PROPORTIONATE ADAPTIVE ALGORITHM FOR NONSPARSE SYSTEMS BASED ON KRYLOV SUBSPACE AND CONSTRAINED OPTIMIZATION

Masahiro Yukawa† *Wolfgang Utschick*‡

†Amari Research Unit, Brain Science Institute, RIKEN, Japan
‡Associate Institute for Signal Processing, Munich University of Technology, Germany

ABSTRACT

In this paper, we propose an efficient design of proportionality factors in the recently established algorithm named Krylov-proportionate normalized least mean-square (KPNLMS), which is an extention of the PNLMS algorithm to nonsparse (or dispersive) unknown systems by means of a Krylov subspace. The designing task takes a form of minimizing the number of iterations that is needed for an upper bound of the system mismatch to reach a specified target value. The minimization is performed under several constraints related to numerical stability, computational requirements, and nonnegativity, and its closed-form solution is derived. Numerical examples demonstrate that the proposed design significantly reduces the number of iterations needed to achieve target values of system mismatch especially when a low level of system mismatch is required.

***Index Terms*—** proportionate adaptive algorithm, Krylov subspace, constrained optimization

1. INTRODUCTION

This paper presents an efficient adaptive algorithm named μ-law Krylov-proportionate normalized least mean-square (MKP-NLMS), originating from the Krylov-proportionate adaptive filtering [1]. The basic concept of the Krylov-proportionate adaptive filtering is to shift the proportionate adaptive filtering paradigm[1] [2–4] from 'exploiting the sparsity of an unknown system' to 'sparsifying a not necessarily sparse unknown-system by means of statistical information'. In [1], the main focus is on proving that it is possible to sparsify nonsparse systems by using a certain Krylov subspace into which the statistical information is incorporated. The proportionality factors given to each basis vector of the subspace are determined based on the idea of an improved version in [4], as it can straightforwardly be extended to the Krylov-proportionate adaptive algorithm.

In [6], another improved version named μ-law proportionate normalized least mean-square (MPNLMS) algorithm has been proposed. The MPNLMS algorithm adjusts the proportionality factors based on an objective criterion (unlike the other versions), and it has been shown to exhibit much faster convergence than the original PNLMS algorithm [2] during the whole adaptation process.

The goal of this paper is to derive an optimal design of the proportionality factors based on objective criteria for the Krylov-proportionate adaptive algorithm, and realize it in a computationally efficient way. For this purpose, we firstly investigate optimal designs for a deterministic steepest descent method with proportionality factors, of which a stochastic approximation is the Krylov-proportionate adaptive algorithm. We formulate the designing problem as optimization problems with two different measures: (i) the so-called *overall convergence* and (ii) what we call ϵ-*convergence*. We present a solution to each problem, and show that the solution to the second one is able to be approximated in the adaptive algorithm with low computational complexity. We thus propose, based on the second one, an efficient design for the proposed stochastic algorithm, followed by numerical examples and conclusion.

2. PROBLEM STATEMENTS

Throughout the paper, we let $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and nonnegative integers, respectively. We consider the following linear system model:

$$d_k := \boldsymbol{u}_k^T \boldsymbol{h}^* + n_k,\ k \in \mathbb{N}, \tag{1}$$

where $\boldsymbol{u}_k := [u_k, u_{k-1}, \cdots, u_{k-N+1}]^T \in \mathbb{R}^N$ is the input vector at time k with the input process $(u_k)_{k\in\mathbb{N}}$, $\boldsymbol{h}^* \in \mathbb{R}^N$ the unknown system, $(d_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ the output process, and $(n_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ the noise process. For convenience, we make the following assumptions.

1. The input is white and stationary.[2]
2. The input and noise processes are statistically orthogonal to each other, i.e., $E\{n_k \boldsymbol{u}_k\} = \boldsymbol{0}$.

By Assumption 1, the autocorrelation matrix is $\boldsymbol{R} := E\{\boldsymbol{u}_k \boldsymbol{u}_k^T\} = \sigma^2 \boldsymbol{I}_N \in \mathbb{R}^{N\times N}$, where σ^2 is the variance of the input signal and $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix for $n \in \mathbb{N}^* := \mathbb{N}\setminus\{0\}$. Moreover, by Assumption 2, the cross-correlation vector is $\boldsymbol{p} := E\{\boldsymbol{u}_k d_k\} = \boldsymbol{R}\boldsymbol{h}^* \in \mathbb{R}^N$. Given a filter $\boldsymbol{h} \in \mathbb{R}^N$, the mean-square error (MSE) of the filter output under Assumption 2 is given as

$$J_{\mathrm{M}}(\boldsymbol{h}) := E\{(d_k - \boldsymbol{u}_k^T \boldsymbol{h})^2\} = \|\boldsymbol{h} - \boldsymbol{h}^*\|_{\boldsymbol{R}}^2 + E\{n_k^2\}. \tag{2}$$

It is silently assumed in (2) that $\boldsymbol{R}$ is positive definite. From (2), it is seen that the stochastic minimum MSE (MMSE) filter, i.e., the minimizer of (2), is given by the deterministic vector $\boldsymbol{h}^*$ under the natural assumption. Also (2) implies under Assumptions 1 and 2 that the MSE is a criterion equivalent to the system mismatch:

$$J_{\mathrm{S}}(\boldsymbol{h}) := \|\boldsymbol{h} - \boldsymbol{h}^*\|^2 \|\boldsymbol{h}^*\|^{-2},\ \boldsymbol{h} \in \mathbb{R}^N. \tag{3}$$

Given available data $(\boldsymbol{u}_k)_{k\in\mathbb{N}}$ and $(d_k)_{k\in\mathbb{N}}$, the problem is to estimate $\boldsymbol{h}^*$.

3. μ-LAW KRYLOV-PROPORTIONATE NLMS ALGORITHM

3.1. Proposed Algorithm

Let $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{p}}$ be estimates of $\boldsymbol{R}$ and $\boldsymbol{p}$; e.g., $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{p}}$ are computed by sample averages with/without exponential window. Given

[1] The proportionate adaptive filtering has been proposed originally in [2], and its improved versions have been proposed, e.g., in [3, 4]. The idea is to assign an individual step size to each filter tap, and the step size is roughly proportional to the absolute value of the current tap weight estimate. In [5], it is shown that the proportionate normalized least mean-square (PNLMS) algorithm projects the current filtering vector *with respect to time-varying metric* onto the same hyperplane as the NLMS algorithm, and the algorithm is generalized by using the set-theoretic adaptive filtering approach.

[2] If the input process is highly colored, the whitening (or preconditioning) will be used with the proposed algorithm so that the input process becomes nearly white. Although perfect whitening is almost impossible in nonstationary environments, it is shown in [1] that KPNLMS performs very well even in highly nonstationary environments.

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

$D \in \{1, 2, \cdots, N\}$, define an orthogonal matrix $\boldsymbol{Q} := [\boldsymbol{Q}_1\ \boldsymbol{Q}_2] \in \mathbb{R}^{N\times N}$ (with $\boldsymbol{Q}_1 \in \mathbb{R}^{N\times D}$ and $\boldsymbol{Q}_2 \in \mathbb{R}^{N\times(N-D)}$) such that $\mathcal{R}\{\boldsymbol{Q}_1\} = \mathcal{K}_D(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{p}}) := \operatorname{span}\{\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{R}}\widehat{\boldsymbol{p}}, \cdots, \widehat{\boldsymbol{R}}^{D-1}\widehat{\boldsymbol{p}}\} \subset \mathbb{R}^N$ and $\mathcal{R}\{\boldsymbol{Q}_2\} = \mathcal{K}_D^{\perp}(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{p}}) \subset \mathbb{R}^N$. Here, $\mathcal{R}\{\cdot\}$ stands for *range* and the superposition $(\cdot)^{\perp}$ the orthogonal complement. Also define

$$\boldsymbol{\Theta}_k := \operatorname{diag}\{\underbrace{\theta_k^{(1)}, \cdots, \theta_k^{(D)}}_{D}, \underbrace{\delta_k, \cdots, \delta_k}_{N-D}\} \in \mathbb{R}^{N\times N}, k \in \mathbb{N}, \quad (4)$$

where $\theta_k^{(n)} > 0$ ($n = 1, 2, \cdots, D$) and $\delta_k > 0$ are the proportionality factors. Then, the proposed algorithm is given as follows:

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - \lambda_k \frac{e_k(\boldsymbol{h}_k)}{\boldsymbol{u}_k^T \boldsymbol{\Omega}_k \boldsymbol{u}_k} \boldsymbol{\Omega}_k \boldsymbol{u}_k,\ k \in \mathbb{N}, \quad (5)$$

where $\boldsymbol{\Omega}_k := \boldsymbol{Q}\boldsymbol{\Theta}_k\boldsymbol{Q}^T$ and $e_k(\boldsymbol{h}) := \boldsymbol{h}^T\boldsymbol{u}_k - d_k$, $\forall \boldsymbol{h} \in \mathbb{R}^N$. Until reasonable estimates of $\boldsymbol{R}$ and $\boldsymbol{p}$ are obtained, we simply let $\boldsymbol{\Omega}_k := \boldsymbol{I}_N$. Because of the special structure of $\boldsymbol{\Theta}_k$, we have [1]

$$\boldsymbol{\Omega}_k\boldsymbol{u}_k = \boldsymbol{Q}_1 \left[\boldsymbol{\Theta}_{k,D} - \delta_k \boldsymbol{I}_D\right] \boldsymbol{Q}_1^T \boldsymbol{u}_k + \delta_k \boldsymbol{u}_k, \quad (6)$$

where $\boldsymbol{\Theta}_{k,D} := \operatorname{diag}\{\theta_k^{(1)}, \cdots, \theta_k^{(D)}\} \in \mathbb{R}^{D\times D}$. (6) suggests that (i) there is no need to compute $\boldsymbol{Q}_2$ and (ii) the update equation (5) requires only $O(N)$ computational complexity (Note: A typical value of D is 4 or 5, which is much smaller than N in many applications).

The focus in this work is on a strategic design of $\boldsymbol{\Theta}_k$, which is explored in the remaining of this section. Defining the Krylov coefficients and its corresponding input vector as $\boldsymbol{w}_k := \boldsymbol{Q}^T\boldsymbol{h}_k$ and $\boldsymbol{v}_k := \boldsymbol{Q}^T\boldsymbol{u}_k$, respectively, (5) can be rewritten as follows:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \lambda_k \frac{\bar{e}_k(\boldsymbol{w}_k)}{\boldsymbol{v}_k^T\boldsymbol{\Theta}_k\boldsymbol{v}_k} \boldsymbol{\Theta}_k\boldsymbol{v}_k,\ k \in \mathbb{N}, \quad (7)$$

where $\bar{e}_k(\boldsymbol{w}) := \boldsymbol{w}^T\boldsymbol{v}_k - d_k, \forall \boldsymbol{w} \in \mathbb{R}^N$. (7) is a transform-domain expression of (5). In the transform domain, the MSE is given under Assumption 2 as

$$\bar{J}_{\mathrm{M}}(\boldsymbol{w}) := E\{(d_k - \boldsymbol{v}_k^T\boldsymbol{w})^2\} = \boldsymbol{w}^T\bar{\boldsymbol{R}}\boldsymbol{w} - 2\boldsymbol{w}^T\bar{\boldsymbol{R}}\boldsymbol{w}^* + E\{d_k^2\},$$

where $\bar{\boldsymbol{R}} := \boldsymbol{Q}^T\boldsymbol{R}\boldsymbol{Q}$ and $\boldsymbol{w}^* := \boldsymbol{Q}^T\boldsymbol{h}^*$. In [1], it has been shown that $\boldsymbol{w}^*$ has a sparse structure when the input signal is fairly uncorrelated, thus the idea of proportionate-type adaptive algorithms can be extended.[3] In [6], a strategic design of the diagonal matrix for the PNLMS algorithm is proposed based on an objective criterion. However, because of the special structure of $\boldsymbol{\Theta}_k$ in (4), an extention of the idea in [6] is not straightforward. In fact, the special structure (coming from computational requirements) imposes an additional constraint on the optimization problem, yielding an interesting difference from the original formulation. In the following subsection, we present two problem formulations for the design of $\boldsymbol{\Theta}_k$.

3.2. Two Formulations for Designing $\boldsymbol{\Theta}_k$

With the gradient $\boldsymbol{\nabla}\bar{J}_{\mathrm{M}}(\boldsymbol{w}) = 2\bar{\boldsymbol{R}}(\boldsymbol{w} - \boldsymbol{w}^*)$, the steepest descent algorithm with proportionality factors for $\bar{J}_{\mathrm{M}}(\boldsymbol{w})$ is given as follows:

$$\boldsymbol{w}_{k+1} := \boldsymbol{w}_k - \frac{\alpha}{2}\boldsymbol{\Theta}_k\boldsymbol{\nabla}\bar{J}_{\mathrm{M}}(\boldsymbol{w}_k) = \boldsymbol{w}_k - \alpha\boldsymbol{\Theta}_k\bar{\boldsymbol{R}}(\boldsymbol{w}_k - \boldsymbol{w}^*), \quad (8)$$

where $\alpha > 0$ is a small constant. We mention that (8) is a deterministic algorithm, and (7) is its stochastic approximation. As (8) is simpler than (7) to investigate an optimal design of $\boldsymbol{\Theta}_k$, we explore an optimal design for (8), and exploit it to design $\boldsymbol{\Theta}_k$ for (7).

[3] In [7], it is shown that near-optimal MSE performance can be achieved by using a low-dimensional Krylov subspace.

Defining the error vector $\widetilde{\boldsymbol{w}}_k := [\widetilde{w}_k^{(1)}, \widetilde{w}_k^{(2)}, \cdots, \widetilde{w}_k^{(N)}]^T := \boldsymbol{w}_k - \boldsymbol{w}^*$, it is verified from (8) that

$$\widetilde{\boldsymbol{w}}_{k+1} = (\boldsymbol{I}_N - \alpha\boldsymbol{\Theta}_k\bar{\boldsymbol{R}})\widetilde{\boldsymbol{w}}_k = \prod_{i=0}^{k}(\boldsymbol{I}_N - \alpha\boldsymbol{\Theta}_i\bar{\boldsymbol{R}})\widetilde{\boldsymbol{w}}_0. \quad (9)$$

By Assumption 1, we have $\bar{\boldsymbol{R}} = \sigma^2\boldsymbol{I}_N$ and (9) is reduced to

$$\widetilde{\boldsymbol{w}}_{k+1} = \prod_{i=0}^{k}(\boldsymbol{I}_N - \xi\boldsymbol{\Theta}_i)\widetilde{\boldsymbol{w}}_0, \quad (10)$$

where $\xi := \alpha\sigma^2 > 0$. We assume that $0 \leq \xi\theta_k^{(n)} \leq x^* \approx 1.2785$, $\forall k \in \mathbb{N}$, $\forall n \in \{1, 2, \cdots, N\}$, where $\theta_k^{(D+1)} := \theta_k^{(D+2)} := \cdots := \theta_k^{(N)} := \delta_k$, $k \in \mathbb{N}$, and x^* is the solution to the equation $x - 1 = e^{-x}$. Then, from (10), for $n = 1, 2, \cdots, N$, we obtain

$$|\widetilde{w}_k^{(n)}| = \left|\prod_{i=0}^{k-1}(1 - \xi\theta_i^{(n)})\widetilde{w}_0^{(n)}\right| \leq \prod_{i=0}^{k-1} e^{-\xi\theta_i^{(n)}}|\widetilde{w}_0^{(n)}|. \quad (11)$$

Instead of the error $|\widetilde{w}_k^{(n)}|$ itself, we consider its upper bound given in (11), as it is more tractable. We remark that the bound is tight, provided that $0 \leq \xi\theta_i^{(n)} \ll 1, \forall i \in \mathbb{N}$. For numerical stability, we introduce the condition [2]: $\sum_{n=1}^{N}\theta_k^{(n)} = N, \forall k \in \mathbb{N}$. In designing $\boldsymbol{\Theta}_k$, we wish to minimize the number of iterations required for convergence. A quantitative measure for convergence, called *overall convergence*, is introduced in [6, Definition 2], according to which and (11) we consider the following condition for a given $\epsilon > 0$.

$$\prod_{i=0}^{k-1} e^{-\xi\theta_i^{(n)}}|\widetilde{w}_0^{(n)}| \leq \epsilon,\ \forall n \in \{1, 2, \cdots, N\} \quad (12)$$

$$\Leftrightarrow \boldsymbol{1}_k^T\boldsymbol{\theta}^{(n)} \geq \frac{1}{\xi}\ln\frac{|\widetilde{w}_0^{(n)}|}{\epsilon},\ \forall n \in \{1, 2, \cdots, N\} \quad (13)$$

$$\Leftrightarrow \boldsymbol{1}_k^T\boldsymbol{\Theta}(k) \geq \frac{1}{\xi}\boldsymbol{c}^T. \quad (14)$$

Here, $\boldsymbol{1}_m := [1, 1, \cdots, 1]^T \in \mathbb{R}^m$ for any $m \in \mathbb{N}^*$, and

$$\boldsymbol{\theta}^{(n)} := [\theta_0^{(n)}, \theta_1^{(n)}, \cdots, \theta_{k-1}^{(n)}]^T \in (0, N)^k \quad (15)$$

$$\boldsymbol{\Theta}(k) := [\boldsymbol{\theta}^{(1)}\ \boldsymbol{\theta}^{(2)}\ \cdots\ \boldsymbol{\theta}^{(N)}] \in (0, N)^{k\times N} \quad (16)$$

$$\boldsymbol{c} := \left[\ln\frac{|\widetilde{w}_0^{(1)}|}{\epsilon}, \ln\frac{|\widetilde{w}_0^{(2)}|}{\epsilon}, \cdots, \ln\frac{|\widetilde{w}_0^{(N)}|}{\epsilon}\right]^T \in \mathbb{R}^N. \quad (17)$$

Note that $\boldsymbol{\Theta}(k)$ is different from the diagonal matrix $\boldsymbol{\Theta}_k$, and the height of $\boldsymbol{\Theta}(k)$ depends on $k \in \mathbb{N}$. The problem is thus formulated as follows.

Problem 1: Given $\epsilon > 0$,

$$\min_{\boldsymbol{\Theta}_i} k \quad \text{s.t. } \boldsymbol{\Theta}(k)\boldsymbol{1}_N = N\boldsymbol{1}_k \quad (18)$$

$$\theta_i^{(D+1)} = \theta_i^{(D+2)} = \cdots = \theta_i^{(N)},\ \forall i \in \mathbb{N} \quad (19)$$

$$\theta_i^{(n)} \geq 0,\ \forall n \in \{1, 2, \cdots, N\},\ \forall i \in \mathbb{N} \quad (20)$$

$$\boldsymbol{1}_k^T\boldsymbol{\Theta}(k) \geq \frac{1}{\xi}\boldsymbol{c}^T. \quad (21)$$

Let us now consider another measure. Define $\bar{J}_{\mathrm{S}}(\boldsymbol{w}) := \|\boldsymbol{w} - \boldsymbol{w}^*\|^2\|\boldsymbol{w}^*\|^{-2} = \|\boldsymbol{h} - \boldsymbol{h}^*\|^2\|\boldsymbol{h}^*\|^{-2}$, $\boldsymbol{w} := \boldsymbol{Q}^T\boldsymbol{h} \in \mathbb{R}^N$. Then, another natural measure for convergence would be defined with either $\bar{J}_{\mathrm{M}}(\cdot)$ or $\bar{J}_{\mathrm{S}}(\cdot)$. As shown in Sec. 2, $J_{\mathrm{M}}(\cdot)$ and $J_{\mathrm{S}}(\cdot)$ are equivalent criteria under Assumptions 1 and 2, thus $\bar{J}_{\mathrm{M}}(\cdot)$ and $\bar{J}_{\mathrm{S}}(\cdot)$ are also equivalent. Adopting $\bar{J}_{\mathrm{S}}(\cdot)$, we define *$\widetilde{\epsilon}$-convergence* as follows: the number of iterations for $\widetilde{\epsilon}$-convergence is the number of iterations that is needed to have $\bar{J}_{\mathrm{S}}(\boldsymbol{w}_k) \leq \widetilde{\epsilon}$. Referring to (11), we have

$$\bar{J}_{\mathrm{S}}(\boldsymbol{w}_k)\,\|\boldsymbol{w}^*\|^2 \leq \sum_{n=1}^{N}\prod_{i=0}^{k-1} e^{-2\xi\theta_i^{(n)}}|\widetilde{w}_0^{(n)}|^2. \tag{22}$$

For tractability and convenience, we consider the upper bound of $\bar{J}_{\mathrm{S}}(\boldsymbol{w}_k)\,\|\boldsymbol{w}^*\|^2$ instead of $\bar{J}_{\mathrm{S}}(\boldsymbol{w}_k)$ in the optimization, leading to the following formulation.

Problem 2: Given $\epsilon' \;(:= \|\boldsymbol{w}^*\|^2\,\widetilde{\epsilon}) \; > 0$,

$$\min_{\boldsymbol{\Theta}_i} k \quad \text{s.t. (18)–(20) and}$$

$$f(k,\boldsymbol{\Theta}(k)) := \sum_{n=1}^{N}\prod_{i=0}^{k-1} e^{-2\xi\theta_i^{(n)}}|\widetilde{w}_0^{(n)}|^2 \leq \epsilon'. \tag{23}$$

In [6], it has been reported that for ξ small the minimum number of iterations in the MSE sense is obtained if and only if the absolute value of each coefficient error becomes equal to ϵ/N after the same number of iterations. This implies under Assumptions 1 and 2 that, as the constraint (19) is not considered in [6], Problems 1 and 2 are essentially equivalent if we remove (19). Interestingly, however, an imposition of (19) makes Problems 1 and 2 essentially different, as clarified below. In what follows, we present a solution to each problem, and discuss which is more suitable for adaptive implementation.

3.3. Closed-Form Solution to Problem 1

Define $n_{\max} \in \arg\max_{n\in\{D+1,\cdots,N\}} |\widetilde{w}_0^{(n)}|$. Then, the following proposition holds.

Proposition 1 *Assume that* $|\widetilde{w}_0^{(n)}| > \epsilon$, $n = 1,2,\cdots,D,n_{\max}$. *Let* $k_{0,\min}$ *be the minimum* k *corresponding to* Problem 1. *Then,*

$$k_{0,\min} = \left\lceil \frac{1}{N\xi}\left((N-D)\ln\frac{|\widetilde{w}_0^{(n_{\max})}|}{\epsilon} + \sum_{n=1}^{D}\ln\frac{|\widetilde{w}_0^{(n)}|}{\epsilon}\right)\right\rceil, \tag{24}$$

which is achieved by (but not only by)

$$\theta_i^{(n)} = \frac{1}{\eta}\ln\frac{|\widetilde{w}_0^{(n)}|}{\epsilon},\; n\in\{1,\cdots,D\},\; \forall i\in\mathbb{N} \tag{25}$$

$$\theta_i^{(n)} = \frac{1}{\eta}\ln\frac{|\widetilde{w}_0^{(n_{\max})}|}{\epsilon},\; n\in\{D+1,\cdots,N\},\; \forall i\in\mathbb{N} \tag{26}$$

$$\eta := \frac{1}{N}\left((N-D)\ln\frac{|\widetilde{w}_0^{(n_{\max})}|}{\epsilon} + \sum_{n=1}^{D}\ln\frac{|\widetilde{w}_0^{(n)}|}{\epsilon}\right). \tag{27}$$

Proof: By (19), the condition (21) is reduced to

$$\boldsymbol{1}_k^T\widetilde{\boldsymbol{\Theta}}(k) \geq \frac{1}{\xi}\widetilde{\boldsymbol{c}}^T, \tag{28}$$

where $\widetilde{\boldsymbol{\Theta}}(k) := [\boldsymbol{\theta}^{(1)}\cdots\boldsymbol{\theta}^{(D)}\;\boldsymbol{\theta}^{(n_{\max})}] \in (0,N)^{k\times(D+1)}$, $\widetilde{\boldsymbol{c}} := \left[\ln\frac{|\widetilde{w}_0^{(1)}|}{\epsilon},\cdots,\ln\frac{|\widetilde{w}_0^{(D)}|}{\epsilon},\ln\frac{|\widetilde{w}_0^{(n_{\max})}|}{\epsilon}\right]^T \in \mathbb{R}^{D+1}$. Moreover, (18) and (19) imply

$$\widetilde{\boldsymbol{\Theta}}(k)\widetilde{\boldsymbol{1}}_{D+1} = N\boldsymbol{1}_k, \tag{29}$$

where $\widetilde{\boldsymbol{1}}_{D+1} := [1,\cdots,1,N-D]^T \in \mathbb{R}^{D+1}$. By (29), we have (28) $\Rightarrow k \geq \frac{1}{N\xi}\widetilde{\boldsymbol{c}}^T\widetilde{\boldsymbol{1}}_{D+1} =: k_{\mathrm{lb}} \in \mathbb{R}$, where k_{lb} is a lower bound of $k_{0,\min}$; i.e., $k_{0,\min} \geq \lceil k_{\mathrm{lb}}\rceil$.

Now, taking a careful look at Problem 1, we see that an arbitrary change of rows in $\boldsymbol{\Theta}(k)$ makes no impact on the conditions. This suggests that there should be multiple solutions in general. We thus introduce another condition: $\theta_1^{(n)} = \theta_2^{(n)} = \cdots = \theta_k^{(n)} (=: \theta^{(n)})$, $\forall n \in \{1,2,\cdots,N\}$. This condition allows us to express $\widetilde{\boldsymbol{\Theta}}(k)$ as $\widetilde{\boldsymbol{\Theta}}(k) = \boldsymbol{1}_k\widetilde{\boldsymbol{\theta}}^T$, where $\widetilde{\boldsymbol{\theta}} := [\theta^{(1)},\cdots,\theta^{(D)},\theta^{(n_{\max})}]^T \in (0,N)^{D+1}$. With the above arguments, Problem 1 is reduced to the following problem:

$$\min_{\widetilde{\boldsymbol{\theta}}} k \quad \text{s.t. } \widetilde{\boldsymbol{\theta}}^T\widetilde{\boldsymbol{1}}_{D+1} = N \tag{30}$$

$$\widetilde{\boldsymbol{\theta}} \geq \boldsymbol{0} \tag{31}$$

$$k\widetilde{\boldsymbol{\theta}} \geq \frac{1}{\xi}\widetilde{\boldsymbol{c}}. \tag{32}$$

The condition (30) implies that (32) $\Rightarrow k \geq \frac{1}{N\xi}\widetilde{\boldsymbol{c}}^T\widetilde{\boldsymbol{1}}_{D+1} (= k_{\mathrm{lb}})$. Let $\widetilde{\boldsymbol{\theta}}^* := \frac{1}{\xi k_{\mathrm{lb}}}\widetilde{\boldsymbol{c}}$. Then, the choice of $(\widetilde{\boldsymbol{\theta}},k) = (\widetilde{\boldsymbol{\theta}}^*,\lceil k_{\mathrm{lb}}\rceil)$ satisfies (30), (31), and (32), meaning that the lower bound $\lceil k_{\mathrm{lb}}\rceil$ is achieved by $\widetilde{\boldsymbol{\theta}}^*$. This verifies that (24)–(26) is a solution to Problem 1. □

3.4. Closed-Form Solution to Problem 2

By (19), it follows that

$$f(k,\boldsymbol{\Theta}(k)) = \sum_{n=1}^{D}\prod_{i=0}^{k-1} e^{-2\xi\theta_i^{(n)}}|\widetilde{w}_0^{(n)}|^2 + \sum_{n=D+1}^{N}\prod_{i=0}^{k-1} e^{-2\xi\delta_i}|\widetilde{w}_0^{(\mathrm{ave})}|^2,$$

where $\widetilde{w}_0^{(\mathrm{ave})} := \left(\frac{1}{N-D}\sum_{n=D+1}^{N}|\widetilde{w}_0^{(n)}|^2\right)^{1/2}$. This implies that, in finding a solution to Problem 2, we can replace $\widetilde{w}_0^{(n)}$, $n \in \{D+1,\cdots,N\}$, by $\widetilde{w}_0^{(\mathrm{ave})}$. Based on this observation, we can prove the following proposition (Proof is omitted due to lack of space).

Proposition 2 *Assume that* $|\widetilde{w}_0^{(n)}| > \epsilon$, $n = 1,2,\cdots,D$, *and* $|\widetilde{w}_0^{(\mathrm{ave})}| > \epsilon$, *where* $\epsilon := \sqrt{\epsilon'/N} (= \|\boldsymbol{w}^*\|\sqrt{\widetilde{\epsilon}/N}) > 0$. *Then, the solution to Problem 2 is given by* (24)–(27) *with the replacement of* $\widetilde{w}_0^{(\max)}$ *by* $\widetilde{w}_0^{(\mathrm{ave})}$.

We should mention that, because we consider an upper bound given in (11) [or (22)], it is guaranteed that the true error (or system mismatch) reaches ϵ (or $\widetilde{\epsilon}$) within $k_{0,\min}$ iterations given in Proposition 1 (or Proposition 2), provided that the assumption $0 \leq \xi\theta_k^{(n)} \leq x^*$ holds.

3.5. Construction of $\boldsymbol{\Theta}_k$ in Proposed Algorithm

In the previous subsections, we show that the solutions to Problems 1 and 2 are different in general. Which formulation is more suitable for the adaptive algorithm from the computational aspect?

In practice, the optimal filter $\boldsymbol{w}^*$ is not available, thus we should approximate $\widetilde{\boldsymbol{w}}_0 (:= \boldsymbol{w}_0 - \boldsymbol{w}^*)$ somehow. A possible candidate would be the current estimate $\widehat{\boldsymbol{w}}_k := [\widehat{w}_k^{(1)},\widehat{w}_k^{(2)},\cdots,\widehat{w}_k^{(N)}]^T := \boldsymbol{w}_0 - \boldsymbol{w}_k (= \boldsymbol{Q}^T(\boldsymbol{h}_0 - \boldsymbol{h}_k))$. If we use Proposition 1, we need all the components of $\widehat{\boldsymbol{w}}_k$, which requires N^2 multiplications (Note: Although only $D+1$ elements of $\widehat{\boldsymbol{w}}_k$ would be involved in the design of $\boldsymbol{\Theta}_k$, the other elements are also required to find $n_{\max}$). On the other hand, if we use Proposition 2, $\widetilde{w}_0^{(\mathrm{ave})}$ is approximated by

$$\widehat{\widetilde{w}}_0^{(\mathrm{ave})} := \left(\frac{\|\widehat{\boldsymbol{w}}_k\|^2 - \|[\widehat{\boldsymbol{w}}_k]_{1:D}\|^2}{N-D}\right)^{1/2}, \tag{33}$$

which can be computed efficiently as explained later; $[\cdot]_{1:D}$ stands for a subvector consisting of the 1st to Dth elements. Therefore, we exploit Proposition 2 in the proposed algorithm.

For convenience, we define $\boldsymbol{y}_k := [y_k^{(1)}, y_k^{(2)}, \cdots, y_k^{(D+1)}]^T := [\widehat{w}_k^{(1)}, \widehat{w}_k^{(2)}, \cdots, \widehat{w}_k^{(D)}, \widehat{\widetilde{w}}_0^{(\mathrm{ave})}]^T \in \mathbb{R}^{D+1}$. Moreover, to ensure $\theta^{(n)} \geq 0$, we use a practical approximation $\ln \frac{|\widetilde{w}_0^{(n)}|}{\epsilon} \approx \ln\left(1 + \frac{|\widetilde{w}_0^{(n)}|}{\epsilon}\right)$, under $\epsilon \ll |\widetilde{w}_0^{(n)}|$, by following the way in [6]. In the proposed algorithm, $\boldsymbol{\Theta}_k$ is constructed as below.

Requirements: $\rho > 0$, $\delta_{\mathrm{p}} > 0$, $\mu := 1/\epsilon$ ($\epsilon = \|\boldsymbol{w}^*\| \sqrt{\widetilde{\epsilon}/N}$).
Construction of $\boldsymbol{\Theta}_k$:

$$
\begin{aligned}
[\boldsymbol{y}_k]_{1:D} :=& \boldsymbol{Q}_1^T(\boldsymbol{h}_0 - \boldsymbol{h}_k) \in \mathbb{R}^D \\
y_k^{(D+1)} =& \sqrt{(\|\boldsymbol{h}_0 - \boldsymbol{h}_k\|^2 - \|[\boldsymbol{y}_k]_{1:D}\|^2)/(N-D)} \\
F(|y_k^{(n)}|) :=& \ln(1 + \mu |y_k^{(n)}|), \ n \in \{1, 2, \cdots, D+1\} \\
\gamma_k^{\min} :=& \rho \max\{\delta_{\mathrm{p}}, F(|y_k^{(1)}|), \cdots, F(|y_k^{(D)}|), F(|y_k^{(D+1)}|)\} \\
\gamma_k^{(n)} :=& \max\{\gamma_k^{\min}, F(|y_k^{(n)}|)\}, \ n \in \{1, 2, \cdots, D+1\} \\
\eta_k :=& (N-D)\gamma_k^{(D+1)} + \sum_{n=1}^{D} \gamma_k^{(n)} \\
\theta_k^{(n)} :=& \gamma_k^{(n)}/\eta_k, \ n \in \{1, \cdots, D\}, \\
\delta_k :=& \gamma_k^{(D+1)}/\eta_k.
\end{aligned}
$$

We call the resulting algorithm μ-law KPNLMS (MKP-NLMS) algorithm, named after the MPNLMS algorithm [6]. Note that, although the optimal design given in Proposition 2 does *not* depend on the time index $k \in \mathbb{N}$, the proposed design for the adaptive algorithm *does* depend on k because $\boldsymbol{w}^*$ is approximated by $\boldsymbol{w}_k$. A convergence analysis for the algorithms having the form of (5) is presented in [8] in a unified manner by extending the framework called *adaptive projected subgradient method*.

Computational complexity: The extra complexity for the proposed design of $\boldsymbol{\Theta}_k$ compared with KPNLMS [1] is no more than $N + 2D$ multiplications plus 1 square-root operation, $D + 1$ logarithmic operations, and $2(D + 1)$ comparisons. Note that $[\boldsymbol{y}_k]_{1:D}$ is common to KPNLMS and the proposed algorithm, and it can be computed in a recursive manner. The computational complexity of KPNLMS per iteration is $(2D + 4)N + 4D$, of which D is for constructing $\boldsymbol{\Theta}_k$. With a typical choice $D = 4$ or $D = 5$, the extra complexity is 7.2% – 10% of the complexity of KPNLMS (For colored inputs, it is about 6.3% – 8.7%). The construction of the matrix $\boldsymbol{Q}_1$ involves $(D - 1)N^2 + D^2 N$ multiplications, however, this computation is required only once.

4. NUMERICAL EXAMPLES

To compare the performance of the proposed algorithm and KPNLMS in the sense of $\widetilde{\epsilon}$-convergence, simulations are conducted with $\boldsymbol{h}^*$ generated randomly for $N = 50$. The input is white with the signal to noise ratio (SNR) $= 60$ dB, where SNR is defined as SNR $:= 10\log_{10}\left(E\{z_k^2\}/E\{n_k^2\}\right)$ dB ($z_k := \boldsymbol{u}_k^T \boldsymbol{h}^*$).[4] The system mismatch is calculated with an arithmetic average over 300 independent runs.

For both algorithms, the step size is set to $\lambda_k = 0.02$ to obtain reasonably small estimation errors. For KPNLMS, the parameters are set to the same values as in [1]. For the proposed algorithm, the parameters are set to $\rho = 0.01$, $\delta_{\mathrm{p}} = 0.01$; μ is determined according to the value of $\widetilde{\epsilon}$ which ranges between 10^{-7} and 10^{-2}. $\widehat{\boldsymbol{R}}$ and $\widehat{\boldsymbol{p}}$ are computed by the simple sample-averages with the initials $\widehat{\boldsymbol{R}}_0 := 0.01\boldsymbol{I}_N$ and $\widehat{\boldsymbol{p}}_0 := \boldsymbol{0}$. The value of D is selected with its maximum $D_{\max} := 8$ in the way presented in [1]; an average value of D in the simulation was 4.0.

[4]Although SNR $= 60$ dB would be impractically high, we intend to show in this numerical example that the lower the required level of system mismatch is, the higher the gain due to the proposed method is.

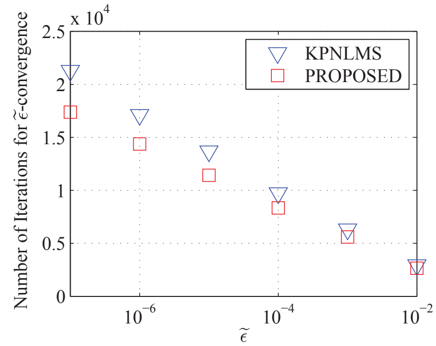

Fig. 1. (a) The required level of system mismatch $\widetilde{\epsilon}$ versus the number of iterations for $\widetilde{\epsilon}$-convergence under SNR $= 60$ dB.

Figure 1 plots $\widetilde{\epsilon}$, which is the required level of system mismatch, against the number of iterations for $\widetilde{\epsilon}$-convergence. It is seen that the proposed algorithm outperforms KPNLMS overall. In particular, the difference is distinctive when a low level of system mismatch is required.

5. CONCLUSION

This paper has presented an efficient adaptive algorithm based on (i) sparsification of an unknown system by the orthogonal transformation with the Krylov subspace basis and (ii) proportionate adaptive filtering that exploits the sparsity of the transformed unknown system. An efficient design of the proportionality factors has been derived based on a constrained optimization. The numerical examples have demonstrated that the proposed algorithm significantly reduces the number of iterations for ϵ-convergence when a low level of system mismatch is required.

6. REFERENCES

[1] M. Yukawa, "Krylov-proportionate adaptive filtering techniques not limited to sparse systems," *IEEE Trans. Signal Processing*, 2009, accepted.

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

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

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

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

[6] H. Deng and M. Doroslovački, "Proportionate adaptive algorithms for network echo cancellation," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1794–1803, May 2006.

[7] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Signal Processing*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[8] M. Yukawa and I. Yamada, "Adaptive parallel variable-metric projection algorithm —An application to acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2007, pp. 1353–1356.