# ADAPTIVE PROXIMAL FORWARD-BACKWARD SPLITTING FOR SPARSE SYSTEM IDENTIFICATION UNDER IMPULSIVE NOISE

*Takayuki Yamamoto, Masao Yamagishi, 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: {takayuki, myamagi, isao}@sp.ss.titech.ac.jp

## ABSTRACT

In this paper, we propose a robust sparsity-aware adaptive filtering algorithm under impulsive noise environment, by using the Huber loss function in the frame of *adaptive proximal forward-backward splitting* (APFBS). The APFBS attempts to suppress a time-varying cost function which is the sum of a smooth function and a nonsmooth function. As the smooth function, we employ the weighted sum of the Huber loss functions of the output residuals. As the nonsmooth function, we employ the weighted $\ell_1$ norm. The use of the Huber loss function robustifies the estimation under impulsive noise and the use of the weighted $\ell_1$ norm effectively exploits the sparsity of the system to be estimated. The resulting algorithm has low computational complexity with order $\mathcal{O}(N)$, where $N$ is the tap length. Numerical examples in sparse system identification demonstrate that the proposed algorithm outperforms conventional algorithms by achieving robustness against impulsive noise.

***Index Terms—*** Sparse system identification, Huber loss function, Robust adaptive filtering algorithm, Parallel projection algorithm

## 1. INTRODUCTION

In real-world adaptive filtering applications, it is well known that an impulsive noise can deteriorate the performance of algorithms significantly, for example, in double-talk situations in echo cancellation [1, 2, 3, 4, 5, 6, 7]. Hence, it is desired that adaptive filtering algorithms can estimate precisely and robustly not only under a Gaussian noise but also under an impulsive noise. On the other hand, recently, there has been an increased interest in developing adaptive filtering algorithms which exploit the sparsity of the unknown system in its estimation [8, 9, 10, 11, 12, 13, 14, 15]. The goal of this paper is to propose a robust sparsity-aware adaptive filtering algorithm.

In order to achieve robustness against impulsive noise in adaptive filtering, the well-known Huber loss function [16, 17] has been utilized [1, 2]. The Huber loss function behaves as a quadratic function for small values and as absolute value function for large values. The absolute value function limits the impact of impulsive noise on the performance of the adaptive filtering compared to a quadratic function which is used in standard least square scenarios. It has been reported that the use of the Huber loss function improves robustness compared to other filters employing a quadratic function in adaptive filtering [1, 2].

Meanwhile, the *Adaptive Proximal Forward-Backward Splitting* (APFBS) scheme has been introduced [12, 13, 14, 15] to develop sparsity-aware adaptive filtering algorithms. The APFBS scheme attempts to suppress a time-varying cost function which is a sum of one smooth convex function (typically used as the data fidelity term) and one nonsmooth convex function (typically used as the sparsity-promoting term). By utilizing a weighted sum of squared distance functions as the data fidelity term, the APFBS reproduces many conventional algorithms, for example, Normalized Least-Mean-Square (NLMS) [18] and Affine Projection Algorithm (APA) [19, 20], to name a few. Moreover, for the sparse system identification, the APFBS with the use of weighted $\ell_1$ norms can effectively promote the sparsity of the system to be estimated and achieve excellent estimation accuracy.

In this paper, we propose the concurrent use of weighted sum of the Huber loss functions of the output residuals as the data fidelity term and the weighted $\ell_1$ norm as the sparsity-promoting term in the frame of the APFBS. A cost function similar to the proposed data fidelity term was used in a robust adaptive filtering scenario [1]. The resulting algorithm simultaneously achieves both of robustness against impulsive noise and excellent estimation accuracy compared with certain conventional algorithms. The proposed algorithm is the composition of the weighted average of the relaxed projections onto certain multiple hyperplanes and the adaptively weighted soft-thresholding which is the proximity operator [24] of the weighted $\ell_1$ norm. The proposed algorithm has low computational complexity with order $\mathcal{O}(N)$, where $N$ is the tap length. We also present an explicit value of the Lipschitz constant of the gradient of the Huber loss function of the output residual. This value is necessary at each iteration of the proposed algorithm. Moreover, the proposed algorithm covers Robust Variable Step-Size NLMS (RVSS-NLMS) [6] as its simplest case. Numerical examples for the sparse system identification demonstrate that the proposed algorithm outperforms the conventional algorithms by achieving robustness

This work was partially supported by SCAT (Support Center for Advanced Telecommunications) and JSPS Grants-in-Aid (09J09539).

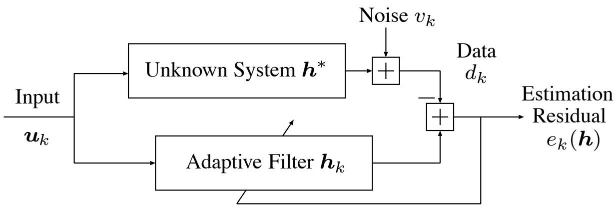


**Fig. 1**. Adaptive filtering scheme.

against impulsive noise.

## 2. PRELIMINARIES

### 2.1. Adaptive filtering problem

Let $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and nonnegative integers, respectively. Denote the set $\mathbb{N} \setminus \{0\}$ by $\mathbb{N}^*$ and transposition of a matrix or a vector by $(\cdot)^T$.

Suppose that we observe an output sequence $(d_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ (i.e., $d_k \in \mathbb{R}, \forall k \in \mathbb{N}$) that obeys the following model:

$$d_k = \boldsymbol{u}_k^T \boldsymbol{h}^* + v_k,$$

where $k \in \mathbb{N}$ denotes the time index, $\boldsymbol{u}_k := [u_k, u_{k-1}, \ldots, u_{k-N+1}]^T \in \mathbb{R}^N$ a known vector defined with the input sequence $(u_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ (where $N \in \mathbb{N}^*$ is the tap length), $\boldsymbol{h}^* \in \mathbb{R}^N$ the unknown system to be estimated (e.g., echo impulse response), and $v_k \in \mathbb{R}$ the noise process. In addition, we define the estimation residual function $e_k : \mathbb{R}^N \to \mathbb{R}$ for $k \in \mathbb{N}$ by $e_k(\boldsymbol{h}) := \boldsymbol{u}_k^T \boldsymbol{h} - d_k,\ \boldsymbol{h} \in \mathbb{R}^N$.

Throughout this paper, we assume that the noise sequence is generated as $v_k = n_k + \eta_k$, where $n_k$ is a Gaussian noise and $\eta_k$ is an impulsive noise. Moreover, we suppose that $\boldsymbol{h}^* \in \mathbb{R}^N$ is *sparse*, i.e., few coefficients are significantly different from zero (active coefficients) and most coefficients are zero or near-zero (inactive coefficients).

A major goal of adaptive sparse system identification is to approximate the unknown system $\boldsymbol{h}^*$ by the adaptive filter $\boldsymbol{h}_k \in \mathbb{R}^N$ with the knowledge on $(\boldsymbol{u}_i, d_i)_{i=0}^k$ together with a priori knowledge on the sparsity of $\boldsymbol{h}^*$.

### 2.2. Adaptive proximal forward-backward splitting [12, 13, 14, 15]

Define the time-varying cost function $\Theta_k : \mathbb{R}^N \to (-\infty, \infty]$ for $k \in \mathbb{N}$ by

$$\Theta_k(\boldsymbol{h}) := \varphi_k(\boldsymbol{h}) + \psi_k(\boldsymbol{h}), \tag{1}$$

where $\psi_k \in \Gamma_0(\mathbb{R}^N)$[1] and $\varphi_k : \mathbb{R}^N \to \mathbb{R}$ is a smooth convex function with its gradient $\nabla\varphi_k$ Lipschitz continuous: there exists some $L_k > 0$ (which is called a Lipschitz constant) such that

$$\|\nabla\varphi_k(\boldsymbol{h}) - \nabla\varphi_k(\boldsymbol{g})\| \leq L_k \|\boldsymbol{h} - \boldsymbol{g}\|, \tag{2}$$

[1] $\Gamma_0(\mathbb{R}^N)$ is the class of all lower semicontinuous convex functions from $\mathbb{R}^N$ to $(-\infty, +\infty]$ that are not identically $+\infty$ [22].

for all $\boldsymbol{h}, \boldsymbol{g} \in \mathbb{R}^N$, where $||\cdot||$ stands for the standard Euclidean norm. Typically, $\varphi_k$ plays the role of a data fidelity term and $\psi_k$ plays the role of a penalty term that exploits the sparsity of $\boldsymbol{h}^*$ in the learning process (e.g. weighted $\ell_1$ norms).

To suppress the time-varying function $\Theta_k$ in an online fashion, we utilize the Adaptive Proximal Forward-Backward Splitting (APFBS) method [12, 13, 14, 15].

**Algorithm 1 (APFBS)** *For an arbitrarily chosen $\boldsymbol{h}_0 \in \mathbb{R}^N$, generate a sequence $(\boldsymbol{h}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^N$ by*

$$\boldsymbol{h}_{k+1} := \mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k}\left(\boldsymbol{h}_k - \frac{\mu_k}{L_k}\nabla\varphi_k(\boldsymbol{h}_k)\right), \tag{3}$$

*where $\mu_k \in (0, 2)$ is the step-size and $\mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k} : \mathbb{R}^N \to \mathbb{R}^N$ defined by*

$$\mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k}(\boldsymbol{h}) := \underset{\boldsymbol{g}\in\mathbb{R}^N}{\arg\min}\left(\psi_k(\boldsymbol{g}) + \frac{L_k}{2\mu_k}||\boldsymbol{h} - \boldsymbol{g}||^2\right)$$

*is called the proximity operator of $\psi_k$ of index $\frac{\mu_k}{L_k} > 0$.*

Note that Algorithm 1 is a time-varying extension of the proximal forward-backward splitting method [22, 23]. Algorithm 1 satisfies the (strictly) monotone approximation property [21]:

$$\left\|\boldsymbol{h}_{k+1} - \boldsymbol{h}^*_{\Theta_k}\right\| < \left\|\boldsymbol{h}_k - \boldsymbol{h}^*_{\Theta_k}\right\| \tag{4}$$

for every $\boldsymbol{h}^*_{\Theta_k} \in \Omega_k := \underset{\boldsymbol{h}\in\mathbb{R}^N}{\arg\min}\,\Theta_k(\boldsymbol{h})$ if $\boldsymbol{h}_k \notin \Omega_k \neq \varnothing$.

## 3. PROPOSED METHOD

We propose a robust sparsity-aware adaptive filtering algorithm in the frame of Algorithm 1 by employing a weighted sum of Huber loss functions as the data fidelity term and a weighted $\ell_1$ norm as the sparsity-promoting term. That is, we propose to use

$$\begin{cases} \varphi_k(\boldsymbol{h}) := \displaystyle\sum_{i\in\mathcal{I}_k} w_i \rho_k^{(i)}(e_i(\boldsymbol{h})), \\ \psi_k(\boldsymbol{h}) := \lambda \displaystyle\sum_{j=1}^{N} \omega_j^{(k)} |h_j| \end{cases} \tag{5}$$

in (1). The function $\varphi_k$ is a weighted sum of the composition of estimation residual functions and Huber loss functions, where $\mathcal{I}_k \subset \{0, 1, \ldots, k-1, k\}$ is the set of indices of the estimation residual functions, the weights $w_i^{(k)} \in (0, 1]$ ($i \in \mathcal{I}_k$) satisfy $\sum_{i\in\mathcal{I}_k} w_i^{(k)} = 1$, and $\rho_k^{(i)} : \mathbb{R} \to \mathbb{R}$ is the well-known Huber loss function defined by

$$\rho_k^{(i)}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \delta_k^{(i)} \\ \delta_k^{(i)}|x| - \frac{1}{2}(\delta_k^{(i)})^2 & \text{otherwise} \end{cases} \tag{6}$$

with the cut-off value $\delta_k^{(i)} > 0$. Note that a cost function similar to $\varphi_k$ was used in a robust adaptive filtering scenario [1].

The function $\psi_k$ is a weighted $\ell_1$-norm with weight $\omega_j^{(k)} > 0$ ($j \in \{1, 2, \ldots, N\}$), where $\lambda > 0$ is the regularization parameter.

The use of $\varphi_k$ and $\psi_k$ of (5) in Algorithm 1 leads to the following adaptive filtering algorithm.

**Algorithm 2 (Proposed robust sparsity-aware algorithm)** *For an arbitrarily chosen $\boldsymbol{h}_0 \in \mathbb{R}^N$, generate a sequence $(\boldsymbol{h}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^N$ by*

$$\boldsymbol{h}_{k+\frac{1}{2}} = \boldsymbol{h}_k - \mu_k \textstyle\sum_{i\in\mathcal{I}_k} \frac{w_i||u_i||^2}{L_k} \min\left[1, \hat{\delta}_k^{(i)}\right] (I - \mathrm{P}_{\Pi_i})(\boldsymbol{h}_k),$$
$$\boldsymbol{h}_{k+1} = \mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k}(\boldsymbol{h}_{k+\frac{1}{2}}),$$

*where $\mu_k \in (0, 2)$ is the step-size,*

$$L_k := \sum_{i\in\mathcal{I}_k} w_i ||u_i||^2 \tag{7}$$

*is a Lipschitz constant of $\nabla\varphi_k$, $P_{\Pi_i}$ is the metric projection*[2] *onto $\Pi_i := \arg\min_{\boldsymbol{h}\in\mathbb{R}^N} |e_i(\boldsymbol{h})|$,*

$$\hat{\delta}_k^{(i)} := \begin{cases} \frac{\delta_k^{(i)}}{||\boldsymbol{u}_i||d(\boldsymbol{h}_k, \Pi_i)} & \textit{if } ||\boldsymbol{u}_i|| \neq 0 \\ 0 & \textit{otherwise}, \end{cases}$$

*and $I$ is the identity operator.*

Note that $L_k$ in (7) satisfies (2) by using the equation $\rho_k^{(i)}(e_i(\boldsymbol{h})) = {}^{\gamma^{(i)}}(||\boldsymbol{u}_i||\delta_k^{(i)} d(\boldsymbol{h}, \Pi_i))$, where $\gamma^{(i)} = \frac{1}{||\boldsymbol{u}_i||^2}$ if $||\boldsymbol{u}_i|| \neq 0$ [3]. The metric projection $P_{\Pi_i}(\boldsymbol{h}_k)$ is expressed as

$$P_{\Pi_i}(\boldsymbol{h}_k) = \begin{cases} \boldsymbol{h}_k - \frac{e_i(\boldsymbol{h}_k)}{||\boldsymbol{u}_i||^2}\boldsymbol{u}_i & \text{if } ||\boldsymbol{u}_i|| \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k}$ for $\psi_k$ in (5) is called *Adaptively Weighted Soft-Thresholding* (AWST) [12] and has a closed form expression:

$$\mathrm{prox}_{\frac{\mu_k}{L_k}\psi_k}(\boldsymbol{h}) = \sum_{j=1}^{N} \mathrm{sgn}(h_j) \max\left\{|h_j| - \frac{\mu_k}{L_k}\lambda\omega_j^{(k)},\ 0\right\} \boldsymbol{e}_j,$$

where $\mathrm{sgn}(\cdot) : \mathbb{R} \to \mathbb{R}$ is the signum function defined by $\mathrm{sgn}(x) := x/|x|$ if $x \neq 0$, $\mathrm{sgn}(x) := 0$ otherwise, and $\{\boldsymbol{e}_j\}_{j=1}^N$ is the standard orthonormal basis of $\mathbb{R}^N$. Intuitively, AWST cuts off the components having smaller absolute values than the threshold $\frac{\mu_k}{L_k}\lambda\omega_j^{(k)}$. The overall complexity of Algorithm 2 is $\mathcal{O}(N)$ if the weight $\omega_j^{(k)}$ and the cut-off value $\delta_k^{(i)}$ can be calculated in order $\mathcal{O}(N)$ (see (8), (9) in Sec. 4). Note that the designs of the weight $\omega_j^{(k)}$ have been proposed in [12, 13, 14]. The cut-off value $\delta_k^{(i)}$ can be designed using many strategies (e.g. theory of robust statistics [17]).

**Remark 1** [A geometrical description of Algorithm 2] Algorithm 2 embodies an intuitive idea: For each $i \in \mathcal{I}_k$, a large distance between $\boldsymbol{h}_k$ and $\Pi_i$ implies that $\Pi_i$ is affected by an impulsive noise and "reliability" of $\Pi_i$ should be low. We consider the situation in which Algorithm 2 updates $\boldsymbol{h}_k$ to $\boldsymbol{h}_{k+\frac{1}{2}}$ at time $k$. Then, $\boldsymbol{h}_{k+\frac{1}{2}}$ is designed by a weighted average of the exact or relaxed projections onto hyperplanes. More precisely, for each $i \in \mathcal{I}_k$, if $d(\boldsymbol{h}_k, \Pi_i) \leq \frac{\delta_k^{(i)}}{||\boldsymbol{u}_i||}$, which implies that "reliability" of $\Pi_i$ is high, then the exact projection $\mathrm{P}_{\Pi_i}(\boldsymbol{h}_k)$ is employed; Otherwise, the relaxed projection $(1 - \hat{\delta}_k^{(i)})\boldsymbol{h}_k + \hat{\delta}_k^{(i)} P_{\Pi_i}(\boldsymbol{h}_k)$ with $\hat{\delta}_k^{(i)} < 1$ is used.

**Remark 2** [Reproducing an existing algorithm] Note that Algorithm 2 reproduces Robust Variable Step-Size Normalized Least-Mean-Square (RVSS-NLMS)[4] [6] if we set

$$\mathcal{I}_k := \{k\}, \quad \lambda = 0, \quad \mu_k = 1.$$

## 4. NUMERICAL EXAMPLES

We examine the efficacy of the proposed algorithm in the context of sparse system identification. We use the sparse echo impulse response $\boldsymbol{h}^*$ of length $N = 512$ initialized according to ITU-T G.168 [25]. The input signal $\boldsymbol{u}_k$ is generated by zero mean white Gaussian with variance 1. The noise $n_k$ is zero mean white Gaussian and signal-to-background-noise ratio (SBNR) is 20(dB), where SBNR $:= 10\log_{10}(E[z_k^2]/E[n_k^2])$ with $z_k := \boldsymbol{u}_k^T\boldsymbol{h}^*$ ($E[\cdot]$ denotes expectation). An impulsive noise is generated by $\eta_k = \xi_k N_k$, where $\xi_k$ is a Bernoulli process with probability of success $P[\xi_k = 1] = p$ and $N_k$ is zero-mean Gaussian with variance $\sigma_N^2 = 1000\sigma_z^2$, and where $\sigma_z^2$ is the variance of $z_k$. For all the simulations, we set the initial vector $\boldsymbol{h}_0 := [0, \ldots, 0]^T$.

We compare Algorithm 2 with NLMS [18], RVSS-NLMS [6], and RZA-NLMS [9][5]. In this section, we call Algorithm 2 with $\mathcal{I}_k = \{k\}$ as Proposed 1, and Algorithm 2 with $\mathcal{I}_k = \{k-9, \ldots, k\}$ as Proposed 2.

Table 1 shows the parameter settings of all the algorithms. Each step-size $\mu_k$ is chosen in such a way that the initial rate of the convergence of all the algorithms is the same under

[2] For a given nonempty closed convex set $C \subset \mathbb{R}^N$, the metric projection of $\boldsymbol{h} \in \mathbb{R}^N$ onto $C$ is defined by $P_C(\boldsymbol{h}) := \arg\min_{\boldsymbol{g}\in C} ||\boldsymbol{h} - \boldsymbol{g}||$ and the distance between an arbitrary point $\boldsymbol{h} \in \mathbb{R}^N$ and a closed convex set $C \subset \mathbb{R}^N$ is defined by $d(\boldsymbol{h}, C) := ||\boldsymbol{h} - P_C(\boldsymbol{h})||$.

[3] For a given $f \in \Gamma_0(\mathbb{R}^N)$, ${}^\gamma f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$ is the Moreau envelope [24] of $f$ of index $\gamma \in (0, +\infty)$ defined by

$${}^\gamma f(\boldsymbol{x}) := \min_{\boldsymbol{y}\in\mathbb{R}^N} \left( f(\boldsymbol{y}) + \frac{1}{2\gamma}||\boldsymbol{x} - \boldsymbol{y}||^2 \right).$$

Then, the gradient of ${}^\gamma f(\boldsymbol{x})$ is $\frac{1}{\gamma}$-Lipschitz continuous.

[4] Algorithm 1 generalizes Robust Variable Step-Size Affine Projection Algorithm (RVSS-APA) [7]: The use of ${}^{\gamma^{(i)}}(\delta_k^{(i)} d(\boldsymbol{h}, S_i))$ with $S_i$ employed in [12, Remark] as $\varphi_k$ in Algorithm 1 leads to the composition of the proximity operator of $\psi_k$ and the update of RVSS-APA.

[5] RZA-NLMS [9] is described by the following equation:

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - \mu\frac{e_k(\boldsymbol{h}_k)}{||\boldsymbol{u}_k||^2}\boldsymbol{u}_k - \lambda\sum_{j=1}^{N} \omega_j^{(k)} \mathrm{sgn}(h_k^{(j)})\boldsymbol{e}_j$$

where $\mu > 0$ is the step-size and $\lambda > 0$ is the regularization parameter.

Gaussian noise environment, i.e., $p = 0$. The regularization parameter $\lambda$ is chosen to obtain the best results in our experiment. We employ the design of the weight $\omega_j^{(k)}$ of the weighted $\ell_1$ norm in [13]:

$$\begin{aligned} \pi_j^{(k)} &= \frac{1}{|h_k^{(j)}|^{1-q}+\upsilon} \\ \omega_j^{(k)} &= \frac{N\pi_j^{(k)}}{\sum_{l=1}^{N}\pi_l^{(k)}} \end{aligned} \qquad (j = 1, 2, \ldots, N), \tag{8}$$

where $\upsilon > 0$ is a small constant for regularization. This technique is based on an approximation of $\ell_q$ quasi-norm $||\cdot||_q$ ($q \in (0, 1)$) to the $q$-th power in the neighborhood of $\boldsymbol{h}_k$,

$$||\boldsymbol{x}||_q^q = \sum_{j=1}^{N} \frac{1}{|x_j|^{1-q}}|x_i| \simeq \sum_{j=1}^{N} \frac{1}{|h_k^{(j)}|^{1-q}}|x_j| \ ,$$

because $||\cdot||_q^q$ evaluates the sparsity of the system more tightly than the $\ell_1$ norm. In [6], [7], the design of $\delta_k^{(i)}$ for RVSS-NLMS was introduced by

$$\begin{aligned} \bar{\delta}_k^{(i)} &= \alpha\bar{\delta}_{k-1}^{(i)} + (1-\alpha)\min\left\{\bar{\delta}_{k-1}^{(i)}, \left(\frac{e_{k-1}(\boldsymbol{h}_{k-1})}{||\boldsymbol{u}_{k-1}||}\right)^2\right\} \\ \delta_k^{(i)} &= ||\boldsymbol{u}_i||\sqrt{\bar{\delta}_k^{(i)}}, \end{aligned} \tag{9}$$

where the forgetting factor $\alpha \in (0, 1)$ is chosen by knowledge of the color of the input signal. Since RVSS-NLMS is the simplest case of Algorithm 2 (see Remark 2), we use the design (9) in this section for a fair comparison with Algorithm 2 and RVSS-NLMS. Note that the computational complexity of the designs in (8) and (9) are $\mathcal{O}(N)$.

We adopt two measures (i) system mismatch defined as $\chi(\boldsymbol{h}_k) := 10\log_{10}\frac{||\boldsymbol{h}^*-\boldsymbol{h}_k||^2}{||\boldsymbol{h}^*||^2}$, and (ii) the approximation of $\ell_0$ norm defined as sparse($\boldsymbol{h}_k$):$= \sum_{j=1}^{N}\big(1 - e^{-\beta|h_k^{(j)}|}\big)(\simeq ||\boldsymbol{h}_k||_0)$, where $\beta$ is a large positive constant (we set $\beta = 1000$). The simulation results are averaged over 300 runs.

First, we consider the case where the output is contaminated by impulsive noise, i.e., $p = 0.01$. From Fig. 2, the two proposed algorithms achieve better performances compared with RVSS-NLMS in terms of system mismatch. Proposed 2 achieves the best steady-state performance and Proposed 1 achieves the best initial rate of the convergence in this example. Fig. 3 shows that the two proposed algorithms successfully estimate sparse($\boldsymbol{h}^*$).

Finally, we examine the case where no impulsive noise is added to the output, i.e., $p = 0$. Even in this case, the two proposed algorithms achieve better steady-state performance compared to all the conventional algorithms in Fig. 4.

## 5. REFERENCES


[1] P. Petrus, "Robust Huber adaptive filter," IEEE Signal Process., vol. 47, no. 1, pp. 1129-1133, Apr. 1999.

[2] T. Gansler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," Trans. Speech Audio Process., vol. 8, no. 6, pp. 656-663, Nov. 2000.

[3] Y. Zou, S. C. Chan, and T. S. Ng, "Least mean M-estimate algorithms for robust adaptive filtering in impulse noise," IEEE Trans. Circuits Syst. II, vol. 47, no. 12, pp. 1564-1569, Dec. 2000.

[4] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," IEEE Signal Process. Lett., vol. 11, no. 1, pp. 56-59, Jan. 2004.

[5] E. V. Papoulis and T. Stathaki, "A normalized robust mixed-norm adaptive algorithm for system identification," IEEE Signal Process. Lett., vol. 4, no. 2, pp. 46-48, Feb. 1997.

[6] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," IEEE Signal Process., vol. 56, no. 5, pp. 1878-1893, May 2008.

[7] L. R. Vega, H. Rey, and J. Benesty, "A robust variable step-size affine projection algorithm," Signal Process, vol. 90, pp. 2806-2810, 2010.

[8] J. Benesty, Y. A. Huang, J. Chen, and P. A. Naylor, "Adaptive algorithm for the identification of sparse impulse responses," chapter 5, pp. 125-153. Springer, 2006.

[9] Y. Chen, Y. Gu, and A. O. Hero, "Regularized least-mean-square algorithms," Available at arxiv.org: http://arxiv.org/pdf/1012.5066.

[10] Y. Kopsinis, K. Slavakis, S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted $\ell_1$ balls," IEEE Signal Process., vol. 59, no. 3, pp. 936-952, Mar. 2010.

[11] I. Yamada, S. Gandy, and M. Yamagishi, "Sparsity-aware adaptive filtering based on a Douglas-Rachford splitting," in Proc. EUSIPCO 2011, pp. 1929-1933, Sep. 2011.

[12] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in Proc. IEEE ICASSP 2010, pp. 3734-3737, Mar. 2010.

[13] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada, "Sparsity-aware adaptive filters based on $\ell_p$-norm inspired soft-thresholding technique," In Proc. IEEE ISCAS, May 2012 (accepted).

[14] M. Yamagishi, M. Yukawa, and I. Yamada, "Sparse system identification by exponentially weighted adaptive parallel projection and generalized soft-thresholding," In Proc. APSIPA ASC 2010, pp. 367-370, Dec. 2010.

[15] M. Yamagishi, M. Yukawa, and I. Yamada, "Acceleration of adaptive proximal forward-backward splitting method and its application to sparse system identification," in Proc. IEEE ICASSP 2011, pp. 4296-4299, 2011.

[16] P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist. 35, pp. 73-101, 1964.

[17] P. J. Huber, "Robust statistics," New York: Wiley, 1981.

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

[19] T. Hinamoto and S. Maekawa,"Extended theory of learning identification," Trans. IEE Japan (in Japanese), vol. 95-C, no. 10, pp. 227-234, 1975.

[20] 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. 67-A, no. 5, pp. 126-132, 1984.

[21] 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 and 8, pp. 593-617, 2004.

[22] B. Mercier, "Lectures on topics in finite element solution of elliptic problems," With notes by G. Vijayasundaram, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, vol 63, Bombay: Tata Institute of Fundamental Research, 1979.

[23] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," Multiscale Model. Simul., vol. 4, no. 4, pp. 1168-1200, 2005.

[24] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," C. R. Acad. Sci. Paris Ser. A Math. 255, pp. 2897-2899, 1962.

[25] Digital Network Echo Cancellers, ITU-T Rec. G.168, 2007.

**Table 1**. Parameter settings of all the algorithms. The cut-off value $\delta_k^{(i)}$ of the Huber loss function in (5) is designed using (9) with the initial value $\bar{\delta}_0$ and the forgetting factor $\alpha$. $\sigma_d^2$ and $\sigma_u^2$ are the variances of the output signal and the input signal. The weight $\omega_j^{(k)}$ of the weighted $\ell_1$ norm in (5) is designed by (8) with parameters $q$ and $\upsilon$.

| | $\mu_k$ | $\delta_0$ | $\alpha$ | $w_k$ | $\lambda$ | $q$ | $\upsilon$ |
|---|---|---|---|---|---|---|---|
| NLMS [18] | $\mu_k = 0.5$ | | | | | | |
| RVSS-NLMS [6] | $\mu_k = 1.0$ | $\bar{\delta}_0 = \frac{\sigma_d^2}{N\sigma_u^2}$ | $\alpha = 0.998$ | | | | |
| RZA-NLMS [9] | $\mu_k = 0.5$ | | | | $\lambda = 2 \times 10^{-5}$ | $q = 0.01$ | $\upsilon = 1.0 \times 10^{-6}$ |
| Proposed 1 | $\mu_k = 0.9$ | $\bar{\delta}_0 = \frac{\sigma_d^2}{N\sigma_u^2}$ | $\alpha = 0.998$ | $w_k = 1$ | $\lambda = 2 \times 10^{-2}$ | $q = 0.01$ | $\upsilon = 1.0 \times 10^{-6}$ |
| Proposed 2 | $\mu_k = 1.0$ | $\bar{\delta}_0 = \frac{\sigma_d^2}{N\sigma_u^2}$ | $\alpha = 0.998$ | $w_k = \frac{1}{10}$ | $\lambda = 2 \times 10^{-2}$ | $q = 0.01$ | $\upsilon = 1.0 \times 10^{-6}$ |

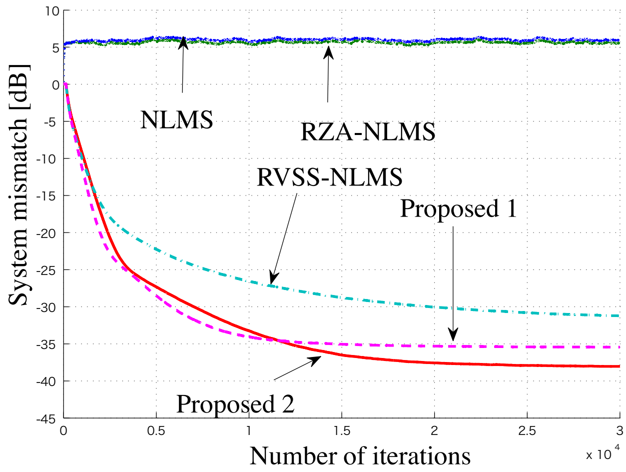


**Fig. 2**. Comparison of system mismatch (under impulsive noise).

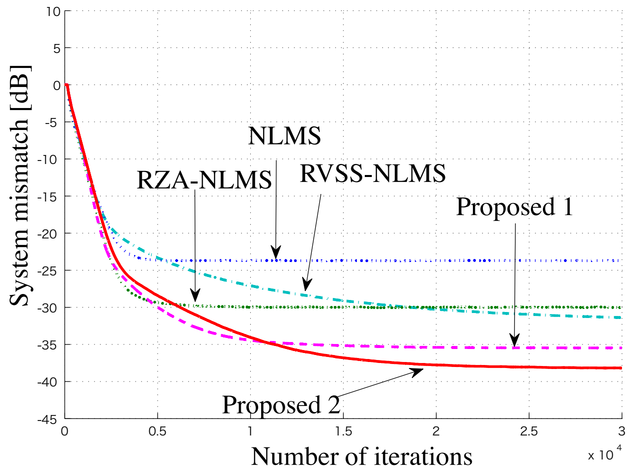


**Fig. 4**. Comparison of system mismatch (no impulsive noise).

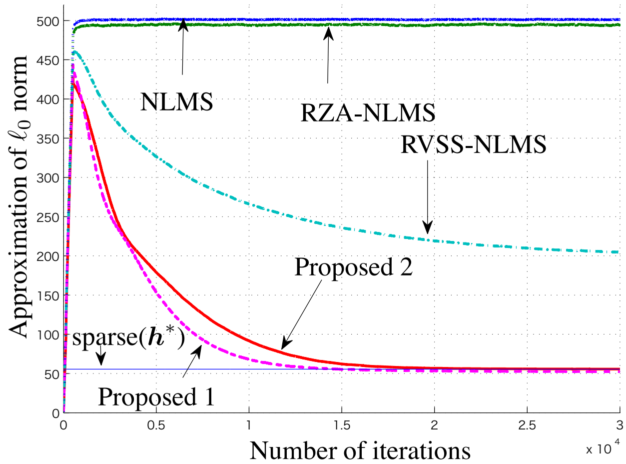


**Fig. 3**. Comparison of sparse($h_k$) (under impulsive noise).

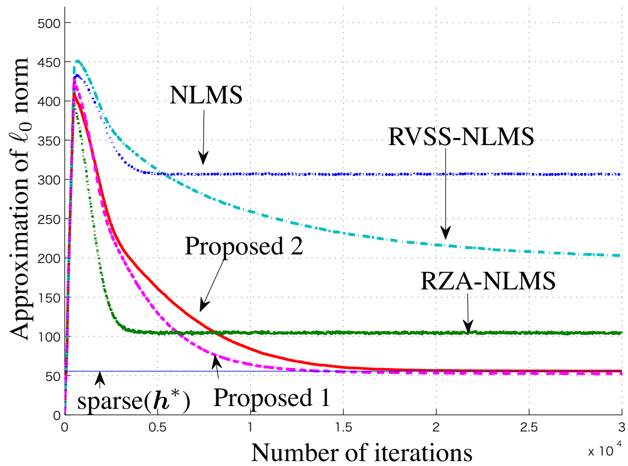


**Fig. 5**. Comparison of sparse($h_k$) (no impulsive noise).