a partial analysis of that of PSP.

2. PRELIMINARIES

2.1. Notation

Throughout the paper, we use the following notation: Let $\mathbb{R}$ denote the set of all real numbers. Vectors and matrices are denoted by bold faces. The Euclidean norm and the transpose of a vector are denoted by $\|\cdot\|$ and $(\cdot)^t$, respectively. The trace of a matrix is denoted by $\mathrm{tr}(\cdot)$. Expectation is denoted by $\mathrm{E}[\,\cdot\,]$.

2.2. Data Model and Hyperslab Projection Algorithm (HSPA)

We consider noisy measurements $d_i \in \mathbb{R}$ $(i = 0, 1, \dots)$ that arise from the following real-valued linear model:

$$d_i = \boldsymbol{u}_i^t \boldsymbol{w}^{\mathrm{o}} + v_i, \quad i = 0, 1, \dots, \tag{1}$$

where $\boldsymbol{w}^{\mathrm{o}} \in \mathbb{R}^N$ is an unknown column vector we wish to estimate, $\{\boldsymbol{u}_i\}_{i\geq 0} \subset \mathbb{R}^N$ is a sequence of input (regressor) vectors, and $\{v_i\}_{i\geq 0} \subset \mathbb{R}$ accounts for measurement noise and modeling errors. Both $\{\boldsymbol{u}_i\}_{i\geq 0}$ and $\{v_i\}_{i\geq 0}$ are stochastic processes, and we assume that $\{v_i\}_{i\geq 0}$ is white and independent of $\{\boldsymbol{u}_i\}_{i\geq 0}$; see (a) in Table 1.

Since v_i is zero-mean, by choosing a parameter $\rho \geq 0$ appropriately, the event $|d_i - \boldsymbol{u}_i^t \boldsymbol{w}^{\mathrm{o}}| \leq \rho$ occurs with high probability. Hence, $\boldsymbol{w}^{\mathrm{o}}$ is highly expected to belong to the following *hyperslabs*:

$$S_i := \{\boldsymbol{w} \in \mathbb{R}^N : |d_i - \boldsymbol{u}_i^t \boldsymbol{w}| \leq \rho\}, \quad i = 0, 1, \dots.$$

The strategy of HSPA is to move its weight vector $\boldsymbol{w}_i \in \mathbb{R}^N$ closer to all points in S_i at each time i. This is achieved by using the metric projection P_{S_i} from $\mathbb{R}^N$ onto S_i as follows:

Algorithm 1 (HSPA, see Fig. 1**).** With an arbitrary initial estimate $\boldsymbol{w}_0 \in \mathbb{R}^N$, generate a sequence $\{\boldsymbol{w}_i\}_{i\geq 0}$ by

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\big(P_{S_i}(\boldsymbol{w}_i) - \boldsymbol{w}_i\big), \quad \forall i = 0, 1, \dots,$$

or equivalently

$$\boldsymbol{w}_{i+1} = \begin{cases} \boldsymbol{w}_i & \text{if } |e_i| \leq \rho, \\ \boldsymbol{w}_i + \mu \dfrac{e_i - \rho}{\|\boldsymbol{u}_i\|^2} \boldsymbol{u}_i & \text{if } e_i > \rho, \\ \boldsymbol{w}_i + \mu \dfrac{e_i + \rho}{\|\boldsymbol{u}_i\|^2} \boldsymbol{u}_i & \text{if } e_i < -\rho, \end{cases} \tag{2}$$

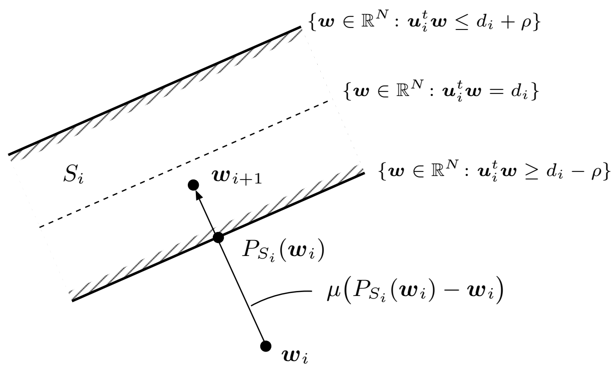

Fig. 1. Geometrical interpretation of HSPA.

where $\mu \in (0, 2)$ is a relaxation parameter (stepsize) and $e_i \in \mathbb{R}$ is the output estimation error at time i, i.e.,

$$e_i := d_i - \boldsymbol{u}_i^t \boldsymbol{w}_i, \quad i = 0, 1, \ldots.$$

Here, introducing the function $f_\rho : \mathbb{R} \to \mathbb{R}$ defined as

$$f_\rho(x) := \frac{|x+\rho| - |x-\rho|}{2}, \quad \forall x \in \mathbb{R},$$

we can express (2) without if-statements as follows:

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \mu\big(e_i - f_\rho(e_i)\big)\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|^2}, \quad \forall i = 0, 1, \ldots. \tag{3}$$

It is simply verified that (2) and (3) are mathematically equivalent. From (3), we can see that HSPA involves the data-normalization [2] and the error-nonlinearity [3]. The function f_ρ characterizes the nonlinearity of HSPA attributed to the parameter ρ.

There is a close relation between HSPA and other projection-type algorithms, which is summarized as follows:

Remark 1. NLMS corresponds to HSPA with $\rho = 0$, while SM-NLMS [4, 6] corresponds to HSPA with $\mu = 1$. Thus, HSPA is a generalization of these algorithms. On the other hand, HSPA is a special example of the adaptive PSP [5], in which multiple hyperslabs are used simultaneously.

Although it has been experimentally shown that HSPA has a desirable performance, how the parameters (μ, ρ) affect its steady-state performance remains an open question. The objective of the paper is to clarify the relation between the parameters and the steady-state performance.

3. STEADY-STATE PERFORMANCE ANALYSIS

We now perform the steady-state performance analysis of HSPA (Algorithm 1). Assumptions used in arguments are listed in Table 1. Due to space limitation, we only highlight main steps.

We are interested in evaluating the steady-state *mean-square error (MSE)*, which is defined as

$$\mathsf{MSE} := \lim_{i\to\infty} \mathrm{E}|e_i|^2.$$

To evaluate this, we introduce the following error measures:

$$\widetilde{\boldsymbol{w}}_i := \boldsymbol{w}^{\mathrm{o}} - \boldsymbol{w}_i, \qquad e_i^{\mathrm{a}} := \boldsymbol{u}_i^t \widetilde{\boldsymbol{w}}_i, \qquad e_i^{\mathrm{p}} := \boldsymbol{u}_i^t \widetilde{\boldsymbol{w}}_{i+1},$$

Table 1. Assumptions used in the steady-state analysis.

(a). The noise $\{v_i\}_{i\ge 0}$ is white with a variance σ^2 and statistically independent of $\{\boldsymbol{u}_i\}_{i\ge 0}$.

(b). For all $i \ge i_{\mathrm{s}}$, the random variable $\|\boldsymbol{u}_i\|^{-2}$ is uncorrelated with all random variables $|e_i^{\mathrm{a}}|^2$, $e_i^{\mathrm{a}} f_\rho(e_i)$, $e_i f_\rho(e_i)$, and $f_\rho^2(e_i)$.

(c). The noise $\{v_i\}_{i\ge 0}$ is Gaussian.

(d). For all $i \ge i_{\mathrm{s}}$, the random variable e_i^{a} is zero-mean Gaussian (i.e, its variance is $\mathrm{E}|e_i^{\mathrm{a}}|^2$).

(e). The sequences $\{\mathrm{E}|e_i|^2\}_{i\ge 0}$ and $\{\mathrm{E}\|\widetilde{\boldsymbol{w}}_i\|^2\}_{i\ge 0}$ are convergent.

(f). The following equalities hold:

$$\mathrm{E}\left[\frac{|e_i^{\mathrm{a}}|^2}{\|\boldsymbol{u}_i\|^2}\right] = \frac{\mathrm{E}|e_i^{\mathrm{a}}|^2}{\mathrm{E}\|\boldsymbol{u}_i\|^2}, \qquad \mathrm{E}\left[\frac{e_i f_\rho(e_i)}{\|\boldsymbol{u}_i\|^2}\right] = \frac{\mathrm{E}\big[e_i f_\rho(e_i)\big]}{\mathrm{E}\|\boldsymbol{u}_i\|^2},$$
$$\mathrm{E}\left[\frac{e_i^{\mathrm{a}} f_\rho(e_i)}{\|\boldsymbol{u}_i\|^2}\right] = \frac{\mathrm{E}\big[e_i^{\mathrm{a}} f_\rho(e_i)\big]}{\mathrm{E}\|\boldsymbol{u}_i\|^2}, \qquad \mathrm{E}\left[\frac{f_\rho^2(e_i)}{\|\boldsymbol{u}_i\|^2}\right] = \frac{\mathrm{E}\big[f_\rho^2(e_i)\big]}{\mathrm{E}\|\boldsymbol{u}_i\|^2}.$$

where $\widetilde{\boldsymbol{w}}_i$ is the weight error vector, e_i^{a} is the a priori error, and e_i^{p} is the a posteriori error. Note that it follows from (1) that

$$e_i = e_i^{\mathrm{a}} + v_i, \quad \forall i = 0, 1, \ldots. \tag{4}$$

Hence, under the assumption (a) in Table 1, we have

$$\mathrm{E}|e_i|^2 = \mathrm{E}|e_i^{\mathrm{a}}|^2 + \sigma^2, \quad \forall i = 0, 1, \ldots. \tag{5}$$

In view of this, MSE can be evaluated through

$$\mathsf{EMSE} := \lim_{i\to\infty} \mathrm{E}|e_i^{\mathrm{a}}|^2,$$

which is called the *excess mean square error (EMSE)*.

Now, subtracting (3) from $\boldsymbol{w}^{\mathrm{o}}$, we have

$$\widetilde{\boldsymbol{w}}_{i+1} = \widetilde{\boldsymbol{w}}_i - \mu\big(e_i - f_\rho(e_i)\big)\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|^2}, \quad \forall i = 0, 1, \ldots. \tag{6}$$

Furthermore, premultiplying (6) by $\boldsymbol{u}_i^t$, we obtain

$$e_i^{\mathrm{p}} = e_i^{\mathrm{a}} - \mu\big(e_i - f_\rho(e_i)\big), \quad \forall i = 0, 1, \ldots. \tag{7}$$

We here apply the *energy conservation relation* [1], which ensures the following equality for all $i = 0, 1, \ldots$:

$$\|\widetilde{\boldsymbol{w}}_{i+1}\|^2 + \big(\|\boldsymbol{u}_i\|^2\big)^\dagger |e_i^{\mathrm{a}}|^2 = \|\widetilde{\boldsymbol{w}}_i\|^2 + \big(\|\boldsymbol{u}_i\|^2\big)^\dagger |e_i^{\mathrm{p}}|^2, \tag{8}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse.[1] It should be noted here that no assumptions or approximations are used to establish (8); see for details [1–3] and references therein. Substituting (7) into (8) and taking the expectation under the assumption (a) in Table 1, we obtain

$$(2-\mu)\,\mathrm{E}\left[\frac{|e_i^{\mathrm{a}}|^2}{\|\boldsymbol{u}_i\|^2}\right] = \mu\sigma^2\,\mathrm{E}\left[\frac{1}{\|\boldsymbol{u}_i\|^2}\right] - 2\mu\,\mathrm{E}\left[\frac{e_i f_\rho(e_i)}{\|\boldsymbol{u}_i\|^2}\right] + 2\,\mathrm{E}\left[\frac{e_i^{\mathrm{a}} f_\rho(e_i)}{\|\boldsymbol{u}_i\|^2}\right] + \mu\,\mathrm{E}\left[\frac{f_\rho^2(e_i)}{\|\boldsymbol{u}_i\|^2}\right] + \delta_i \tag{9}$$

[1] The pseudo-inverse of a scalar x is defined as follows: $x^\dagger = x^{-1}$ if $x \ne 0$ and $x^\dagger = 0$ if $x = 0$.

for all $i = 0, 1, \ldots$, where

$$\delta_i := \frac{1}{\mu}\big(\mathrm{E}\|\widetilde{\boldsymbol{w}}_i\|^2 - \mathrm{E}\|\widetilde{\boldsymbol{w}}_{i+1}\|^2\big), \quad \forall i = 0, 1, \ldots.$$

So far, we have established (9) only under the assumption (a) in Table 1, which could be natural in practice. However, in general, it is hard to evaluate the expectations in (9). To overcome this difficulty, we use two different assumptions.

3.1. Uncorrelation Assumption

In what follows, we focus on the steady-state, which is regarded as time i such that $i \geq i_{\mathrm{s}}$ for a sufficiently large i_{s}. We first use the assumption (b) in Table 1. Although in general this assumption might not hold, it helps us to make (9) much simpler as follows:

$$\begin{aligned}(2-\mu)\,\mathrm{E}|e_i^{\mathrm{a}}|^2 = \mu\sigma^2 - 2\mu\,\mathrm{E}\big[e_i f_\rho(e_i)\big] + 2\,\mathrm{E}\big[e_i^{\mathrm{a}} f_\rho(e_i)\big] \\ + \mu\,\mathrm{E}\big[f_\rho^2(e_i)\big] + \delta_i\,\mathrm{E}\left[\frac{1}{\|\boldsymbol{u}_i\|^2}\right]^{-1}\end{aligned} \quad (10)$$

for all $i \geq i_{\mathrm{s}}$. To evaluate the expectations on the right hand side, we further assume Gaussianities of v_i and e_i^{a}; see (c) and (d) in Table 1. Note that, in view of (4), e_i is also Gaussian. Hence, $\mathrm{E}[e_i f_\rho(e_i)]$ and $\mathrm{E}[f_\rho^2(e_i)]$ can be evaluated by direct calculation. On the other hand, since e_i^{a} and e_i are jointly Gaussian, $\mathrm{E}[e_i^{\mathrm{a}} f_\rho(e_i)]$ can be evaluated by using *Price's theorem*; see [1, 3] for details. Calculation results are as follows:

$$\left.\begin{aligned}\mathrm{E}\big[e_i f_\rho(e_i)\big] &= \mathrm{E}|e_i|^2\,\mathrm{erf}\left(\frac{\rho}{\sqrt{2\,\mathrm{E}|e_i|^2}}\right)\\ \mathrm{E}\big[e_i^{\mathrm{a}} f_\rho(e_i)\big] &= \mathrm{E}|e_i^{\mathrm{a}}|^2\,\mathrm{erf}\left(\frac{\rho}{\sqrt{2\,\mathrm{E}|e_i|^2}}\right)\\ \mathrm{E}\big[f_\rho^2(e_i)\big] &= (\mathrm{E}|e_i|^2 - \rho^2)\,\mathrm{erf}\left(\frac{\rho}{\sqrt{2\,\mathrm{E}|e_i|^2}}\right)\\ &\quad - \rho\sqrt{\frac{2\,\mathrm{E}|e_i|^2}{\pi}}\exp\left(-\frac{\rho^2}{2\,\mathrm{E}|e_i|^2}\right) + \rho^2\end{aligned}\right\} \quad (11)$$

where $\mathrm{erf}\colon \mathbb{R} \to \mathbb{R}$ is the *error function* defined as

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt, \quad \forall x \in \mathbb{R}.$$

Finally, we use the assumption (e) in Table 1, which is natural if the filter reaches its steady-state. Note that $\delta_i \to 0$ as $i \to \infty$ under this assumption. Combining (5), (10), and (11) and letting $i \to \infty$, we obtain the following result:

Theorem 1. *Under the assumptions* (a)–(e) *in Table 1, the steady-state MSE, say ξ, of HSPA is given as a solution of the following system:*

$$\begin{cases}\xi = \dfrac{2\sigma^2 + \mu\rho^2}{2-\mu} - \dfrac{\mu\rho}{2-\mu}\sqrt{\dfrac{2\xi}{\pi}}\,\dfrac{\exp\left(-\frac{\rho^2}{2\xi}\right)}{1-\mathrm{erf}\left(\frac{\rho}{\sqrt{2\xi}}\right)};\\ \xi \geq \sigma^2.\end{cases}$$

The steady-state EMSE of HSPA is given by $\xi - \sigma^2$.

3.2. Separation Assumption

We next use the assumption (f) instead of (b) in Table 1. Similar assumptions have often been used; see e.g. [1]. In a way similar to the previous subsection, applying the assumption (f) to (9) and using (11), we obtain the following result:

Theorem 2. *Let $\boldsymbol{R} := \mathrm{E}[\boldsymbol{u}_i\boldsymbol{u}_i^t]$ be a correlation matrix of the input vector and let*

$$C := \mathrm{tr}(\boldsymbol{R})\,\mathrm{E}\left[\frac{1}{\|\boldsymbol{u}_i\|^2}\right].$$

Then, under the assumptions (a) *and* (c)–(f) *in Table 1, the steady-state MSE, say ξ, of HSPA is given as a solution of the following system:*

$$\begin{cases}\xi = \dfrac{2\sigma^2 + \mu\rho^2}{2-\mu} - \dfrac{\mu}{2-\mu}\cdot\dfrac{(1-C)\sigma^2 + \rho\sqrt{\frac{2\xi}{\pi}}\exp\left(-\frac{\rho^2}{2\xi}\right)}{1-\mathrm{erf}\left(\frac{\rho}{\sqrt{2\xi}}\right)};\\ \xi \geq \sigma^2.\end{cases}$$

The steady-state EMSE of HSPA is given by $\xi - \sigma^2$.

This result reflects the statistics of $\{\boldsymbol{u}_i\}_{i\geq 0}$, while Theorem 1 is independent of it. Finally, we mention the following:

Remark 2. When $\rho = 0$ (i.e., HSPA corresponds to NLMS), Theorems 1 and 2 reproduce well-known analysis results of NLMS; see e.g. [1]. Thus, our results are natural extensions of these results.

4. SIMULATION RESULTS

In this section, we compare experimental steady-state MSEs of HSPA with Theorems 1 and 2 in the following settings: The unknown vector $\boldsymbol{w}^{\mathrm{o}}$ has 16 taps and is randomly generated. The input vector has a shift structure, i.e.,

$$\boldsymbol{u}_i := (u_i, u_{i-1}, \ldots, u_{i-15})^t,$$

where we set $u_i = 0$ if $i < 0$. For the input $\{u_i\}$, we consider three cases: (i) colored Gaussian, (ii) correlated uniform, and (iii) binary signals. The colored Gaussian and correlated uniform signals are generated by passing white Gaussian and uniform processes $\{x_i\}_{i\geq 0}$ through a first-order autoregressive (AR(1)) model:

$$u_i = a u_{i-1} + x_i, \quad \forall i = 0, 1, \ldots,$$

where $a \in [0, 1)$ is a pole of the AR(1) model. For the colored Gaussian input, $\{x_i\}_{i\geq 0}$ is white Gaussian with unit variance and a is set to 0.8, while for the correlated uniform input the process $\{x_i\}_{i\geq 0}$ is iid uniformly distributed on $[-1, 1]$ and a is set to 0.5. For the binary input, $\{u_i\}_{i\geq 0}$ is iid uniformly distributed on $\{\pm 1\}$. The noise $\{v_i\}_{i\geq 0}$ is white Gaussian and its variance σ^2 is set so that the signal-to-noise ratio (SNR) is 10 dB.

For the parameters of HSPA, a wide range of the stepsize μ and $\rho \in \{0, \sigma, 2\sigma\}$ are tested for each input. All results are obtained by ensemble averaging 100 independent trials, and the steady-state MSE of each trial is obtained by averaging last 10^4 samples of $|e_i|^2$ after 10^5 iterations. Figs. 2, 3, and 4 show results for the colored Gaussian, correlated uniform, and binary inputs, respectively. These figures show good matches between Theorem 1 and experiments. However, big mismatches between Theorem 2 and the experiments are observed when ρ is large except for the binary input. This is

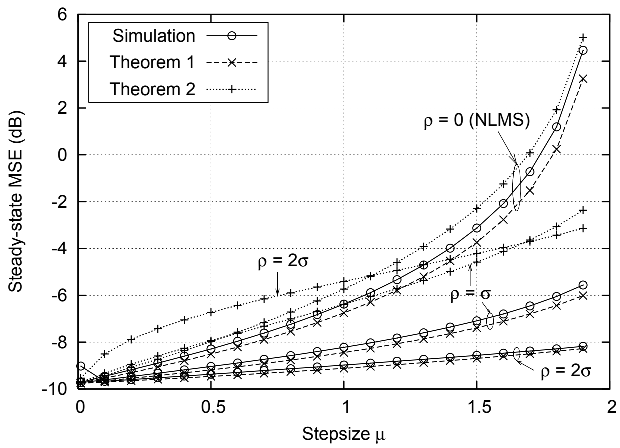

Fig. 2. Comparison between the theories and experiments for the Gaussian AR(1) with pole at $a = 0.8$.

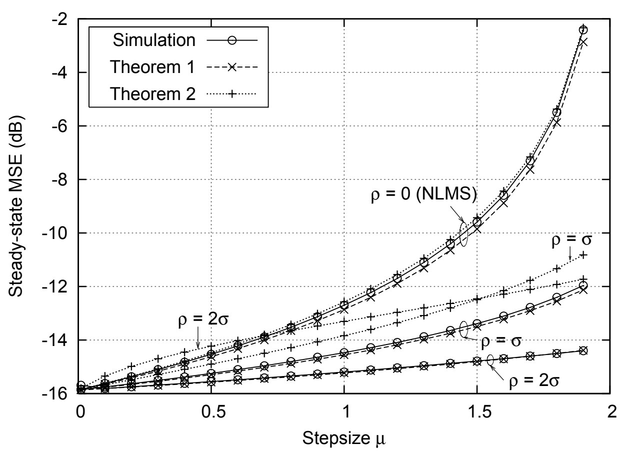

Fig. 3. Comparison between the theories and experiments for the uniform AR(1) with pole at $a = 0.5$.

because the approximation error of (f) in Table 1 increases as ρ becomes larger. For the binary input, we can see that Theorems 1 and 2 are almost perfect. This is because the assumptions (b) and (f) in Table 1 hold exactly.

5. CONCLUDING REMARKS

We have analyzed the steady-state MSE and EMSE of HSPA (Algorithm 1) by using the energy conservation relation and presented two results based on different assumptions; see Theorems 1 and 2. These results have been natural extensions of well-known analyses of NLMS. We have also shown in simulations good matches between theoretical and experimental MSEs for three types of input signals. Let us finally mention that, while this paper have focused on the statistical analysis of HSPA, a deterministic convergence analysis of HSPA was studied in the framework of the *adaptive projected subgradient method (APSM)*; see for details [8].

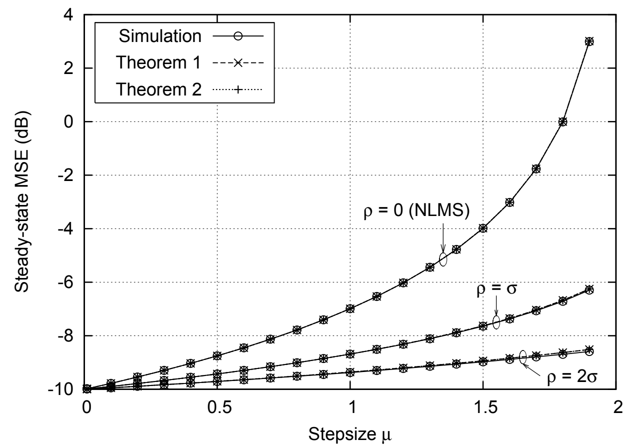

Fig. 4. Comparison between the theories and experiments for the binary input.

ACKNOWLEDGMENT

The authors would like to acknowledge the continuing guidance and encouragement of Prof. K. Sakaniwa at Tokyo Institute of Technology, Japan. This work was supported in part by Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-IEEE Press, June 2003.

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

[3] T. Y. Al-Naffouri and A. H. Sayed, “Transient analysis of adaptive filters with error nonlinearities,” *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 653–663, Mar. 2003.

[4] 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 Processing Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.

[5] 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.

[6] P. S. R. Diniz and S. Werner, “Set-membership binormalized data-reusing LMS algorithms,” *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 124–134, Jan. 2003.

[7] A. Carini and G. L. Sicuranza, “Analysis of a multichannel filtered-x set-membership affine projection algorithm,” in *Proc. IEEE ICASSP*, Toulouse, France, May 2006.

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