AN ALGEBRAIC REAL TRANSLATION OF HYPERCOMPLEX LINEAR SYSTEMS AND ITS APPLICATION TO ADAPTIVE FILTERING

Takehiko Mizoguchi[†] *and Isao Yamada*[†]

[†]Department of Communications and Integrated Systems, Tokyo Institute of Technology, Japan
{mizoguchi, isao}@sp.ss.titech.ac.jp

ABSTRACT

The hypercomplex (e.g., complex, quaternion) valued linear model often arises in the signal processing field and attract increasing attention recently. In this paper, we present an algebraic translation of a hypercomplex valued linear systems into a real valued linear model. This translation is designed by taking advantage of isomorphism between hypercomplex numbers and multi-dimensional real vectors and enables us to straightforwardly apply real valued optimization frameworks to various estimation problems for the hypercomplex linear model. We also clarify the useful algebraic properties of the translation. As an application to hypercomplex valued adaptive filtering problems, we derived $\mathbb{A}_m$-adaptive projected subgradient method ($\mathbb{A}_m$-APSM) for hypercomplex valued system identification problems, and show that many hypercomplex adaptive filtering algorithms can be viewed as a special case of this algorithm. Numerical example shows that a new algorithm derived from proposed algorithm outperforms existing hypercomplex adaptive algorithms.

Index Terms— hypercomplex number, Cayley-Dickson procedure, linear system, adaptive filtering, parallel projection

1. INTRODUCTION

Hypercomplex (e.g., complex, quaternion) signals arise naturally in many areas of engineering and science such as communications, wind forecasting [1, 2, 3] computer graphics [4] and robotics [5]. In the statistical signal processing field, effective utilization of hypercomplex number system, e.g., complex and quaternion, have been investigated extensively.

For design of hypercomplex valued adaptive learning algorithms, we often need to evaluate the derivative of a cost function for an computational strategy in the hypercomplex optimization framework. However, even if the system is represented by hypercomplex number system, the cost functions should be real valued, hence not analytic as a univariate hypercomplex valued function, e.g., Cauchy-Riemann equation is not satisfied in the complex domain. As a systematic use of real differentiability of the cost function in the hypercomplex domain, special calculuses have been established and applied in the existing optimization framework to the hypercomplex domain.

In the adaptive filtering field, for example, the Wirtinger calculus ($\mathbb{CR}$-calculus) [6, 7] has been used for design of complex adaptive algorithms [8, 9, 10] and the $\mathbb{HR}$-calculus [11] was proposed specially to design quaternion adaptive algorithms [11, 12]. These special calculuses give us insight for extending relatively simple gradient descent-type adaptive filtering algorithms e.g., *normalized least mean square* (NLMS) [13], *affine projection algorithm* (APA) [14], to the hypercomplex domain [9, 10, 12]. However, observing $\mathbb{CR}$-calculus and $\mathbb{HR}$-calculus, we see that the complexity for the special calculuses tends to increase w.r.t. the dimension of $\mathbb{A}_m$ (see, (1)). Moreover, such a special calculus has not yet been established for the general hypercomplex number systems. This situation could be a burden to create further advanced algorithms in hypercomplex number systems.

In this paper, to clarify the relation between hypercomplex and real vector valued linear system, we propose an algebraic real translation of hypercomplex linear models and show that any hypercomplex linear model can be translated into a real one by taking advantage of the isomorphism between hypercomplex numbers and multi-dimensional real vectors. We also clarify useful algebraic properties of this translation. Thanks to these properties, the proposed translation enable us to immediately obtain isomorphic real models to hypercomplex linear models as well as obtain an optimal solution without passing through hypercomplex optimization method which requires such special calculuses.

As an application to hypercomplex valued adaptive filtering problem, we present a hypercomplex adaptive algorithm named $\mathbb{A}_m$-APSM. This algorithm is based on the *adaptive projected subgradient method* (APSM) [15, 16], which has been proposed as an efficient algorithm for asymptotic minimization of a certain sequence of nonnegative convex functions. The proposed adaptive algorithm, $\mathbb{A}_m$-APSM is an extension of the APSM to the hypercomplex domain by using the proposed translation. Similar to the real valued case, $\mathbb{A}_m$-APSM covers a wide range of the hypercomplex valued projection based adaptive filtering algorithms. Indeed, by designing a certain sequence of convex objectives, a variety of hypercomplex valued adaptive filtering algorithms such as $\mathbb{A}_m$-*normalized least mean square* ($\mathbb{A}_m$-NLMS), $\mathbb{A}_m$-*affine projection algorithm* ($\mathbb{A}_m$-APA) and $\mathbb{A}_m$-*adaptive parallel subgradient projection* ($\mathbb{A}_m$-APSP) are derived in a unified manner as simple examples of the $\mathbb{A}_m$-APSM. Numerical examples show that an algorithm derived as a special case of proposed adaptive algorithm outperforms existing adaptive algorithms.

2. HYPERCOMPLEX NUMBERS

Let $\mathbb{N}$ and $\mathbb{R}$ be respectively the set of all nonnegative integers and the set of all real numbers. Define an m-dimensional *hypercomplex* number $\mathbb{A}_m$ $(m \in \mathbb{N} \setminus \{0\})$ expanded on the real vector space [17]

$$a := a_1 i_1 + a_2 i_2 + \cdots + a_m i_m \in \mathbb{A}_m,\ a_1, \ldots, a_m \in \mathbb{R} \quad (1)$$

based on imaginary units $i_1, \ldots, i_m$, where $i_1 = 1$ represents the vector identity element. A *multiplication table* defines the products of any imaginary unit with each other or with itself. Any hypercomplex number is expressed uniquely in the form of (1). We also define

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

the *conjugate* of hypercomplex number a as

$$a^* := a_1 i_1 - a_2 i_2 - \cdots - a_m i_m. \quad (2)$$

In this paper, we consider the hypercomplex number systems which are constructed recursively by the *Cayley-Dickson doubling procedure* (C-D procedure) [17]. The C-D procedure is a well-known method for extending a number system. This method is used to extend the real number into the complex number, and the complex number into quaternion number. By using the C-D procedure, an m-dimensional hypercomplex number $\mathbb{A}_m$ is extended to $\mathbb{A}_{2m}$ by [17, 18]

$$z := x + y i_{m+1} \in \mathbb{A}_{2m}, \quad x, y \in \mathbb{A}_m, \quad (3)$$

where $i_{m+1} \notin \mathbb{A}_m$ is the additional imaginary unit for doubling the dimension of $\mathbb{A}_m$ satisfying $i_{m+1}^2 = -1$, $i_1 i_{m+1} = i_{m+1} i_1 = i_{m+1}$ and $i_v i_{m+1} = -i_{m+1} i_v =: i_{m+v}$ for all $v = 2, \ldots, m$. For example, the real number system ($\mathbb{A}_1 :=$) $\mathbb{R}$ is extended into complex number system $\mathbb{C}$ ($= \mathbb{A}_2$) by the C-D procedure. Note that the value of m is restricted to the form of 2^n ($n \in \mathbb{N}$). The hypercomplex number systems constructed inductively from the real number by the C-D procedure are called *Cayley-Dickson number system*.

Fact 2.1. According to this procedure, imaginary units of Cayley-Dickson number system have the following properties:

1. $i_\alpha^2 = -1$ for all $\alpha \in \{2, \ldots, m\}$.
2. $i_\alpha i_\beta = -i_\beta i_\alpha$ for all $\alpha, \beta \in \{2, \ldots, m\}$.
3. There exist $\gamma \in \{1, \ldots, m\}$ s.t. $i_\alpha i_\beta = i_\gamma$ or $-i_\gamma$ for all $\alpha, \beta \in \{1, \ldots, m\}$.

Note that Fact 2.1 ensures $aa^* = \sum_{\ell=1}^m a_\ell^2 \geq 0$ for any $a \in \mathbb{A}_m$ in (1) and $a^* \in \mathbb{A}_m$ in (2).

A representative example of hypercomplex number is the *quaternion* $\mathbb{H}$. The quaternion number system is constructed from the complex number system by using the C-D procedure. A quaternion number is a 4-dimensional hypercomplex which is defined as

$$q = q_1 + q_2 \imath + q_3 \jmath + q_4 \kappa \in \mathbb{H}, \quad q_1, q_2, q_3, q_4 \in \mathbb{R} \quad (4)$$

with the multiplication table:

$$\begin{gathered} \imath\jmath = -\jmath\imath = \kappa,\ \jmath\kappa = -\kappa\jmath = \imath,\ \kappa\imath = -\imath\kappa = \jmath, \\ \imath^2 = \jmath^2 = \kappa^2 = -1 \end{gathered} \quad (5)$$

by letting $M = 4$, $i_1 = 1$, $i_2 = \imath$, $i_3 = \jmath$ and $i_4 = \kappa$. From (5), quaternions are not *commutative*, that is, $pq = qp$ for $p, q \in \mathbb{H}$ does not hold in general.

Remark 2.1. The octonion $\mathbb{O}$ can be constructed from quaternion by the C-D procedure. Note that the octonions are neither commutative nor *associative*, that is, $pq = qp$ or $p(qr) = (pq)r$ for $p, q, r \in \mathbb{O}$ does not hold in general.

We also define the hypercomplex space $\mathbb{A}_m^N$, $\forall N \in \mathbb{N} \setminus \{0\}$ equipped with the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{A}_m} := \boldsymbol{x}^H \boldsymbol{y} \in \mathbb{A}_m, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_m^N$ and its induced norm $\|\boldsymbol{x}\|_{\mathbb{A}_m} := \langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbb{A}_m}^{1/2}$, $\forall \boldsymbol{x} \in \mathbb{A}_m^N$, where $(\cdot)^H$ denotes the *Hermitian transpose* of vectors or matrices (e.g., $\boldsymbol{x}^H := [x_1^*, \ldots, x_N^*]$ for $\boldsymbol{x} := [x_1, \ldots, x_N]^\top$, where $x_1 \ldots, x_N \in \mathbb{A}_m$ and $(\cdot)^\top$ stands for the transpose). For any nonempty closed convex set[1] $C \subset \mathbb{A}_m^N$, the *projection operator* $P_C^{\mathbb{A}_m} : \mathbb{A}_m^N \to C$ assigns a vector $\boldsymbol{x} \in \mathbb{A}_m^N$ to the unique vector $P_C^{\mathbb{A}_m}(\boldsymbol{x}) \in C$ s.t. $d_{\mathbb{A}_m}(\boldsymbol{x}, C) := \left\|\boldsymbol{x} - P_C^{\mathbb{A}_m}(\boldsymbol{x})\right\|_{\mathbb{A}_m} = \min_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{A}_m}$.

[1]A set $C \subset \mathbb{A}_m^N$ is said to be *convex* provided that $\forall \boldsymbol{x}, \boldsymbol{y} \in C, \forall \nu \in (0, 1)$, $\nu \boldsymbol{x} + (1 - \nu)\boldsymbol{y} \in C$.

3. PROPOSED ISOMORPHIC TRANSLATION

In this section, we propose an algebraic translation of hypercomplex valued linear systems into real systems. Suppose that the hypercomplex valued linear systems are expressed as

$$\boldsymbol{y} := \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad (6)$$

where $\boldsymbol{y}, \boldsymbol{b} \in \mathbb{A}_m^M$, $\boldsymbol{x} := \sum_{\ell=1}^m \boldsymbol{x}_\ell i_\ell \in \mathbb{A}_m^N$ ($\boldsymbol{x}_\ell \in \mathbb{R}^N$) and $\boldsymbol{A} := \sum_{\ell=1}^m \boldsymbol{A}_\ell i_\ell \in \mathbb{A}_m^{M \times N}$ ($\boldsymbol{A}_\ell \in \mathbb{R}^{M \times N}$). A trivial correspondence (mapping) of hypercomplex vectors[2] or matrices to real ones is

$$\widehat{(\cdot)} : \mathbb{A}_m^{M \times N} \to \mathbb{R}^{mM \times N} : \\ \boldsymbol{A} \mapsto \widehat{\boldsymbol{A}} := \left[\boldsymbol{A}_1^\top, \boldsymbol{A}_2^\top, \ldots, \boldsymbol{A}_m^\top\right]^\top. \quad (7)$$

This correspondence is just concatenating a real part and all imaginary parts in the hypercomplex vectors or matrices. Obviously, this mapping is invertible and thus we can also define

$$\breve{(\cdot)} : \mathbb{R}^{mM \times N} \to \mathbb{A}_m^{M \times N} : \widehat{\boldsymbol{A}} \mapsto \boldsymbol{A}. \quad (8)$$

Unfortunately, it is hard to obtain the correspondence of $\boldsymbol{A}\boldsymbol{x}$ only in terms of the mappings $\widehat{(\cdot)}$ and $\breve{(\cdot)}$, so we propose the following nontrivial mapping:

$$\widetilde{(\cdot)} : \mathbb{A}_m^{M \times N} \to \mathbb{R}^{mM \times mN} : \\ \boldsymbol{A} \mapsto \widetilde{\boldsymbol{A}} := \left[\boldsymbol{L}_M^{(1)\top}\widehat{\boldsymbol{A}}, \boldsymbol{L}_M^{(2)\top}\widehat{\boldsymbol{A}}, \ldots, \boldsymbol{L}_M^{(m)\top}\widehat{\boldsymbol{A}}\right], \quad (9)$$

where the real valued matrix $\boldsymbol{L}_M^{(\ell)} \in \mathbb{R}^{mM \times mM}$ ($\ell = 1, \ldots m$) is defined for the m-dimensional hypercomplex number $\mathbb{A}_m$ as

$$\boldsymbol{L}_M^{(\ell)} = \begin{bmatrix} \delta_{1,1}^{(\ell)}\boldsymbol{I}_M & \delta_{1,2}^{(\ell)}\boldsymbol{I}_M & \cdots & \delta_{1,m}^{(\ell)}\boldsymbol{I}_M \\ -\delta_{2,1}^{(\ell)}\boldsymbol{I}_M & -\delta_{2,2}^{(\ell)}\boldsymbol{I}_M & \cdots & -\delta_{2,m}^{(\ell)}\boldsymbol{I}_M \\ \vdots & \vdots & \ddots & \vdots \\ -\delta_{m,1}^{(\ell)}\boldsymbol{I}_M & -\delta_{m,2}^{(\ell)}\boldsymbol{I}_M & \cdots & -\delta_{m,m}^{(\ell)}\boldsymbol{I}_M \end{bmatrix}, \quad (10)$$

$\boldsymbol{I}_M$ is the M-dimensional identity matrix and

$$\delta_{\alpha,\beta}^{(\gamma)} := \begin{cases} 1 & (\text{if } i_\alpha i_\beta = i_\gamma), \\ -1 & (\text{if } i_\alpha i_\beta = -i_\gamma), \\ 0 & (\text{otherwise}). \end{cases}$$

Lemma 3.1. $\delta_{\alpha,\beta}^{(\gamma)}$ satisfies the following:

1. $\delta_{\alpha,1}^{(\beta)} = \delta_{1,\alpha}^{(\beta)}$ for all $\alpha, \beta \in \{1, \ldots, m\}$.
2. $\delta_{\alpha,1}^{(\beta)} = 1 \iff \alpha = \beta$ for all $\alpha, \beta \in \{1, \ldots, m\}$.
3. $\delta_{1,1}^{(1)} = 1$, $\delta_{\alpha,\alpha}^{(1)} = -1$ for all $\alpha \in \{2, \ldots, m\}$.
4. $\delta_{\alpha,\beta}^{(\gamma)} = -\delta_{\gamma,\beta}^{(\alpha)} = -\delta_{\alpha,\gamma}^{(\beta)} = -\delta_{\beta,\alpha}^{(\gamma)}$ if $\alpha, \beta, \gamma \in \{2, \ldots, m\}$ are distinct.
5. There exists $\gamma \in \{1, \ldots, m\}$ s.t. $\delta_{\alpha,\beta}^{(\gamma)} = 1$ or -1 for all $\alpha, \beta \in \{1, \ldots, m\}$.

Lemma 3.2. $\boldsymbol{L}_M^{(\ell)}$ is an orthogonal matrix i.e., $\boldsymbol{L}_M^{(\ell)\top} = \boldsymbol{L}_M^{(\ell)-1}$ for all $\ell = 1, \ldots, m$.

For the mappings introduced above, we establish the following theorem.

[2]All vectors in this paper are column vectors, that is, $\mathbb{A}_m^N = \mathbb{A}_m^{N \times 1}$.

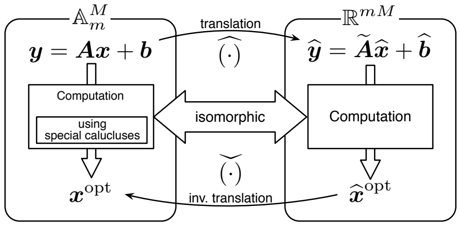

Fig. 1. Concept of proposed translation

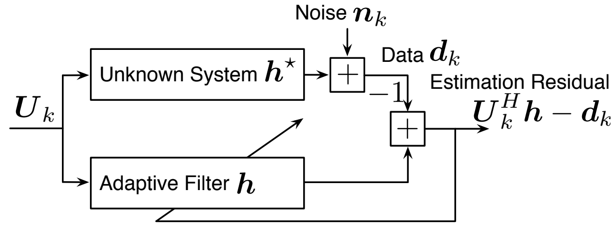

Fig. 2. Adaptive filtering scheme

Theorem 3.1 (Algebraic correspondence between real and hypercomplex vectors/matrices)**.** If the hypercomplex number system $\mathbb{A}_m$ is constructed by the C-D procedure, the following relations hold true:

1. $(\widehat{\boldsymbol{A}+\boldsymbol{B}}) = \widehat{\boldsymbol{A}} + \widehat{\boldsymbol{B}}$ for all $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{A}_m^{M\times N}$
2. $(\widetilde{\boldsymbol{A}+\boldsymbol{B}}) = \widetilde{\boldsymbol{A}} + \widetilde{\boldsymbol{B}}$ for all $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{A}_m^{M\times N}$
3. $(\widehat{\boldsymbol{A}\boldsymbol{B}}) = \widetilde{\boldsymbol{A}}\widehat{\boldsymbol{B}}$ for all $\boldsymbol{A} \in \mathbb{A}_m^{M\times N}$ and $\boldsymbol{B} \in \mathbb{A}_m^{N\times L}$.
4. $(\widetilde{\boldsymbol{A}^H}) = \widetilde{\boldsymbol{A}}^\top$ for all $\boldsymbol{A} \in \mathbb{A}_m^{M\times N}$.
5. $(\widehat{\boldsymbol{A}\boldsymbol{x}}) = \widetilde{\boldsymbol{A}}\widehat{\boldsymbol{x}}$ for all $\boldsymbol{A} \in \mathbb{A}_m^{M\times N}$ and $\boldsymbol{x} \in \mathbb{A}_m^N$.
6. $(\widehat{\boldsymbol{x}^H\boldsymbol{y}}) = \widetilde{\boldsymbol{x}}^\top\widehat{\boldsymbol{y}}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_m^N$.
7. $\|\boldsymbol{x}\|_{\mathbb{A}_m} = \|\widehat{\boldsymbol{x}}\|_{\mathbb{R}}$ for all $\boldsymbol{x} \in \mathbb{A}_m^N$.
8. $\widehat{P_C^{\mathbb{A}_m}(\boldsymbol{x})} = P_{\widehat{C}}^{\mathbb{R}}(\widehat{\boldsymbol{x}})$ for any closed convex set $C \subset \mathbb{A}_m^N$ and any point $\boldsymbol{x} \in \mathbb{A}_m^N$, where $\widehat{C} := \{\widehat{\boldsymbol{x}} \in \mathbb{R}^{mN} | \boldsymbol{x} \in C\} \subset \mathbb{R}^{mN}$. □

Based on Theorem 3.1, we propose an algebraic translation of hypercomplex linear model (6) into the following real vector valued linear model:

$$\widehat{\boldsymbol{y}} = \widetilde{\boldsymbol{A}}\widehat{\boldsymbol{x}} + \widehat{\boldsymbol{b}}. \tag{11}$$

Fig. 1 illustrates the concept of the proposed translation. A hypercomplex linear system (6) is once translated into an equivalent real vector valued linear model (11) by the mapping $\widehat{(\cdot)}$. Suppose that we obtain a real vector $\widehat{\boldsymbol{x}}^{\rm opt} \in \mathbb{R}^{mM}$ by applying some computations to (11). Then we also obtain the corresponding hypercomplex vector $\boldsymbol{x}^{\rm opt} \in \mathbb{A}_m^M$ by applying the inverse mapping $\breve{(\cdot)}$ to $\widehat{\boldsymbol{x}}^{\rm opt}$. By these steps, the proposed translation enables us to obtain solution $\boldsymbol{x}^{\rm opt}$ without using the calculuses designed specially for each hypercomplex number systems.

3.1. Applications to adaptive filtering

In this section, we propose a new adaptive algorithm in the general hypercomplex domain. Let us elaborate on the following hypercomplex adaptive filtering estimation problem. Let $\boldsymbol{u}_k := [u_k, u_{k-1}, \ldots, u_{k-N+1}]^\top \in \mathbb{A}_m^N$ be the input vector and $\boldsymbol{U}_k := [\boldsymbol{u}_k, \boldsymbol{u}_{k-1}, \ldots, \boldsymbol{u}_{k-r+1}] \in \mathbb{A}_m^{N\times r}$ be the input matrix at time k. Let also $n_k \in \mathbb{A}_m$ denote the noise process. If $\boldsymbol{h}^\star := [h_1^\star, h_2^\star, \ldots, h_N^\star]^\top \in \mathbb{A}_m^N$ be the unknown system to be estimated, and $\boldsymbol{n}_k = [n_k, n_{k-1}, \ldots, n_{n-r+1}]^\top \in \mathbb{A}_m^r$ be the noise at time k, we introduce the following hypercomplex linear model for the data process $\boldsymbol{d}_k \in \mathbb{A}_m^r$:

$$\boldsymbol{d}_k := \boldsymbol{U}_k^H\boldsymbol{h}^\star + \boldsymbol{n}_k. \tag{12}$$

By using Theorem 3.1, we immediately obtain the following real valued data process $\widehat{\boldsymbol{d}}_k \in \mathbb{R}^{mr}$:

$$\widehat{\boldsymbol{d}}_k := \widetilde{\boldsymbol{U}}_k^\top\widehat{\boldsymbol{h}}^\star + \widehat{\boldsymbol{n}}_k. \tag{13}$$

Hence the goal of hypercomplex adaptive filtering problem is reduced to approximating the real valued unknown system $\widehat{\boldsymbol{h}}^\star \in \mathbb{R}^{mN}$ by the real valued adaptive filter $\widehat{\boldsymbol{h}}_n \in \mathbb{R}^{mN}$ with the knowledge on $(\widetilde{\boldsymbol{U}}_k, \widehat{\boldsymbol{d}}_k) \in \mathbb{R}^{mN\times mr} \times \mathbb{R}^{mr}, \forall k < n$. Note that at any time we can obtain the corresponding hypercomplex adaptive filter $\boldsymbol{h}_n \in \mathbb{A}_m^N$ to $\widehat{\boldsymbol{h}}_n$ by (8). This is completely the same form as the real valued adaptive filtering problem, so we can directly apply fairly general methods established in the real domain. In this paper, we propose adaptive algorithms based on the *adaptive projected subgradient method* (APSM).

$\mathbb{A}_m$-adaptive projected subgradient method ($\mathbb{A}_m$-APSM). Let $\Theta_k : \mathbb{A}_m^L \to [0, \infty)$ $(k \in \mathbb{N})$ be a sequence of continuous convex[3] functions and $\partial\Theta_k(\boldsymbol{y})$ be the *subdifferential*[4] of Θ_k at $\boldsymbol{y} \in \mathbb{A}_m^L$. The $\mathbb{A}_m$-APSM provides a vector sequence which minimizes asymptotically the sequence of functions Θ_k over a closed convex set $K \subset \mathbb{A}_m^L$. For an arbitrarily given $\boldsymbol{h}_0 \in K$, the $\mathbb{A}_m$-APSM produces a sequence $(\boldsymbol{h}_k)_{k\in\mathbb{N}}$ by

$$\boldsymbol{h}_{k+1} = \begin{cases} P_K^{\mathbb{A}_m}\left(\boldsymbol{h}_k - \lambda_k \frac{\Theta_k(\boldsymbol{h}_k)}{\|\Theta_k'(\boldsymbol{h}_k)\|_{\mathbb{A}_m}^2}\Theta_k'(\boldsymbol{h}_k)\right) & \\ \qquad\qquad\qquad\qquad (\text{if } \Theta_k'(\boldsymbol{h}_k) \neq \boldsymbol{0}), & \\ P_K^{\mathbb{A}_m}(\boldsymbol{h}_k) \qquad\qquad\quad (\text{otherwise}), & \end{cases} \tag{14}$$

where $\Theta_k'(\boldsymbol{h}_k) \in \partial\Theta_k(\boldsymbol{h}_k)$, $0 \leq \lambda_k \leq 2$.

Theorem 3.2 (Properties of $\mathbb{A}_m$-APSM)**.** Similar to the real valued case [15, 16], the sequence $(\boldsymbol{h}_k)_{k\in\mathbb{N}}$ produced by $\mathbb{A}_m$-APSM satisfies the following properties:

(1) (Monotone approximation)

$$\|\boldsymbol{h}_{k+1} - \boldsymbol{h}^{\star(k)}\|_{\mathbb{A}_m} \leq \|\boldsymbol{h}_k - \boldsymbol{h}^{\star(k)}\|_{\mathbb{A}_m}, \tag{15}$$
$$\boldsymbol{h}^{\star(k)} \in \Omega_k := \{\boldsymbol{h} \in K | \Theta_k(\boldsymbol{h}) = \inf_{\boldsymbol{x}\in K}\Theta_k(\boldsymbol{x})\}.$$

(2) (Asymptotic optimality) Suppose

$$\exists N_0 \in \mathbb{N} \text{ s.t. } \begin{cases} \inf_{\boldsymbol{x}}\Theta(\boldsymbol{x}), \forall k \geq N_0 \text{ and} \\ \Omega := \bigcap_{k\geq N_0}\Omega_k \neq \varnothing. \end{cases} \tag{16}$$

Then $(\boldsymbol{h}_k)_{k\in\mathbb{N}}$ is bounded. Moreover, if we use $\lambda_n \in [\varepsilon_1, 2-\varepsilon_2] \subset (0, 2)$ for all k, we have

$$\lim_{k\to\infty}\Theta_k(\boldsymbol{h}_k) = 0 \tag{17}$$

provided that $(\Theta_k')_{k\in\mathbb{N}}$ is bounded. □

As a class of the $\mathbb{A}_m$-APSM, we present the following algorithm.

---

[3] A function $\Theta : \mathbb{A}_m^L \to \mathbb{R}$ is said to be *convex* if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_m^L$ and $\forall\nu \in (0, 1)$, $\Theta(\nu\boldsymbol{x} + (1-\nu)\boldsymbol{y}) \leq \nu\Theta(\boldsymbol{x}) + (1-\nu)\Theta(\boldsymbol{y})$.

[4] The *subdifferential* of Θ at $\boldsymbol{y}$ is the set of all the *subgradient* of Θ at $\boldsymbol{y}$; $\partial\Theta(\boldsymbol{y}) := \{\boldsymbol{s} \in \mathbb{A}_m^L | \langle\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}}, \widehat{\boldsymbol{s}}\rangle_{\mathbb{R}} + \Theta(\boldsymbol{y}) \leq \Theta(\boldsymbol{x}), \forall\boldsymbol{x} \in \mathbb{A}_m^L\}$

Algorithm 3.1. Let $S_i^{(k)} \subset \mathbb{A}_m^L := \mathbb{A}_m^N$, $i \in \mathcal{I}_k \subset \mathbb{Z}$ be closed convex sets. Define the sequence of continuous convex function by $\Theta_k(\boldsymbol{x}) = \frac{1}{L_k}\sum_{i\in\mathcal{I}_k}\omega_i^{(k)} d_{\mathbb{A}_m}(\boldsymbol{h}_k, S_i^{(k)}) d_{\mathbb{A}_m}(\boldsymbol{x}, S_i^{(k)})$, where $\sum_{i\in\mathcal{I}_k}\omega_i^{(k)} = 1$, $\{\omega_i^{(k)}\}_{i\in\mathcal{I}_k} \subset (0,1]$ if $L_k := \sum_{i\in\mathcal{I}_k}\omega_i^{(k)} d_{\mathbb{A}_m}(\boldsymbol{h}_k, S_i^{(k)}) \neq 0$, and $\Theta_k(\boldsymbol{x}) = 0$ otherwise. In this case we have $\Theta_k'(\boldsymbol{x}) = \frac{1}{L_k}\sum_{i\in\mathcal{I}_k}\omega_i^{(k)}(\boldsymbol{x} - P^{\mathbb{A}_m}_{S_i^{(k)}}(\boldsymbol{x})) \in \partial\Theta_k(\boldsymbol{x})$ if $L_n \neq 0$, and $\Theta_k'(\boldsymbol{x}) = \boldsymbol{0} \in \partial\Theta_k(\boldsymbol{x})$ otherwise. By applying (14) to Θ_n with $K \subset \mathbb{A}_m^N$, we deduce a following scheme:

$$\boldsymbol{h}_{k+1} = P_K^{\mathbb{A}_m}\left(\boldsymbol{h}_k + \mu_k\left(\sum_{i\in\mathcal{I}_k}\omega_i^{(k)} P^{\mathbb{A}_m}_{S_i^{(k)}}(\boldsymbol{h}_k) - \boldsymbol{h}_k\right)\right), \quad (18)$$

where $\boldsymbol{h}_0 \in K$, $\mu_k \in [0, 2\mathcal{M}_k]$ and

$$\mathcal{M}_k := \begin{cases} \dfrac{\sum_{i\in\mathcal{I}_k}\omega_i^{(k)}\left\|P^{\mathbb{A}_m}_{S_i^{(k)}}(\boldsymbol{h}_k) - \boldsymbol{h}_k\right\|^2_{\mathbb{A}_m}}{\left\|\sum_{i\in\mathcal{I}_k}\omega_i^{(k)} P^{\mathbb{A}_m}_{S_i^{(k)}}(\boldsymbol{h}_k) - \boldsymbol{h}_k\right\|^2_{\mathbb{A}_m}} & (\text{if } \boldsymbol{h}_k \notin \bigcap_{i\in\mathcal{I}_k} S_i^{(k)}), \\ 1 & (\text{otherwise}). \end{cases}$$

This is a generalization of Algorithm 1 in [19], hence it includes many useful adaptive algorithms shown as the following examples.

Example 3.1. Algorithm 3.1 reproduces the $\mathbb{A}_m$*-affine projection algorithm* ($\mathbb{A}_m$-APA, in particular, $\mathbb{A}_m$-NLMS if $r = 1$) if we set $\mathcal{I}_k = \{k\}$, $K = \mathbb{A}_m^N$, and

$$S_i^{(k)} = V_k := \underset{\boldsymbol{h}\in\mathbb{A}_m^N}{\arg\min}\left\|\boldsymbol{U}_k^H\boldsymbol{h} - \boldsymbol{d}_k\right\|_{\mathbb{A}_m}.$$

This is a hypercomplex extension of the APA [14] (NLMS [13]). As the simplest examples, consider the complex case, i.e., the case where $\mathbb{A}_m = \mathbb{C}$. Then we obtain an algorithm which agrees with the *complex affine projection algorithm* ($\mathbb{C}$-APA) [9]. Note that the complex widely linear model [20] can be expressed in the form of (12), so this algorithm also covers *widely linear complex affine projection algorithm* (WL-$\mathbb{C}$-APA) [10] for noncircular input signals.

Example 3.2. Algorithm 3.1 reproduces the $\mathbb{A}_m$*-adaptive parallel subgradient projection* ($\mathbb{A}_m$-APSP) if we set $\mathcal{I}_k = \{k, k-1, \ldots, k-q+1\}$, $K = \mathbb{A}_m^N$, and

$$\begin{aligned} S_i^{(k)} &= H_i^-(\boldsymbol{h}_k) \\ &:= \{\boldsymbol{x} \in \mathbb{A}_m^N | (\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{h}}_k)^\top \nabla g_i(\boldsymbol{h}_k) + g_i(\boldsymbol{h}_k) \leq 0\}, \quad (19) \end{aligned}$$

where q is the number of parallel processors and $g_i(\boldsymbol{x}) = \|\boldsymbol{U}_i^H\boldsymbol{x} - \boldsymbol{d}_i\|^2_{\mathbb{A}_m} - \rho$, $\forall\rho \geq 0$. Similar to the real valued case [19], the $\mathbb{A}_m$-APSP uses weighted average of multiple subgradient projection to keep low computational cost of $\mathbb{A}_m$-NLMS as well as to achieve fast and stable convergence even in severely noisy environment.

4. NUMERICAL EXAMPLES

We examine the efficiency of the new algorithm derived from the $\mathbb{A}_m$-APSM in the context of a simple system identification problem. Note that the proposed algorithm is designed in the unified way for each hypercomplex number systems. In this paper, we perform the comparison in the complex case as a simple example. We use a complex system $\boldsymbol{h}^\star \in \mathbb{C}^{200}$ with coefficients $h_k^\star = \alpha\left(1 + \cos\left(\frac{2\pi(n-100)}{200}\right) - j\left[1 + \cos\left(\frac{2\pi(n-100)}{400}\right)\right]\right)$, $(k = 1, \ldots, 200)$, where $\alpha = 0.0684$ to ensure unit weight norm. This setting is based on [8]. The input signal u_k is generated by $u_k := \sqrt{1-\beta^2}z_k + j\beta z_k$, where $\beta = 0.1$ and z_k is i.i.d. from real valued Gaussian distribution with mean 0 and variance 1. The noise n_k is zero mean complex circular white Gaussian and signal-to-noise ratio (SNR) $=$ 30dB, where $\text{SNR} := 10\log_{10}(\mathbb{E}\|\boldsymbol{u}_k^H\boldsymbol{h}^\star\|^2_{\mathbb{C}}/\mathbb{E}\|n_k\|^2_{\mathbb{C}})$ and $\mathbb{E}(\cdot)$ denotes expectation. We compare the existing $\mathbb{C}$-LMS [21], $\mathbb{C}$-APA [9] and $\mathbb{A}$-APSP in the complex domain, that is, $\mathbb{C}$-APSP. We set the parameters λ(step-size) $= 0.06$ for $\mathbb{C}$-LMS, $(\mu_k, r, q) = (2, 2, 10), \forall k$ for the $\mathbb{C}$-APSP and $(\mu_k, r) = (1, 2), \forall k$ for the $\mathbb{C}$-APA. The step-sizes of these methods are fixed so that their initial convergence speed are the same. Fig. 3. depicts a comparison of these three methods in the sense of system-mismatch $10\log_{10}(\|\boldsymbol{h}^\star - \boldsymbol{h}_k\|^2_{\mathbb{C}} / \|\boldsymbol{h}^\star\|^2_{\mathbb{C}})$ averaged over 300 trials. It shows that $\mathbb{C}$-APSP achieves better steady-state behavior than $\mathbb{C}$-LMS and $\mathbb{C}$-APA.

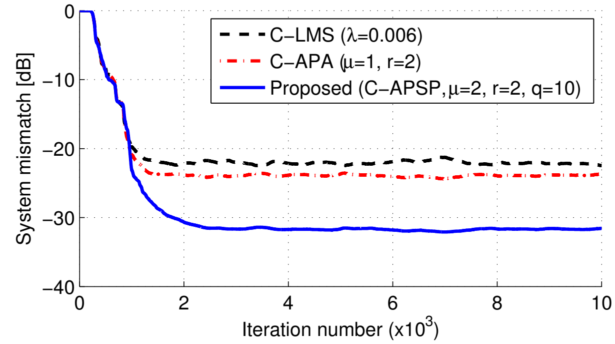

Fig. 3. Comparison of system mismatch

5. CONCLUSIONS WITH RELATION TO PRIOR WORK

We have proposed an algebraic real translation of the hypercomplex valued linear systems and show that this translation enables us to immediately obtain isomorphic real models to hypercomplex linear models. The proposed translation is designed to take advantage of the isomorphism between $\mathbb{A}_m$ and $\mathbb{R}^m$. We have clarified the algebraic properties of the translation. We have also proposed a new hypercomplex adaptive algorithm named $\mathbb{A}_m$-APSM based on the APSM as an application of the proposed translation to adaptive filtering problems. The proposed algorithm covers wide range of hypercomplex adaptive filtering algorithms. Numerical example shows that $\mathbb{C}$-APSP derived from the proposed algorithm achieves better steady-state behavior than some existing adaptive algorithms.

We finally present a short review of related prior work. This work focused on the optimization frameworks in hypercomplex valued signal processing. In general, the cost function is real valued hence not analytic as a univariate hypercomplex valued function due to its strong requirement for the analyticity. Wirtinger [6] presented a special calculus which relaxes this strong requirement, and many prior works employ this derivative in their optimization techniques [8, 9, 10, 22, 23]. Recently, Mandic et al. [11] extended this derivative to the quaternion domain. This work clears the relation between hypercomplex and real systems as well as enables us to design hypercomplex optimization technique without passing through such calculuses.

6. REFERENCES

[1] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, John Wiley and Sons Ltd, 2009.

[2] S. L. Goh, M. Chen, D. H. Popovi, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31, pp. 1733–1750, 2006.

[3] C. Cheong-Took, D. P. Mandic, and K. Aihara, "Quaternion-valued short term forecasting of wind profile," in *Proc. IJCNN*, 2010.

[4] S. B. Choe and J. J. Faraway, "Modeling head and hand orientation during motion using quaternions," *Journal of Aerospace*, vol. 113, no. 1, pp. 186–192, 2004.

[5] J. C. K. Chou, "Quaternion kinematic and dynamic differential equations," *IEEE Trans. Robot. Autom.*, vol. 8, no. 1, pp. 53–64, Feb. 1992.

[6] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veranderlichen," *Math. Ann.*, vol. 97, pp. 357–375, 1927.

[7] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," 2009, [Online]. Available: http://dsp.ucsd.edu/˜kreutz/PEI05.html, Course ECE275A Lecture Supplement on Complex Vector Calculus.

[8] T. Adalı and S. Haykin, Eds., *Adaptive Signal Processing: Next Generation Solutions*, Wiley, 2009.

[9] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, Berlin, 3rd edition, 2008.

[10] Y. Xia, C. C. Took, and D. P. Mandic, "An augmented affine projection algorithm for the filtering of noncircular complex signals," *Signal Processing*, vol. 90, pp. 1788–1799, 2010.

[11] D. P. Mandic, C. Jahanchahi, and C. Cheong Took, "A quaternion gradient operator and its applications," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 47–50, Jan. 2011.

[12] C. Jahanchahi, C. C. Took, and D. P. Mandic, "On gradient calculation in quaternion adaptive filtering," in *Proc. IEEE ICASSP*, 2012.

[13] J. Nagumo and J. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, Mar. 1967.

[14] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *IEICE Trans.*, vol. 67-A, no. 5, pp. 126–132, 1984.

[15] I. Yamada, "Adaptive projected subgradient method: a unified view for projection based adaptive algorithms," *The Journal of IEICE (in japanese)*, vol. 86, no. 8, pp. 654–658, May. 2003.

[16] 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.

[17] I. Kantor and A. Solodovnikov, *Hypercomplex numbers, An Elementary Introduction to Algebras*, Springer-Verlag, New York, 1989.

[18] D. Alfsmann, "On families of 2^N-dimenisonal hypercomplex algebras suitable for digital signal processing," in *Proc. EUSIPCO*, Florence, Italy, 2006.

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

[20] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[21] B. Widrow, J. Cook, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, vol. 63, pp. 719–720, 1975.

[22] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*, Cambridge University Press, 2011.

[23] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients," *IEEE Trans. on Neural Net. and Learning Sys.*, vol. 23, no. 3, pp. 425–438, Mar. 2012.