Proceedings of 2009 APSIPA Annual Summit and Conference, Sapporo, Japan, October 4-7, 2009

An Efficient Parallel Variable-metric Projection Algorithm Based on Set-theoretic and Krylov-proportionate Adaptive Filtering Techniques

Masahiro Yukawa* and Isao Yamada†
* RIKEN Brain Science Institute, Wako Saitama Japan E-mail: myukawa@riken.jp Tel: +81-48-462-1111
† Tokyo Institute of Technology, Ookayama Tokyo Japan E-mail: isao@comm.ss.titech.ac.jp Tel: +81-3-5734-2503

***Abstract*—We blend two adaptive filtering techniques for further efficiency: the set-theoretic adaptive filter (STAF, Yamada *et al.* 2002) and the Krylov-proportionate adaptive filter (KPAF, Yukawa 2009). Although the ideas behind these techniques are quite different from each other, there is a way to blend them together by noticing that KPAF can be seen as a sort of 'variable-metric' projection algorithm. We propose a blended algorithm named *set-theoretic Krylov-proportionate adaptive filter (SKAF)*, which features *iterative parallel variable-metric projection* onto well-designed closed convex sets. We present comparisons in complexity and mean square error (MSE) performance, showing significant advantages of the proposed algorithm over the existing algorithms.**

I. Introduction to Set-theoretic and Krylov-proportionate Adaptive Filtering

We consider the following linear model:

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

where $\boldsymbol{u}_k := [u_k, u_{k-1}, \cdots, u_{k-N+1}]^T \in \mathbb{R}^N$ (T: transposition) is the length-N input vector at time $k \in \mathbb{N}$, $\boldsymbol{h}^* \in \mathbb{R}^N$ the unknown system, and $(n_k)_{k\in\mathbb{N}}$ the noise process. An adaptive filter $(\boldsymbol{h}_k)_{k\in\mathbb{N}}$ is controlled by an iterative algorithm to estimate $\boldsymbol{h}^*$; note that the minimum mean square error (MMSE) filter coincides with $\boldsymbol{h}^*$ under the natural assumption $E\{\boldsymbol{u}_k n_k\} = \boldsymbol{0}$, $\forall k \in \mathbb{N}$ (see, e.g., [1]).

In this paper, we propose an efficient adaptive algorithm based on the two different concepts of adaptive filtering: (i) set-theoretic and (ii) Krylov-proportionate. We briefly summarize their general ideas before going into the detail.

A. General Idea of Set-theoretic Adaptive Filtering

A typical optimization task is to minimize, or maximize, a given cost function (under possible constraints). In real engineering problems, however, such a cost function depends usually on measurable data, which are stuck in uncertainty due to the presence of noise. In adaptive signal processing, moreover, the nature of data may change dynamically due to, e.g., time variation of $\boldsymbol{h}^*$, nonstationarity of u_k etc. Therefore, it makes little sense to define a 'fixed' cost function based on such unassured data and optimize it.

Set-theoretic adaptive filter (STAF) [2], [3] takes a different approach from the conventional optimization. It is derived from the adaptive projected subgradient method (APSM) [4]–[6], which minimizes 'time-varying' cost functions in an asymptotic sense. APSM has extensively been used to derive efficient algorithms for a wide range of engineering problems (see, e.g, [7]–[10]). At each time instant, STAF moves the filter closer to the set of filtering vectors that are consistent with recently measured data by means of *parallel subgradient projection*. Specifically, at time $k \in \mathbb{N}$, the set is characterized as the intersection of the following q closed convex sets: $C_\ell(\rho) := \left\{\boldsymbol{x} \in \mathbb{R}^N : \sum_{\iota=0}^{r-1} e_{\ell-\iota}^2(\boldsymbol{x}) \le \rho\right\}$, $\ell \in \mathcal{I}_k := \{k, k-1, \cdots, k-q+1\}$, where $e_k(\boldsymbol{h}) := \boldsymbol{u}_k^T\boldsymbol{h} - d_k$, $r \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$, and $\rho \ge 0$.[1]

B. General Idea of Krylov-proportionate Adaptive Filtering

Krylov-proportionate adaptive filter (KPAF) [1], [11] was inspired by the successful previous works of the proportionate adaptive filter (PAF) [12]–[15] and the Krylov-subspace-based reduced-rank adaptive filters [16]–[21]. While those reduced-rank filters are suboptimal due to the restriction of their search to the subspace, KPAF is free from such a restriction and thus provides optimal steady-state performance. The PAF paradigm has been shifted from "exploiting the sparsity of $\boldsymbol{h}^*$" to "sparsifying $\boldsymbol{h}^*$ as $\boldsymbol{Q}^T\boldsymbol{h}^*$"; i.e., most components of the vector $\boldsymbol{Q}^T\boldsymbol{h}^*$ are nearly zero. Here, $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is the orthonormalized Krylov-matrix associated with estimates of $\boldsymbol{R} := E\{\boldsymbol{u}_k\boldsymbol{u}_k^T\}$ and $\boldsymbol{p} := E\{\boldsymbol{u}_k d_k\}$. The sparsification allows us to extend the idea of PAF (which exploits the sparsity to improve the convergence rate) to dispersive systems.

In [22], it has been shown that PAF can be interpreted as an iterative orthogonal projection method onto the hyperplanes $H_k := \left\{\boldsymbol{x} \in \mathbb{R}^N : e_k(\boldsymbol{x}) = 0\right\}$, $k \in \mathbb{N}$, *with time-varying metrics*. Analogously, KPAF can be seen as a sort of *variable-metric projection method*; the metric is defined with the inverse of a certain positive definite matrix $\boldsymbol{\Omega}_k \in \mathbb{R}^{N\times N}$. For $k < K_0$ ($K_0 \in \mathbb{N}$: training period), $\boldsymbol{R}$ and $\boldsymbol{p}$ are estimated and $\boldsymbol{Q}$ is not yet available. Therefore, we let $\boldsymbol{\Omega}_k := \boldsymbol{I}$ for $k < K_0$. At $k = K_0$, the matrix $\boldsymbol{Q}$ is constructed; in fact, not the whole matrix $\boldsymbol{Q}$ but only its sub-matrix with the first D columns, say $\boldsymbol{Q}_1 \in \mathbb{R}^{N\times D}$, needs to be computed in practice [1] (see Table I; $\|\cdot\|_2$ denotes the Euclidean norm). For $k \ge K_0$, $\boldsymbol{\Omega}_k$ is constructed as $\boldsymbol{\Omega}_k := \boldsymbol{Q}\boldsymbol{\Theta}_k\boldsymbol{Q}^T$, where $\boldsymbol{\Theta}_k$ is a diagonal positive-definite matrix whose diagonal entries are determined based on the sparsity (see Table II). We emphasize that, although $\boldsymbol{\Omega}_k^{-1}$-metric is employed virtually, only the sub-matrix $\boldsymbol{Q}_1$ is used in actual computation.

[1] The index set $\mathcal{I}_k$ is more general in [2].

TABLE I
CONSTRUCTION OF $\boldsymbol{Q}_1$ AT $k = K_0$.

Requirements: $D_{\max} \in \mathbb{N},\ D := D_{\max},\ \delta \in (0,1)$
Training for $\widehat{\boldsymbol{R}}$ **and** $\widehat{\boldsymbol{p}}$ (for $k < K_0$):
Initialization: $\widehat{\boldsymbol{p}}_0 \in \mathbb{R}^N$, $\widehat{\boldsymbol{r}}_0 \in \mathbb{R}^N$
for $k = 0 : K_0 - 1$
  $\widehat{\boldsymbol{p}}_{k+1} := \widehat{\boldsymbol{p}}_k + d_k \boldsymbol{u}_k$
  $\widehat{\boldsymbol{r}}_{k+1} := \widehat{\boldsymbol{r}}_k + u_k \boldsymbol{u}_k$
end
Construction of $\boldsymbol{Q}_1$ (at $k = K_0$):
$\widehat{\boldsymbol{R}}$: use $\widehat{\boldsymbol{r}}_{K_0}$ with the symmetric and Toeplitz structure of $\boldsymbol{R}$
$\boldsymbol{t}_1 := \widehat{\boldsymbol{p}}_{K_0}$, $\boldsymbol{M}_1 := \boldsymbol{t}_1 / \|\boldsymbol{t}_1\|_2$
for $i = 2 : D_{\max}$
  if $D = D_{\max}$
    $\boldsymbol{t}_i := \widehat{\boldsymbol{R}} \boldsymbol{t}_{i-1}$
    $\boldsymbol{n}_i := \boldsymbol{t}_i / \|\boldsymbol{t}_i\|_2 - \boldsymbol{M}_{i-1} \boldsymbol{M}_{i-1}^T \boldsymbol{t}_i / \|\boldsymbol{t}_i\|_2$
    if $\|\boldsymbol{n}_i\|_2 < \delta$
      $D := i - 1$ (redefine)
      $\boldsymbol{Q}_1 := \boldsymbol{M}_{i-1}$
    else
      $\boldsymbol{m}_i := \boldsymbol{n}_i / \|\boldsymbol{n}_i\|_2$
      $\boldsymbol{M}_i := [\boldsymbol{M}_{i-1}\ \boldsymbol{m}_i]$
    end
  end
end

TABLE II
CONSTRUCTION OF $\boldsymbol{\Theta}_k$ [11] FOR $k \geq K_0$.

Requirements: $\xi > 0,\ \delta_{\mathrm{p}} > 0,\ \mu := \|\boldsymbol{h}^*\|_2^{-1} \sqrt{N/\epsilon}$
(ϵ: the target level of system mismatch)
Construction of $\boldsymbol{\Theta}_k$**:**
$\bar{\boldsymbol{h}}_{k,D} := [\bar{h}_k^1, \bar{h}_k^2, \cdots, \bar{h}_k^D]^T := \boldsymbol{Q}_1^T \boldsymbol{h}_k \in \mathbb{R}^D$
$\bar{h}_k^{(D+1)} := \sqrt{(\|\boldsymbol{h}_k\|_2^2 - \left\|\bar{\boldsymbol{h}}_{k,D}\right\|_2^2)/(N-D)}$
$F(\bar{h}_k^{(n)}) := \ln(1 + \mu |\bar{h}_k^{(n)}|),\ n \in \{1, 2, \cdots, D+1\}$
$\gamma_k^{\min} := \xi \max\{\delta_{\mathrm{p}}, F(\bar{h}_k^{(1)}), \cdots, F(\bar{h}_k^{(D)}), F(\bar{h}_k^{(D+1)})\}$
$\gamma_k^{(n)} := \max\{\gamma_k^{\min}, F(\bar{h}_k^{(n)})\},\ n \in \{1, \cdots, D+1\}$
$\eta_k := (N-D)\gamma_k^{(D+1)} + \sum_{n=1}^{D} \gamma_k^{(n)}$
$\theta_k^{(n)} := \gamma_k^{(n)} / \eta_k > 0,\ n \in \{1, \cdots, D\}$
$\delta_k := \gamma_k^{(D+1)} / \eta_k > 0$
$\boldsymbol{\Theta}_k := \mathrm{diag}(\theta_k^{(1)}, \theta_k^{(2)}, \cdots, \theta_k^{(D)}, \underbrace{\delta_k, \cdots \delta_k}_{N-D})$

II. PROPOSED SET-THEORETIC KRYLOV-PROPORTIONATE ADAPTIVE FILTERING

The projection-based interpretation of KPAF guides us to unifying STAF and KPAF in a natural way. In this section, we present the set-theoretic Krylov-proportionate adaptive filter (SKAF), which is a special example of *the adaptive parallel variable-metric projection algorithm* [22] with a specific design of closed convex sets and metric. With the $\boldsymbol{\Omega}_k^{-1}$-metric employed, the inner product is defined as $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_{\boldsymbol{\Omega}_k^{-1}} := \boldsymbol{a}^T \boldsymbol{\Omega}_k^{-1} \boldsymbol{b}$, $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, and its induced norm as $\|\boldsymbol{a}\|_{\boldsymbol{\Omega}_k^{-1}} := \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle_{\boldsymbol{\Omega}_k^{-1}}}$, $\forall \boldsymbol{a} \in \mathbb{R}^N$. Moreover, we denote by $P_C^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{x})$ the metric projection of an $\boldsymbol{x} \in \mathbb{R}^N$ onto a given closed convex set C w.r.t. the $\boldsymbol{\Omega}_k^{-1}$-metric; $P_C^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{x})$ is characterized as $\left\| P_C^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{x}) - \boldsymbol{x} \right\|_{\boldsymbol{\Omega}_k^{-1}} = \min_{\boldsymbol{y} \in C} \|\boldsymbol{y} - \boldsymbol{x}\|_{\boldsymbol{\Omega}_k^{-1}}$.

A. Set Design

Using the $\boldsymbol{\Omega}_k^{-1}$-metric, the stochastic property set $C_k(\rho)$ for $r = 1$ can be expressed as follows:

$$C_k^{(\iota)}(\rho) := \left\{ \boldsymbol{x} \in \mathbb{R}^N : (\langle \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}, \boldsymbol{x} \rangle_{\boldsymbol{\Omega}_k^{-1}} - d_{k-\iota})^2 \leq \rho \right\}, \quad \iota \in \mathcal{I} := \{0, 1, \cdots, q-1\},\ k \in \mathbb{N} \tag{2}$$

where $\rho \geq 0$ controls the set-membership probability that $\boldsymbol{h}^* \in C_k^{(\iota)}(\rho)$ [2]. The set $C_k^{(\iota)}(\rho)$ is sandwiched between two hyperplanes whose normal vectors are both $\boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}$; such a set (which is apparently closed and convex) is called *hyperslab*. The design of ρ is related to the probabilistic theory and here we simply give the resultant examples [2]: $\rho_1 := (1+\sqrt{2})\sigma_n^2$, $\rho_2 := \sigma_n^2$, or $\rho_3 := 0$, where $\sigma_n^2 := E\{n_k^2\}$. The projection onto $C_k^{(\iota)}(\rho)$ w.r.t. the $\boldsymbol{\Omega}_k^{-1}$-metric is given as follows:

$$P_{C_k^{(\iota)}(\rho)}^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{h}) = \boldsymbol{h} - \frac{\alpha_k^{(\iota)} \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}}{\boldsymbol{u}_{k-\iota}^T \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}} \tag{3}$$

where note that $\boldsymbol{u}_{k-\iota}^T \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota} = \|\boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}\|_{\boldsymbol{\Omega}_k^{-1}}^2$ and

$$\alpha_k^{(\iota)} := \begin{cases} e_{k-\iota}(\boldsymbol{h}) - \sqrt{\rho} & \text{if } e_{k-\iota}(\boldsymbol{h}) > \sqrt{\rho} \\ e_{k-\iota}(\boldsymbol{h}) + \sqrt{\rho} & \text{if } e_{k-\iota}(\boldsymbol{h}) < -\sqrt{\rho} \\ 0 & \text{if } -\sqrt{\rho} \leq e_{k-\iota}(\boldsymbol{h}) \leq \sqrt{\rho}. \end{cases} \tag{4}$$

B. Set-theoretic Krylov-proportionate Adaptive Filtering

Let $w_k^{(\iota)} > 0$, $\iota \in \mathcal{I}$, $k \in \mathbb{N}$, be the weight to $C_k^{(\iota)}(\rho)$, respectively, satisfying $\sum_{\iota \in \mathcal{I}} w_k^{(\iota)} = 1$, $\forall k \in \mathbb{N}$. Employing the sets $C_k^{(\iota)}(\rho)$ and the $\boldsymbol{\Omega}_k^{-1}$-metric in the adaptive parallel variable-metric projection algorithm [22], the set-theoretic Krylov-proportionate adaptive filter (SKAF) algorithm is given as follows ($\boldsymbol{h}_0 := \boldsymbol{0}$):[2]

$$\boldsymbol{h}_{k+1} := \boldsymbol{h}_k + \lambda_k \mathcal{M}_k \left(\sum_{\iota \in \mathcal{I}} w_k^{(\iota)} P_{C_k^{(\iota)}(\rho)}^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right) \tag{5}$$

$$= \boldsymbol{h}_k - \lambda_k \mathcal{M}_k \left(\sum_{\iota \in \mathcal{I}} w_k^{(\iota)} \frac{\alpha_k^{(\iota)} \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}}{\boldsymbol{u}_{k-\iota}^T \boldsymbol{\Omega}_k \boldsymbol{u}_{k-\iota}} \right),\ k \in \mathbb{N} \tag{6}$$

where $\lambda_k \in [0, 2]$ is the step size and $\mathcal{M}_k$ is the *extrapolation coefficient* defined as

$$\mathcal{M}_k := \begin{cases} \dfrac{\sum_{\iota \in \mathcal{I}} w_k^{(\iota)} \left\| P_{C_k^{(\iota)}(\rho)}^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|_{\boldsymbol{\Omega}_k^{-1}}^2}{\left\| \sum_{\iota \in \mathcal{I}} w_k^{(\iota)} P_{C_k^{(\iota)}(\rho)}^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{h}_k) - \boldsymbol{h}_k \right\|_{\boldsymbol{\Omega}_k^{-1}}^2} & \text{if } \boldsymbol{h}_k \notin \bigcap_{\iota \in \mathcal{I}} C_k^{(\iota)}(\rho) \\ 1 & \text{otherwise.} \end{cases}$$

Remark 1 (Inherent parallelism): As inherited from STAF, SKAF is suitable for parallel implementation due to its *inherently parallel structure* [2], [3], [23]–[26]. Namely, the

[2] If $\boldsymbol{h}_0 \neq \boldsymbol{0}$, we can construct $\boldsymbol{Q}_1$ by using the initial residual $\widehat{\boldsymbol{p}}_{K_0} - \widehat{\boldsymbol{R}} \boldsymbol{h}_0$ in place of $\widehat{\boldsymbol{p}}_{K_0}$. In this case, the initial error vector $\boldsymbol{h}^* - \boldsymbol{h}_0$ is sparsified by $\boldsymbol{Q}^T$-transformation.

projections in (5) are independent from each other, meaning that they can naturally be computed in parallel by means of q concurrent processors (In addition, it has a *fault tolerance* nature [3]).

C. Metric Design

We show that (i) the Ω_k^{-1}-metric is effective in SKAF and (ii) the low-complexity recursive formula for $\bar{\boldsymbol{h}}_{k,D}$ can be extended. Let us define $J_{\rm M}(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^T\boldsymbol{R}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{p}$, $\boldsymbol{x} \in \mathbb{R}^N$. Then, if we adopt the Euclidean metric (i.e., $\boldsymbol{\Omega}_k^{-1} = \boldsymbol{I}$) and neglect the effects of ρ (or simply let $\rho = 0$) in (6), $\sum_{\iota\in\mathcal{I}} w_k^{(\iota)} \dfrac{\alpha_k^{(\iota)}\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota}}{\boldsymbol{u}_{k-\iota}^T\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota}} = \sum_{\iota\in\mathcal{I}} w_k^{(\iota)} \dfrac{e_{k-\iota}(\boldsymbol{h}_k)\boldsymbol{u}_{k-\iota}}{\boldsymbol{u}_{k-\iota}^T\boldsymbol{u}_{k-\iota}}$ is an approximation of the normalized version of the gradient $J'_{\rm M}(\boldsymbol{h}_k) = \boldsymbol{R}\boldsymbol{h}_k - \boldsymbol{p} = E\{\boldsymbol{u}_k e_k(\boldsymbol{h}_k)\}$. The same applies to the domain transformed by $\boldsymbol{Q}^T$. To be precise, left-multiply both sides of (6) by $\boldsymbol{Q}^T$, define $\bar{\boldsymbol{h}}_k := [\bar{h}_k^{(1)}, \bar{h}_k^{(2)}, \cdots, \bar{h}_k^{(N)}]^T := \boldsymbol{Q}^T\boldsymbol{h}_k$ and $\bar{\boldsymbol{u}}_k := \boldsymbol{Q}^T\boldsymbol{u}_k$, and let $\boldsymbol{\Theta}_k := \boldsymbol{I}$. Then, the transform-domain algorithm approximates the normalized gradient of the function $\bar{J}_{\rm M}(\bar{\boldsymbol{x}}) := \frac{1}{2}\bar{\boldsymbol{x}}^T\bar{\boldsymbol{R}}\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}^T\bar{\boldsymbol{p}}$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, from q sample data, where $\bar{\boldsymbol{R}} := \boldsymbol{Q}^T\boldsymbol{R}\boldsymbol{Q}$ and $\bar{\boldsymbol{p}} := \boldsymbol{Q}^T\boldsymbol{p}$. Therefore, we can exploit the approach in [11] which is based on constrained optimization on $\boldsymbol{\Theta}_k$ regarding the (deterministic) gradient method for $\bar{J}_{\rm M}(\bar{\boldsymbol{x}})$, as it is simpler and more tractable than the adaptive algorithm. The objective of the optimization is to minimize —under several constraints related to computational requirements etc.— the number of iterations required to reduce the system mismatch $\|\boldsymbol{h}^* - \boldsymbol{h}_k\|_2^2 \|\boldsymbol{h}^*\|_2^{-2}$ to a target value (corresponding to ϵ in Table II). Referring to the obtained results in Table II, we can see that $\bar{\boldsymbol{h}}_{k,D} := \boldsymbol{Q}_1^T\boldsymbol{h}_k \in \mathbb{R}^D$, $k \in \mathbb{N}$, needs to be computed at each iteration. Left-multiplying both sides of (6) by $\boldsymbol{Q}_1^T$ yields the following recursion:

$$\bar{\boldsymbol{h}}_{k+1,D} = \bar{\boldsymbol{h}}_{k,D} - \lambda_k\mathcal{M}_k\left(\sum_{\iota\in\mathcal{I}} w_k^{(\iota)}\frac{\alpha_k^{(\iota)}\bar{\boldsymbol{\beta}}_{k,D}^{(\iota)}}{\boldsymbol{u}_{k-\iota}^T\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota}}\right),\ k \in \mathbb{N}. \tag{7}$$

Here, $\bar{\boldsymbol{\beta}}_{k,D}^{(\iota)} := \boldsymbol{\Theta}_{k,D}\bar{\boldsymbol{u}}_{k-\iota,D}$, where $\boldsymbol{\Theta}_{k,D} := \mathrm{diag}(\theta_k^{(1)}, \theta_k^{(2)}, \cdots, \theta_k^{(D)})$ and $\bar{\boldsymbol{u}}_{k-\iota,D} := \boldsymbol{Q}_1^T\boldsymbol{u}_{k-\iota}$. Fortunately, the recursion in (7) does *not* cause any severe computational burden. Indeed, in computing (6), $\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota}$ is efficiently computed as follows:

$$\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota} = \boldsymbol{Q}_1(\bar{\boldsymbol{\beta}}_{k,D}^{(\iota)} - \delta_k\bar{\boldsymbol{u}}_{k-\iota,D}) + \delta_k\boldsymbol{u}_{k-\iota}. \tag{8}$$

Therefore, after computing (6), all the quantities $\mathcal{M}_k \in \mathbb{R}$, $\alpha_k^{(\iota)} \in \mathbb{R}$, $\bar{\boldsymbol{\beta}}_{k,D}^{(\iota)} \in \mathbb{R}^D$, and $\boldsymbol{u}_{k-\iota}^T\boldsymbol{\Omega}_k\boldsymbol{u}_{k-\iota} \in \mathbb{R}$ are available and can be used to compute (7).

III. Comparisons to Conventional Algorithms — Complexity and MSE-Performance

We compare the proposed algorithm with NLMS, KPAF [11], STAF, and RLS algorithms in computational complexity and mean square error (MSE) performance in Sections III-A and III-B, respectively. All the results are discussed in Section III-C.

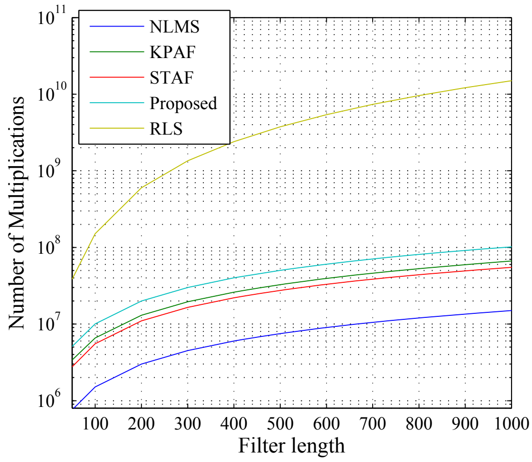

Fig. 1. Complexity of the proposed algorithm, NLMS, KPAF, STAF, and RLS for 5000 iterations.

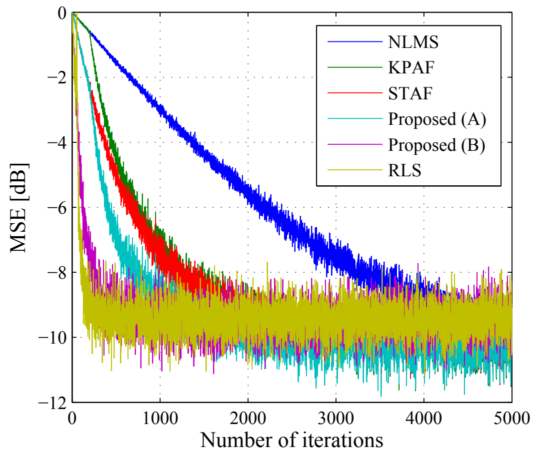

Fig. 2. Proposed algorithm versus NLMS, KPAF, STAF, and RLS under SNR 10 dB. For the proposed algorithm, (A) $\lambda_k = 0.02$ and (B) $\lambda_k = 1.0$.

A. Computational Complexity

Referring to (3) and (4), one can see that $P_{C_k^{(\iota)}(\rho)}^{(\boldsymbol{\Omega}_k^{-1})}(\boldsymbol{h}_k) = \boldsymbol{h}_k$ in the case of $|e_{k-\iota}(\boldsymbol{h}_k)| \le \sqrt{\rho}$, implying that in such a case no computation is required to obtain the projection. Therefore, the computational complexity of the proposed algorithm depends on the frequency at which each projection is computed. In fact, the frequency is a function of time. At the initial phase of adaptation, the error is large in general, indicating that the frequency tends to be high. In contrast, at the steady state, the error is small, thus the frequency tends to be low; the same applies to STAF.

We compute average frequencies for the proposed and STAF algorithms by simulations under several different conditions, and according to the frequencies we plot the complexity (the total number of multiplications required for 5000 iterations) in Fig. 1 as a function of the filter length N. For comparisons, we also plot the complexity of NLMS, KPAF, and RLS.

B. MSE Performance

We compare the MSE performance of the proposed, NLMS, KPAF, STAF, and RLS algorithms by simulations. We use the white input signals $(u_k)_{k\in\mathbb{N}}$ and randomly generated unknown systems $\boldsymbol{h}^* \in \mathbb{R}^N$ for $N = 50$ under the signal to noise ratio (SNR) 10 dB, where SNR $:= 10\log_{10}\left(E\left\{z_k^2\right\}/E\left\{n_k^2\right\}\right)$ with $z_k := \boldsymbol{u}_k^T\boldsymbol{h}^*$.

For the RLS algorithm, the initial estimate of the autocorrelation matrix is set to $\widehat{\boldsymbol{R}}_0^{\mathrm{RLS}} := 0.01\boldsymbol{I}$ and the forgetting factor is set to $\gamma^{\mathrm{RLS}} := 1 - 1/N$. For all the other algorithms, we set the step size $\lambda_k := 0.02$ to attain good steady-state performance. To compete with RLS, we also use $\lambda_k = 1.0$ for the proposed algorithm. For the metric design of KPAF and the proposed algorithm, we set $D_{\max} := 8$, $\delta := 0.01$, $K_0 := 200$, $\widehat{\boldsymbol{R}}_0 := \boldsymbol{O}$, $\widehat{\boldsymbol{p}}_0 := \boldsymbol{0}$, $\xi := \delta_{\mathrm{p}} := 0.01$, and $\epsilon := 10^{-3}$; the average value of D was 4.0. For STAF and the proposed algorithm, we set $q := 8$ and $\rho := \rho_1$.

C. Discussion

From Figs. 1 and 2, it is seen that "Proposed (A)" improves the MSE performance compared to STAF and KPAF with a fairly small increase of computational complexity; "Proposed (A)", STAF, KPAF, and NLMS attain approximately the same steady-state performance. Moreover, "Proposed (B)" achieves the MSE performance comparable to RLS with much lower complexity. When compared to "Proposed (A)", STAF, KPAF, and NLMS, "Proposed (B)" exhibits slightly inferior steady-state performance. The achieved fast convergence is due to the use of parallel projection and the effective (variable) metric design, and the low complexity is due mainly to a low rate of computing the projections. It should be mentioned that, if q parallel processors are engaged for the proposed and STAF algorithms, the computational complexity for each processor is even lower (see Remark 1). Finally, although we focus on the case of $K_0 := 200$ in the present numerical studies, the use of larger K_0 will make $\boldsymbol{Q}^T\boldsymbol{h}^*$ sparser, thus is expected to yield higher gain.

IV. Conclusion

This paper has presented an efficient algorithm named *set-theoretic Krylov-proportionate adaptive filter (SKAF)*. The proposed algorithm is based on (i) the use of variable metric reflecting the sparsity realized by the orthogonal transformation related to the Krylov subspace and (ii) the simultaneous use of multiple closed convex sets containing the optimal filter with high reliability. A low-complexity recursive formula for the metric design has been derived. Significant advantages of the proposed algorithm over the existing algorithms have been demonstrated through the comparisons in complexity and MSE-performance. A convergence analysis of the proposed algorithm will be presented elsewhere.

Acknowledgment: This work was partially supported by SCAT grant-in-aid.

References

[1] M. Yukawa, "Krylov-proportionate adaptive filtering techniques not limited to sparse systems," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 927–943, Mar. 2009.

[2] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1091–1101, May 2002.

[3] M. Yukawa and I. Yamada, "Pairwise optimal weight realization — Acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4557–4571, Dec. 2006.

[4] I. Yamada, "Adaptive projected subgradient method: A unified view for projection based adaptive algorithms," *The Journal of IEICE*, vol. 86, no. 8, pp. 654–658, Aug. 2003, in Japanese.

[5] 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&8, pp. 593–617, 2004.

[6] K. Slavakis, I. Yamada, and N. Ogura, "Adaptive projected subgradient method over the fixed point set of strongly attracting non-expansive mappings," *Numer. Funct. Anal. Optim.*, vol. 27, no. 7&8, pp. 905–930, 2006.

[7] M. Yukawa, "A study of efficient adaptive filtering algorithms and their applications to acoustic and communication systems," Ph.D. dissertation, Tokyo Institute of Technology, Tokyo: Japan, Sept. 2006.

[8] R. L. G. Cavalcante, "A study of adaptive projected subgradient method for multiple access interference and peak-to-average power reduction," Ph.D. dissertation, Tokyo Institute of Technology, Tokyo: Japan, Mar. 2008.

[9] K. Slavakis, S. Theodoridis, and I. Yamada, "Online kernel-based classification using adaptive projection algorithms," *IEEE Trans. Signal Processing*, vol. 56, no. 7, Part 1, pp. 2781–2796, Jul. 2008.

[10] R. L. G. Cavalcante, I. Yamada, and B. Mulgrew, "An adaptive projected subgradient approach to learning in diffusion networks," *IEEE Trans. Signal Processing*, 2009, in Press.

[11] M. Yukawa and W. Utschick, "Proportionate adaptive algorithm for non-sparse systems based on Krylov subspace and constrained optimization," in *Proc. IEEE ICASSP*, 2009, pp. 3121–3124.

[12] 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.

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

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

[15] 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.

[16] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for DS-CDMA systems," *International Journal of Wireless Information Networks*, vol. 3, no. 1, pp. 1–17, 1996.

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

[18] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on multistage Wiener filter," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.

[19] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Selected Areas in Commun.*, vol. 16, no. 9, pp. 1765–1773, Dec. 1998.

[20] S. Chowdhury and M. D. Zoltowski, "Application of conjugate gradient methods in MMSE equalization for the forward link of DS-CDMA," in *Proc. IEEE VTC 2001-Fall*, Oct. 2001, pp. 2434–2438.

[21] S. Burykh and K. Abed-Meraim, "Reduced-rank adaptive filtering using Krylov subspace," *EURASIP J. Appl. Signal Processing*, no. 12, pp. 1387–1400, Dec. 2002.

[22] 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.

[23] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. of IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[24] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, vol. 38, no. 3, pp. 367–426, 1996.

[25] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithm, and Optimization*. Oxford University Press, 1997.

[26] D. Butnariu, Y. Censor, and S. Reich, Eds., *Inherently parallel algorithms in feasibility and optimization and their applications*. New York: Elsevier, 2001.