HYPERCOMPLEX LOW RANK MATRIX COMPLETION WITH NON-NEGATIVE CONSTRAINTS VIA CONVEX OPTIMIZATION

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

[†]Department of Information and Communications Engineering, Tokyo Institute of Technology, Japan
Email: {mizoguchi, isao}@sp.ce.titech.ac.jp

ABSTRACT

Expressing multidimensional information as a value in hypercomplex number systems (e.g., quaternion, octonion, etc.) has great potential, in data sciences, e.g., signal processing, to enjoy their nontrivial algebraic benefits which are not available in standard real or complex vector systems. Strategic utilizations of such benefits would include, e.g., hypercomplex singular value decomposition (SVD) and low rank approximation of matrices. In real world applications, e.g., representing color images, of hypercomplex number systems, all attributes are often restricted to be non-negative. In this paper, we formulate non-negative matrix completion problem in hypercomplex domain as a convex optimization problem in real domain. These formulation is based on algebraic translations of Cayley-Dickson (C-D) linear systems. We then derive an algorithmic solution to hypercomplex low rank matrix completion with non-negative constraint based on a proximal splitting technique. Numerical experiments are performed in a scenario of high-dimensional hypercomplex matrix completion problem and show that the proposed algorithm recovers much more faithfully the original information, masked randomly by noise, than a part-wise state-of-art algorithms.

Index Terms— Cayley-Dickson number system, hypercomplex matrix completion, non-negative constraint, convex optimization, proximal splitting

1. INTRODUCTION

Multidimensional information arises naturally in many areas of engineering and science since almost all observations have many attributes. Utilizing hypercomplex number system for representing such multidimensional information is one of the most effective ways because we can express multidimensional information not in terms of vectors but in terms of numbers among which we can define the four basic arithmetic operators. Indeed, it has been used in many areas such as computer graphics [1] and robotics [2, 3] wind forecasting [4, 5, 6] and noise reduction in acoustic systems [7]. In the statistical signal processing field, effective utilization of the m-dimensional Cayley-Dickson number system (C-D number system) [8, 9], which is a standard class of hypercomplex number systems [10], including, e.g., real $\mathbb{R}$, complex $\mathbb{C}$, quaternion $\mathbb{H}$, octonion $\mathbb{O}$ and sedenion $\mathbb{S}$ etc., have been investigated.

A hypercomplex number has one real part and many imaginary parts, and it can represent multidimensional data as a number for which the four arithmetic operations including multiplication and division are available. It can fulfill the four arithmetic operations for multidimensional information, which are not available for ordinary real multidimensional vectors. Moreover, thanks to the nontrivial algebraic structure, the multiplication of hypercomplex numbers can enjoy algebraically interactions among real and imaginary parts. Hypercomplex vectors, matrices and tensors can also enjoy these benefits. For example, in 3D object modeling, each point in 3-dimensional space can have multiple attribute such as color, material, intensity, etc., and each attribute may have correlation with other attributes. Modeling of the correlations among attributes in multidimensional data and will be more and more important by the popularization of 3D printer [11], virtual reality, medical imaging etc. Algebraically natural operations in hypercomplex number system has great potential for such modelings of various correlations (see e.g., [12, 13, 14, 15, 16, 17] for color image processing applications). However, because of the "singularity" of higher dimensional C-D number systems, few mathematical tools have been maintained [18, 19, 20, 21]. To overcome this situation, in our previous works [22, 23], we have proposed several useful mathematical tools for designing advanced algorithm for optimization, learning and low rank approximation in hypercomplex domain. In [22] we proposed an algebraic real translation for clarifying the relation between C-D linear system and real vector valued linear systems, and successfully designed some online learning algorithms which are available in general C-D domain. Moreover, in [23, 24], we also proposed useful tools C-D singular value decomposition, rank, low rank approximation technique and a sparsity measure.

In real world applications, the range of each attribute is often restricted. For example, the observations of color images are intensities of RGB color spaces and thus they are always non-negative. In an audio context, it is typical to consider only the non-negative magnitudes without taking care about phases [25]. Moreover, in both cases, the low rank assumption is typically justified by the fact that natural images have certain patterns and audio signals are superpositions of relatively few component signals. In hypercomplex domains, matrix completion itself has been proposed at most in the quaternion domain [26] without regarding non-negativeness explicitly despite its application is mainly for RGB color images. Under these situations, therefore, recovering hypercomplex low-rank matrix with non-negative constraint is needed and expected to play an important role in many applications.

In this paper, to establish a matrix completion framework with input restriction, first, we formulate hypercomplex matrix completion problem with non-negative constraint. To achieve it, we introduce a *part-wise non-negativeness* of hypercomplex numbers. Thanks to the simple definition of general hypercomplex non-negativeness, the hypercomplex non-negative matrix completion problem can be recasted to equivalent to a structured convex optimization problem in real domain by utilizing algebraic translations proposed in [22]. We then propose an algorithmic solution to hypercomplex non-negative low rank completion completion algorithm based on a proximal splitting method, *Douglas-Rachford splitting (DRS)* [27]. The proposed algorithm is a C-D generalization of the dual of non-negative matrix completion algorithm proposed in [25] and can be applied to general C-D domains.

Numerical experiments including a scenario of high-dimensional hypercomplex non-negative matrix completion problem show that the proposed algorithm successfully utilizes algebraically natural correlations of each attribute to recover much more faithfully the original information, masked randomly by noise, than a part-wise state-of-art non-negative matrix completion algorithm.

2. PRELIMINARIES

2.1. Hypercomplex Number System

Let $\mathbb{N}$ and $\mathbb{R}$ be respectively the set of all non-negative 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 [8]

$$a := a_1\mathbf{i}_1 + a_2\mathbf{i}_2 + \cdots + a_m\mathbf{i}_m \in \mathbb{A}_m,\ a_1, \ldots, a_m \in \mathbb{R} \qquad (1)$$

based on imaginary units $\mathbf{i}_1, \ldots, \mathbf{i}_m$, where $\mathbf{i}_1 = 1$ represents the vector identity element. Any hypercomplex number is expressed uniquely in the form of (1). The coefficient of each imaginary unit a_ℓ ($\ell = 1, \ldots m$) is called *ℓth imaginary part* and represented as $a_\ell = \Im_\ell(a)$. A *multiplication table* defines the products of any imaginary unit with each other or with itself (e.g., $\mathbf{i}_1^2 = 1, \mathbf{i}_2^2 = -1$ and $\mathbf{i}_1\mathbf{i}_2 = \mathbf{i}_2\mathbf{i}_1 = \mathbf{i}_2$ for $\mathbb{A}_2(=: \mathbb{C})$). We also define the *conjugate* of hypercomplex number a as

$$a^* := a_1\mathbf{i}_1 - a_2\mathbf{i}_2 - \cdots - a_m\mathbf{i}_m. \qquad (2)$$

In this paper, we consider the hypercomplex number systems which are constructed recursively by the *Cayley-Dickson construction* (*C-D construction* or *C-D (doubling) procedure*) [8]. The C-D construction is a standard method for extending a number system. This method has been used in extending $\mathbb{R}$ to $\mathbb{C}$, $\mathbb{C}$ to $\mathbb{H}$ and $\mathbb{H}$ to $\mathbb{O}$. By using the C-D construction, an m-dimensional hypercomplex number $\mathbb{A}_m$ is extended to $\mathbb{A}_{2m}$ [8, 9] as

$$z := x + y\mathbf{i}_{m+1} \in \mathbb{A}_{2m}, \quad x, y \in \mathbb{A}_m,$$

where $i_{m+1} \notin \mathbb{A}_m$ is the additional imaginary unit for doubling the dimension of $\mathbb{A}_m$ satisfying $\mathbf{i}_{m+1}^2 = -1$, $\mathbf{i}_1\mathbf{i}_{m+1} = \mathbf{i}_{m+1}\mathbf{i}_1 = \mathbf{i}_{m+1}$ and $\mathbf{i}_v\mathbf{i}_{m+1} = -\mathbf{i}_{m+1}\mathbf{i}_v =: \mathbf{i}_{m+v}$ 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 construction. 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 construction are called *Cayley-Dickson number system* (*C-D number system*). The imaginary units appeared in the C-D number systems have many characteristic properties [22] such as $\mathbf{i}_\alpha^2 = -1$ and $\mathbf{i}_\alpha\mathbf{i}_\beta = -\mathbf{i}_\beta\mathbf{i}_\alpha (\alpha \neq \beta)$ for all $\alpha, \beta \in \{2, \ldots, m\}$. These properties 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) and enable us to define the absolute values of C-D number a as $|a| := \sqrt{aa^*}$.

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 construction. A quaternion number is a 4-dimensional hypercomplex which is defined as

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

with the multiplication table:

$$\begin{gathered} \imath\jmath = -\jmath\imath = \kappa,\ \jmath\kappa = -\kappa\jmath = \imath,\ \kappa\imath = -\imath\kappa = \jmath, \\ \imath^2 = \jmath^2 = \kappa^2 = -1 \end{gathered} \qquad (3)$$

by letting $m = 4$, $\mathbf{i}_1 = 1$, $\mathbf{i}_2 = \imath$, $\mathbf{i}_3 = \jmath$ and $\mathbf{i}_4 = \kappa$. From (3), quaternions are not *commutative*, i.e., $pq \neq qp$ for $p, q \in \mathbb{H}$ in general.

The octonion $\mathbb{O}$ can be constructed from the quaternion $\mathbb{H}$ by the C-D construction. Note that the multiplication in $\mathbb{O}$ is neither commutative nor *associative*, i.e., $pq \neq qp$ and $(pq)r \neq p(qr)$ for $p, q, r \in \mathbb{O}$ in general [10]. For the octonion multiplication table, see, e.g., [10].

We also define $\mathbb{A}_m^N := \{[x_1, \ldots, x_N]^\top | x_i \in \mathbb{A}_m\,(i = 1, \ldots, N)\}$ for $\forall N \in \mathbb{N} \setminus \{0\}$, where $(\cdot)^\top$ stands for the transpose. Define $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{A}_m^N} := \boldsymbol{x}^\mathsf{H}\boldsymbol{y} \in \mathbb{A}_m, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_m^N$ and $\|\boldsymbol{x}\|_{\mathbb{A}_m^N} := \langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbb{A}_m^N}^{1/2}$, $\forall \boldsymbol{x} \in \mathbb{A}_m^N$, where $(\cdot)^\mathsf{H}$ denotes the *Hermitian transpose* of vectors or matrices (e.g., $\boldsymbol{x}^\mathsf{H} := [x_1^*, \ldots, x_N^*]$ for $\boldsymbol{x} := [x_1, \ldots, x_N]^\top \in \mathbb{A}_m^N$, where $x_1 \ldots, x_N \in \mathbb{A}_m$). We also define the *addition* of two hypercomplex vectors $\boldsymbol{x} + \boldsymbol{y} := [x_1+y_1, \cdots, x_N+y_N]^\top \in \mathbb{A}_m^N$ for $\boldsymbol{x}, \boldsymbol{y} (:= [y_1, \ldots, y_N]^\top) \in \mathbb{A}_m^N$. Let $\mathcal{S} := \mathbb{R}$, $\mathcal{S} := \mathbb{C}$ or $\mathcal{S} := \mathbb{A}_m$ ($m \geq 4$), and call the element of $\mathcal{S}$ *scalar*. If we define the *left scalar multiplication* as $\alpha\boldsymbol{x} := [\alpha x_1, \ldots, \alpha x_N]^\top \in \mathbb{A}_m^N$ for $\alpha \in \mathcal{S}$ and $\boldsymbol{x} \in \mathbb{A}_m^N$, we have $\alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathbb{A}_m^N, \forall \alpha, \beta \in \mathcal{S}, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_m^N$. We can also define the *right scalar multiplication* $\boldsymbol{x}\alpha \in \mathbb{A}_m^N$ in a similar way.

2.2. Algebraic Real Translations of C-D Linear Systems

We briefly review the algebraic translation of C-D valued vectors and matrices proposed in [22]. A trivial correspondence (mapping) of hypercomplex vectors or matrices to real ones is

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

This correspondence is just concatenating a real and all imaginary parts in the hypercomplex vectors or matrices. Obviously, this mapping is invertible and thus we can also define $\widecheck{(\cdot)} : \mathbb{R}^{mM \times N} \to \mathbb{A}_m^{M \times N} : \widehat{\boldsymbol{A}} \mapsto \boldsymbol{A}$. Only in terms of the mappings $\widehat{(\cdot)}$ and $\widecheck{(\cdot)}$, it is hard to obtain the correspondence of matrix-vector product $\boldsymbol{A}\boldsymbol{x}$, so we also introduce the following non-trivial mapping:

$$\begin{aligned} &\widetilde{(\cdot)} : \mathbb{A}_m^{M \times N} \to \mathbb{R}^{mM \times mN} : \\ &\qquad \boldsymbol{A} \mapsto \widetilde{\boldsymbol{A}} := \left[\boldsymbol{L}_M^{(1)\top}\widehat{\boldsymbol{A}}, \boldsymbol{L}_M^{(2)\top}\widehat{\boldsymbol{A}}, \ldots, \boldsymbol{L}_M^{(m)\top}\widehat{\boldsymbol{A}}\right], \end{aligned} \qquad (4)$$

where the 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

$$\boldsymbol{L}_M^{(\ell)} := \begin{bmatrix} \delta_{1,1}^{(\ell)}\boldsymbol{I}_M & \cdots & \delta_{1,m}^{(\ell)}\boldsymbol{I}_M \\ -\delta_{2,1}^{(\ell)}\boldsymbol{I}_M & \cdots & -\delta_{2,m}^{(\ell)}\boldsymbol{I}_M \\ \vdots & \ddots & \vdots \\ -\delta_{m,1}^{(\ell)}\boldsymbol{I}_M & \cdots & -\delta_{m,m}^{(\ell)}\boldsymbol{I}_M \end{bmatrix}, \delta_{\alpha,\beta}^{(\gamma)} := \begin{cases} 1 & (\text{if } \mathbf{i}_\alpha\mathbf{i}_\beta = \mathbf{i}_\gamma), \\ -1 & (\text{if } \mathbf{i}_\alpha\mathbf{i}_\beta = -\mathbf{i}_\gamma), \\ 0 & (\text{otherwise}), \end{cases}$$

and $\boldsymbol{I}_M$ is the M-dimensional identity matrix. Obvious from (4), the degree of freedom of $\widetilde{\boldsymbol{A}}$ is at most that of $\widehat{\boldsymbol{A}} \in \mathbb{R}^{mM \times N}$. More precisely, $\widetilde{(\cdot)}$ is a mapping onto

$$\begin{aligned} \mathfrak{S}_{\mathbb{A}_m}(M, N) &:= \{\widetilde{\boldsymbol{A}} \in \mathbb{R}^{mM \times mN} | \boldsymbol{A} \in \mathbb{A}_m^{M \times N}\} \\ &= \left\{ \left[\boldsymbol{L}_M^{(1)\top}\boldsymbol{B}, \ldots, \boldsymbol{L}_M^{(m)\top}\boldsymbol{B}\right] \,\middle|\, \boldsymbol{B} \in \mathbb{R}^{mM \times N} \right\}. \end{aligned} \qquad (5)$$

$\mathfrak{S}_{\mathbb{A}_m}(M, N)$ represents exactly all C-D matrices in $\mathbb{A}_m^{M \times N}$ in terms of real matrices. Similar to the trivial mapping, $\widetilde{(\cdot)}$ is also invertible and thus we define $\underset{\smile}{(\cdot)} : \mathfrak{S}_{\mathbb{A}_m}(M, N) \to \mathbb{A}_m^{M \times N} : \widetilde{\boldsymbol{A}} \mapsto \boldsymbol{A}$. These translations have many useful algebraic properties. For detail, see [22].

3. NON-NEGATIVE MATRIX COMPLETION IN HYPERCOMPLEX DOMAIN

In this section, we first formulate non-negative hypercomplex matrix completion problem and then propose an algorithmic solution to it.

3.1. Formulation

First of all we have to define the non-negativeness of hypercomplex number. For simplicity, in this paper, we consider the following non-negativeness:

$$\mathbb{A}_{m+} = \{a \in \mathbb{A}_m \,|\, \Im_\ell(a) \geq 0, \forall \ell = 1, \ldots, m\} \subset \mathbb{A}_m. \quad (6)$$

We call an element in $\mathbb{A}_{m+}$ a *part-wise non-negative* C-D number.

By using this definition, the non-negative low rank hypercomplex matrix completion can be formulated as the following optimization problem:

$$\underset{\boldsymbol{X} \in \mathbb{A}_{m+}^{M\times N}}{\text{minimize}} \quad \text{rank}(\widetilde{\boldsymbol{X}}) \quad \text{s.t.} \quad \boldsymbol{X}_\Omega = \boldsymbol{Y}_\Omega, \quad (7)$$

where $\boldsymbol{X}_\Omega$ denotes the restriction of the matrix on the entries given by Ω and $\boldsymbol{Y}_\Omega$ contains the values of those entries of $\boldsymbol{X}$. With the sampling operator $\mathcal{L}_\Omega : \mathbb{A}_m^{M\times N} \to \mathbb{A}_m^p$ extracting p observed entries into a vector $\boldsymbol{b} \in \mathbb{A}_m^p$ and the convex relaxation with the nuclear norm and Lagrange multiplier, we obtain the following unconstrained formulation:

$$\underset{\boldsymbol{X} \in \mathbb{A}_{m+}^{M\times N}}{\text{minimize}} \quad \left\|\widetilde{\boldsymbol{X}}\right\|_* + \frac{\lambda}{2} \left\|\mathcal{L}_\Omega(\boldsymbol{X}) - \boldsymbol{b}\right\|_{\mathbb{A}_m^p}^2, \quad (8)$$

where $\|\cdot\|_*$ is the nuclear norm of real matrices i.e., the sum of positive singular values. In this paper, we call the problem (8) *Cayley-Dickson non-negative matrix completion (C-D NNMC).*

3.2. Algorithm based on Douglas-Rachford Splitting

In this section, we derive a new algorithm based on the Douglas-Rachford splitting technique [27] to solve the C-D NNMC (8) efficiently. The DRS is briefly summarized in Appendix. Denote the 2-fold Cartesian product of the spaces of real matrices by $\mathcal{H}_0 := \mathbb{R}^{mM\times mN} \times \mathbb{R}^{mM\times mN}$. Define the inner product $\langle \mathcal{X}, \mathcal{Y}\rangle_{\mathcal{H}_0} := \frac{1}{2}\operatorname{tr}(\boldsymbol{X}_1^\top \boldsymbol{Y}_1) + \frac{1}{2}\operatorname{tr}(\boldsymbol{X}_2^\top \boldsymbol{Y}_2)$, where $\mathcal{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2] \in \mathcal{H}_0$ and $\mathcal{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2] \in \mathcal{H}_0$ ($\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}_1, \boldsymbol{Y}_2 \in \mathbb{R}^{mM\times mN}$) and induced norm $\|\mathcal{X}\|_{\mathcal{H}_0} := \sqrt{\langle \mathcal{X}, \mathcal{X}\rangle_{\mathcal{H}_0}}$, then $\mathcal{H}_0$ becomes a real Hilbert space.

We recast the problem (8) into an unconstrained minimization of the sum of two functions f and g:

$$\underset{\mathcal{Z} \in \mathcal{H}_0}{\text{minimize}} \quad f(\mathcal{Z}) + g(\mathcal{Z}), \quad (9)$$

where

$$\begin{cases} f(\mathcal{Z}) := f_1(\boldsymbol{Z}_1) + f_2(\boldsymbol{Z}_2) = \|\boldsymbol{Z}_1\|_* + \left\|\widehat{\mathcal{L}}_\Omega(\boldsymbol{Z}_2) - \widehat{\boldsymbol{b}}\right\|_2^2, \\ g(\mathcal{Z}) := \iota_D(\mathcal{Z}) = \begin{cases} 0 & (\text{if } \mathcal{Z} \in D), \\ +\infty & (\text{otherwise}), \end{cases} \end{cases}$$

$\mathcal{Z} := [\boldsymbol{Z}_1, \boldsymbol{Z}_2] \in \mathcal{H}_0, D := \{[\boldsymbol{Z}, \boldsymbol{Z}] \in D_1 \,|\, \boldsymbol{Z}_{i,j} \geq 0, \forall (i,j) \in \mathcal{I}\} \subset D_1$, $D_1 := \{[\boldsymbol{Z}_1, \boldsymbol{Z}_2] \in D_2 \,|\, \boldsymbol{Z}_1 = \boldsymbol{Z}_2\} \subset D_2$, $D_2 = \mathfrak{S}_{\mathbb{A}_m}(M,N) \times \mathfrak{S}_{\mathbb{A}_m}(M,N) \subset \mathcal{H}_0, \mathcal{I} := \{1, \ldots, M\} \times \{1, \ldots, N\}$ and $\widehat{\mathcal{L}}_\Omega$ satisfies $\widehat{\mathcal{L}}_\Omega(\widetilde{\boldsymbol{X}}) = \widehat{\mathcal{L}_\Omega(\boldsymbol{X})}$ for all $\boldsymbol{X} \in \mathbb{A}_m^{M\times N}$. Apparently this formulation (9) is equivalent to (8), so we only have to provide the concrete calculation of the proximity operators of f and g. The proximity operator of f is given by

$$\text{prox}_{\gamma f}(\mathcal{X}) = [\text{prox}_{2\gamma f_1}(\boldsymbol{X}_1), \text{prox}_{2\gamma f_2}(\boldsymbol{X}_2)].$$

The proximity operator of f_1 with index $\tau := 2\gamma$ is given by

$$\text{prox}_{\tau f_1}(\boldsymbol{X}_1) = \text{shrink}(\boldsymbol{X}_1, \tau)$$

and

$$\left[\text{prox}_{\tau f_2}(\boldsymbol{X}_2)\right]_{i,j} = \begin{cases} \left[\frac{\tau}{\lambda\tau+1}\{\lambda \widehat{\mathcal{L}}_\Omega^*(\widehat{\boldsymbol{b}})\} + \frac{1}{\tau}\boldsymbol{X}_2\right]_{i,j} & (\text{if } (i,j) \in \Omega), \\ [\boldsymbol{X}_2]_{i,j} & (\text{otherwise}), \end{cases}$$

where $\widehat{\mathcal{L}}_\Omega^* : \mathbb{R}^{mp} \to \mathbb{R}^{mM\times mN}$ is the adjoint operator of $\widehat{\mathcal{L}}_\Omega$ satisfying $\langle \widehat{\mathcal{L}}_\Omega(\boldsymbol{X}), \boldsymbol{v}\rangle_{\mathbb{R}^{mp}} = \langle \boldsymbol{X}, \widehat{\mathcal{L}}_\Omega^*(\boldsymbol{v})\rangle_{\mathbb{R}^{mM\times mN}}$ for all $\boldsymbol{X} \in \mathbb{R}^{mM\times mN}$ and $\boldsymbol{v} \in \mathbb{R}^{mp}$. Note that the shrinkage operator $\text{shrink}(\cdot)$ is the soft-thresholding w.r.t. singular value vector.

For g, the proximity operator of the indicator function ι_{M_1} is the orthogonal projection P_D onto the subspace D, i.e.,

$$\text{prox}_{\gamma g}(\mathcal{X}) = P_D(\mathcal{X}) := \underset{\mathcal{Y} \in D}{\arg\min} \|\mathcal{X} - \mathcal{Y}\|_{\mathcal{H}_0}.$$

Since $D \subset D_1 \subset D_2 \subset \mathcal{H}_0$, we have by [28, 5.14, Reduction principle]

$$P_D(\mathcal{X}) = P_D|D_1 \circ P_{D_1}(\mathcal{X}) = P_D|D_1 \circ P_{D_1}|D_2 \circ P_{D_2}(\mathcal{X}).$$

Note that '$|D_i$' ($i = 1, 2$) stands for the restriction of the domain to the subspace D_i. The orthogonal projections $P_D|D_1 : D_1 \to D$, $P_{D_1}|D_2 : D_2 \to D_1$ and $P_{D_2} : \mathcal{H}_0 \to D_2$ respectively can be calculated as

$$\begin{aligned} P_{D_2}(\mathcal{X}) &= [P_{\mathfrak{S}}(\boldsymbol{X}_1), P_{\mathfrak{S}}(\boldsymbol{X}_2)], \\ P_{D_1}|D_2(\mathcal{X}) &= \frac{1}{2}[\boldsymbol{X}_1 + \boldsymbol{X}_2, \boldsymbol{X}_1 + \boldsymbol{X}_2], \\ P_D|D_1(\mathcal{X}) &= [\max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}_1, 0), \max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}_1, 0)], \end{aligned}$$

where $\mathfrak{S} := \mathfrak{S}_{\mathbb{A}_m}(M,N)$ and

$$[\max^{\mathbb{A}_m}(\boldsymbol{A}, 0)]_{i,j} := \sum_{\ell=1}^m \max(\Im_\ell(\boldsymbol{A}_{i,j}), 0)\mathbf{i}_\ell.$$

For $P_{\mathfrak{S}}(\boldsymbol{X}_i)$ ($i = 1, 2$), let $\boldsymbol{E}_{p,q,\ell} := \boldsymbol{E}_{p,q}\mathbf{i}_\ell \in \mathbb{A}_m^{M\times N}$ ($\ell = 1, \ldots, m$), where $\boldsymbol{E}_{p,q} \in \mathbb{R}^{M\times N}$ is the matrix only whose (p,q)-th entry ($p = 1, \ldots, M$, $q = 1, \ldots, N$) is 1 and all other entries are 0. Then, we can easily verify that

$$\langle \widetilde{\boldsymbol{E}}_{p,q,\ell}, \widetilde{\boldsymbol{E}}_{p',q',\ell'}\rangle_{\mathbb{R}^{mM\times mN}} = \begin{cases} m & (\text{if } (p,q,\ell) = (p',q',\ell')), \\ 0 & (\text{otherwise}). \end{cases}$$

and therefore, $\{\frac{1}{\sqrt{m}}\widetilde{\boldsymbol{E}}_{p,q,\ell}\}_{p=1,q=1,\ell=1}^{M,N,m}$ is an orthonormal basis of $\mathfrak{S}$ and thus $P_{\mathfrak{S}}(\boldsymbol{X}_i)$ can be easily calculated as:

$$P_{\mathfrak{S}}(\boldsymbol{X}_i) = \frac{1}{m}\sum_{p=1}^M \sum_{q=1}^N \sum_{\ell=1}^m \langle \boldsymbol{X}_i, \widetilde{\boldsymbol{E}}_{p,q,\ell}\rangle_{\mathbb{R}^{mM\times mN}} \widetilde{\boldsymbol{E}}_{p,q,\ell}.$$

Now, we can calculate

$$\begin{aligned} \text{prox}_{\gamma g}(\mathcal{X}) &= P_D|D_1 \circ P_{D_1}|D_2 \circ P_{D_2}(\mathcal{X}) \\ &= P_D|D_1 \circ P_{D_1}|D_2[P_{\mathfrak{S}}(\boldsymbol{X}_1), P_{\mathfrak{S}}(\boldsymbol{X}_2)] \\ &= P_D|D_1[\boldsymbol{X}^\star, \boldsymbol{X}^\star] \\ &= [\max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}^\star, 0), \max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}^\star, 0)], \end{aligned}$$

where $\boldsymbol{X}^\star := \frac{1}{2}\{P_{\mathfrak{S}}(\boldsymbol{X}_1) + P_{\mathfrak{S}}(\boldsymbol{X}_2)\}$. Since all ingredients are identified, we can summarize the proposed matrix completion algorithm in Algorithm 1. Here, $(t_k)_{k\geq 0} \subset [0,2]$ satisfied $\sum_{k\geq 0} t_k(2 - t_k) = +\infty$, $\gamma \in (0, +\infty)$. Note that the shrinkage operator $\text{shrink}(\cdot)$ does not keep the special structure of $\widetilde{(\cdot)}$, i.e.,

Algorithm 1: $\mathbb{A}_m$-Douglas-Rachford splitting for hypercomplex non-negative matrix completion ($\mathbb{A}_m$-DRS-NNMC)

Input : $\boldsymbol{M}, t_k, \lambda$

Output: Recovered matrix $\boldsymbol{X} \in \mathbb{A}_{m+}^{M \times N}$

$k \leftarrow 0$, $\boldsymbol{X}_i^{(0)} \leftarrow \boldsymbol{0}$ ($\forall i = 1, 2$);

while *not converged* **do**

  $\boldsymbol{X}^{\star} \leftarrow \frac{1}{2}\{P_{\mathfrak{S}}(\boldsymbol{X}_1^{(k)}) + P_{\mathfrak{S}}(\boldsymbol{X}_2^{(k)})\}$;

  $\boldsymbol{X}_+^{\star} \leftarrow \max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}^{\star}, 0)$;

  $\mathrm{prox}_{2\gamma f_1}(2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_1^{(k)}) \leftarrow \mathrm{shrink}(2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_1^{(k)}, 2\gamma)$;

  $$\left[\mathrm{prox}_{2\gamma f_0}(2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_2^{(k)})\right]_{i,j} \leftarrow \begin{cases} \left[\frac{2\gamma}{2\lambda\gamma+1}\{\lambda\widehat{\mathcal{L}}_{\Omega}^{*}(\widehat{\boldsymbol{b}}) + \frac{1}{2\gamma}(2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_2^{(k)})\}\right]_{i,j} & (\text{if } (i,j) \in \Omega) \\ [2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_2^{(k)}]_{i,j} & (\text{otherwise}) \end{cases};$$

  for $i = 1, 2$ **do**

    $\boldsymbol{X}_i^{(k+1)} \leftarrow \boldsymbol{X}_i^{(k)} + t_k \left[\mathrm{prox}_{2\gamma f_i}(2\widetilde{\boldsymbol{X}}_+^{\star} - \boldsymbol{X}_i^{(k)}) - \widetilde{\boldsymbol{X}}_+^{\star}\right]$;

  $k \leftarrow k + 1$

$\boldsymbol{X}^{\star} \leftarrow \frac{1}{2}\{P_{\mathfrak{S}}(\boldsymbol{X}_1^{(k)}) + P_{\mathfrak{S}}(\boldsymbol{X}_2^{(k)})\}$;

$\boldsymbol{X} \leftarrow \max^{\mathbb{A}_m}(\underset{\sim}{\boldsymbol{X}}^{\star}, 0)$;

$\mathrm{shrink}(\widetilde{\boldsymbol{A}}, 2\gamma) \notin \mathfrak{S}$ in general, so we need the projection onto the structure $P_{\mathfrak{S}}$. However, in complex and quaternion domain, it keeps the structure [24]. Especially if $m = 1$ (i.e., $\mathbb{A}_m = \mathbb{R}$), Algorithm 1 is identical to the dual of the non-negative matrix completion in real domain proposed in [25]. Lastly, we state the convergence of the proposed algorithm.

Theorem 1 (Convergence of $\mathbb{A}_m$-DRS-NNMC)**.** *Let parameters of Algorithm 1 be chosen so that $\gamma \in (0, +\infty)$, $(t_k)_{k \geq 0} \subset [0, 2]$ satisfying $\sum_{k \geq 0} t_k(2 - t_k) = +\infty$. Then, the output of Algorithm 1 converges to a minimizer of (8).*

Remark 1. In this paper, we employ the DRS for solving (8) but it can be also solved by other advanced convex optimization techniques such as the *alternating direction method of multipliers (ADMM)* [29] and the *primal-dual splitting (PDS)* [30, 31].

4. NUMERICAL EXAMPLES

In this section, we perform some numerical experiments for examining the effectiveness of the proposed method. Following the settings in [32, 33], we randomly generate part-wise non-negative C-D matrices as follows: $\boldsymbol{X} := \boldsymbol{X}_L \boldsymbol{X}_R^{\mathsf{H}} \in \mathbb{A}_m^{M \times N}$, where $\boldsymbol{X}_L \in \mathbb{A}_m^{M \times r}$ and $\boldsymbol{X}_R \in \mathbb{A}_m^{N \times r}$ ($r < \min(M, N)$) with all real and imaginary parts of each entry of $\boldsymbol{X}_L, \boldsymbol{X}_R$ being i.i.d. from $\mathcal{U}(0, 1)$. Only with this procedure, $\boldsymbol{X}$ is not always part-wise non-negative, so we add an absolute value of minimum negative value for each imaginary part. Note that r is not always agree to $m\mathrm{rank}(\widetilde{\boldsymbol{X}})$ (for detail, see [23]). In these experiments, we fix $r = 2$ and $\mathrm{rank}(\widetilde{\boldsymbol{X}})$ becomes 66 (of full rank $8 \times 32 = 256$). For investigating the limitation of recovery, we try various percentage ρ of the entries to be known and randomly chose the support of the known entries. The value and the locations of the known entries of $\mathcal{X}_0$ are used as inputs for the algorithms. For the parameters, we set $\lambda = 2$ and $t_k = 1$. We perform experiments in the case where $\mathbb{A}_m = \mathbb{O}$ ($m = 8$). Since hypercomplex non-negative matrix completion is itself completely a new, so we compare the proposed method $\mathbb{A}_m$-DRS-NNMC and three quaternion part-wise methods, $\mathbb{H}^2$-DRS-NNMC, $\mathbb{C}^4$-DRS-NNMC and $\mathbb{R}^4$-DRS-NNMC. These part-wise methods split $\mathbb{O}$ into $\mathbb{H}^2$, $\mathbb{C}^4$ etc. and estimate separately. Table 1 shows the performance comparisons

Table 1. Performance comparison

$\boldsymbol{X} \in \mathbb{O}^{32 \times 32}, \rho = 0.4, \mathrm{rank}(\widetilde{\boldsymbol{X}}) = 66$		
Algorithm	error	# iter.
$\mathbb{A}_m$-DRS-NNMC	3.0e-2	1,628
$\mathbb{H}^2$-DRS-NNMC	5.0e-1	1,530
$\mathbb{C}^4$-DRS-NNMC	1.1e+1	1,272
$\mathbb{R}^8$-DRS-NNMC	7.8e-1	1,155
$\boldsymbol{X} \in \mathbb{O}^{32 \times 32}, \rho = 0.1, \mathrm{rank}(\widetilde{\boldsymbol{X}}) = 66$		
Algorithm	error	# iter.
$\mathbb{A}_m$-DRS-NNMC	1.3	44,867
$\mathbb{H}^2$-DRS-NNMC	1.0e+1	43,327
$\mathbb{C}^4$-DRS-NNMC	2.6e+1	64,137
$\mathbb{R}^8$-DRS-NNMC	1.1e+1	42,340

of all four algorithms. It shows that the proposed method $\mathbb{A}_m$-DRS-NNMC outperforms part-wise methods by exploiting all correlations among real and imaginary parts for both case. In the case where $\rho = 0.4$, $\mathbb{R}^4$-DRS-NNMC outperforms $\mathbb{C}^4$-DRS-NNMC. It indicates that the performance can be worse if we consider wrong correlations. If ρ is less than 0.1, even $\mathbb{A}_m$-DRS-NNMC cannot recover the original matrix accurately so the recovery limitation is around here but it still better than the part-wise method.

5. CONCLUSIONS

In this paper, we have proposed an algorithmic solution to hypercomplex non-negative low rank matrix completion based on a proximal splitting technique. This solution utilizes both part-wise non-negativeness and special algebraic structure of Cayley-Dickson matrices. Numerical experiments show that the proposed algorithm fully utilizes the correlation among all imaginary parts and recovers much more faithfully than part-wise algorithms.

APPENDIX

Douglas-Rachford Splitting

The *Douglas-Rachford splitting (DRS)* [34, 27, 35] is a well-defined proximal splitting method that solves the minimization of the sum of two functions

$$f(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{10}$$

where f and g are assumed to be elements of the class, denoted by $\Gamma_0(\mathcal{H})$, of proper lower semicontinuous convex functions from a real Hilbert space $\mathcal{H}$ to $\mathbb{R} \cup \{+\infty\}$. For given $\gamma \in (0, +\infty)$, the DRS approximates a minimizer of (10) with $\left(\mathrm{prox}_{\gamma g}(x_k)\right)_{k \geq 0}$ by generating the following sequence $(\boldsymbol{x}_k)_{k \geq 0}$:

$$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + t_k\{\mathrm{prox}_{\gamma f}[2\,\mathrm{prox}_{\gamma g}(\boldsymbol{x}_k) - \boldsymbol{x}_k] - \mathrm{prox}_{\gamma g}(\boldsymbol{x}_k)\}, \tag{11}$$

where $(t_k)_{k \geq 0} \subset [0, 2]$ satisfies $\sum_{k \geq 0} t_k(2 - t_k) = +\infty$ and the proximity operator [36] of index γ of $f \in \Gamma_0(\mathcal{H})$ is defined as

$$\mathrm{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H} : \boldsymbol{x} \mapsto \underset{\boldsymbol{y} \in \mathcal{H}}{\arg\min} \left\{ f(\boldsymbol{y}) + \frac{1}{2\gamma} \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{H}}^2 \right\}$$

with the norm on $\mathcal{H}$ denoted by $\|\cdot\|_{\mathcal{H}}$. Indeed, if $\dim(\mathcal{H}) < \infty$, $\left(\mathrm{prox}_{\gamma g}(x_k)\right)_{k \geq 0}$ converges to a minimizer of (10) (see e.g., [37]).

REFERENCES

[1] 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.

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

[3] J. S. Yuan, "Closed-loop manipulator control using quaternion feedback," *IEEE Trans. Robot. Autom.*, vol. 4, no. 4, pp. 434–440, Aug. 1988.

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

[5] 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.

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

[7] J. Benesty, J. Chen, and Y. Huang, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 469–472, May 2010.

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

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

[10] J. C. Baez, "The octonions," *Bulletin of the American Mathematical Society*, vol. 39, no. 2, pp. 145–205, 2001.

[11] W. E. Frazier, "Metal additive manufacturing: A review," *Journal of Materials Engineering and Performance*, vol. 23, no. 6, pp. 1917–1928, Jun. 2014.

[12] T. Adalı, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process*, vol. 59, no. 11, pp. 5101–5125, 2011.

[13] S. J. Sangwine and T. A. Ell, "Hypercomplex auto- and cross-correlation of color images," in *Proc. IEEE ICIP*, 1999, pp. 319–322.

[14] N. L. Bihan and S. J. Sangwine, "Quaternion principal component analysis of color images," in *Proc. IEEE ICIP*, 2003, pp. 809–812.

[15] S. Miron, N. L. Bihan, and J. I. Mars, "Quaternion-MUSIC for vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1218–1229, Apr. 2006.

[16] T. A. Ell and S. J. Sangwine, "Hypercomplex Fourier transforms of color images," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 22–35, Jan. 2007.

[17] A. M. Grigoryan and S. S. Agaian, *Quaternion and Octonion Color Image Processing with MATLAB*. SPIE Press Monographs, 2018.

[18] L. A. Wolf, "Similarity of matrices in which the elements are real quaternions," *Bulletin of the American Mathematical Society*, vol. 42, no. 10, pp. 737–743, 1936.

[19] N. A. Wiegmann, "Some theorems on matrices with real quaternion elements," *Canadian Journal of Mathematics*, vol. 7, pp. 191–201, 1955.

[20] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra and Its Applications*, vol. 251, pp. 21–57, 1997.

[21] T. Dray and C. A. Manogue, "The octonionic eigenvalue problem," *Advances in Applied Clifford Algebras*, vol. 8, no. 2, pp. 341–364, 1998.

[22] T. Mizoguchi and I. Yamada, "An algebraic translation of Cayley-Dickson linear systems and its applications to online learning," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1438–1453, Mar. 2014.

[23] ——, "Hypercomplex tensor completion with Cayley-Dickson singular value decomposition," in *Proc. IEEE ICASSP*, Calgary, Alberta, Canada, 2018.

[24] ——, "Hypercomplex principal component pursuit via convex optimization," in *Proc. APSIPA ASC*, Honolulu, HI, USA, 2018.

[25] D. L. Sun and R. Mazumder, "Non-negative matrix completion for bandwidth extension: A convex optimization approach," in *Proc. IEEE MLSP*, Southampton, UK, 2018.

[26] X. Han, J. Wu, L. Yan, L. Senhadji, and H. Shu, "Color image recovery via quaternion matrix completion," in *Proc. 6th International Congress on Image and Signal Processing (CISP)*, Hangzhou, China, 2013.

[27] P. L. Combettes and J. C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[28] F. R. Deutsch, *Best Approximation in Inner Product Spaces*. Springer-Verlag, New York, 2001.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 2, pp. 1–122, 2011.

[30] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[31] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

[32] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," UIUC Technical Report UILU-ENG-09-2214, Tech. Rep., Jul. 2009.

[33] S. Gandy and I. Yamada, "Convex optimization techniques for the efficient recovery of a sparsely corrupted low-rank matrix," *Journal of Math-for-Industry*, vol. 2, pp. 147–156, 2010.

[34] J. Douglas and H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Trans. of the American Mathematical Society*, vol. 82, pp. 421–439, 1956.

[35] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, pp. 293–318, Apr. 1992.

[36] J. J. Moreau, "Fonctions convexes duales et points proximaux dans unespace hilbertien," *Comptes Rendus de l'Académie des Sciences (Paris), Série A*, vol. 255, pp. 2897–2899, 1962.

[37] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Baischke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer, 2011, ch. 17, pp. 345–390.