# A DETERMINISTIC ANALYSIS OF LINEARLY CONSTRAINED ADAPTIVE FILTERING ALGORITHMS

*Masahiro Yukawa*† *and Isao Yamada*‡

† Department of Electrical and Electronic Engineering, Niigata University
2-8050 Ikarashi, Nishi-ku, Niigata, 950-2181 Japan
phone/fax: +81 (25) 262 7549, email: yukawa@eng.niigata-u.ac.jp
‡ Department of Communications and Integrated Systems, Tokyo Institute of Technology
S3-60 Tokyo, 152-8550 Japan
phone: +81 (3) 5734 2503, fax: +81 (3) 5734 2905, email: isao@sp.ss.titech.ac.jp

## ABSTRACT

This paper presents a mathematically rigorous analysis of linearly constrained adaptive filtering algorithms based on the adaptive projected subgradient method. We provide the novel concept of *constraint-embedding functions* that enables to analyze certain classes of linearly constrained adaptive algorithms in a unified manner. Trajectories of the linearly constrained adaptive filters always lie in the affine constraint set, a translation of a closed subspace. Based on this fact, we translate all the points on the constraint set to its underlying subspace — *which we regard as a Hilbert space* — thereby making the analysis feasible. Derivations of the linearly constrained adaptive filtering algorithms are finally presented in connection with the analysis.

## 1. INTRODUCTION

In signal processing applications, we often encounter situations in which an estimate is required to satisfy a system of linear equations. In the context of adaptive filtering, such linear constraints appear in adaptive beamforming, blind multiple access interference suppression in wireless communication systems, etc [1, 2]. A significant amount of effort has been devoted to developing efficient algorithms for each application [3–7].

*Metric projection* has been proven a powerful tool in a bunch of signal processing applications including adaptive filtering/learning [8, 9]. *The adaptive projected subgradient method (APSM)* [8] serves as a unified guiding principle for various projection-based adaptive filtering algorithms. It encompasses the normalized least mean square (NLMS) algorithm, the affine projection algorithm (APA), the adaptive parallel subgradient projection algorithm, and their *convexly-constrained* versions. Also it offers a *deterministic* analysis (which is much more challenging than a stochastic analysis) for those algorithms, proving the strong convergence of a vector sequence generated by APSM in a Hilbert space. The analysis was however built under the (implicit) assumption that the constraint set has an interior point, which does not hold true in the case of linear constraints. Therefore, the analysis cannot directly be applied to this important case. It would thus be of theoretical interest to develop a mathematically rigorous analysis for linearly-constrained adaptive filtering algorithms.

In this paper, we present a deterministic analysis of linearly-constrained adaptive algorithms. To resolve the conflict regarding the assumption of the existence of an interior point, we shed light on the fact that the trajectory of the vector sequence lies in the affine constraint set. We translate the affine set to its underlying subspace — which we regard as a Hilbert space — thereby making the analysis feasible. We introduce the new concept of *constraint-embedding functions*. This leads to a unified analysis for three classes of linearly constrained adaptive algorithm, each of which includes the projected APA, the embedded-constraint APA, and the constrained APA, respectively.

## 2. PRELIMINARIES

Following mathematical background, some examples of affine projection type algorithms for linearly constrained adaptive filtering are presented to facilitate the access to the main body of the present work.

### 2.1 Mathematical Tools

Throughout, $\mathbb{R}$, $\mathbb{N}$, and $\mathbb{N}^*$ denote the sets of all real numbers, nonnegative integers, and positive integers, respectively. Let $\mathcal{H}$ be a real Hilbert space[1] equipped with the inner product $\langle \cdot,\cdot \rangle_{\mathcal{H}} : \mathcal{H}^2 \to \mathbb{R}$ and its induced norm $\|\cdot\|_{\mathcal{H}} : \mathcal{H} \to [0,\infty)$. A set $C \subset \mathcal{H}$ is said to be *convex* if $\alpha x + (1-\alpha)y \in C$, $\forall x,y \in C$, $\forall \alpha \in (0,1)$. A function $f : \mathcal{H} \to \mathbb{R}$ is said to be *convex* if $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$, $\forall x,y \in \mathcal{H}$, $\forall \alpha \in (0,1)$. Given a nonempty closed convex set $C \subset \mathcal{H}$ and an arbitrary point $x \in \mathcal{H}$, there exists a unique point $y^* \in C$ closest to $x$. In this case, $d(x,C) := \min_{y \in C} \|x-y\|_{\mathcal{H}} = \|x-y^*\|_{\mathcal{H}}$ is called a metric distance function, and $P_C(x) := y^*$ is called the metric projection of $x$ onto C; i.e., $P_C : \mathcal{H} \to C$ is an operator that maps $x \in \mathcal{H}$ to the unique vector $P_C(x) \in C$ satisfying $\|x - P_C(x)\|_{\mathcal{H}} = d(x,C)$.

For any continuous (possibly nondifferentiable) convex function $f : \mathcal{H} \to \mathbb{R}$ and an arbitrary point $x \in \mathcal{H}$, there always exists $\hat{x} \in \mathcal{H}$ satisfying $\langle z-x, \hat{x} \rangle_{\mathcal{H}} + f(x) \leq f(z), \forall z \in \mathcal{H}$. Such $\hat{x}$ is called a subgradient of $f$ at $x$. If in particular $f$ is differentiable, there exists a unique subgradient that coincides with the gradient. The set of all subgradients of $f$ at $x \in \mathcal{H}$ is called the subdifferential of $f$ at $x$. In other words, the subdifferential is a set-valued function defined as $\partial f : \mathcal{H} \to 2^{\mathcal{H}}, x \mapsto \{\hat{x} \in \mathcal{H} : \langle z-x, \hat{x} \rangle_{\mathcal{H}} + f(x) \leq f(z),\ \forall z \in \mathcal{H}\}$. Assume $\mathrm{lev}_{\leq 0} f := \{x \in \mathcal{H} : f(x) \leq 0\} \neq \emptyset$. Then, the mapping $T_{\mathrm{sp}(f)} : \mathcal{H} \to \mathcal{H}$ defined as

$$T_{\mathrm{sp}(f)} : x \mapsto \begin{cases} x - \dfrac{f(x)}{\|f'(x)\|_{\mathcal{H}}^2} f'(x) & \text{if } f(x) > 0 \\ x & \text{otherwise} \end{cases}$$

is called a subgradient projection relative to $f$, where $f'(x) \in \partial f(x)$, $\forall x \in \mathcal{H}$.

### 2.2 Three Examples of Affine Projection Type Algorithms for Linearly Constrained Adaptive Filtering

Let $u \in \mathcal{H}$ be the estimator, or the adaptive filter, on which a linear constraint is imposed as follows: $u \in V$, where $V$ is a linear variety in $\mathcal{H}$ with its underlying closed subspace $M \subset \mathcal{H}$. Define $v := P_V(0)$, where $0 \in \mathcal{H}$ denotes the null vector. It holds then that $V \cap M^{\perp} = \{v\}$ [11], where $M^{\perp}$ denotes the orthogonal complement

[1] We present our analysis in a real Hilbert space to cover general cases where the space is non-Euclidean; e.g., in the case of kernel adaptive filters the space consists of functions and has a possibly infinite dimension [10]. The reader who is interested solely in linear adaptive filters may regard $\mathcal{H}$ simply as a Euclidean space equipped with the standard inner product.

 

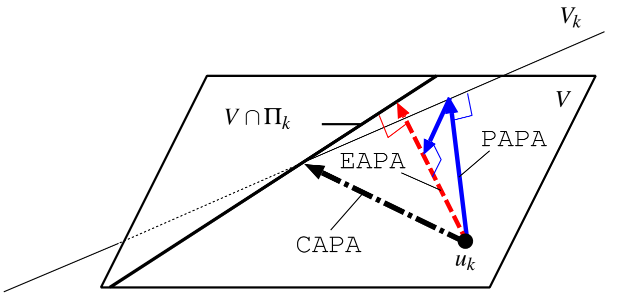


Figure 1: A geometric interpretation of PAPA, EAPA, and CAPA for $\lambda_k = 1$. We define $\Pi_k := \{x \in \mathscr{H} : \langle P_{V_k}(u_k) - u_k, P_{V_k}(u_k) - x \rangle_{\mathscr{H}} = 0\}$.

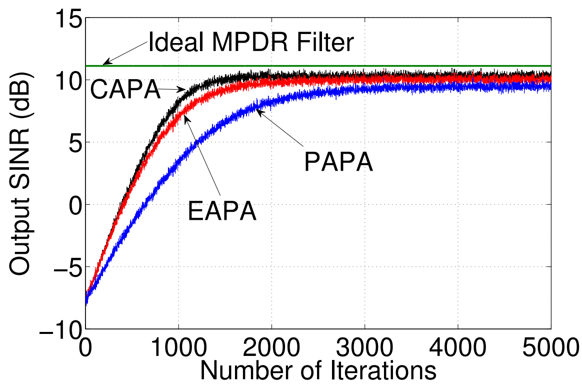


Figure 2: A numerical example in adaptive beamforming.

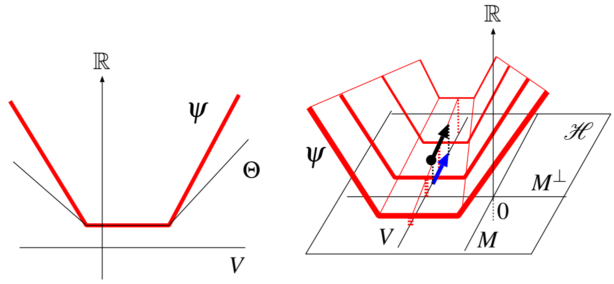


Figure 3: A constraint-embedding function $\psi$ for $(\Theta, V)$.

of $M$ ($\mathscr{H} = M \oplus M^{\perp}$). The linear variety can be expressed[2] as $V = M + v := \{x + v : x \in M\}$.

Let $V_k$ be a data-dependent linear variety in $\mathscr{H}$ at time instant $k \in \mathbb{N}$. We present a specific example in the case of linear adaptive filters. Given an input process $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}$, suppose that the output process $(y_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ is generated as $y_k := \mathbf{x}_k^{\mathsf{T}} \mathbf{u}^* + n_k$, where $\mathbf{x}_k := [x_k, x_{k-1}, \cdots, x_{k-N+1}]^{\mathsf{T}}$ for some $N \in \mathbb{N}^*$, $\mathbf{u}^* \in \mathbb{R}^N$ is an unknown linear system, and $(n_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ is the noise process. Here $(\cdot)^{\mathsf{T}}$ denotes *transpose*. Define $\mathbf{X}_k := [\mathbf{x}_k \ \mathbf{x}_{k-1} \ \cdots \ \mathbf{x}_{k-r+1}]$ and $\mathbf{y}_k := [y_k, y_{k-1}, \cdots, y_{k-r+1}]^{\mathsf{T}}$ for some $r \in \mathbb{N}^*$. Then $V_k$ can be defined as $V_k := \operatorname{argmin}_{\mathbf{u}\in\mathbb{R}^N} \left\| \mathbf{X}_k^{\mathsf{T}} \mathbf{u} - \mathbf{y}_k \right\|_{\mathbb{R}^r}$, where $\|\cdot\|_{\mathbb{R}^r}$ denotes the Euclidean norm in $\mathbb{R}^r$.

In a general case, affine projection type algorithms for linearly constrained adaptive filtering are given as follows:

$$u_{k+1} := P_V\left(u_k + \lambda_k (P_{V_k}(u_k) - u_k)\right), \tag{1}$$

$$u_{k+1} := \begin{cases} u_k + \eta_k P_M\left(P_{V_k}(u_k) - u_k\right) & \text{if } P_{V_k}(u_k) - u_k \notin M^{\perp} \\ u_k & \text{otherwise,} \end{cases}$$

$$\eta_k := \lambda_k \frac{\|P_{V_k}(u_k) - u_k\|_{\mathscr{H}}^2}{\|P_M\left(P_{V_k}(u_k) - u_k\right)\|_{\mathscr{H}}^2} \geq \lambda_k, \tag{2}$$

$$u_{k+1} := \begin{cases} u_k + \lambda_k (P_{V_k \cap V}(u_k) - u_k) & \text{if } V_k \cap V \neq \emptyset \\ u_k & \text{otherwise,} \end{cases} \tag{3}$$

where $\lambda_k \in [0,2]$ is the step size parameter of each algorithm. We refer to the algorithms in (1), (2), and (3) respectively as the projected APA (PAPA), the embedded-constraint APA (EAPA) [8], and the constrained APA (CAPA) [6]. A geometric interpretation of each algorithm is presented in Fig. 1. If in particular $V_k$ is a hyperplane (i.e., if $r = 1$ in the previous paragraph), then (1) is reduced to the projected NLMS algorithm [3], and both (2) and (3) are reduced to the constrained NLMS algorithm [4] (see also [5, 7]). For reference, the performance of the algorithms in an adaptive beamforming application [12] is depicted in Fig. 2, although we omit to show the details about the simulation conditions (The conditions are exactly the same as in [13] except the algorithms tested; MPDR stands for minimum power distortionless response). It should be mentioned that the algorithms in (1) and (2) do *not* have their generalized sidelobe canceller (GSC) counterparts for $r \geq 2$ and hence one cannot follow the typical way of analyzing a linearly constrained adaptive algorithm through the analysis of its GSC counterpart.

## 3. A DETERMINISTIC ANALYSIS

Let $(\Theta_k)_{k\in\mathbb{N}}$ be a sequence of cost functions, where $\Theta_k : \mathscr{H} \to [0,\infty)$, $k \in \mathbb{N}$, is an instantaneous cost function that reflects the data observed at each time instant $k$. We assume $\Theta_k$ to be continuous and convex. A simple example is $\Theta_k(x) := d(x, V_k)$, $x \in \mathscr{H}$ (see Section 2). Many of the problems encountered in adaptive signal processing fall into the linearly constrained asymptotic minimization problem shown below.

**Problem 1** *Minimize the sequence of cost functions $(\Theta_k)_{k\in\mathbb{N}}$ over $V$ in an asymptotic sense.*

Problem 1 is a subclass of the problem addressed in [8] where any closed convex set can be used in place of $V$. Applied to Problem 1, APSM iteratively generates the vector sequence $(u_k)_{k\in\mathbb{N}}$, for any initial point $u_0 \in \mathscr{H}$, by

$$u_{k+1} := \begin{cases} P_V\left(u_k - \lambda_k \dfrac{\Theta_k(u_k)}{\|\Theta_k'(u_k)\|_{\mathscr{H}}^2} \Theta_k'(u_k)\right) & \text{if } \Theta_k'(u_k) \neq 0 \\ u_k & \text{otherwise,} \end{cases} \quad k \in \mathbb{N}, \tag{4}$$

where $\lambda_k \in [0,2]$ is the step size. Unfortunately, a large part of the analysis presented in [8] cannot directly be applied to this case, because $V$ has no interior point in $\mathscr{H}$.

We present a deterministic analysis of a family of linearly constrained adaptive filtering algorithms (including PAPA, EAPA, and CAPA) in a unified way. To this end, we address Problem 1 via considering an alternative problem of minimizing a sequence $(\psi_k)_{k\in\mathbb{N}}$ of *constraint-embedding functions* over $V$ asymptotically. The key is the following: the function $\psi_k$ (i) inherits the information about the solution (i.e., $\psi_k$ and $\Theta_k$ share the same minimizers and the same minimum value over $V$), and (ii) enjoys a certain favorable structure which we call *constraint embedding*.

### 3.1 Constraint-Embedding Functions — Definition and Idea

We define the constraint-embedding functions as follows.

**Definition 1 (Constraint-embedding functions)** *Let $\Theta : \mathscr{H} \to [0,\infty)$ be a continuous convex function having a minimum over a linear variety $V$. Then, another continuous convex function $\psi : \mathscr{H} \to [0,\infty)$ is said to be constraint-embedding for $(\Theta, V)$ if the following conditions are satisfied.*

[2] $V$ is a translation of $M$ by $v$. Although not necessarily required, we let $v \in M^{\perp}$ to simplify the discussion.

(C1) *Preservation of the minimizers:*
$\operatorname{argmin}_{x\in V}\psi(x)=\operatorname{argmin}_{x\in V}\Theta(x)$.
(C2) *Preservation of the minimum:*
$\psi(x)=\Theta(x),\ \forall x\in\operatorname{argmin}_{x\in V}\psi(x)\subset V$.
(C3) *Lower boundedness by* $\Theta$*:*
$\psi(x)\geq\Theta(x),\ \forall x\in V$.
(C4) *Constraint-embedding structure:*
$\psi(u)\leq\psi(u+y),\ \forall(u,y)\in V\times M^{\perp}$.

Figure 3 illustrates a constraint-embedding function; the left panel describes (C1)–(C3) and the right one describes (C4). To make the idea of constraint-embedding function clear, let us consider the problem of minimizing $\Theta$ over $V$, which is a static counterpart of Problem 1. Due to (C1) and (C2) in Definition 1, we may obtain a solution to this problem indirectly by solving the alternative problem of minimizing a constraint-embedding function $\psi$ for $(\Theta,V)$ over $V$. *With an initial point on V, we can apply the subgradient method to the alternative problem without performing the metric projection onto V to enforce the constraint.* This is the key for the unified analysis to be presented in Section 3.2, and it is supported by the following lemma.

**Lemma 1** *Let $\psi:\mathscr{H}\to[0,\infty)$ be a continuous convex function satisfying* (C4) *in Definition 1. Then, $P_M(\partial\psi(u))\subset\partial\psi(u),\ \forall u\in V$.*

Proof: Fix $u\in V$. Any subgradient $\psi'(u)\in\partial\psi(u)$ is uniquely decomposed as $\psi'(u)=\psi'_M(u)+\psi'_{M^\perp}(u)$ with $(\psi'_M(u),\psi'_{M^\perp}(u))\in M\times M^\perp$. Because $\psi'_M(u)=P_M(\psi'(u))$, it is sufficient to show that $\psi'_M(u)\in\partial\psi(u)$. Any $x\in\mathscr{H}$ has a unique decomposition as follows: $x=x_M+x_{M^\perp}=(x_M+v)+(x_{M^\perp}-v)$ with $(x_M,x_{M^\perp})\in M\times M^\perp$. Let $x_V:=x_M+v\in V$. We can then verify

$$\begin{aligned}
&\psi(u)+\left\langle\psi'_M(u),x-u\right\rangle_{\mathscr{H}}\\
=&\psi(u)+\left\langle\psi'_M(u),x_V+(x_{M^\perp}-v)-u\right\rangle_{\mathscr{H}}\\
=&\psi(u)+\left\langle\psi'_M(u),x_V-u\right\rangle_{\mathscr{H}}+\left\langle\psi'_M(u),x_{M^\perp}-v\right\rangle_{\mathscr{H}}\\
=&\psi(u)+\left\langle\psi'_M(u)+\psi'_{M^\perp}(u),x_V-u\right\rangle_{\mathscr{H}}\\
\leq&\psi(x_V)\leq\psi(x_V+(x_{M^\perp}-v))=\psi(x).
\end{aligned}\tag{5}$$

Here, the third equality holds because $x_V-u\in M$ and $x_{M^\perp}-v\in M^\perp$, and the second inequality comes from (C4). The equation (5) implies $\psi'_M(u)\in\partial\psi(u)$, which verifies the claim. □

Lemma 1 guarantees the existence of a subgradient of $\psi$ in $M$; i.e., $\partial\psi(u)\cap M\neq\emptyset,\ \forall u\in V$. Such a subgradient is indicated by the blue arrow in Fig. 3. *As M is parallel to V, one can search for a solution without stepping away from V by using a subgradient $\psi'(u)\in\partial\psi(u)\cap M$, $u\in\mathscr{H}$, hence there is no need to perform the metric projection onto V to enforce the constraint.* This implies that the information regarding the linear constraint $V$ is embedded into $\psi$ due to (C4). The condition (C3) is a technical one required for the discussion in Section 3.2. Two examples of constraint-embedding function are given below.

**Example 1 (Constraint-embedding functions)**

1. *Given any continuous convex function $\Theta:\mathscr{H}\to[0,\infty)$, the function $\psi:\mathscr{H}\to[0,\infty),\ x\mapsto\Theta(P_V(x))$, is constraint-embedding for $(\Theta,V)$. In this case, the equality holds for* (C3) *and* (C4) *in Definition 1, and $P_M(\Theta'(P_V(x)))\in\partial\psi(x),\ \forall x\in\mathscr{H}$, where $\Theta'(x)\in\partial\Theta(x)$* [8, Example 5] *(cf. Lemma 1 below). The function $\psi$ is convex because the convexity of $\Theta$ is preserved under the composition with the affine mapping $P_V$* [14]. *The continuity of $\psi$ is readily verified by that of $P_V$.*
2. *Let $C_i\subset\mathscr{H}$, $i=1,2,\cdots,q$, be closed convex sets satisfying $C:=V\cap\left(\bigcap_{i=1}^q C_i\right)\neq\emptyset$. Define an average distance function $\Theta:\mathscr{H}\to[0,\infty),x\mapsto\sum_{i=1}^q w_i d(x,C_i)$, where $w_i>0$ satisfies $\sum_{i=1}^q w_i=1$. Then the function $\psi:\mathscr{H}\to[0,\infty),x\mapsto\sum_{i=1}^q w_i d(x,C_i\cap V)$ is constraint-embedding for $(\Theta,V)$; this type of function is used in* [7]. *For* (C1) *in Definition 1, $\operatorname{argmin}_{x\in V}\psi(x)=\operatorname{argmin}_{x\in V}\Theta(x)=C$. The fact $d(x,C_i\cap V)\geq d(x,C_i)(\Leftarrow C_i\cap V\subset C_i)$ implies* (C3). *For* (C4)*, note that $d(u,C_i\cap V)\leq d(u+y,C_i\cap V),\ \forall(u,y)\in V\times M^\perp$, for each $i=1,2,\cdots,q$. It is clear that $\psi$ is continuous and convex.*

### 3.2 Convergence Analysis of APSM under Linear Constraints via Constraint-Embedding Functions

For each $k\in\mathbb{N}$, let $\psi_k:\mathscr{H}\to[0,\infty)$ be a continuous convex function that is constraint-embedding for $(\Theta_k,V)$, where $\Theta_k:\mathscr{H}\to[0,\infty)$ is a continuous convex function having a minimum over $V$. Consider the following problem.

**Problem 2** *Minimize $(\psi_k)_{k\in\mathbb{N}}$ over V in an asymptotic sense.*

Due to (C1) and (C2), Problems 1 and 2 share the same minimum and minimizers. For Problem 2, we have an observation similar to the one presented in Section 3.1 for the static cost function $\psi$. Starting at some initial point $u_0\in V$ and adopting a subgradient $\psi'_k(u_k)\in\partial\psi_k(u_k)\cap M,\ \forall k\in\mathbb{N}$, the following recursion generates a vector sequence $(u_k)_{k\in\mathbb{N}}\subset V$:

$$u_{k+1}:=\begin{cases}u_k-\lambda_k\dfrac{\psi_k(u_k)}{\left\|\psi'_k(u_k)\right\|^2_{\mathscr{H}}}\psi'_k(u_k) & \text{if } \psi'_k(u_k)\neq 0\\ u_k & \text{otherwise,}\end{cases}\tag{6}$$

where $\lambda_k\in[0,2]$. Note here that $\psi_k(u_k)>\inf_{x\in\mathscr{H}}\psi_k(x)\Leftrightarrow 0\notin\partial\psi_k(u_k)$. Lemma 1 with the assumption $u_k\in V$, $k\in\mathbb{N}$, guarantees the existence of $\psi'_k(u_k)\in\partial\psi_k(u_k)\cap M$, and the recursion (6) moves (if it moves) $u_k$ in the direction of $\psi'_k(u_k)\in M$ along with $V$, which ensures $u_{k+1}\in V$.

For the same reason as stated under (4), there is a difficulty in the convergence analysis of (6) in $\mathscr{H}$. To eliminate the difficulty, we focus on the fact that the trajectory of $u_k$ lies in $V$, and shift the stage of analysis from $\mathscr{H}$ to $M$. To be precise, we turn our attention to the real Hilbert space $(M,\langle\cdot,\cdot\rangle_M)$, where $\langle x,y\rangle_M:=\langle x,y\rangle_{\mathscr{H}},\ \forall(x,y)\in M^2$. Accordingly the induced norm is defined by $\|x\|_M:=\sqrt{\langle x,x\rangle_M}=\|x\|_{\mathscr{H}},\ x\in M$.

Define a function $\varphi_k:M\to[0,\infty),x\mapsto\psi_k(x+v)$, $k\in\mathbb{N}$; $\varphi_k$ is a translation of $\psi_k$. We then obtain $\psi_k(u_k)=\psi_k(w_k+v)=\varphi_k(w_k)$, where $w_k:=u_k-v\in M$, $k\in\mathbb{N}$. Moreover, $M\supset\partial\varphi_k(w_k)=\partial\psi_k(w_k+v)\cap M=\partial\psi_k(u_k)\cap M$ [15, Theorem 4.2.1], hence we let $\varphi'_k(w_k):=\psi'_k(u_k)\in\partial\varphi_k(w_k)\subset M$. Subtracting $v\ (=P_V(0))$ from both sides of (6) and letting $w_k=u_k-v$, (6) can thus be rewritten as

$$w_{k+1}:=\begin{cases}w_k-\lambda_k\dfrac{\varphi_k(w_k)}{\left\|\varphi'_k(w_k)\right\|^2_{\mathscr{H}}}\varphi'_k(w_k) & \text{if } \varphi'_k(w_k)\neq 0\\ w_k & \text{otherwise.}\end{cases}\tag{7}$$

The sequence $(u_k)_{k\in\mathbb{N}}$ on $V$ is translated to $(w_k)_{k\in\mathbb{N}}$ on $M$. We emphasize that, without using the constraint embedding functions, the existence of a common subgradient is not guaranteed, and thus we cannot analyze properties of $(u_k)_{k\in\mathbb{N}}$ through an analysis of $(w_k)_{k\in\mathbb{N}}$. The analysis in [8] can directly be applied to $(w_k)_{k\in\mathbb{N}}$ in $(M,\langle\cdot,\cdot\rangle_M)$, yielding the following theorem on the analysis of $(u_k)_{k\in\mathbb{N}}$ in $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathscr{H}})$.

**Theorem 1** *Let $M\subset\mathscr{H}$ be a closed subspace and $V:=M+v$ for $v\in M^\perp$. Also let $\psi_k$, $k\in\mathbb{N}$, be a continuous convex function that is constraint-embedding for $(\Theta_k,V)$. Then, the sequence $(u_k)_{k\in\mathbb{N}}\subset V$ generated by* (6) *for an arbitrary $u_0\in V$ and $\psi'_k(u_k)\in\partial\psi_k(u_k)\cap M$, $\forall k\in\mathbb{N}$, satisfies the following.*

*(a)* **Monotone approximation :** *Assume that $u_k\notin\Omega_k:=\operatorname{argmin}_{x\in V}\psi_k(x)\neq\emptyset$; note that $\Omega_k=V\cap\operatorname{argmin}_{x\in\mathscr{H}}\psi_k(x)$ and $\Omega_k=\emptyset\Leftrightarrow\operatorname{argmin}_{x\in\mathscr{H}}\psi_k(x)=\emptyset$. Then, $\forall\lambda_k\in\left(0,2\left(1-\frac{\psi_k^*}{\psi_k(u_k)}\right)\right)$, where $\psi_k^*:=\min_{x\in\mathscr{H}}\psi_k(x)=\min_{x\in V}\Theta_k(x)$, $k\in\mathbb{N}$,*

$$\left\|u_{k+1}-u^*_{(k)}\right\|_{\mathscr{H}}<\left\|u_k-u^*_{(k)}\right\|_{\mathscr{H}},\ \forall u^*_{(k)}\in\Omega_k.\tag{8}$$

(b) **Asymptotic optimality of the sequence** $(u_k)_{k\in\mathbb{N}}$ **:** *Assume (i)* $\exists K_0 \in \mathbb{N}$ *s.t.* $\begin{cases} \psi_k^* = 0, \ \forall k \geq K_0, \text{ and} \\ \Omega := \bigcap_{k\geq K_0} \Omega_k \neq \emptyset. \end{cases}$ *Then,* $(u_k)_{k\in\mathbb{N}}$ *is bounded. Moreover, if we specially use* $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$ *and* $(\psi_k'(u_k))_{k\in\mathbb{N}}$ *is bounded, then*

$$\lim_{k\to\infty} \Theta_k(u_k) = 0. \tag{9}$$

(c) **Convergence & asymptotic optimality of the limit point :** *Assume (b-i) and that* $\Omega$ *has a relative interior w.r.t.* $V \cap H \neq \emptyset$ *for a hyperplane* $H \subset \mathcal{H}$*. Then, by using* $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$*,* $\forall k \in \mathbb{N}$*,* $(u_k)_{k\in\mathbb{N}}$ *converges strongly to some* $\hat{u} \in V$*, i.e.,*

$$\lim_{k\to\infty} \|u_k - \hat{u}\|_{\mathcal{H}} = 0. \tag{10}$$

*Moreover,* $\lim_{k\to\infty} \Theta_k(\hat{u}) = 0$*, if (i)* $(\psi_k'(u_k))_{k\in\mathbb{N}}$ *is bounded and (ii) there exists bounded* $(\psi_k'(\hat{u}))_{k\in\mathbb{N}}$ *where* $\psi_k'(\hat{u}) \in \partial\psi_k(\hat{u})$*,* $\forall k \in \mathbb{N}$*.*

(d) **A characterization of the limit point :** *Assume (b-i), (c-i), and (c-ii). Assume also that* $\Omega$ *has a relative interior* $u_{\mathrm{ri}} \in \Omega$ *w.r.t.* $V$*. With* $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$*,* $\forall k \in \mathbb{N}$*, let* $\hat{u} := \lim_{k\to\infty} u_k \in V$*. Then, it follows that* $\hat{u} \in \overline{\liminf_{k\to\infty} \Omega_k}$*, provided that*[3] *(i) there exists* $\delta > 0$ *such that* $\inf_{u_k \in \Gamma_k(\varepsilon,\rho), k \geq K_0} \psi_k(u_k) \geq \delta$*,* $\forall \varepsilon > 0$*,* $\forall \rho > 0$*, where* $\Gamma_k(\varepsilon,\rho) := \{u \in \mathcal{H} : d(u, lev_{\leq 0}\psi_k) \geq \varepsilon, \|u - u_{\mathrm{ri}}\|_{\mathcal{H}} \leq \rho\}$*.*

*Proof:* The claims (a)–(d) are proved by translating the analysis of $(w_k)_{k\in\mathbb{N}}$ to $(u_k)_{k\in\mathbb{N}}$, as shown below. Define $\Gamma_k := \operatorname{argmin}_{x\in M} \varphi_k(x)$. Then, recalling $\varphi_k(x) = \psi_k(x+v)$, $\forall x \in M$, we have $\Omega_k = \Gamma_k + v := \{x + v : x \in \Gamma_k\}$ hence $\Gamma_k \neq \emptyset (\Leftrightarrow \Omega_k \neq \emptyset)$; note that $v = P_V(0)$. Let $\varphi_k^* := \min_{x\in M} \varphi_k(x) = \psi_k^*$, $k \in \mathbb{N}$.

*Proof of* (a) : By the assumption and $u_k = w_k + v$, we obtain $w_k + v \notin \Gamma_k + v$ $(= \Omega_k \neq \emptyset)$, hence $w_k \notin \Gamma_k \neq \emptyset$. Therefore, [8, Theorem 1(a)] tells us that, $\forall \lambda_k \in \left(0, 2\left(1 - \frac{\varphi_k^*}{\varphi_k(w_k)}\right)\right) = \left(0, 2\left(1 - \frac{\psi_k^*}{\psi_k(u_k)}\right)\right)$, $w_{(k)}^* := u_{(k)}^* - v \in \Gamma_k$ for any $u_{(k)}^* \in \Omega_k$ satisfies

$$\left\|w_{k+1} - w_{(k)}^*\right\|_M < \left\|w_k - w_{(k)}^*\right\|_M. \tag{11}$$

By $w_k - w_{(k)}^* = u_k - u_{(k)}^*$ and $w_{k+1} - w_{(k)}^* = u_{k+1} - u_{(k)}^*$, we can verify the claim.

*Proof of* (b) : By the assumption, we have

$$\exists K_0 \in \mathbb{N} \text{ s.t. } \begin{cases} \varphi_k^* (= \psi_k^*) = 0, \ \forall k \geq K_0, \text{ and} \\ \Gamma := \bigcap_{k\geq K_0} \Gamma_k (= \Omega - v) \neq \emptyset. \end{cases} \tag{12}$$

Therefore, [8, Theorem 1(b)] tells us that, $(w_k)_{k\in\mathbb{N}}$ is bounded, and, for $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$, $\lim_{k\to\infty} \varphi_k(w_k) = 0$ if $(\varphi_k'(w_k))_{k\in\mathbb{N}}$ is bounded. The boundedness of $(u_k)_{k\in\mathbb{N}}$ is verified by noting $\|u_k\|_{\mathcal{H}}^2 = \|w_k + v\|_{\mathcal{H}}^2 = \|w_k\|_{\mathcal{H}}^2 + \|v\|_{\mathcal{H}}^2 = \|w_k\|_M^2 + \|v\|_{\mathcal{H}}^2$. The rest of the claim is proved by noting $0 \leq \Theta_k(u_k) \leq \psi_k(u_k) = \varphi_k(w_k)$ and $\psi_k'(u_k) = \varphi_k'(w_k)$.

*Proof of* (c) : The assumption *(b-i)* implies (12). Moreover, the existence of a relative interior of $\Omega$ w.r.t. $V \cap H$ ensures the existence of a relative interior of $\Gamma$ w.r.t. $\tilde{H} := \{x \in M : \langle x, a\rangle_M = b\} \neq \emptyset$ for some $(a,b) \in M \setminus \{0\} \times \mathbb{R}$. Note here that $\tilde{H}$ is a hyperplane in $(M, \langle\cdot,\cdot\rangle_M)$. Therefore, [8, Theorem 1(c)] tells us that by using $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$, $\forall k \in \mathbb{N}$, $(w_k)_{k\in\mathbb{N}}$ converges strongly to some $\hat{w} \in M$, and $\lim_{k\to\infty} \varphi_k(\hat{w}) = 0$ if (i) $(\varphi_k'(w_k))_{k\in\mathbb{N}}$ is bounded and (ii) there exists bounded $(\varphi_k'(\hat{w}))_{k\in\mathbb{N}}$ where $\varphi_k'(\hat{w}) \in \partial\varphi_k(\hat{w})$, $\forall k \in \mathbb{N}$. Letting $\hat{u} := \hat{w} + v$ and noting $0 \leq \Theta_k(\hat{u}) \leq \psi_k(\hat{u}) = \varphi_k(\hat{w})$, $\partial\psi_k(\hat{u}) = \partial\varphi_k(\hat{w})$, and $\psi_k'(u_k) = \varphi_k'(w_k)$, we can verify the claim.

*Proof of* (d) : The assumptions *(b-i)* and *(c-i)* imply (12) and the boundedness of $(\varphi_k'(w_k))_{k\in\mathbb{N}}$, respectively. The assumption *(c-ii)* automatically ensures the existence of a bounded sequence $(\psi_k'(\hat{u}))_{k\in\mathbb{N}}$ where $\psi_k'(\hat{u}) \in \partial\psi_k(\hat{u}) \cap M$, $\forall k \in \mathbb{N}$, because $P_M(\partial\psi_k(\hat{u})) \subset \partial\psi_k(\hat{u})$ by Lemma 1 and $\|P_M(\psi_k'(\hat{u}))\|_{\mathcal{H}} \leq \|\psi_k'(\hat{u})\|_{\mathcal{H}}$. This surely implies the existence of $(\varphi_k'(\hat{w}))_{k\in\mathbb{N}}$ where $\varphi_k'(\hat{w}) \in \partial\varphi_k(\hat{w})$, $\forall k \in \mathbb{N}$. Moreover, the existence of a relative interior of $\Omega$ w.r.t. $V$ ensures the existence of an interior $w_{\mathrm{int}} := u_{\mathrm{ri}} - v$ of $\Gamma$ in $(M, \langle\cdot,\cdot\rangle_M)$. With $\lambda_k \in [\varepsilon_1, 2-\varepsilon_2] \subset (0,2)$, $\forall k \in \mathbb{N}$, let $\hat{w} := \hat{u} - v = \lim_{k\to\infty} w_k \in M$. The condition *(d-i)* implies

$$\inf_{\substack{d(w_k, \mathrm{lev}_{\leq 0}\varphi_k) \geq \varepsilon \\ \|w_k - w_{\mathrm{int}}\|_M \leq r \\ k \geq K_0}} \varphi_k(w_k) \geq \delta, \ \forall \varepsilon > 0, \ \forall r > 0, \ \exists \delta > 0, \tag{13}$$

because $d(u_k, \mathrm{lev}_{\leq 0}\psi_k) = d(w_k, \mathrm{lev}_{\leq 0}\varphi_k)$, $\|u_k - u_{\mathrm{ri}}\|_{\mathcal{H}} = \|w_k - w_{\mathrm{int}}\|_M$, and $\psi_k(u_k) = \varphi_k(w_k)$. Therefore, [8, Theorem 1(d)] tells us that $\hat{w} \in \overline{\liminf_{k\to\infty} \Gamma_k}$. Recalling $\Omega_k = \Gamma_k + v$, it follows that $\hat{u} \in \overline{\liminf_{k\to\infty} \Omega_k}$, which completes the proof. □

## 4. DERIVATIONS OF LINEARLY CONSTRAINED ADAPTIVE ALGORITHMS

This section presents some examples of linearly constrained adaptive algorithms. First we consider the general case of Problem 2 in Section 4.1, and then narrow down our focus to the set-theoretic problem formulation. PAPA, EAPA, and CAPA are obtained as particular examples.

### 4.1 A Generic Approach to Problem 1

Given a continuous convex function $\Theta_k : \mathcal{H} \to [0,\infty)$, it can be verified that

$$\psi_k : \mathcal{H} \to [0,\infty), x \mapsto \Theta_k(P_V(x)), \ k \in \mathbb{N}, \tag{14}$$

is constraint-embedding for $(\Theta_k, V)$; see Example 1.1. It holds that $\psi_k'(x) := P_M\left(\Theta_k'(P_V(x))\right) \in \partial\psi_k(x)$, $x \in \mathcal{H}$, where $\Theta_k'(x) \in \partial\Theta_k(x)$, $x \in V$ [8]. An application of (6) to $(\psi_k)_{k\in\mathbb{N}}$ in (14) yields the following recursion:

$$u_{k+1} := \begin{cases} u_k - \lambda_k \dfrac{\Theta_k(u_k)}{\left\|P_M(\Theta_k'(u_k))\right\|_{\mathcal{H}}^2} P_M(\Theta_k'(u_k)) & \text{if } \Theta_k'(u_k) \notin M^\perp \\ u_k & \text{otherwise.} \end{cases} \tag{15}$$

We repeat here that, starting at some $u_0 \in V$, all the sequence $(u_k)_{k\in\mathbb{N}}$ generated by (6) lies on $V$ (i.e., $P_V(u_k) = u_k$, $\forall k \in \mathbb{N}$) because of the constraint-embedding structure of $\psi_k$. We mention that $\Theta_k'(u_k) \notin M^\perp \Leftarrow 0 \notin \partial\psi_k(u_k) \Leftrightarrow \psi_k(u_k) > \inf_{x\in\mathcal{H}} \psi_k(x) = \inf_{x\in V} \psi_k(x) \Leftrightarrow \Theta_k(u_k) > \inf_{x\in V} \Theta_k(x) \geq \inf_{x\in\mathcal{H}} \Theta_k(x)$; see Definition 1. Assume $\Theta_k'(u_k) \notin M^\perp$ $(\Leftrightarrow P_M(\Theta_k'(u_k)) \neq 0)$. Then, by (i) the relation[4] $P_V(x) = P_M(x) + v$ for any $x \in \mathcal{H}$ and (ii) the linearity of $P_M$, we can verify that (15) becomes

$$u_{k+1} = P_V\left(u_k - \tilde{\lambda}_k \frac{\Theta_k(u_k)}{\left\|\Theta_k'(u_k)\right\|_{\mathcal{H}}^2} \Theta_k'(u_k)\right), \tag{16}$$

where $\tilde{\lambda}_k := \lambda_k \dfrac{\left\|\Theta_k'(u_k)\right\|_{\mathcal{H}}^2}{\left\|P_M(\Theta_k'(u_k))\right\|_{\mathcal{H}}^2} \geq \lambda_k$, $k \in \mathbb{N}$. It is seen that, with the range of $\tilde{\lambda}_k$ restricted to $[0,2]$, the recursion (16) coincides with the original APSM presented in (4).

**Remark 1** *The derivation above brings the following insight: the original APSM for the linear constraint* $V$*, i.e.* (16) *for* $\tilde{\lambda}_k \in [0,2]$*, needs the sequence* $\left(\dfrac{\left\|P_M(\Theta_k'(u_k))\right\|_{\mathcal{H}}^2}{\left\|\Theta_k'(u_k)\right\|_{\mathcal{H}}^2}\right)_{k\in\mathbb{N}} \subset (0,1]$ *to be*

[3] $\liminf_{k\to\infty} \Omega_k := \bigcup_{k=0}^{\infty} \bigcap_{n\geq k} \Omega_n$. The overline denotes *closure*.

[4] Given any closed convex set $K \subset \mathcal{H}$ and any $v \in \mathcal{H}$, $P_{K+v}(x) = P_K(x - v) + v$ for $x \in \mathcal{H}$ [16]. Letting $K = M$ yields $P_V(x) = P_M(x - v) + v = P_M(x) + v$, $x \in \mathcal{H}$: note the linearity of $P_M$ and $P_M(v) = 0$ (as $v = P_V(0) \in M^\perp$).

*bounded below by a positive constant. The condition is required for ensuring that* $(\lambda_k)_{k\in\mathbb{N}}$ *is bounded below to apply Theorem 1. On the other hand,* (15) *does not require such a boundedness condition.*

### 4.2 Three Classes of Algorithm for Set-Theoretic Problem Formulation

We present three classes of algorithm with a set-theoretic problem formulation. Exploiting a set of available data at each iteration, we construct nonempty closed convex sets, say $C_k^{(i)} \subset \mathscr{H}$, $i \in \mathscr{I}_k \subset \mathbb{N}^*$, which contain the optimal filter (e.g., the MPDR filter in beamforming applications) with high reliability. The reader may refer to [7, 9, 13] and the references therein for particular designs of $C_k^{(i)}$. In this case, the goal is to find a common point $x \in V$ of the convex sets $C_k^{(i)}$ for all $k \in \mathbb{N}$, if such $x$ exists, otherwise find $x \in V$ that is 'closest' (in some sense) to all $C_k^{(i)}$. To specify a criterion for 'closest', define $\omega_k^{(i)} > 0$, $i \in \mathscr{I}_k$, $k \in \mathbb{N}$, satisfying $\sum_{i\in\mathscr{I}_k} w_k^{(i)} = 1$, and let $L_k := \sum_{i\in\mathscr{I}_k} \omega_k^{(i)} d(u_k, C_k^{(i)})$. If $L_k \neq 0 \left(\Leftrightarrow u_k \notin \bigcap_{i\in\mathscr{I}_k} C_k^{(i)}\right)$, the weight to each $C_k^{(i)}$ is given by $\nu_k^{(i)} := \dfrac{\omega_k^{(i)} d(u_k, C_k^{(i)})}{L_k}$; it holds that $\sum_{i\in\mathscr{I}_k} \nu_k^{(i)} = 1$, $k \in \mathbb{N}$. The factor $d(u_k, C_k^{(i)})$ appearing in the definition of $\nu_k^{(i)}$ assigns a larger weight to such $C_k^{(i)}$ that is more distant than the other sets from the current estimate $u_k$, while the user designing parameter $\omega_k^{(i)}$ may reflect the priority of each set; e.g., a large $\omega_k^{(i)}$ could be assigned to such $C_k^{(i)}$ that is associated with a recently measured datum. The cost function $\Theta_k : \mathscr{H} \to [0,\infty)$ is then defined as follows:

$$\Theta_k : x \mapsto \begin{cases} \sum_{i\in\mathscr{I}_k} \nu_k^{(i)} d(x, C_k^{(i)}) & \text{if } L_k \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

The set-theoretic problem formulation is given as follows: minimize $(\Theta_k)_{k\in\mathbb{N}}$ in (17) over $V$ in an asymptotic sense. Our first algorithm for this problem — which we refer to as *projected type* — is obtained by applying (16) with $\tilde{\lambda}_k \in [0,2]$ to $(\Theta_k)_{k\in\mathbb{N}}$ defined as in (17). The resultant algorithm is the adaptive parallel projection algorithm [8, 9] composed with the orthogonal projection operator $P_V$. Our second algorithm — which we refer to as *embedded-constraint type* — is obtained by applying (15) to $(\Theta_k)_{k\in\mathbb{N}}$ defined as in (17), yielding a general form of the blind constrained parallel projection algorithm [7]. To obtain our third algorithm — which we refer to as *constrained type* — we define $\mathscr{L}_k := \left\{i \in \mathscr{I}_k : C_k^{(i)} \cap V \neq \emptyset\right\}$, $N_k := \sum_{i\in\mathscr{L}_k} \omega_k^{(i)} d(u_k, C_k^{(i)} \cap V)$ with $\omega_k^{(i)} > 0$ satisfying $\sum_{i\in\mathscr{L}_k} \omega_k^{(i)} = 1$, and $\mu_k^{(i)} := \dfrac{\omega_k^{(i)} d(u_k, C_k^{(i)} \cap V)}{N_k}$, $i \in \mathscr{L}_k$, for $N_k \neq 0$. The algorithm is then derived by applying (6) to $(\psi_k)_{k\in\mathbb{N}}$ with $\psi_k : \mathscr{H} \to [0,\infty)$ defined as follows:

$$\psi_k : x \mapsto \begin{cases} \sum_{i\in\mathscr{L}_k} \mu_k^{(i)} d(x, C_k^{(i)} \cap V) & \text{if } \mathscr{L}_k \neq \emptyset, N_k \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

In the particular case that we only have a single linear variety $C_k^{(1)} := V_k$ ($\mathscr{I}_k := \{1\}$) at each time $k \in \mathbb{N}$, the three algorithms described above are reduced respectively to PAPA, EAPA, and CAPA. Most of the conditions in Theorem 1 are automatically satisfied for the algorithms described in this subsection; cf. [8].

## 5. CONCLUSION

We have presented a deterministic analysis of the linearly constrained adaptive filtering algorithms with the adaptive projected subgradient method. Based on the fact that the trajectory of instantaneous estimate lies in the affine constraint set, all the points on the constraint set have been translated to its underlying subspace, which has been regarded as a Hilbert space. The constraint-embedding functions introduced has enabled to analyze the three classes of linearly constrained adaptive algorithms in a unified way. The general framework with the constraint-embedding functions will be useful in deriving efficient algorithms for a wide range of adaptive signal processing applications involving linear constraints.

**Acknowledgment:** This work was partially supported by KAKENHI (20760252).

## REFERENCES

[1] O. L. III Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. of IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[2] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 944–960, July 1995.

[3] S. C. Park and J. F. Doherty, "Generalized projection algorithm for blind interference suppression in DS/CDMA communications," *IEEE Trans. Circuits and Systems–II*, vol. 44, no. 6, pp. 453–460, June 1997.

[4] J. A. Apolinário, Jr., S. Werner, P. S. R. Diniz, and T. I. Laakso, "Constrained normalized adaptive filters for CDMA mobile communications," in *Proc. EUSIPCO*, 1998, vol. IV, pp. 2053–2056.

[5] M. L. R. de Campos, S. Werner, and J. A. Apolinário, Jr., "Constrained adaptive algorithms employing Householder transformation," *IEEE Trans. Signal Processing*, pp. 2187–2195, Sept. 2002.

[6] S. Werner, J. A. Apolinário Jr., M. L. R. de Campos, and P. S. R. Diniz, "Low-complexity constrained affine-projection algorithms," *IEEE Trans. Signal Processing*, vol. 53, no. 12, pp. 4545–4555, Dec. 2005.

[7] M. Yukawa, R. L. G. Cavalcante, and I. Yamada, "Efficient blind MAI suppression in DS/CDMA systems by embedded constraint parallel projection techniques," *IEICE Trans. Fundamentals*, vol. E88-A, no. 8, pp. 2062–2071, Aug. 2005.

[8] 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.

[9] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections: a unifying framework for linear and nonlinear classification and regression tasks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 97–123, Jan. 2011.

[10] W. Liu, J. Príncipe, and S. Haykin, *Kernel Adaptive Filtering*, Wiley, New Jersey, 2010.

[11] I. Yamada, *Kougaku no Tameno Kansu Kaiseki (Functional Analysis for Engineering)*, Suurikougaku-Sha / Saiensu-Sha, Tokyo, May 2009, in Japanese.

[12] H. L. Van Trees, *Optimum Array Processing —Part IV of Detection, Estimation, and Modulation Theory*, New York: Wiley, 2002.

[13] M. Yukawa and I. Yamada, "Adaptive beamforming by constrained parallel projection in the presence of spatially-correlated interferences," in *Proc. IEEE ICASSP*, 2006, pp. 1009–1012.

[14] S. Boyd and L.Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[15] J. B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, Berlin, 2001.

[16] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer, 2001.