

SHRINKAGE TUNING BASED ON AN UNBIASED MSE ESTIMATE FOR SPARSITY-AWARE ADAPTIVE FILTERING

Masao Yamagishi[†]Masahiro Yukawa[‡]Isao Yamada[†]

[†]Department of Communications and Computer Engineering
Tokyo Institute of Technology, 2-12-1-S3-60 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

[‡]Department of Electronics and Electrical Engineering
Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
Email: {myamagi, isao}@sp.ce.titech.ac.jp, yukawa@elec.keio.ac.jp

ABSTRACT

Effective utilization of *sparsity* of the system to be estimated is a key to achieve excellent adaptive filtering performances. This can be realized by the adaptive proximal forward-backward splitting (APFBS) with carefully chosen parameters. In this paper, we propose a systematic parameter tuning based on a minimization principle of an unbiased MSE estimate. Thanks to the piecewise quadratic structure of the proposed MSE estimate, we can obtain its minimizer with low computational load. A numerical example demonstrates the efficacy of the proposed parameter tuning by its excellent performance over a broader range of SNR than a heuristic parameter tuning of the APFBS.

Index Terms— Shrinkage parameter tuning, sparsity-aware adaptive filtering, Mallor's C_p statistic, Stein's lemma, proximity operator

1. INTRODUCTION

In adaptive filtering, exploiting *sparsity* of the system to be estimated is a key technique to achieve excellent performance [1, 2]; the sparsity implies that many coefficients of the system are zero. Such a property has been observed and exploited in many applications including network/acoustic echo cancellation and active noise control (e.g. see [1–17] and references therein).

Many sparsity-aware adaptive filtering algorithms are derived by introducing a sparsity promoting term in their implicit/explicit cost function. One of these is the adaptive proximal forward-backward splitting (APFBS) scheme [8, 9], which is a principle to adaptively suppress the sum of a smooth convex function and a nonsmooth convex function. A typical choice of the nonsmooth convex term to promote the sparsity is a weighted ℓ_1 norm with a regularization parameter (which we refer to as *shrinkage parameter* in this paper). In this technique, careful adaptive tuning of the weight and the shrinkage parameter is required to achieve excellent performance. For a fixed parameter, several weight designs have been proposed and examined their improved performance and robustness against environmental changes (e.g. [14]).

In this paper, we propose a systematic parameter tuning based on a minimization principle of an unbiased estimate of the Mean Squared Error (MSE): our idea is an extension of the spirit of the

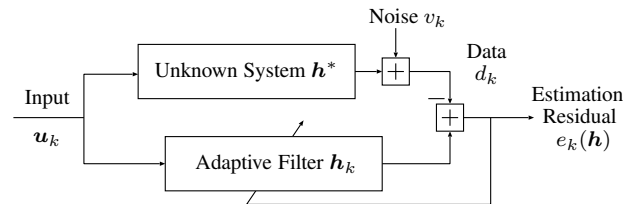


Fig. 1. Adaptive filtering scheme.

Mallor's C_p -type cost [18]¹ to the case of adaptive filtering scenario. Since the MSE is an inherent criterion in adaptive filtering, the shrinkage parameter minimizing the MSE is a natural choice, but the exact information of the MSE cannot be available in practice. To alleviate this difficulty, we derive an unbiased MSE estimate, which can be computed without the complete knowledge of the MSE with the help of the well-known Stein's lemma [19].² Thanks to the piecewise quadratic structure of the unbiased MSE estimate, we obtain a closed form expression of a minimizer, which results in an efficient computation for tuning the parameter. A numerical example demonstrates its efficacy from excellent performance over a broader range of SNR than a heuristic parameter tuning of the APFBS.

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 (see Fig. 1):

$$d_k = \mathbf{u}_k^t \mathbf{h}_* + v_k, \quad (1)$$

where $k \in \mathbb{N}$ denotes the time index, $\mathbf{u}_k := [u_k, u_{k-1}, \dots, u_{k-N+1}]^t \in \mathbb{R}^N$ 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), $\mathbf{h}_* \in \mathbb{R}^N$ the unknown system to be estimated (e.g., echo impulse response), and

¹In the context of linear regression, [18] introduced an unbiased estimate of MSE as a criterion for parameter tuning.

²The Stein's lemma has been utilized in several signal processing problems, e.g., [20–22]. Recently, in the context of ill-conditioned linear regression models, it was utilized for the rank selection of an innovative extension of a reduced rank estimator [23].

This work was supported in part by JSPS Grants-in-Aid (24800022, 24760292, B-21300091).

$v_k \in \mathbb{R}$ the noise process. In this paper, we suppose that \mathbf{h}_* is sparse.

The major goal of adaptive system identification is to approximate the unknown system \mathbf{h}_* by the adaptive filter $\mathbf{h}_k := [h_k^{(1)}, h_k^{(2)}, \dots, h_k^{(N)}]^t \in \mathbb{R}^N$ with $(\mathbf{u}_i, d_i)_{i=0}^k$ together with a priori knowledge, e.g., the sparsity, on \mathbf{h}_* .

2.2. Adaptive proximal forward-backward splitting

Define a time-varying cost function³ $\Theta_k \in \Gamma_0(\mathbb{R}^N)$ for $k \in \mathbb{N}$ by

$$\Theta_k(\mathbf{h}) := \varphi_k(\mathbf{h}) + \psi_k(\mathbf{h}), \quad (2)$$

where $\psi_k \in \Gamma_0(\mathbb{R}^N)$ and $\varphi_k: \mathbb{R}^N \rightarrow \mathbb{R}$ is a smooth convex function with its gradient $\nabla \varphi_k$ Lipschitz continuous; i.e., there exists a some $L_k > 0$ (which is called a Lipschitz constant) such that

$$\|\nabla \varphi_k(\mathbf{h}) - \nabla \varphi_k(\mathbf{g})\| \leq L_k \|\mathbf{h} - \mathbf{g}\| \quad (3)$$

for all $\mathbf{h}, \mathbf{g} \in \mathbb{R}^N$, where $\|\cdot\|$ stands for the standard Euclidean norm. Typically, φ_k plays the role of a data fidelity term and ψ_k plays the role of a penalty term that exploits the sparsity of \mathbf{h}_* in the learning process (e.g. weighted ℓ_1 norms are adopted as ψ_k).

To suppress the time-varying function Θ_k in an online fashion, the Adaptive Proximal Forward-Backward Splitting (APFBS) method [8, 9] has been proposed.

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

$$\mathbf{h}_{k+1} := \text{prox}_{\frac{\mu_k}{L_k} \psi_k} \left(\mathbf{h}_k - \frac{\mu_k}{L_k} \nabla \varphi_k(\mathbf{h}_k) \right), \quad (4)$$

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

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

is called the proximity operator of ψ_k of index $\frac{\mu_k}{L_k} > 0$ [25].

Note that Algorithm 1 is a time-varying extension of the proximal forward-backward splitting method [26–29] (see also [30–32]) and Algorithm 1 satisfies the (strictly) monotone approximation property [33]:

$$\|\mathbf{h}_{k+1} - \mathbf{h}_{\Theta_k}^*\| < \|\mathbf{h}_k - \mathbf{h}_{\Theta_k}^*\| \quad (5)$$

for every $\mathbf{h}_{\Theta_k}^* \in \Omega_k := \arg \min_{\mathbf{h} \in \mathbb{R}^N} \Theta_k(\mathbf{h})$ if $\mathbf{h}_k \notin \Omega_k \neq \emptyset$. An acceleration of the APFBS has been proposed in [11].

Here we show a simple sparsity-aware adaptive filtering algorithm in the frame of the APFBS.

Example 1 Define the smooth term φ_k of the objective function Θ_k as follows:

$$\varphi_k(\mathbf{h}) := \frac{1}{2} d^2(\mathbf{h}, S_k). \quad (6)$$

Here, $S_k := \arg \min_{\mathbf{h} \in \mathbb{R}^N} |d_k - \mathbf{u}_k^t \mathbf{h}|$ is a closed convex set⁴, of which the elements are consistent with the data available at time

³ $\Gamma_0(\mathbb{R}^N)$ is the class of all lower semicontinuous convex functions from \mathbb{R}^N to $(-\infty, +\infty]$ that are not identically $+\infty$ [24].

⁴The distance between an arbitrary point $\mathbf{x} \in \mathbb{R}^N$ and a closed convex set $C \subset \mathbb{R}^N$ is defined by $d(\mathbf{x}, C) := \min_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$.

k . In this case, the Lipschitz constant of the gradient of φ_k is one. Moreover, in order to exploit the sparsity of the unknown system, we adopt a weighted ℓ_1 -norm as the nonsmooth term, i.e., $\psi_k = \lambda_k \|\cdot\|_1^{\omega_k}$ with

$$\|\mathbf{h}\|_1^{\omega_k} := \sum_{j=1}^N \omega_k^{(j)} |h_j|, \quad \mathbf{h} := [h_1, h_2, \dots, h_N]^t \in \mathbb{R}^N,$$

where $\lambda_k > 0$ is the regularization parameter, and $\omega_k^{(j)} > 0$, $j \in \{1, 2, \dots, N\}$, the weights of the ℓ_1 norm defined with available knowledge. Then the update equation becomes the following:

$$\mathbf{h}_{k+1} := \text{prox}_{\mu_k \lambda_k \|\cdot\|_1^{\omega_k}} \left(\mathbf{h}_k + \mu_k \frac{d_k - \mathbf{u}_k^t \mathbf{h}_k}{\|\mathbf{u}_k\|^2} \mathbf{u}_k \right), \quad (7)$$

where $(\mu_k)_{k \in \mathbb{N}} \subset (0, 2)$ is the step-size,

$$\text{prox}_{\mu_k \lambda_k \|\cdot\|_1^{\omega_k}}(\mathbf{h}) = \sum_{j=1}^N \text{sgn}(h_j) \max \left\{ |h_j| - \mu_k \lambda_k \omega_k^{(j)}, 0 \right\} \mathbf{e}_j,$$

$\text{sgn}(\cdot)$ is the signum function defined by

$$\text{sgn}(x) := \begin{cases} x/|x| & \text{if } x \neq 0, \\ 0 & \text{otherwise,} \end{cases} \quad \forall x \in \mathbb{R}$$

and $\{\mathbf{e}_j\}_{j=1}^N$ is the standard orthonormal basis of \mathbb{R}^N (i.e., $\mathbf{e}_j := [0, \dots, 0, 1, 0, \dots, 0]^t$, $j \in \{1, 2, \dots, N\}$, with the value 1 assigned to the j -th position).

3. PROPOSED METHOD

We propose a systematic tuning of the shrinkage parameter λ_k in (7): select a candidate of the adaptive filter at time $k+1$ parametrized by $\lambda \in \mathbb{R}_+ := \{r \geq 0\}$, i.e.,

$$\hat{\mathbf{h}}_{k+1}(\lambda) = \text{prox}_{\lambda \|\cdot\|_1^{\omega_k}} \left(\mathbf{h}_k + \mu_k \frac{d_k - \mathbf{u}_k^t \mathbf{h}_k}{\|\mathbf{u}_k\|^2} \mathbf{u}_k \right),$$

by minimizing the unbiased estimate $J(\lambda)$ (in Prop. 1 below) of

$$E \left[(\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - \mathbf{u}_k^t \mathbf{h}_*)^2 \right] \quad (8)$$

which is the MSE of the system output.

Proposition 1 Assume that the additive noise is according to zero mean Gaussian noise

$$p(v_k) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left(-\frac{v_k^2}{2\sigma^2} \right).$$

Define⁵ $J: \mathbb{R}_+ \rightarrow \mathbb{R}$,

$$J(\lambda) := (\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - d_k)^2 + 2\sigma^2 \frac{\text{tr}[\mathbf{A}_k(\lambda) \mathbf{u}_k \mathbf{u}_k^t]}{\|\mathbf{u}_k\|^2} - \sigma^2$$

⁵For a given square matrix \mathbf{A} , $\text{tr}[\mathbf{A}]$ indicates its trace. For a vector $\mathbf{x} \in \mathbb{R}^N$, $\text{diag}(\mathbf{x}) \in \mathbb{R}^{N \times N}$ denotes the diagonal matrix whose diagonal entries are given by \mathbf{x} .

with a diagonal matrix $\mathbf{A}_k(\lambda)$ whose diagonal entries indicate the support of the adaptive filter $\hat{\mathbf{h}}_{k+1}(\lambda)$, i.e.,

$$\begin{aligned}\mathbf{A}_k(\lambda) &:= \text{diag}(a_k^{(1)}(\lambda), a_k^{(2)}(\lambda), \dots, a_k^{(N)}(\lambda)) \in \mathbb{R}^{N \times N}, \\ a_k^{(j)}(\lambda) &:= \begin{cases} 1 & |g_k^{(j)}| > \lambda \omega_k^{(j)} \\ 0 & \text{otherwise,} \end{cases} \\ \mathbf{g}_k &:= (g_k^{(1)}, g_k^{(2)}, \dots, g_k^{(N)})^t := \mathbf{h}_k + \mu_k \frac{d_k - \mathbf{u}_k^t \mathbf{h}_k}{\|\mathbf{u}_k\|^2} \mathbf{u}_k.\end{aligned}$$

Then J satisfies that

$$E[J(\lambda)] = E \left[(\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - \mathbf{u}_k^t \mathbf{h}_*)^2 \right].$$

Note that the unbiased MSE estimate J can be computed without the exact knowledge on \mathbf{h}_* and its derivation utilizes the idea of [19].⁶

We shall show an efficient computation for the minimizer of J . Thanks to the fact that J is a quadratic function over the range that $\mathbf{A}_k(\lambda)$ is invariant, we can compute the local minimizer of J in each invariant range and then, by evaluating the value of J at all local minima, find the global minimizer. This idea is embodied by the following proposition (see also the resulting algorithm in Algorithm 2).

Proposition 2 (i) Let $(\hat{\lambda}_j)_{j=0}^N$ be a sorted sequence of all entries of

$$\left(0, \frac{|g_k^{(1)}|}{\omega_k^{(1)}}, \frac{|g_k^{(2)}|}{\omega_k^{(2)}}, \dots, \frac{|g_k^{(N)}|}{\omega_k^{(N)}} \right) \quad (9)$$

in nondecreasing order.⁷ The function J is quadratic over $[\hat{\lambda}_j, \hat{\lambda}_{j+1})$ because

$$\mathbf{A}_k(\lambda) = \mathbf{A}_k(\hat{\lambda}_j), \quad \forall \lambda \in [\hat{\lambda}_j, \hat{\lambda}_{j+1}).$$

(ii) J is lower semicontinuous and

$$\inf_{\lambda \in \mathbb{R}_+} J(\lambda) = \min_{\lambda \in \{\hat{\lambda}_j^*\}_{j=0}^N} J(\lambda),$$

where ⁸

$$\hat{\lambda}_j^* := P_{[\hat{\lambda}_j, \hat{\lambda}_{j+1}]} \left[\frac{\text{tr}[\mathbf{A}_k(\hat{\lambda}_j) \mathbf{g}_k \mathbf{u}_k^t] - d_k}{\text{tr}[\mathbf{A}_k(\hat{\lambda}_j) \text{diag}(\boldsymbol{\omega}_k) \text{sgn}(\mathbf{g}_k) \mathbf{u}_k^t]} \right] \quad (10)$$

is a minimizer of $J_j : [\hat{\lambda}_j, \hat{\lambda}_{j+1}] \rightarrow \mathbb{R}$,

$$J_j(\lambda) := (\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - d_k)^2$$

except that $\hat{\lambda}_N^* := \hat{\lambda}_N$.

⁶A rough sketch of the proof of Proposition 1: In the MSE (8), we can eliminate \mathbf{h}_* by substituting (1) and then expand the quadratic term as (MSE)

$$= E \left[(\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - d_k)^2 \right] + 2E \left[v_k (\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - d_k) \right] + \sigma^2.$$

In the second term, by using the fact that $\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\lambda) - d_k$ is piecewise linear w.r.t. v_k , we decompose the expectation into the sum of the integrals over each line segment. In addition, by applying the Stein's lemma [19] to each integral, we complete the proof.

⁷In this paper, for simplicity, we assume that the vector (9) has no overlapping entries. This assumption can be relaxed easily.

⁸For $a, b \in \mathbb{R} : a < b$, the projection onto $[a, b]$ is given as

$$P_{[a,b]} : \mathbb{R} \rightarrow \mathbb{R}, P_{[a,b]}(r) = \begin{cases} a & \text{if } r < a \\ r & \text{if } r \in [a, b] \\ b & \text{if } r > b. \end{cases}$$

Algorithm 2 Sparsity-aware adaptive filtering algorithm with the proposed parameter tuning

Require: A nonnegative integer $N_{\max} < N$, $\mathbf{h}_0 \in \mathbb{R}^N$, $\mu_k > 0$. Repeat the following step:

1. Compute

$$\mathbf{g}_k = \mathbf{h}_k + \mu_k \frac{d_k - \mathbf{u}_k^t \mathbf{h}_k}{\|\mathbf{u}_k\|^2} \mathbf{u}_k.$$

2. Calculate the weight $\boldsymbol{\omega}_k$ (e.g. (12)).

3. Sort

$$\left(0, \frac{|g_k^{(1)}|}{\omega_k^{(1)}}, \frac{|g_k^{(2)}|}{\omega_k^{(2)}}, \dots, \frac{|g_k^{(N)}|}{\omega_k^{(N)}} \right)$$

into $(\hat{\lambda}_j)_{j=0}^N$ by nondecreasing order.

4. Compute $(\hat{\lambda}_j^*)_{j=0}^{N_{\max}}$ by

$$\hat{\lambda}_j^* := P_{[\hat{\lambda}_j, \hat{\lambda}_{j+1}]} \left[\frac{\text{tr}[\mathbf{A}_k(\hat{\lambda}_j) \text{diag}(\boldsymbol{\xi}_k)] - d_k}{\text{tr}[\mathbf{A}_k(\hat{\lambda}_j) \text{diag}(\boldsymbol{\zeta}_k)]} \right],$$

where $\mathbf{A}_k(\hat{\lambda}_j) = \text{diag}(a_k^{(1)}(\hat{\lambda}_j), \dots, a_k^{(N)}(\hat{\lambda}_j))$,

$$a_k^{(i)}(\hat{\lambda}_j) = \begin{cases} 1 & |g_k^{(i)}| > \hat{\lambda}_j \omega_k^{(i)} \\ 0 & \text{otherwise,} \end{cases}$$

and vectors $\boldsymbol{\xi}_k$ and $\boldsymbol{\zeta}_k$ contain the diagonal entries of $\mathbf{g}_k \mathbf{u}_k^t$ and $\text{diag}(\boldsymbol{\omega}_k) \text{sgn}(\mathbf{g}_k) \mathbf{u}_k^t$, respectively.

5. Find $\lambda^* \in \arg \min \{J(\lambda) \mid \lambda \in \{\hat{\lambda}_j^*\}_{j=0}^{N_{\max}}\}$.

6. Update $\mathbf{h}_{k+1} = \text{prox}_{\lambda^* \|\cdot\|_1}^{\boldsymbol{\omega}_k}(\mathbf{g}_k)$.

Remark 1: (a) (**Numerical Stability**) Computation of $\hat{\lambda}_j^*$ is numerically unstable because the denominator in (10)

$$\text{tr}[\mathbf{A}_k(\hat{\lambda}_j) \text{diag}(\boldsymbol{\omega}_k) \text{sgn}(\mathbf{g}_k) \mathbf{u}_k^t]$$

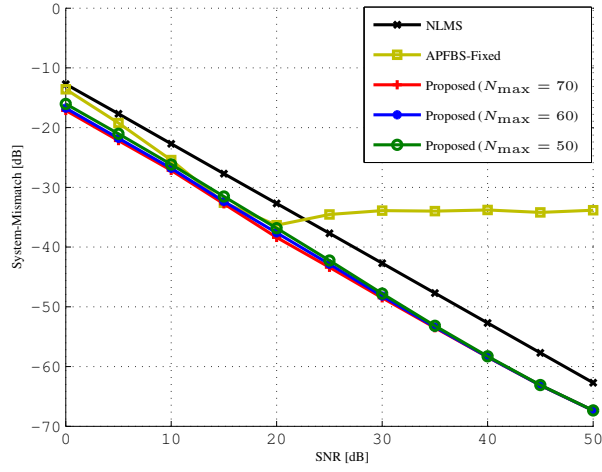
becomes significantly small for a large j and it expands influence of noise contained in d_k . Hence we limit the range of j to ensure numerical stability, e.g.,

$$\lambda^* \in \arg \min \{J(\lambda) \mid \lambda \in \{\hat{\lambda}_j^*\}_{j=0}^{N_{\max}}\} \quad (11)$$

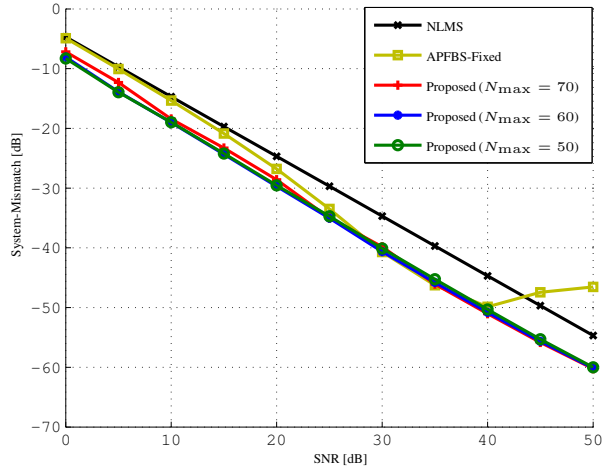
with a predefined nonnegative integer $N_{\max} (< N)$.

(b) (**Computational Cost**) The computational cost of Algorithm 2 is $\mathcal{O}(N \log N)$ comparisons and $\mathcal{O}(N)$ multiplications. In step 3, the sorting requires $\mathcal{O}(N \log N)$ comparisons. Step 1 and step 6 are $\mathcal{O}(N)$ multiplications. Step 2 depends on the weight design but typically has $\mathcal{O}(N)$. In step 4, the most demanding part is $\mathcal{O}(N)$ multiplications for $\boldsymbol{\xi}_k$ and $\boldsymbol{\zeta}_k$. In step 5, the second term of J can be computed with $\mathcal{O}(N)$ multiplications for the diagonal entries of $\mathbf{u}_k \mathbf{u}_k^t$. The first term of J can be computed with $\mathcal{O}(N)$ multiplications by using the following recursive form of $(\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\hat{\lambda}_j^*))_{j=0}^N$:

$$\begin{aligned}\mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\hat{\lambda}_N^*) &= 0, \\ \mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\hat{\lambda}_j^*) &= \mathbf{u}_k^t \hat{\mathbf{h}}_{k+1}(\hat{\lambda}_{j+1}^*) \\ &\quad - (\hat{\lambda}_{j+1} - \hat{\lambda}_{j+1}^*) \text{tr}[\mathbf{A}_k(\hat{\lambda}_{j+1}^*) \text{diag}(\boldsymbol{\zeta}_k)] \\ &\quad - (\hat{\lambda}_j^* - \hat{\lambda}_{j+1}) \text{tr}[\mathbf{A}_k(\hat{\lambda}_j^*) \text{diag}(\boldsymbol{\zeta}_k)]\end{aligned}$$



(a) The stepsize is set to $\mu_k = 0.1$.



(b) The stepsize is set to $\mu_k = 0.5$.

Fig. 2. Steady-state performance in system-mismatch averaged over 300 trials. The horizontal and vertical axes indicate the SNR and the system mismatch, respectively.

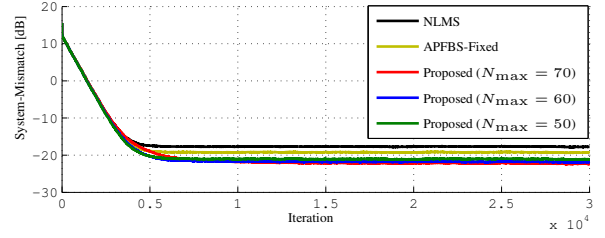
for each nonnegative integer $j \leq N-1$. Hence step 5 can be handled by $\mathcal{O}(N)$ multiplications.

4. NUMERICAL EXAMPLE

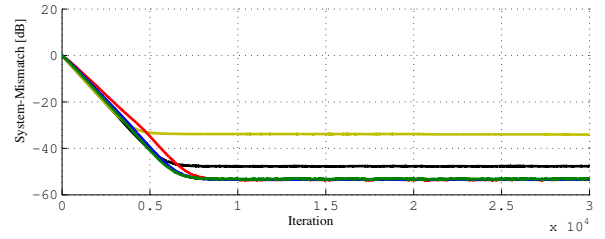
We examine the efficacy of the proposed technique in an example of adaptive filtering setting. The unknown system \mathbf{h}_* of length $N = 100$ is generated artificially to be sparse (it has only $N_A = 30$ active coefficients). The additive noise $(v_k)_{k \geq 0}$ is drawn from the zero mean Gaussian noise with unit variance. The input signal $(u_k)_{k \geq 0}$ is also generated from the zero mean Gaussian noise, and the SNR is varied in 5dB increments from 0dB to 50dB. We adopt the system-mismatch

$$(\text{system-mismatch}) = 20 \log_{10} \left(\frac{\|\mathbf{h}_k - \mathbf{h}_*\|}{\|\mathbf{h}_*\|} \right) \quad [\text{dB}]$$

as a performance measure.



(a) The SNR is set to 5dB.



(b) The SNR is set to 35dB.

Fig. 3. Learning curve averaged over 300 trials. The stepsize is chosen as $\mu_k = 0.1$.

Three adaptive filtering algorithms are examined: The normalized least mean square (NLMS) [34], the APFBS (7) with adaptively weighted soft-thresholding of fixed parameter (labeled as APFBS-Fixed) [8], and Algorithm 2 (referred to as Proposed). The step-size parameter of the algorithms is chosen as $\mu_k = 0.1$ or 0.5 . We adopt as the weight design [14] of the APFBS

$$\omega_k^{(j)} := \frac{1}{|h_k^{(j)}|^{1-p} + \nu}, \quad (12)$$

where $\nu > 0$ is a small positive constant. In this experiment, we set $p = 0$ and $\nu = 10^{-5}$ (see [14] for a superior performance of the choice $p = 0$ compared with different choices). The parameter $\lambda_k = 4.5 \times 10^{-4}$ of APFBS-Fixed is chosen to minimize the system mismatch at 15dB for $\mu_k = 0.1$. The interest region of λ in (11) of Algorithm 2 is limited as $N_{\max} = 70, 60$, or 50 . All the algorithms are terminated 30000 iterations and the system mismatch is averaged over the last 10000 iterations.

Figures 2 and 3 depict the resulting system-mismatch behavior. Figure 2 shows that the proposed method achieves excellent steady-state performance over all the observed SNR for all the selected N_{\max} , while APFBS-Fixed deteriorates the performance in high SNR. Moreover, Figure 3 illustrates that the proposed tuning technique does not affect significantly their convergence speed in early iterations. These demonstrate the efficacy of the proposed technique as well as its robustness against selection of N_{\max} .

5. CONCLUDING REMARKS

We have proposed a parameter tuning for a sparsity-aware variant of the APFBS by utilizing a minimization of an unbiased MSE estimate, which robustly achieves excellent performance against the SNR environmental change in the sense of the system-mismatch. Thanks to the piecewise quadratic structure of the unbiased MSE estimate, we can efficiently obtain the minimizer of the unbiased MSE estimate.

Future work includes an extension of the proposed parameter tuning technique to various adaptive filtering algorithms, e.g., [12, 16, 17, 35], and an extension of the MSE estimate under various noise distributions.

6. REFERENCES

- [1] S. Makino and Y. Kaneda, "Exponentially weighted step-size projection algorithm for acoustic echo cancellers," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. 75, no. 11, pp. 1500–1508, 1992.
- [2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.
- [3] A. W. H. Khong and P. A. Naylor, "Efficient use of sparse adaptive filters," in *Proc. IEEE ACSSC*, 2006, pp. 1375–1379.
- [4] Y. Gu, J. Jin, and S. Mei, " ℓ_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, pp. 774–777, 2009.
- [5] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE ICASSP*, 2009, pp. 3125–3128.
- [6] C. Paleologu, J. Benesty, and S. Ciochina, *Sparse Adaptive Filters for Echo Cancellation*, Morgan and Claypool Publishers, 2010.
- [7] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the ℓ_1 -norm," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3436–3447, 2010.
- [8] 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.
- [9] 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.
- [10] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 936–952, 2011.
- [11] 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.
- [12] I. Yamada, S. Gandy, and M. Yamagishi, "Sparsity-aware adaptive filtering based on a Douglas-Rachford splitting," in *Proc. EUSIPCO*, 2011, pp. 1929–1933.
- [13] A. Asaei, M. J. Taghizadeh, H. Bourlard, and V. Cevher, "Multi-party speech recovery exploiting structured sparsity models," in *Proc. INTERSPEECH*, 2011, pp. 185–188.
- [14] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada, "Sparsity-aware adaptive filters based on ℓ_p -norm inspired soft-thresholding technique," in *Proc. IEEE ISCAS*, 2012, pp. 2749–2752.
- [15] T. Yamamoto, M. Yamagishi, and I. Yamada, "Adaptive proximal forward-backward splitting for sparse system identification under impulsive noise," in *Proc. EUSIPCO*, 2012, pp. 2620–2624.
- [16] S. Ono, M. Yamagishi, and I. Yamada, "A sparse system identification by using adaptively-weighted total variation via a primal-dual splitting approach," in *Proc. IEEE ICASSP*, 2013, pp. 6029–6033.
- [17] M. Yamagishi and I. Yamada, "Exploiting sparsity in feed-forward active noise control with adaptive Douglas-Rachford splitting," in *Proc. APSIPA ASC*, 2013, 6pp.
- [18] C. L. Mallows, "Some comments on C_p ," *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.
- [19] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, pp. 1135–1151, 1981.
- [20] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.
- [21] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 471–481, 2009.
- [22] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Processing*, vol. 57, no. 12, pp. 4616–4632, 2009.
- [23] M. Yamagishi and I. Yamada, "A rank selection of MV-PURE with an unbiased predicted-MSE criterion and its efficient implementation in image restoration," in *Proc. IEEE ICASSP*, 2013, pp. 1573–1577.
- [24] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, vol. 1–2, Springer-Verlag, 1993.
- [25] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Sér.*, vol. 255, pp. 2897–2899, 1962.
- [26] B. Mercier, *Lectures on Topics in Finite Element Solutions of Elliptic Problems*, vol. 63, Springer, 1979.
- [27] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [28] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, 2011.
- [29] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 345–390. Springer, 2011.
- [30] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [31] M. Yamagishi and I. Yamada, "Over-relaxation of the fast iterative shrinkage-thresholding algorithm with variable stepsize," *Inverse Problems*, vol. 27, no. 10, 15pp., 2011.
- [32] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1199–1226, 2013.
- [33] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical Functional Analysis and Optimization*, pp. 593–617, 2004.
- [34] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, 1967.
- [35] H. Kuroda, S. Ono, M. Yamagishi, and I. Yamada, "Exploiting group sparsity in nonlinear acoustic echo cancellation by adaptive proximal forward-backward splitting," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E96-A, no. 10, pp. 1918–1927, 2013.