SUMMARY We propose a novel adaptive filtering scheme named metric-combining normalized least mean square (MC-NLMS). The proposed scheme is based on iterative metric projections with a metric designed by combining multiple metric-matrices convexly in an adaptive manner, thereby taking advantages of the metrics which rely on multiple pieces of information. We compare the improved PNLMS (IPNLMS) algorithm with the natural proportionate NLMS (NPNLMS) algorithm, which is a special case of MC-NLMS, and it is shown that the performance of NPNLMS is controllable with the combination coefficient as opposed to IPNLMS. We also present an application to an acoustic echo cancellation problem and show the efficacy of the proposed scheme.

key words: adaptive filter, proportionate NLMS, transform domain adaptive filter, metric projection

1. Introduction

The goal of this paper is to provide a simple but effective adaptive filtering algorithm based on a simultaneous use of multiple pieces of information. Indeed, a significant amount of researches has been done to improve the convergence behavior of the classical normalized least mean square (NLMS) algorithm by using available information about input signals and unknown systems to be estimated [1]–[12]. For instance, Narayan et al. have proposed the transform-domain adaptive filter (TDAF) which improves the convergence behavior for colored inputs, and this approach can be regarded as using information of the second-order statistics of input signals [1], [2]. Meanwhile, Dutweiler has proposed the proportionate NLMS (PNLMS) algorithm which accelerates the convergence speed for sparse impulse responses, and this approach exploits the a priori information that unknown systems tend to be sparse [3]. Here, sparse means that most of the coefficients are (close to) zero. An efficient adaptive filtering algorithm with high convergence speed is demanded for advanced communication systems [13]. Acoustic echo cancellation (AEC), for instance, is still an important application of adaptive filtering due to the rapid spread of Internet conference. In this application, the unknown systems are typically sparse and the speech input signals have high autocorrelation. This is a case in which multiple pieces of information are available for adaptation. As yet however, no techniques have been developed to exploit multiple pieces of information simultaneously so that better performance can be attained than using only one of them.

In this paper, we propose an efficient adaptive filtering scheme based on iterative use of metric projection with the metric changing in time according to incoming data. The variable-metric projection technique has been proposed by Yukawa and Yamada in 2009 [14]. From the variable-metric projection viewpoint, the TDAF and PNLMS algorithms can be regarded as ones using time-varying metrics designed with available information about the input statistics and the system sparsity, respectively. Here, the metric measures the distance between two points in the Euclidean space of adaptive filters and is given by a positive definite matrix. The projection has been used extensively as a backbone in adaptive filtering algorithms [15], [16], and is defined as the closest point, in the sense of minimizing the metric distance, in a closed convex set from the current adaptive filter. The proposed scheme combines multiple metric-matrices convexly in an adaptive manner for taking advantages of the metrics, dubbed the Metric-Combining Normalized Least Mean Square (MC-NLMS) algorithm. In terms of metric, it turns out that the improved PNLMS (IPNLMS) algorithm (Benesty and Gay, 2002) [5] combines two metrics as a convex combination of the inverses of the metric-matrices. To see that metric is an intrinsic factor that governs the performance of iterative projection methods, we study the performance of the proposed scheme and the IPNLMS algorithm using the same metric-matrices. Numerical examples show that the performance of the proposed scheme changes gradually as the convex-combination coefficients slide from zero to one, whereas that of IPNLMS changes drastically by a slight change of the coefficients. From a practical point of view, the proposed scheme is user-friendly compared to IPNLMS since an assignment of the coefficient leads to expectable performance, or in other words, the performance of the scheme is quite controllable. This supports the significance of considering the metric used for the computation of
the projection.

The above example exploits two metrics: the Euclidean metric (which is used, for example, in NLMS) and the metric based on sparsity information. As an example to exploit multiple pieces of information, we apply the proposed scheme to the AEC problem. As already mentioned, it is possible to use the information about the system sparsity and the coloredness of input signals. More specifically, the sparsity is exploited to design a metric-matrix by following the way of the $\mu$-law proportionate NLMS (MPNLMS) algorithm (Deng and Doroslavci, 2005). Roughly speaking, the aim of using these two metrics is taking the advantage of TDAF (which is the uniform rate of convergence) and the advantage of MPNLMS (which is the fast initial convergence) and the adaptive algorithm (Deng and Doroslavci, 2005). Roughly speaking, the aim of using these two metrics is taking the advantage of TDAF (which is the uniform rate of convergence) and the advantage of MPNLMS (which is the fast initial convergence) and the adaptive algorithm (Deng and Doroslavci, 2005).

2. Preliminaries

2.1 A System Model

Throughout this paper, the following notations are used. Let $\mathbb{R}$ and $\mathbb{N}$ denote the sets of all real numbers and non-negative integers, respectively. We denote by $(\cdot)^T$ the transpose of a vector/matrix. Also we denote by $\| \cdot \|$ and $\| \cdot \|_2$ the $\ell_1$ and $\ell_2$ norms, respectively. We consider the following linear system model:

$$ d_k := z_k + n_k = h^T u_k + n_k, \quad k \in \mathbb{N}, $$

where $u_k := [u_k, u_{k-1}, \ldots, u_{k-N+1}]^T \in \mathbb{R}^N$ is the input vector of length $N$ at time $k$ with the input process $(u_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$, $h^* := [h_1^*, h_2^*, \ldots, h_N^*]^T \in \mathbb{R}^N$ is the unknown system, $z_k := h^T u_k \in \mathbb{R}$, and $(n_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}$ is the noise process. The adaptive filter at time $k$ is denoted by $h_k := [h_k^{(1)}, h_k^{(2)}, \ldots, h_k^{(N)}]^T \in \mathbb{R}^N$. The residual-error function for each pair of input-output data $(u_k, d_k)$ is given by $e_k(h) := h^T u_k - d_k$, $h \in \mathbb{R}^N$. The signal to noise ratio (SNR) is defined as

$$ \text{SNR} := 10 \log_{10} \frac{E\left[ z_k^2 \right]}{E\left[ n_k^2 \right]} \text{ (dB)}, $$

where $E[\cdot]$ stands for expectation. The goal is to build an adaptive algorithm that diminishes the error $e_k(h_k)$ in a small number of iterations. In the following subsection, we introduce the variable-metric projection (time-varying metric projection) framework which is a fundamental technique to derive the proposed scheme.

2.2 Variable-Metric Projection Framework

We first show that the projection depends on metric design. Given any positive definite matrix $(\mathbb{R}^{N \times N} \ni) A > 0$, we define an inner product and its induced norm, respectively, as $(x, y)_A := x^T A y$, $\forall x, y \in \mathbb{R}^N$, and $\| x \|_A := (\sqrt{(x, x)_A})$, $\forall x \in \mathbb{R}^N$. Hereafter, we regard $A$ as a metric which determines the metric-distance between $x$ and $y$ by $\| x - y \|_A$.

**Definition 1:**

(a) **A-orthogonal:** Vectors $x$ and $y$ are said to be $A$-orthogonal, if $(x, y)_A = 0$.

(b) **A-projection:** Given a closed convex set $C \subseteq \mathbb{R}$, the metric projection of $x$ onto $C$ in terms of the metric $A$ is defined as

$$ P^A_C(x) := \arg\min_{y \in C} \| x - y \|_A, $$

which is referred to as the $A$-projection of $x$ onto $C$.

The set of filters $h$ that nullify the instantaneous error $e_k(h)$ is the following hyperplane:

$$ H_k := \{ h \in \mathbb{R}^N : h^T u_k = (h, A^{-1} u_k)_A = d_k \}, \quad k \in \mathbb{N}. $$

In the Hilbert space equipped with the inner product $(\cdot, \cdot)_A$, the normal vector of $H_k$ is $A^{-1} u_k$, and we can verify the $A$-projection of $x$ onto the hyperplane $H_k$ can be computed as

$$ P^A_{H_k}(x) = x - \frac{e_k(x)}{u_k^T A^{-1} u_k} A^{-1} u_k. $$

Figure 1 illustrates the projections onto $H_k$ in the case of $d_k = 0$ based on three different metrics: the Euclidean metric (i.e., the identity matrix $I$) and some other metrics $A > 0$ and $B > 0$. It is clear that the input vector $u_k$ is $I$-orthogonal to any $h \in H_k$ since $(h, u_k)_I = h^T u_k = 0$. Moreover, the vector $A^{-1} u_k$ is $A$-orthogonal to any $h \in H_k$ since $(h, A^{-1} u_k)_A = h^T A A^{-1} u_k = 0$. The same applies to the metric $B$. For any $x \notin H_k$, the projections $P^I_{H_k}(x)$ and $P^B_{H_k}(x)$ are different from each other (unless $A$ is proportional to $I$), implying that the orthogonal projection depends on the metric design. In fact, the metric design affects the performance of an adaptive algorithm, as will be shown in Sect.3.

Let us explain now that the use of variable metric is beneficial for improving the performance of an adaptive algorithm. Indeed,

- an adequate metric may change in time due to the nonstationarity of data or may depend on the phase of

![Fig. 1](image-url) A geometric interpretation of the orthogonal projections onto the hyperplane $H_k$ in terms of three different metrics.
3. The Proposed Adaptive Algorithm

The proposed adaptation process is shown. The proposed adaptive algorithm is closely related to IPNLMS. In this section, the proposed algorithm is first presented. Finally, a particular case of MC-NLMS (which is referred to as the metric-combining NLMS (MC-NLMS) algorithm. Several specific metric designs are then presented in Sect. 3.3 that the normalization by \( \|A_{k,m}\|_2 \) is reasonable. The proposed adaptive algorithm is given as follows (see (5)):

\[
A_k := \sum_{m=1}^{M} \eta_{k,m} A_{k,m},
\]

where \( \eta_{k,m} \in [0, 1] \) is the convex combination coefficients satisfying \( \sum_{m=1}^{M} \eta_{k,m} = 1, \forall k \in \mathbb{N} \). We let the maximal eigenvalue of \( A_{k,m} \) to be normalized, or equivalently \( \|A_{k,m}\|_2 = 1 \). Although there are other ways for normalizing the metrics \( A_{k,m} \), it will become clear in Sect. 3.3 that the normalization by \( \|A_{k,m}\|_2 \) is reasonable. The proposed adaptive algorithm is given as follows (see (5)):

\[
h_{k+1} := h_k + \lambda_k p_{H_k}^A(h_k) - h_k = h_k - \lambda_k \frac{p_{H_k}^A(h_k)}{u_k A_k^{-1} u_k} A_k u_k, k \in \mathbb{N},
\]

where \( \lambda_k \in (0, 2) \) is the step size. We call (7) together with (6) the MC-NLMS algorithm. The performance of MC-NLMS is governed by the design of the metrics \( A_{k,m} \) and the combination coefficients \( \eta_{k,m} \). Each \( A_{k,m} \) is designed based on each piece of available information (such as the sparsity information), and the proposed algorithm exploits the whole available information by designing the metric \( A_k \) in the form of (6). Furthermore, \( \eta_{k,m} \) is controlled dynamically to increase the gain coming from the combined metric \( A_k \) at each iteration.

3.2 Examples of Metric Design

We present some special cases of the MC-NLMS algorithm with different metric designs.

1) Case of \( M = 1 \):

a) Letting \( A_k = I \) yields the NLMS algorithm. In this case, the metric, and the norm, become Euclidean.

b) Letting

\[
A_k := \left( \sum_{i=1}^{N} \gamma_k^{(i)} \right) \text{diag}^{-1}(\gamma_k^{(1)}, \gamma_k^{(2)}, \cdots, \gamma_k^{(N)})
\]

yields the PNLMS algorithm [3], where \( \text{diag}^{-1} \) is the inverse of diagonal matrix and

\[
\gamma_k^{(i)} := \max \left\{ \gamma_k^{(i)}, \|h_k^{(i)}\| > 0, \gamma_k^{(i)} := \rho \max \left\{ \sigma, |h_k^{(1)}|, |h_k^{(2)}|, \cdots, |h_k^{(N)}| \right\} > 0 \right. \]

(10)

Here, \( \rho > 0 \) and \( \sigma > 0 \) are positive constants.

Given \( M \) different metrics \( A_{k,m} (m = 1, 2, \cdots, M) \), we can consider the \( M \) different Hilbert spaces \( (\mathbb{R}^N, \langle \cdot, \cdot \rangle_{A_{k,m}}) \), \( m = 1, 2, \cdots, M \). What if we do the learning in the Cartesian product of those spaces? This fundamental question has been investigated in [18]. The metric-combining idea presented in (6) comes from the algorithm that is obtained as a particular example of the adaptive projected subgradient method (APSM) [19] in the product space.
Note that the inverse $A_k^{-1}$ of the metric matrix, rather than $A_k$, is used in (7).

3) Other examples of MC-NLMS for $M = 1$ include KPNLMS [20], TDAF [1, 2], and QNAF [21, 22].

2) Case of $M = 2$:

a) Letting

$$A_k := [\eta A_{k,1} + (1 - \eta)A_{k,2}]^{-1},$$

$$A_{k,1}^{-1} := \frac{1}{N} I,$$

$$A_{k,2}^{-1} := \frac{1}{||h_k||_2 + \epsilon} \text{diag}(|h_k|),$$

yields the IPNLMS algorithm [5], where $\eta \in [0, 1]$, $\epsilon \geq 0$ is the regularization parameter, and $|$ is the elementwise absolute-value operation. Note here that this is not a special case of MC-NLMS because $A_k$ is not given in the form of (6). (Here, $A_{k,1}$ is employed by NLMS and $A_{k,2}$ for $\epsilon = 0$ is identical to $A_k$ in (8) with $\gamma_k^{\max} := 0$.)

b) The Natural-PNLMS (NPNLMS) algorithm: Let

$$A_k := \eta A_{k,1} + (1 - \eta)A_{k,2},$$

$$A_{k,1} := I,$$

$$A_{k,2} := \frac{1}{||G_k||_2} \text{diag}(G_k),$$

where $G_k := \text{diag}(g_{k,1}^{(1)}, g_{k,2}^{(2)}, \ldots, g_{k,\tau}^{(\tau)})$, $g_{k,\tau}^{(i)} := \left(\max\{\tau, |h_k^{(i)}|\}\right)^{-1}$, $\tau > 0$ is the regularization parameter, $\eta_{k,1} := \eta \in [0, 1]$, and $\eta_{k,2} := 1 - \eta$.

We refer to this special case of MC-NLMS as NPNLMS.

NPNLMS exploits basically the same information as IPNLMS but in a natural way. The following subsection will show the significant advantages of NPNLMS over IPNLMS.

3.3 Advantage of NPNLMS over IPNLMS

Computer simulations are conducted to show how much impact the change of the convex combination coefficient $\eta$ gives on the performance of IPNLMS and NPNLMS. We use white input signals, additive white Gaussian noise with SNR = 30 dB, and impulse responses $h_1^*$ and $h_2^*$ which have different sparsity levels for $N = 256$ (see Figs. 3(a) and 3(b)). The normalized mean squared error (NMSE) is defined as follows:

$$J_M(h) := \frac{E\{e_k^2(h)\}}{E\{e_k^2\}}, h \in \mathbb{R}^N,$$

which is approximately computed in our experiments by taking an arithmetic average over 300 independent trials. For IPNLMS, $A_k = 0.4$, $\forall k \in \mathbb{N}$, and $\epsilon = 10^{-4}$. For NPNLMS, $A_k = 0.4$, $\forall k \in \mathbb{N}$, and $\tau = 10^{-4}$.

Figures 4 and 5 show the results for $h_1^*$ and $h_2^*$, respectively. For both IPNLMS and NPNLMS, $\eta = 0$ and $\eta = 1.0$ correspond respectively to PNLMS and NLMS. It can be seen that the performance of the proposed scheme (NPNLMS) is well controlled by the combination coefficient. In contrast, the performance of IPNLMS is not well controlled by the combination coefficient, and $\eta = 0.2, 0.4, \text{and} 0.6$ yield nearly the same performance on average. As a result, the best performance of IPNLMS is inferior to that of NPNLMS for both $h_1^*$ and $h_2^*$. This implies that the performance of the proposed scheme is quite controllable with the combination coefficients.

To analyze the phenomena observed in Figs. 4 and 5, we plot in Figs. 6 and 7 the diagonal elements of $A_k$ for $h_1^*$ and $h_2^*$, respectively, as a function of $\eta$. By definition, the metric projection is unchanged under a multiplication of the metric by a positive scalar (see (3)). Therefore, for each $\eta$, the metric matrix $A_k$ for IPNLMS is divided by its 2-norm (i.e., its spectrum radius) so that its maximum eigenvalue becomes unity. It can be seen that the ratios between the maximum eigenvalue and the other eigenvalues drop linearly for NPNLMS. For IPNLMS, on the other hand, the ra-
Fig. 4  MSE learning curves for the high-sparse impulse response \( h^*_1 \). \( \eta = 0 \) corresponds to PNLMS, and \( \eta = 1 \) corresponds to NLMS.

Fig. 5  MSE learning curves for the semi-sparse impulse response \( h^*_2 \). \( \eta = 0 \) corresponds to PNLMS, and \( \eta = 1 \) corresponds to NLMS.

4. Application to Echo Cancellation
The PNLMS algorithm exploits the sparsity information for accelerating the convergence speed, while the TDAF algorithm exploits the second-order statistics of input signals to decorrelate the signals and to improve the convergence behavior. It is known that the coloredness of input signals deteriorates the performance of PNLMS. This motivates us to use the information about both of the sparsity and the input statistics simultaneously for the AEC problem, as speech input signals are known to be highly colored. What would be an effective approach to exploiting the sparsity and statistics simultaneously? A straightforward approach is to whiten the signal first and then apply the proportionate adaptation to the whitened input signal. This Whitening and Then Proportionate (WTP) approach is discussed in Sect. 4.1 and will turn out to be disadvantageous. We then show in Sect. 4.2 how to apply the MC-NLMS algorithm to the AEC problem. Finally, we present the simulation results in Sect. 4.3.

4.1 A Straightforward Approach
The WTP approach is illustrated in Fig. 8(a). It can be regarded as a serial approach composed of two stages: the colored signals are whitened in the first stage, and then the whitened signals are processed by the PNLMS algorithm in the second stage. The WTP approach is given by:

\[
\hat{w}_{k+1} = \hat{w}_k - \lambda_k \frac{\hat{e}_k(\hat{w}_k)}{\hat{u}^T_k A_k^{-1} \hat{u}_k} A_k^{-1} \hat{u}_k.
\]

Here, \( \hat{w}_k \) is the adaptive filter in the transform domain, \( \hat{e}_k(\hat{w}) := \hat{w}^T \hat{v}_k - \hat{d}_k, \hat{w} \in \mathbb{R}^N, \hat{v} \) is the whitened input vector defined as \( \hat{v}_k := \Delta_k^{-1/2} v_k := \Delta_k^{-1/2} Q u_k \in \mathbb{R}^N \), where \( Q \in \mathbb{R}^{N \times N} \) is an appropriate orthogonal matrix such as the discrete cosine transform (DCT) matrix, \( \Delta_k \in \mathbb{R}^{N \times N} \) is a diagonal matrix that normalizes the power of each element of the transform-domain input vector, and \( A_k \) is a proportionate-type metric. Unfortunately, this approach is ineffective for the following reason. To obtain the same response as the pair of unknown system and input \( (h^*, u_k) \), the companion of \( \hat{v}_k \) is \( \tilde{w} := \Delta^{1/2} Q h^* \).

The filter-input pairs \( (\tilde{w}, \tilde{u}_k) \), \( (w, v_k) \), and \( (h, u_k) \) should have the same response, namely \( \tilde{w}^T \tilde{u}_k = w^T v_k = h^T u_k, \forall u_k \Rightarrow \tilde{w}^T \Delta^{1/2} Q u_k = w^T Q u_k = h^T u_k, \forall u_k \Rightarrow Q \Delta^{1/2} \tilde{w} = Q^* w = h. \) Thus, the adaptive filter for the modified inputs \( \tilde{v}_k \) is \( \tilde{w}_k := \Delta_k^{1/2} Q h^* \).
that the sparseness of $\hat{h}$ is preserved under the transformation; i.e., $\hat{w}$ is no longer sparse in general as illustrated in Fig. 3(c).

4.2 The Proposed Approach Using MC-NLMS

To exploit the sparsity and statistical information simultaneously by MC-NLMS, we let

$$A_k := \eta_k A_{k,1} + (1 - \eta_k) A_{k,2},$$

$$A_{k,1} = \frac{1}{\|G_k^{MP}\|_2} G_k^{MP},$$

$$A_{k,2} = Q^T \left( \frac{1}{\|A_k\|_2} A_k \right) Q.$$  \hfill (21)

Here, $G_k^{MP}$ is the metric of MPNLMS [23] which improves the convergence speed of PNLMS by designing the metric optimally for white inputs, and $A_{k,2}$ is the metric of TDAF in the time domain since $\langle h, u_k \rangle_{A_{k,2}} = h^T A_{k,2}^{-1} u_k = h^T Q^T A_k^{-1} Q u_k = w^T A_k^{-1} v_k = \langle w, v \rangle_{A_k}$. The MC-NLMS algorithm can take the advantages of both algorithms by controlling $\eta_k \in [0, 1]$ dynamically. If $\eta_k = 1$ for all $k \in \mathbb{N}$, MC-NLMS is reduced to MPNLMS. On the other hand, if $\eta_k = 0$ for all $k \in \mathbb{N}$, it is reduced to the transform-domain NLMS (TD-NLMS) algorithm which is a normalized version of TDAF in the time-domain.

Our controlling strategy of $\eta_k$ is the following.

1) Initial phase: Assign a large weight to $A_{k,1}$-metric to attain fast initial convergence. (Take the advantage of MPNLMS.)

2) From middle phase to steady state: Assign a large weight to $A_{k,2}$-metric to gain the benefit of the constant rate of convergence. (Take the advantage of TD-NLMS.)

3) According to the analysis in [14], the fluctuations of the metric $A_k$ should be sufficiently small for the algorithm to converge, implying that the parameters $\eta_k$ should be changed gradually.

Under the above ideas, we recursively define $\eta_k$ for an initial value $\eta_0 := 1$ as follows:\footnote{We conducted experiments by letting $\eta_k = 1$ in the initial phase and $\eta_k = 0$ after that. The point here is when to switch the weight $\eta_k$ from 1 to 0. When we set the switching timing appropriately, the result was as good as the performance of the proposed algorithm presented in Sect. 4.3. However, when the timing was inappropriate, the performance was degraded seriously. In contrast to the sensitivity to the choice of the timing behind this switching strategy, our proposed technique in (22) and (23) shifts the weight $\eta_k$ gradually, and thus it is fairly insensitive to the choice of $\beta$.}

$$\eta_{k+1} = \min \{1, \alpha_k \},$$ \hfill (22)
\[ \alpha_{k+1} = (1 - \beta) \alpha_k + \beta \frac{||h_{k+1} - h_k||_2}{||h_k||_2}, \]  
(23)

where \( \beta \in (0, 1) \) and \( \alpha_k \) is initialized as \( \alpha_0 = 1 \). A typical choice is \( \beta = 10^{5-\log_2 N} \) for \( 128 \leq N \leq 512 \).

If the unknown system \( h^* \) changes into \( \tilde{h}^* \in \mathbb{R}^N \) at a time instant \( k \) abruptly, one may detect the change, reinitialize the \( \alpha_k \) parameter as \( \alpha_k = 1 \), and then track the new unknown system \( \tilde{h}^* \) by adapting \( \tilde{h}_k \) with a large weight reassigned to \( A_{k,1} \). Note here that the metric \( A_{k,1} \) in (20) should reflect the sparsity of the difference vector \( \tilde{h}^* - \tilde{h}_k \) rather than that of \( \tilde{h}^* \) [24]. The \( A_{k,1} \) is designed based on \( h_k - h_k \) in practice.

4.3 Numerical Examples

We show the efficacy of the proposed scheme for the AEC application. The proposed scheme is compared with NLMS, MPNLMS, TD-NLMS, and WTP. The simulation settings are given as follows: the sparse unknown system is defined as \( h^* := [h^T_2, 0, \cdots, 0]^T \in \mathbb{R}^N \) for \( N = 512 \). We use two types of input signals for \( u_k \in \mathbb{N} \): one is the zero-mean white Gaussian signal with the unit variance and the other is the USASI signal (which is speech-like wide sense stationary). Here, the USASI signal is modeled on the auto-regressive moving average (ARMA) process [25] and is characterized by the rational function:

\[ H(z) := \frac{1 - \zeta^2}{1 - 1.70223\zeta^{-1} + 0.71902\zeta^{-2}}, \quad \zeta \in \mathbb{C}, \]  
(24)

where \( \mathbb{C} \) denotes the set of all complex numbers. The noise \( n_k \) is additive white Gaussian with SNR = 30 dB. For all adaptive algorithms, we set \( \lambda_k = 0.4 \). For MPNLMS [23], \( \rho = 0.01, \sigma = 0.01, \) and \( \mu = 1000 \). For TD-NLMS, the forgetting factor is set to \( \nu = 10^{-3} \). For MC-NLMS, \( \beta = 5.0 \times 10^{-4} \) which gives \( \eta_k \) depicted in Fig. 9.

Figure 10 shows the NMSE learning curves. It can be seen that the proposed scheme outperforms the other algorithms for both white and highly-colored input signals.
In particular, the good overall performance in the highly-colored input case comes from taking the advantages of MPNLMS and TD-NLMS by controlling the combination coefficient $\eta$ adequately. The MPNLMS algorithm attains good performance in white input case, while its convergence slows down gradually since its metric is no longer optimal for colored inputs. The TD-NLMS algorithm keeps its convergence speed to be fairly constant. As explained in Sect. 4.1, the WTP serial approach loses the sparsity of $h^*$ by its whitening process, resulting in poor performance.

5. Conclusion

We presented an efficient adaptive filtering scheme, MC-NLMS, which incorporates multiple pieces of information into the metric and improves the convergence behavior. The proposed scheme is based on the variable metric projection concept. We compared the IPNLMS algorithm with the NPNLMS algorithm, which is a special case of MC-NLMS, and it was shown that the performance of NPNLMS was controllable with the combination coefficient as opposed to IPNLMS. We also presented an application to an echo cancellation problem and showed the efficacy of the proposed scheme. We applied the MC-NLMS algorithm to the acoustic echo cancellation (AEC) application. The proposed scheme outperformed the other algorithms for both white and highly-colored input signals. The present study provides an attractive approach to exploiting multiple pieces of information available in potential applications of adaptive filters.

Acknowledgment

A part of this work was conducted under a contract of research and development for radio resource enhancement, organized by the Ministry of Internal Affairs and Communications, Japan. This work was partially supported by JSPS Grants-in-Aid (24760292).

References

Osamu Toda received the B.E. and M.E. degrees from Niigata University, Niigata, Japan, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree in Department of Electronics and Electrical Engineering at Keio University, Kanagawa, Japan. His research interests are in mathematical adaptive signal processing, nonlinear adaptive filtering, and their applications to sensor network systems.

Masahiro Yukawa received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology in 2002, 2004, and 2006, respectively. After studying as a Visiting Researcher at the University of York, U.K., from October 2006 to March 2007, he worked as a Special Postdoctoral Researcher for RIKEN, Saitama, Japan, from April 2007 to March 2010. From August to November 2008, he was a Guest Researcher at the Associate Institute for Signal Processing, the Technical University of Munich, Germany. He was an Associate Professor at the Department of Electrical and Electronic Engineering, Niigata University, Japan, from April 2010 to March 2013. He is currently an Assistant Professor at the Department of Electronics and Electrical Engineering, Keio University, Japan. His current research interests are in mathematical signal processing, machine learning, and convex/sparse optimization. Masahiro Yukawa is a member of the Institute of Electrical and Electronics Engineers (IEEE). He has been an Associate Editor for several journals, including IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences from 2009 to 2013, and the Journal of Multidimensional Systems and Signal Processing, Springer since 2012. From April 2005 to March 2007, he was a recipient of the Research Fellowship of the Japan Society for the Promotion of Science (JSPS). He received the Excellent Paper Award and the Young Researcher Award from the IEICE in 2006 and in 2010, respectively, the Yasujiro Niwa Outstanding Paper Award from Tokyo Denki University in 2007, and the Ericsson Young Scientist Award from Nippon Ericsson in 2009.

Shigenobu Sasaki received B.E., M.E. and Ph.D. degrees from Nagaoka University of Technology, Nagaoka, Japan, in 1987, 1989 and 1993, respectively. Since 1992, he has been with Niigata University, where he is now a Professor at the Department of Electrical and Electronic Engineering. From 1999 to 2000, he was a visiting scholar at the Department of Electrical and Computer Engineering, University of California, San Diego. From 2003 to 2006, he was with the UWB technology institute, National Institute of Information and Communication Technology (NICT) as an Expert Researcher. Currently, he is serving the Chairman of the Technical Committee on Wideband Systems, IEICE. His research interests are in the area of digital communications with special emphasis on spread spectrum communication systems, ultra-wideband systems, and cognitive radio technology. He is a member of IEEE.

Hisakazu Kikuchi received B.E. and M.E. degrees from Niigata University, Niigata, in 1974 and 1976, respectively, and Dr. Eng. degree in electrical and electronic engineering from Tokyo Institute of Technology, Tokyo, in 1988. From 1976 to 1979 he worked at Information Processing Systems Laboratory, Fujitsu Ltd., Tokyo. Since 1979, he has been with Niigata University, where he is a Professor in the Department of Electrical and Electronic Engineering. During the 1992 academic year, he was a visiting scholar in the Electrical Engineering Department, University of California, Los Angeles, USA. He holds a visiting professorship at Chongqing University of Posts and Telecommunications and Nanjing University of Information Science and Technology, both in China, since 2002 and 2005, respectively. His research interests include digital signal processing, image processing, and video coding. He is a Fellow of IEICE and a Member of ITE, IIEE, and IEEE. He served the general co-chair of ITC-CSCC 2011 in Korea, the chair of Circuits and Systems Group, IEICE, in 2000 and the general chair of Digital Signal Processing Symposium, IEICE, in 1998 and Karuizawa Workshop on Circuits and Systems, IEICE, in 1996, respectively.