# Hypercomplex Principal Component Pursuit via Convex Optimization 

Takehiko Mizoguchi* ${ }^{\dagger}$ and Isao Yamada*<br>${ }^{*}$ Department of Information and Communications Engineering, Tokyo Institute of Technology, Tokyo, Japan<br>${ }^{\dagger}$ NEC Corporation, Kanagawa, Japan<br>E-mail: \{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 signal processing applications, to enjoy their nontrivial algebraic benefits which are not available in standard real or complex vector systems. Exploiting sparsity of hypercomplex matrices or vectors would play an important role in utilizing such benefits. In this paper, we propose a new sparsity measure for evaluating sparsity of hypercomplex matrices. With utilizing this measure, the hypercomplex robust principal component analysis can be relaxed into the hypercomplex robust principal component pursuit and it can be reduced to a real convex optimization problem. We then derive an algorithmic solution to the hypercomplex principal component pursuit based on a proximal splitting technique. Numerical experiments are performed in the octonion domain and show that the proposed algorithm outperforms a part-wise state-of-art real, complex and quaternion algorithm.


## I. 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]-[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 operation 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 modelingof various correlations (see e.g., [12]-[16] for color image processing applications). However, because of the "singularity" of higher dimensional C-D number systems (see e.g., Example 1), few mathematical tools had been maintained [17]-[20]. To overcome this situation, in our previous works [21]-[23], we have proposed several useful mathematical tools for designing advanced algorithm for optimization, learning and low rank approximation in hypercomplex domain. In [21] 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, we also proposed in [22], [23] useful tools C-D singular value decomposition, $\mathbb{R}$-rank (see Section II-C) and low rank approximation technique and proposed an algorithmic solution for low rank hypercomplex tensor completion.

One of the other standard approximation methods utilizing simpler structure of matrices is the robust principal component analysis (RPCA) [24], which separates an input matrix into a low-rank and sparse ones:

$$
\begin{equation*}
\underset{\boldsymbol{L}, \boldsymbol{S} \in \mathbb{R}^{M \times N}}{\operatorname{minimize}} \quad \operatorname{rank}(\boldsymbol{L})+\lambda\|\boldsymbol{S}\|_{0} \quad \text { s.t. } \quad \boldsymbol{M}=\boldsymbol{L}+\boldsymbol{S} \tag{1}
\end{equation*}
$$

where $\|\cdot\|_{0}$ denotes the $\ell_{0}$-norm, which counts the number of non-zero entries of $\boldsymbol{S}$. The RPCA has been successfully used for various signal processing applications such as source separation [25], face recognition [26] and so on. Unfortunately, the formulation (1) is NP-hard, so its convex relaxation called the principal component pursuit [24] is often considered in stead of the RPCA. Recently, the PCP is extended to simpler hypercomplex domains $\mathbb{C}$ and $\mathbb{H}$ using the complex matrix isomorphism [27]. However, it can be only applied up to four dimensional data and hard to be generalized to octonion and higher dimensional $\mathrm{C}-\mathrm{D}$ domain only with this isomorphism.

In this paper, to establish a RPCA framework in hypercomplex domain, first, we present Cayley-Dickson principal component analysis (C-D PCP) as a convex relaxation of the C-D extension of RPCA. This relaxation is based on the
$\mathbb{R}$-rank proposed in [22] and a new sparsity measure, $\ell_{1}$ norm of C-D matrices. This sparsity measure can be interpreted as evaluating a group sparsity of real matrices, so the proximity operator can be easily calculated. Hence, the C-D PCP is a convex optimization in real domain which can be solved by applying proximal splitting techniques to a certain structured convex optimization problem. We finally propose an hypercomplex PCP algorithm $\mathbb{A}_{m}$ based on DouglasRachford splitting (DRS) [28]. The proposed algorithm is a C-D generalization of a PCP algorithm (DR-PCP) proposed in [29] and can be applied to general C-D domains.

Numerical experiments are performed in the context of recovering sparsely corrupted low rank matrices in octonion domain and show that the proposed algorithm successfully utilizes algebraically natural correlations of real and all imaginary parts to recover much more faithfully the original matrices, corrupted randomly by noise, than a part-wise real, complex and quaternion PCP algorithms.

## II. Preliminaries

## A. 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 $\mathbb{A}_{m}(m \in \mathbb{N} \backslash\{0\})$ expanded on the real vector space [8]

$$
\begin{equation*}
a:=a_{1} \mathbf{i}_{1}+a_{2} \mathbf{i}_{2}+\cdots+a_{m} \mathbf{i}_{m} \in \mathbb{A}_{m}, a_{1}, \ldots, a_{m} \in \mathbb{R} \tag{2}
\end{equation*}
$$

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 (2). The coefficient of each imaginary unit $a_{\ell}(\ell=1, \ldots, m)$ is 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 $\mathbb{A}_{2}(=: \mathbb{C})$ ). The addition and the subtraction of two hypercomplex numbers are defined as commutative binomial operations

$$
a \pm b:=\left(a_{1} \pm b_{1}\right) \mathbf{i}_{1}+\left(a_{2} \pm b_{2}\right) \mathbf{i}_{2}+\cdots+\left(a_{m} \pm b_{m}\right) \mathbf{i}_{m}
$$

for $a, b \in \mathbb{A}_{m}$, where $b:=b_{1} \mathbf{i}_{1}+b_{2} \mathbf{i}_{2}+\cdots+b_{m} \mathbf{i}_{m}$, $b_{1}, \ldots, b_{m} \in \mathbb{R}$. From the unique expression of (2), the multiplication of two hypercomplex numbers

$$
\begin{aligned}
a b & =\left(a_{1} \mathbf{i}_{1}+a_{2} \mathbf{i}_{2}+\cdots+a_{m} \mathbf{i}_{m}\right)\left(b_{1} \mathbf{i}_{1}+b_{2} \mathbf{i}_{2}+\cdots+b_{m} \mathbf{i}_{m}\right) \\
& :=\sum_{k=1}^{m} \sum_{\ell=1}^{m} a_{k} b_{\ell} \mathbf{i}_{k} \mathbf{i}_{\ell} \in \mathbb{A}_{m}
\end{aligned}
$$

is determined uniquely according to the multiplication table. We also define the conjugate of hypercomplex number $a$ as

$$
\begin{equation*}
a^{*}:=a_{1} \mathbf{i}_{1}-a_{2} \mathbf{i}_{2}-\cdots-a_{m} \mathbf{i}_{m} \tag{3}
\end{equation*}
$$

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 $\mathrm{C}-\mathrm{D}$ 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 $\mathbb{A}_{2 m}$ [8], [9] as

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

where $\mathbf{i}_{m+1} \notin \mathbb{A}_{m}$ 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 $\left(\mathbb{A}_{1}:=\right) \mathbb{R}$ is extended into complex number system $\mathbb{C}\left(=\mathbb{A}_{2}\right)$ 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 [21] 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 $a a^{*}=\sum_{\ell=1}^{m} a_{\ell}^{2} \geq 0$ for any $a \in \mathbb{A}_{m}$ in (2) and $a^{*} \in \mathbb{A}_{m}$ in (3) and enable us to define the absolute values of C-D number $a$ as $|a|:=\sqrt{a a^{*}}$.

Example 1. 1) 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:

$$
\begin{gather*}
\imath \jmath=-\jmath \imath=\kappa, \jmath \kappa=-\kappa \jmath=\imath, \kappa \imath=-\imath \kappa=\jmath, \\
\imath^{2}=\jmath^{2}=\kappa^{2}=-1 \tag{4}
\end{gather*}
$$

by letting $m=4, \mathbf{i}_{1}=1, \mathbf{i}_{2}=\imath, \mathbf{i}_{3}=\jmath$ and $\mathbf{i}_{4}=\kappa$. From (4), quaternions are not commutative, i.e., $p q \neq q p$ for $p, q \in \mathbb{H}$ in general.
2) 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., $p q \neq q p$ and $(p q) r \neq p(q r)$ for $p, q, r \in \mathbb{O}$ in general [10]. For the octonion multiplication table, see, e.g., [10].

The C-D number system can be seen as an algebraically natural higher dimensional generalization of our familiar fields, i.e., $\mathbb{R}$ and $\mathbb{C}$.

We also define $\mathbb{A}_{m}^{N}:=\left\{\left[x_{1}, \ldots, x_{N}\right]^{\top} \mid x_{i} \in \mathbb{A}_{m}(i=\right.$ $1, \ldots, N)\}$ for $\forall N \in \mathbb{N} \backslash\{0\}$, where $(\cdot)^{\top}$ stands for the transpose. Define $\langle\boldsymbol{x}, \boldsymbol{y}\rangle_{\mathbb{A}_{m}^{N}}:=\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y} \in \mathbb{A}_{m}, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{A}_{m}^{N}$ 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 $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose of vectors or matrices (e.g., $\boldsymbol{x}^{\mathrm{H}}:=\left[x_{1}^{*}, \ldots, x_{N}^{*}\right]$ for $\boldsymbol{x}:=\left[x_{1}, \ldots, x_{N}\right]^{\top} \in \mathbb{A}_{m}^{N}$, where $x_{1} \ldots, x_{N} \in \mathbb{A}_{m}$ ). We also define the addition of two hypercomplex vectors $\boldsymbol{x}+\boldsymbol{y}:=\left[x_{1}+y_{1}, \cdots, x_{N}+y_{N}\right]^{\top} \in \mathbb{A}_{m}^{N}$ for $\boldsymbol{x}, \boldsymbol{y}\left(:=\left[y_{1}, \ldots, y_{N}\right]^{\top}\right) \in \mathbb{A}_{m}^{N}$. 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 $\alpha \boldsymbol{x}:=\left[\alpha x_{1}, \ldots, \alpha x_{N}\right]^{\top} \in \mathbb{A}_{m}^{N}$ 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.

## B. Algebraic Translations

In this section, we introduce algebraic translation of C-D valued vectors and matrices proposed in [21]. A trivial correspondence (mapping) of hypercomplex vectors or matrices to real ones is

$$
\widehat{(\cdot)}: \mathbb{A}_{m}^{M \times N} \rightarrow \mathbb{R}^{m M \times N}: \boldsymbol{A} \mapsto \widehat{\boldsymbol{A}}:=\left[\begin{array}{c}
\boldsymbol{A}_{1}  \tag{5}\\
\vdots \\
\boldsymbol{A}_{m}
\end{array}\right]
$$

where $\boldsymbol{A}=\boldsymbol{A}_{1} \mathbf{i}_{1}+\cdots+\boldsymbol{A}_{m} \mathbf{i}_{m} \in \mathbb{A}_{m}^{M \times N}$ and $\boldsymbol{A}_{1} \ldots, \boldsymbol{A}_{m} \in$ $\mathbb{R}^{M \times N}$. This correspondence is just concatenating a real and all imaginary parts in the hypercomplex matrices. Obviously, this mapping is invertible and thus we can also define

$$
\begin{equation*}
\widetilde{(\cdot)}: \mathbb{R}^{m M \times N} \rightarrow \mathbb{A}_{m}^{M \times N}: \widehat{\boldsymbol{A}} \mapsto \boldsymbol{A} . \tag{6}
\end{equation*}
$$

Only in terms of the mappings $\widehat{(\cdot)}$ and $\widetilde{(\cdot)}$, it is hard to obtain the correspondence of matrix-vector product $\boldsymbol{A x}$, so we also introduce the following non-trivial mapping:

$$
\begin{align*}
\widetilde{(\cdot)}: \mathbb{A}_{m}^{M \times N} \rightarrow & \mathfrak{S}_{\mathbb{A}_{m}}(M, N): \\
& \boldsymbol{A} \mapsto \widetilde{\boldsymbol{A}}:=\left[\boldsymbol{L}_{M}^{(1) \top} \widehat{\boldsymbol{A}}, \ldots, \boldsymbol{L}_{M}^{(m) \top} \widehat{\boldsymbol{A}}\right], \tag{7}
\end{align*}
$$

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

$$
\boldsymbol{L}_{M}^{(\ell)}=\left[\begin{array}{cccc}
\delta_{1,1}^{(\ell)} \boldsymbol{I}_{M} & \delta_{1,2}^{(\ell)} \boldsymbol{I}_{M} & \cdots & \delta_{1, m}^{(\ell)} \boldsymbol{I}_{M} \\
-\delta_{2,1}^{(\ell)} \boldsymbol{I}_{M} & -\delta_{2,2}^{(\ell)} \boldsymbol{I}_{M} & \cdots & -\delta_{2, m}^{(\ell)} \boldsymbol{I}_{M} \\
\vdots & \vdots & \ddots & \vdots \\
-\delta_{m, 1}^{(\ell)} \boldsymbol{I}_{M} & -\delta_{m, 2}^{(\ell)} \boldsymbol{I}_{M} & \cdots & -\delta_{m, m}^{(\ell)} \boldsymbol{I}_{M}
\end{array}\right],
$$

with the $M$-dimensional identity matrix $\boldsymbol{I}_{M}$ and

$$
\delta_{\alpha, \beta}^{(\gamma)}:=\left\{\begin{aligned}
1 & \left(\text { if } \mathbf{i}_{\alpha} \mathbf{i}_{\beta}=\mathbf{i}_{\gamma}\right) \\
-1 & \text { (if } \left.\mathbf{i}_{\alpha} \mathbf{i}_{\beta}=-\mathbf{i}_{\gamma}\right) \\
0 & \text { (otherwise) }
\end{aligned}\right.
$$

$\mathfrak{S}_{\mathbb{A}_{m}}(M, N) \subset \mathbb{R}^{m M \times m N}$ is the restriction of $\widetilde{\boldsymbol{A}}$ and represented as

$$
\begin{align*}
\mathfrak{S}_{\mathbb{A}_{m}}(M, N) & :=\left\{\widetilde{\boldsymbol{A}} \mid \boldsymbol{A} \in \mathbb{A}_{m}^{M \times N}\right\} \subset \mathbb{R}^{m M \times m N} \\
& =\left\{\left[\boldsymbol{L}_{M}^{(1) \top} \boldsymbol{A}^{\prime}, \ldots, \boldsymbol{L}_{M}^{(m) \top} \boldsymbol{A}^{\prime}\right] \mid \boldsymbol{A}^{\prime} \in \mathbb{R}^{m M \times N}\right\} . \tag{8}
\end{align*}
$$

Using the imaginary unit vector $\boldsymbol{i}:=\left[\mathbf{i}_{1}, \mathbf{i}_{2}, \ldots, \mathbf{i}_{m}\right]^{\top} \in \mathbb{A}_{m}^{m}$, $\boldsymbol{L}_{M}^{(\ell)}$ is also compactly represented as

$$
\boldsymbol{L}_{M}^{(\ell)}=\Im_{\ell}\left(i i^{\mathrm{H}} \otimes \boldsymbol{I}_{M}\right)
$$

where ' $\otimes$ ' is the Kronecker product. Similar to the trivial mapping, $(\cdot)$ is also invertible and thus we can define

$$
(\cdot): \mathfrak{S}_{\mathbb{A}_{m}}(M, N) \rightarrow \mathbb{A}_{m}^{M \times N}: \widetilde{\boldsymbol{A}} \mapsto \boldsymbol{A} .
$$

These mappings have the following useful algebraic properties:

Fact 1 (Algebraic correspondence between real and C-D vectors and matrices [21]). For all $\boldsymbol{A}, \boldsymbol{A}^{\prime} \in \mathbb{A}_{m}^{M \times N}, \boldsymbol{B} \in \mathbb{A}_{m}^{N \times L}$ and $\boldsymbol{x} \in \mathbb{A}_{m}^{N}$,

1) $\left(\widehat{\boldsymbol{A + A}}{ }^{\prime}\right)=\widehat{\boldsymbol{A}}+{\widehat{\boldsymbol{A}^{\prime}}}^{\prime}(\widehat{\alpha \boldsymbol{A}})=\alpha \widehat{\boldsymbol{A}}$,

$$
\left(\widetilde{\boldsymbol{A}+\boldsymbol{A}^{\prime}}\right)=\widetilde{\boldsymbol{A}}+\widetilde{\boldsymbol{A}^{\prime}},(\widetilde{\alpha \boldsymbol{A})}=\alpha \widetilde{\boldsymbol{A}} \text { for all } \alpha \in \mathbb{R}
$$

2) $\left(\widetilde{\boldsymbol{A}^{\mathrm{H}}}\right)=\widetilde{\boldsymbol{A}}^{\top}$,
3) $\|\boldsymbol{x}\|_{\mathbb{A}_{m}^{N}}=\|\widehat{\boldsymbol{x}}\|_{\mathbb{R}^{m N}}$,
4) $(\widehat{\boldsymbol{A B}})=\underset{\widetilde{\boldsymbol{A}}}{\widetilde{\boldsymbol{B}}}$ and $(\widehat{\boldsymbol{A x}})=\widetilde{\boldsymbol{A}} \widehat{\boldsymbol{x}}$,
5) $(\widetilde{\boldsymbol{A B}})=\widetilde{\boldsymbol{A}} \widetilde{\boldsymbol{B}}$ if $m \leq 4$, i.e., if $\mathbb{A}_{m}=\mathbb{R}$ or $\mathbb{C}$ or $\mathbb{H}$.

Example 2. For a quaternion matrix $\boldsymbol{A}:=\boldsymbol{A}_{1}+\boldsymbol{A}_{2} \imath+\boldsymbol{A}_{3} \jmath+$ $\boldsymbol{A}_{4} \kappa \in \mathbb{H}^{M \times N}, \widehat{\boldsymbol{A}} \in \mathbb{R}^{4 M \times N}$ and $\widetilde{\boldsymbol{A}} \in \mathbb{R}^{4 M \times 4 N}$ are given as

$$
\widehat{\boldsymbol{A}}=\left[\begin{array}{l}
\boldsymbol{A}_{1} \\
\boldsymbol{A}_{2} \\
\boldsymbol{A}_{3} \\
\boldsymbol{A}_{4}
\end{array}\right], \widetilde{\boldsymbol{A}}=\left[\begin{array}{cccc}
\boldsymbol{A}_{1} & -\boldsymbol{A}_{2} & -\boldsymbol{A}_{3} & -\boldsymbol{A}_{4} \\
\boldsymbol{A}_{2} & \boldsymbol{A}_{1} & -\boldsymbol{A}_{4} & \boldsymbol{A}_{3} \\
\boldsymbol{A}_{3} & \boldsymbol{A}_{4} & \boldsymbol{A}_{1} & -\boldsymbol{A}_{2} \\
\boldsymbol{A}_{4} & -\boldsymbol{A}_{3} & \boldsymbol{A}_{2} & \boldsymbol{A}_{1}
\end{array}\right] .
$$

Remark 1. Obviously, $\mathfrak{S}_{\mathbb{A}_{m}}(M, N)$ is an $m M N$-dimensional real vector space and the non-trivial mapping $\widetilde{(\cdot)}$ is guaranteed to be an isomorphism between $\mathbb{A}_{m}^{M \times N}$ and $\mathfrak{S}_{\mathbb{A}_{m}}(M, N)$ regarding $\mathbb{A}_{m}^{M \times N}$ as a vector space over $\mathbb{R}$ (see Section II-A).

## C. Singular Value Decomposition, Rank, and Low Rank Approximation

In this section, we introduce useful notions $C-D$ singular value decomposition ( $C$-D SVD) and $\mathbb{R}$-rank for C-D matrices originally proposed in [22]. For any C-D matrix $\boldsymbol{A} \in \mathbb{A}_{m}^{M \times N}$, there exist orthogonal real matrices $\boldsymbol{U} \in \mathbb{R}^{m M \times m M}$ and $\boldsymbol{V} \in$ $\mathbb{R}^{m N \times m N}$ such that

$$
\begin{equation*}
\widetilde{\boldsymbol{A}}=\boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^{\top} \tag{9}
\end{equation*}
$$

where $\boldsymbol{\Sigma}:=\operatorname{diag}\left(\sigma_{1}, \ldots, \sigma_{r}, 0, \ldots, 0\right) \in \mathbb{R}^{m M \times m N}$ is a rectangular diagonal matrix with positive singular values $\sigma_{1} \geq \cdots \geq \sigma_{r}(>0)$ of $\widetilde{\boldsymbol{A}}$ on the diagonal We call it $C$-D singular value decomposition ( $C-D$ SVD) and call $\operatorname{rank}^{\mathbb{R}}(\boldsymbol{A}):=r=\operatorname{rank}(\widetilde{\boldsymbol{A}}) \leq \max (m M, m N) \mathbb{R}$-rank of $\boldsymbol{A}$. The $\mathbb{R}$-rank has very strong relation to well-established original ranks [19] in C-D domain.

Fact 2 ( $\mathbb{R}$-rank and original ranks in C-D domain [22]). For complex ( $m=2$ ) or quaternion ( $m=4$ ) cases,

$$
\operatorname{rank}^{\mathbb{R}}(\boldsymbol{A})=\operatorname{mrank}(\boldsymbol{A})
$$

holds for all $\boldsymbol{A} \in \mathbb{A}_{m}^{M \times N}$.
Originally, the rank of matrices is defined as the maximum number of column vectors of them which are linearly independent. In quaternion and higher dimensional C-D domain, we can consider left and right linearly independence since left and right scalar multiplications are distinct, so we have to define carefully the rank over $\mathbb{A}_{m}$. In quaternion domain, by convention, the rank is defined as the maximum number of columns which are right linearly independent [19], since if we define so, the rank becomes equal to the number of positive singular values with the right eigenvalues. For octonion and higher dimensional C-D domain, the concrete definition of the
rank has not yet been established to the best of the author's knowledge. However, the $\mathbb{R}$-rank is explicitly available for general C-D matrices keeping the consistency with the known results (e.g., [19]) from Fact 2. Therefore, the $\mathbb{R}$-rank is a natural generalization of the rank to hypercomplex domain.

Finally, we discuss a low rank approximation of matrices in C-D domain. By passing through the Schmidt-Eckart-Young theorem [30] we can also perform a low rank approximation of C-D matrices in terms of $\mathbb{R}$-rank minimization with truncating C-D SVD. Note that this approximation does not always provide a low $\mathbb{R}$-rank matrix in $\mathfrak{S}_{\mathbb{A}_{m}}(M, N)$, i.e., the existence the corresponding C-D matrix of the low $\mathbb{R}$ rank approximation is not guaranteed (for details, see [22]). However, there are several cases where the corresponding CD matrices of low $\mathbb{R}$-rank approximation is guaranteed. The following fact is one of the examples (another examples can be seen in [22]).

Fact 3 (Inheritance of special structure of non-trivial mapping with the shrinkage operator). Let $\operatorname{shrink}(\widetilde{\boldsymbol{A}}, \tau)$ be the shrinkage operator given by

$$
\begin{equation*}
\operatorname{shrink}(\widetilde{\boldsymbol{A}}, \tau)=\boldsymbol{U} \boldsymbol{\Sigma}_{\tau} \boldsymbol{V}^{\top} \tag{10}
\end{equation*}
$$

for (9) and the shrunk diagonal matrix $\boldsymbol{\Sigma}_{\tau}:=\operatorname{diag}\left(\max \left\{\sigma_{1}-\right.\right.$ $\left.\tau, 0\}, \ldots, \max \left\{\sigma_{r}-\tau, 0\right\}, 0, \ldots, 0\right)$. For any $\boldsymbol{A} \in \mathbb{A}_{m}^{M \times N}$ and $\tau>0$,

1) If $m \leq 4$, the shrinkage operator keeps the special structure of non-trivial mapping $\widetilde{(\cdot)}$, i.e., $\operatorname{shrink}(\widetilde{\boldsymbol{A}}, \tau) \in$ $\mathfrak{S}_{\mathbb{A}_{m}}(M, N)$.
2) If $m>4$, the shrinkage operator does not always keep this structure, i.e., $\operatorname{shrink}(\widetilde{\boldsymbol{A}}, \tau) \notin \mathfrak{S}_{\mathbb{A}_{m}}(M, N)$ in general.

The shrinkage operator will be used in Section III-B and it often appears as the proximity operator (see Appendix A) of nuclear norm regularization in many context of matrix low rank approximation via proximal splitting methods including the principal component pursuit [24] and low rank matrix (or tensor) completion [31], [32].

## III. Hypercomplex Principal Component Pursuit

## A. Convex Relaxation of Hypercomplex Robust Principal Component Analysis

In this section, we formulate the robust principal component analysis (RPCA) in C-D domain. Since the $\mathbb{R}$-rank defined in II-C is available for general C-D domain, we can formulate it in C-D domain as follows:

$$
\begin{equation*}
\underset{\boldsymbol{L}, \boldsymbol{S} \in \mathbb{A}_{m}^{M \times N}}{\operatorname{minimize}} \quad \operatorname{rank}^{\mathbb{R}}(\boldsymbol{L})+\lambda\|\boldsymbol{S}\|_{0, \mathbb{A}_{m}} \quad \text { s.t. } \quad \boldsymbol{M}=\boldsymbol{L}+\boldsymbol{S}, \tag{11}
\end{equation*}
$$

where $\lambda>0$ and $\|\cdot\|_{0, \mathbb{A}_{m}}$ is the number of non-zero entries in $\boldsymbol{A} \in \mathbb{A}^{M \times N}$. Obviously from Fact 2 , this is a C-D generalization of RPCA in real domain. Similar to the real case [24], (11) is NP-hard. For relaxing (11) to a convex optimization problem, we first introduce newly the $\ell_{1}$-norm
of a C-D matrix as follows:

$$
\begin{equation*}
\|\boldsymbol{A}\|_{1, \mathbb{A}_{m}}:=\sum_{i, j=1}^{M, N}\left|\boldsymbol{A}_{i, j}\right|, \quad \boldsymbol{A} \in \mathbb{A}_{m}^{M \times N} \tag{12}
\end{equation*}
$$

For any C-D matrix $\boldsymbol{A} \in \mathbb{A}_{m}^{M \times N}$, we can consider the following real matrix:

$$
\widehat{\boldsymbol{A}}=\left[\begin{array}{ccc}
\widehat{\boldsymbol{A}}_{1,1} & \cdots & \widehat{\boldsymbol{A}}_{1, N}  \tag{13}\\
\vdots & \ddots & \vdots \\
\widehat{\boldsymbol{A}}_{M, 1} & \cdots & \widehat{\boldsymbol{A}}_{M, N}
\end{array}\right] \in \mathbb{R}^{m M \times N}
$$

with the mapping $\widehat{(\cdot)}: \mathbb{A}_{m}^{M \times N} \rightarrow \mathbb{R}^{m M \times N}$. Note that $\widehat{(\cdot)}$ is just a permutation of $\widehat{(\cdot)}$ in (5) and we can define its inverse $\widehat{(\cdot)}: \mathbb{R}^{m M \times N} \rightarrow \mathbb{A}_{m}^{M \times N}: \widehat{\boldsymbol{A}} \mapsto \boldsymbol{A}$. Then, we have

$$
\|\boldsymbol{A}\|_{1, \mathbb{A}_{m}}=\sum_{i, j=1}^{M, N}\left\|\widehat{\boldsymbol{A}}_{i, j}\right\|_{2}=:\|\widehat{\boldsymbol{A}}\|_{1}^{\mathbb{A}_{m}}
$$

where $\|\cdot\|_{2}$ is the $\ell_{2}$-norm of real vectors and note that $\widehat{\boldsymbol{A}}_{i, j} \in$ $\mathbb{R}^{m}(i=1, \ldots, M, j=1, \ldots, N)$. This implies that the $\ell_{1}-$ norm of a C-D matrix $\boldsymbol{A}$ can be regarded as a convex function $\|\cdot\|_{1}^{\mathbb{A}_{m}}$ of the real matrix $\widehat{\boldsymbol{A}} \in \mathbb{R}^{m M \times N}$ and it evaluates the group sparsity (structured sparsity) [33] of $\boldsymbol{A}^{\prime}$. Therefore, we have the following theorem:
Theorem 1 (Proximity operator of $\|\cdot\|_{1}^{\mathbb{A}_{m}}$ ). For any $\boldsymbol{A} \in$ $\mathbb{A}_{m}^{M \times N}$, the proximity operator (see Appendix $A$ ) of $\|\cdot\|_{1}^{\mathbb{A}_{m}}$ with index $\tau>0$ can be easily calculated group-wise as

$$
\begin{align*}
{\left[\operatorname{prox}_{\gamma\|\cdot\|_{1}^{\hat{A}_{m}}}(\widehat{\widehat{\boldsymbol{A}}})\right]_{i, j} } & =\frac{\widehat{\boldsymbol{A}}_{i, j}}{\left\|\widehat{\boldsymbol{A}}_{i, j}\right\|_{2}} \max \left(0,\left\|\widehat{\boldsymbol{A}}_{i, j}\right\|_{2}-\tau\right)  \tag{14}\\
& =: \widehat{\widehat{\operatorname{ST}}(\widehat{\widehat{\boldsymbol{A}}}, \tau)]_{i, j}} \tag{15}
\end{align*}
$$

where the indices $[\cdot]_{i, j}(i=1, \ldots, M, j=1, \ldots, N)$ stand for the $(i, j)$-th group of size $m \times 1$ in $\operatorname{prox}_{\gamma\|\cdot\|_{1}^{A_{m}}}\left(\boldsymbol{A}^{\prime}\right) \in \mathbb{R}^{m M \times N}$.

If we note that $\left\|\widehat{\boldsymbol{A}}_{i, j}\right\|_{2}=\left|\boldsymbol{A}_{i, j}\right|$ and by applying $\widetilde{(\cdot)}$ defined in (6) to the right hand side of (14), we formally obtain the following entry-wise soft-thresholding function of C-D matrices:

$$
[\operatorname{ST}(\boldsymbol{A}, \tau)]_{i, j}:=\frac{\boldsymbol{A}_{i, j}}{\left|\boldsymbol{A}_{i, j}\right|} \max \left(0,\left|\boldsymbol{A}_{i, j}\right|-\tau\right)
$$

This is obviously equivalent to (15) and a C-D generalization of real, complex and quaternion soft-thresholding functions.
By using $\ell_{1}$ norm we have discussed above and approximating $\mathbb{R}$-rank with nuclear norm, we have the following convex optimization problem:

$$
\begin{equation*}
\underset{\boldsymbol{L}, \boldsymbol{S} \in \mathbb{A}_{m}^{M \times N}}{\operatorname{minimize}}\|\widetilde{\boldsymbol{L}}\|_{*}+\lambda\|\boldsymbol{S}\|_{1, \mathbb{A}_{m}} \quad \text { s.t. } \quad \boldsymbol{M}=\boldsymbol{L}+\boldsymbol{S}, \tag{16}
\end{equation*}
$$

where $\|\cdot\|_{*}$ is the nuclear norm of real matrices i.e., the sum of positive singular values. In this paper, we call the problem (16) Cayley-Dickson principal component pursuit ( $C$ -
$D P C P)$. Obviously, if we set $\mathbb{A}_{m}=\mathbb{R}(m=1)$ and $\mathbb{A}_{m}=\mathbb{C}$ or $\mathbb{H}\left(\mathbb{A}_{m}=2\right.$ or 4$)$, (16) is respectively identical to the original PCP in real domain [24], and the complex and quaternionic PCP in [27]. Therefore, the C-D PCP is a natural generalization of these problems. Moreover, the C-D PCP can be regarded as a convex optimization problem in real domain since the $\ell_{1}$-norm of C-D matrices $\boldsymbol{A}$ can be regarded as a convex function of real matrices, and can be solved by proximal splitting techniques.

## B. Hypercomplex Principal Component Pursuit via Convex Optimization

In this section, we derive a new algorithm based on the Douglas-Rachford splitting technique [28] to solve the C-D PCP (16) efficiently. Denote the 2 -fold Cartesian product of the spaces of real matrices by $\mathcal{H}_{0}:=\mathbb{R}^{m M \times m N} \times \mathbb{R}^{m M \times N}$. By defining the inner product $\langle\mathcal{X}, \mathcal{Y}\rangle_{\mathcal{H}_{0}}:=\frac{1}{2} \operatorname{tr}\left(\boldsymbol{X}_{1}^{\top} \boldsymbol{Y}_{1}\right)+$ $\frac{1}{2} \operatorname{tr}\left(\boldsymbol{X}_{2}^{\top} \boldsymbol{Y}_{2}\right)=: \frac{1}{2}\left\langle\boldsymbol{X}_{1}, \boldsymbol{Y}_{1}\right\rangle_{\mathbb{R}^{m M \times m N}}+\frac{1}{2}\left\langle\boldsymbol{X}_{2}, \boldsymbol{Y}_{2}\right\rangle_{\mathbb{R}^{m M \times N}}$, where $\mathcal{X}:=\left[\boldsymbol{X}_{1}, \boldsymbol{X}_{2}\right] \in \mathcal{H}_{0}$ and $\mathcal{Y}:=\left[\boldsymbol{Y}_{1}, \boldsymbol{Y}_{2}\right] \in \mathcal{H}_{0}$, $\left(\boldsymbol{X}_{1}, \boldsymbol{Y}_{1} \in \mathbb{R}^{m M \times m N}, \boldsymbol{X}_{2}, \boldsymbol{Y}_{2} \in \mathbb{R}^{m M \times N}\right)$ and induced norm $\|\mathcal{X}\|_{\mathcal{H}_{0}}:=\sqrt{\langle\mathcal{X}, \mathcal{X}\rangle_{\mathcal{H}_{0}}}, \mathcal{H}_{0}$ becomes a real Hilbert space. First, we reformulate the problem (16) as an unconstrained the sum of two functions as follows:

$$
\begin{equation*}
\underset{\mathcal{Z} \in \mathcal{H}_{0}}{\operatorname{minimize}} \quad f(\mathcal{Z})+g(\mathcal{Z}) \tag{17}
\end{equation*}
$$

where

$$
\begin{aligned}
& \left\{\begin{array}{l}
f(\mathcal{Z}):=f_{1}\left(\boldsymbol{Z}_{1}\right)+f_{2}\left(\boldsymbol{Z}_{2}\right)=\left\|\boldsymbol{Z}_{1}\right\|_{*}+\left\|\boldsymbol{Z}_{2}\right\|_{1}^{\mathbb{A}_{m}}, \\
g(\mathcal{Z}):=\iota_{D_{1}}(\mathcal{Z})= \begin{cases}0 & \text { (if } \left.\mathcal{Z} \in D_{1}\right), \\
+\infty & \text { (otherwise) },\end{cases}
\end{array}\right. \\
& \mathcal{Z}:=\left[\boldsymbol{Z}_{1}, \boldsymbol{Z}_{2}\right] \in \mathcal{H}_{0}, \\
& D_{1}:=\left\{\left[\boldsymbol{Z}_{1}, \boldsymbol{Z}_{1}\right] \in D_{2} \mid \boldsymbol{M}=\underline{\boldsymbol{Z}}_{1}+\check{\check{\boldsymbol{Z}}}_{2}\right\} \subset D_{2} \text {, } \\
& D_{2}:=\mathfrak{S} \times \mathbb{R}^{m M \times N} \subset \mathcal{H}_{0} \text {, } \\
& \mathfrak{S}:=\mathfrak{S}_{\mathbb{A}_{m}}(M, N) \subset \mathbb{R}^{m M \times m N} .
\end{aligned}
$$

Note that the subspace $D_{1}$ represents the constraint that the observation $\boldsymbol{M}$ is from the sum of low rank and sparse matrices. This requests that both $Z_{1}$ belong to $S$, so we need the subspace $D_{2}$.

Apparently this reformulation (17) is equivalent to (16), so all we need is to identify the concrete calculation of the proximity operators of $f$ and $g$. In the same way as [29], the proximity operator of $f$ is given by

$$
\operatorname{prox}_{\gamma f}(\mathcal{X})=\left[\operatorname{prox}_{2 \gamma f}\left(\boldsymbol{X}_{1}\right), \operatorname{prox}_{2 \gamma f}\left(\boldsymbol{X}_{2}\right)\right]
$$

The proximity operator of $f_{1}$, i.e., the nuclear norm with index $2 \gamma$ is given by

$$
\operatorname{prox}_{2 \gamma f_{1}}\left(\boldsymbol{X}_{1}\right)=\operatorname{shrink}\left(\boldsymbol{X}_{1}, 2 \gamma\right) .
$$

By Theorem 1, the proximity operator of $f_{2}$, reduces to the group-wise soft-thresholding (15) of a real matrix:

$$
\begin{equation*}
\operatorname{prox}_{2 \gamma f_{2}}\left(\boldsymbol{X}_{2}\right)=\widehat{\mathrm{ST}}\left(\boldsymbol{X}_{2}, 2 \gamma \lambda\right) . \tag{18}
\end{equation*}
$$

For the function $g$, the proximity operator of the indicator function $\iota_{D_{1}}$ is the orthogonal projection $P_{D_{1}}$ onto the subspace $D_{1}$, i.e.,

$$
\operatorname{prox}_{\gamma g}(\mathcal{X})=P_{D_{1}}(\mathcal{X}):=\underset{\mathcal{Y} \in D_{1}}{\arg \min }\|\mathcal{X}-\mathcal{Y}\|_{\mathcal{H}_{0}} .
$$

Since $D_{1} \subset D_{2} \subset \mathcal{H}_{0}$, we have by $[34,5.14$, Reduction principle]

$$
P_{D_{1}}(\mathcal{X})=P_{D_{1}} \mid D_{2} \circ P_{D_{2}}(\mathcal{X}) .
$$

Note that ' $\mid D_{2}$ ' in $P_{D_{1}} \mid D_{2}$ stands for the restriction of the domain to the subspace $D_{2}$. The orthogonal projection $P_{D_{2}}: \mathcal{H}_{0} \rightarrow D_{2}$ and $P_{D_{1}} \mid D_{2}: D_{2} \rightarrow D_{1}$ respectively can be calculated as

$$
P_{D_{2}}(\mathcal{X})=\left[P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right), \boldsymbol{X}_{2}\right]
$$

and

$$
P_{D_{1}} \left\lvert\, D_{2}(\mathcal{X})=\frac{1}{2}\left[\widetilde{\boldsymbol{M}}+\boldsymbol{X}_{1}-\widetilde{\boldsymbol{X}}_{2}^{\star}, \widehat{\boldsymbol{M}}-\widehat{\boldsymbol{X}}_{1}^{\star}+\boldsymbol{X}_{2}\right]\right.
$$

where $\boldsymbol{X}_{1}^{\star}:=\underline{\boldsymbol{X}}_{1} \in \mathbb{A}_{m}^{M \times N}$ and $\boldsymbol{X}_{2}^{\star}:=\check{\boldsymbol{X}}_{2} \in \mathbb{A}_{m}^{M \times N}$. For $P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right)$, 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
$\left\langle\widetilde{\boldsymbol{E}}_{p, q, \ell}, \widetilde{\boldsymbol{E}}_{p^{\prime}, q^{\prime}, \ell^{\prime}}\right\rangle_{\mathbb{R}^{m M \times m N}}= \begin{cases}m & \left(\text { if }(p, q, \ell)=\left(p^{\prime}, q^{\prime}, \ell^{\prime}\right)\right), \\ 0 & \text { (otherwise) } .\end{cases}$ and therefore, $\left\{\frac{1}{\sqrt{m}} \widetilde{\boldsymbol{E}}_{p, q, \ell}\right\}_{p=1, q=1, \ell=1}^{M, N, m}$ is an orthonormal basis of $\mathfrak{S}$ and thus $P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right)$ can be easily calculated as:

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

Now, we can calculate

$$
\begin{aligned}
\operatorname{prox}_{\gamma g}(\mathcal{X}) & =P_{D_{1}} \mid D_{2} \circ P_{D_{2}}(\mathcal{X}) \\
& =P_{D_{1}} \mid D_{2}\left[P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right), \boldsymbol{X}_{2}\right] \\
& =\frac{1}{2}\left[\widetilde{\boldsymbol{M}}+P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right)-\widetilde{\boldsymbol{X}}_{2}^{\star}, \widehat{\boldsymbol{M}}-\widehat{\boldsymbol{X}}_{1}^{\star \star}+\boldsymbol{X}_{2}\right],
\end{aligned}
$$

where $\boldsymbol{X}_{1}^{\star \star}:=P_{\mathfrak{S}}\left(\boldsymbol{X}_{1}\right) \in \mathbb{A}_{m}^{M \times N}$. Since all ingredients are identified, we can summarize the proposed hypercomplex principal component pursuit algorithm in Algorithm 1. Here, $\left(t_{k}\right)_{k \geq 0} \subset[0,2]$ satisfied $\sum_{k \geq 0} t_{k}\left(2-t_{k}\right)=+\infty$, $\gamma \in(0,+\infty)$. Note that the shrinkage operator does not keep the special structure of $\widetilde{(\cdot)}$, i.e., $\operatorname{shrink}(\widetilde{\boldsymbol{A}}, 2 \gamma) \notin \mathfrak{S}$ in general, so we need the projection onto the structure $P_{\mathfrak{G}}$. However, in complex and quaternion domain, it keeps the structure as shown in Fact 3, so $\boldsymbol{L}^{(k)} \in \mathfrak{S}$ and thus $P_{\mathfrak{S}}\left(\boldsymbol{L}^{(k)}\right)=\boldsymbol{L}^{(k)}$ for all $k \geq 0$. Especially if $m=1$ (i.e., $\mathbb{A}_{m}=\mathbb{R}$ ), Algorithm 1 is identical to the original DRS for the PCP (DR-PCP) proposed in [29]. Lastly, we state the convergence of the proposed algorithm.

Theorem 2 (Convergence of $\mathbb{A}_{m}$-DRS-PCP). Let parameters of Algorithm 1 be chosen so that $\gamma \in(0,+\infty),\left(t_{k}\right)_{k \geq 0} \subset$

```
Algorithm 1: \(\mathbb{A}_{m}\)-Douglas-Rachford splitting for hy-
percomplex principal component pursuit ( \(\mathbb{A}_{m}\)-DRS-
PCP)
    Input : \(\boldsymbol{M}, t_{k}, \lambda\)
    Output: Low \(\mathbb{R}\)-rank \(\boldsymbol{L}\) and sparse \(\boldsymbol{S}\)
    Initialize \(k \leftarrow 0, \boldsymbol{L}^{(k)} \leftarrow \mathbf{0}, \boldsymbol{S}^{(k)} \leftarrow \mathbf{0}\);
    repeat
        \(\boldsymbol{L}^{\star \star} \leftarrow P_{\mathfrak{S}}\left(\boldsymbol{L}^{(k)}\right), \boldsymbol{S}^{\star \star} \leftarrow \stackrel{\check{\boldsymbol{S}}}{ }^{(k)} ;\)
        \(\boldsymbol{L}^{\star} \leftarrow\left(\widetilde{\boldsymbol{M}}+P_{\mathfrak{S}}\left(\boldsymbol{L}^{(k)}\right)-\widetilde{\boldsymbol{S}}^{\star \star}\right) / 2\);
        \(\boldsymbol{S}^{\star} \leftarrow\left(\widehat{\boldsymbol{M}}-\widehat{\boldsymbol{L}}^{\star \star}+\boldsymbol{S}^{(k)}\right) / 2 ;\)
        \(\boldsymbol{L}^{(k+1)} \leftarrow\)
            \(\boldsymbol{L}^{(k)}+t_{k}\left(\operatorname{shrink}\left(2 \boldsymbol{L}^{\star}-\boldsymbol{L}^{(k)}, 2 \gamma\right)-\boldsymbol{L}^{\star}\right) ;\)
        \(\boldsymbol{S}^{(k+1)} \leftarrow \boldsymbol{S}^{(k)}+t_{k}\left(\widehat{\mathrm{ST}}\left(2 \boldsymbol{S}^{\star}-\boldsymbol{S}^{(k)}, 2 \gamma \lambda\right)-\boldsymbol{S}^{\star}\right) ;\)
        \(k \leftarrow k+1 ;\)
    until convergence;
    \(\boldsymbol{L}^{\star \star} \leftarrow P_{\mathfrak{S}}\left(\boldsymbol{L}^{(k)}\right), \boldsymbol{S}^{\star \star} \leftarrow \widetilde{\boldsymbol{S}}^{(k)} ;\)
    \(\boldsymbol{L}^{\star} \leftarrow\left(\widetilde{\boldsymbol{M}}+P_{\mathfrak{S}}\left(\boldsymbol{L}^{(k)}\right)-\widetilde{\boldsymbol{S}}^{\star \star}\right) / 2 ;\)
    \(\boldsymbol{S}^{\star} \leftarrow\left(\widehat{\boldsymbol{M}}-\widehat{\boldsymbol{L}}^{\star \star}+\boldsymbol{S}^{(k)}\right) / 2\);
    \([\boldsymbol{L}, \boldsymbol{S}] \leftarrow\left[\boldsymbol{L}^{\star}, \breve{S}^{\star}\right] ;\)
```

$[0,2]$ satisfying $\sum_{k \geq 0} t_{k}\left(2-t_{k}\right)=+\infty$. Then, the output of Algorithm 1 converges to a minimizer of (16).

Remark 2. In this paper, we employ the DRS for solving (16) but it can be also solved by other advanced convex optimization techniques such as the alternating direction method of multipliers (ADMM) [35] and the primal-dual splitting (PDS) [36], [37].

## IV. Numerical Examples

In this section, we perform some numerical experiments for examining the effectiveness of the proposed method. Following the settings in [29], [38], we randomly generate an input pairs $(\boldsymbol{L}, \boldsymbol{S})$ as follows: $\boldsymbol{L}:=\boldsymbol{X}_{L} \boldsymbol{X}_{R}^{\mathrm{H}} \in \mathbb{A}_{m}^{M \times N}$, 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 the all real and imaginary parts of each entry of $\boldsymbol{X}_{L}, \boldsymbol{X}_{R}$ being i.i.d from $\mathcal{N}(0,1)$. Note that $r$ is not always agree to $m \operatorname{rank}^{\mathbb{R}}$ since Fact 2 does not hold for $m>4$ in general. We choose the support set of $S$ uniformly at random from all support set of size $\rho M N(\rho \in(0,1))$. All real and imaginary parts of the non-zero entries are independently drawn form $\mathcal{U}(-256,256)$. We fixed $\lambda=1 / \sqrt{\max (M, N)}$ in the experiments. We perform experiments in the case where $\mathbb{A}_{m}=\mathbb{O}(m=8)$. We compare the proposed method $\mathbb{A}_{m}$-DRS-PCP and three partwise DRS-PCP method, $\mathbb{H}^{2}$-DRS-PCP, $\mathbb{C}^{4}$-DRS-PCP and $\mathbb{R}^{8}$ -DRS-PCP. These part-wise methods split $\mathbb{O}$ into $\mathbb{H}^{2}, \mathbb{C}^{4}$ and $\mathbb{R}^{8}$ estimate all parts separately. TABLE I shows the performance comparisons of all four algorithms. Fig. 1 and Fig. 2 show the differences of all real and parts between the estimated low rank matrices and original matrices for the right case of TABLE I for $\mathbb{A}_{m}$-DRS-PCP (Fig. 1) and $\mathbb{H}^{2}$ -DRS-PCP (Fig. 2). They show that the proposed method $\mathbb{A}_{m^{-}}$

TABLE I
Performance comparison

| $\boldsymbol{L}, \boldsymbol{S} \in \mathbb{O}^{32 \times 32}, \rho=0.2, \operatorname{rank}^{\mathbb{R}}(\boldsymbol{L})=29$ |  |  |
| :--- | :---: | :--- |
| Algorithm |  | error |
| $\mathbb{A}_{m}$-DRS-PCP | $2.0 \mathrm{e}-6$ | \# iter. |
| $\mathbb{H}^{2}$-DRS-PCP | 1.0 | 3,044 |
| $\mathbb{C}^{4}$-DRS-PCP | $8.2 \mathrm{e}-1$ | 2,265 |
| $\mathbb{R}^{8}$-DRS-PCP | 37.5 | 2,990 |
| $\boldsymbol{L}, \boldsymbol{S} \in \mathbb{O}^{32 \times 32}, \rho=0.2$, rank $^{\mathbb{R}}(\boldsymbol{L})=58$ |  |  |
| Algorithm |  | error |
| $\mathbb{A}_{m}$-DRS-PCP | \# iter. |  |
| $\mathbb{H}^{2}$-DRS-PCP | 11.9 | 2,971 |
| $\mathbb{C}^{4}$-DRS-PCP | 9.6 | 2,553 |
| $\mathbb{R}^{8}$-DRS-PCP | 78.7 | 1,894 |

DRS-PCP outperforms all part-wise methods by exploiting all correlations among real and imaginary parts. $\mathbb{H}$-DRS-PCP and $\mathbb{C}$-DRS-PCP much better than $\mathbb{R}$-DRS-PCP since it may utilize these correlations in part. This experiment also shows that $\mathbb{C}^{4}$ -DRS-PCP is a little bit better than $\mathbb{H}^{2}$-DRS-PCP. This may be because the exploiting correlations with quaternion is not so important in this example.

## V. Conclusions

In this paper, we have proposed an algorithmic solution to hypercomplex principal component pursuit based on a proximal splitting technique. This solution solves the hypercomplex principal component pursuit, which is a convex relaxation of hypercomplex robust principal component analysis with a new sparsity measure of C-D matrices, and utilizes a useful mathematical tools including C-D SVD and $\mathbb{R}$-rank based on algebraic translations of C-D number systems. Numerical experiments show that the proposed algorithm separates the observed matrices into the sum of low rank and sparse ones much more faithfully than existing algorithms.

## Appendix A <br> Douglas-Rachford Splitting

The Douglas-Rachford splitting (DRS) [28], [39], [40] is a well-defined proximal splitting method that solves the minimization of the sum of two functions

$$
\begin{equation*}
f(\boldsymbol{x})+g(\boldsymbol{x}) \tag{19}
\end{equation*}
$$

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 (19) with $\left(\operatorname{prox}_{\gamma g}\left(x_{k}\right)\right)_{k \geq 0}$ by generating the following sequence $\left(\boldsymbol{x}_{k}\right)_{k \geq 0}$ :
$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_{k}+t_{k}\left\{\operatorname{prox}_{\gamma f}\left[2 \operatorname{prox}_{\gamma g}\left(\boldsymbol{x}_{k}\right)-\boldsymbol{x}_{k}\right]-\operatorname{prox}_{\gamma g}\left(\boldsymbol{x}_{k}\right)\right\}$,
where $\left(t_{k}\right)_{k \geq 0} \subset[0,2]$ satisfies $\sum_{k \geq 0} t_{k}\left(2-t_{k}\right)=+\infty$ and the proximity operator [41] of index $\gamma$ of $f \in \Gamma_{0}(\mathcal{H})$ is defined


Fig. 1. Difference between the original matrix $\boldsymbol{L}^{\text {opt }}$ and the estimated low rank matrix $\boldsymbol{L}$ with $\mathbb{A}_{m}$-DRS-PCP


Fig. 2. Difference between the original matrix $L^{\text {opt }}$ and the estimated low rank matrix $\boldsymbol{L}$ with $\mathbb{H}^{2}$-DRS-PCP
as

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

with the norm on $\mathcal{H}$ denoted by $\|\cdot\|_{\mathcal{H}}$. Indeed, if $\operatorname{dim}(\mathcal{H})<\infty$, $\left(\operatorname{prox}_{\gamma g}\left(x_{k}\right)\right)_{k \geq 0}$ converges to a minimizer of (19) (see e.g., [42]).

## 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] S. J. Sangwine and T. A. Ell, "Hypercomplex auto- and cross-correlation of color images," in Proc. IEEE ICIP, 1999, pp. 319-322.
[13] N. L. Bihan and S. J. Sangwine, "Quaternion principal component analysis of color images," in Proc. IEEE ICIP, 2003, pp. 809-812.
[14] S. Miron, N. L. Bihan, and J. I. Mars, "Quaternion-MUSIC for vectorsensor array processing," IEEE Trans. Signal Process., vol. 54, no. 4, pp. 1218-1229, Apr. 2006.
[15] 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.
[16] A. M. Grigoryan and S. S. Agaian, Quaternion and Octonion Color Image Processing with MATLAB. SPIE Press Monographs, 2018.
[17] 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.
[18] N. A. Wiegmann, "Some theorems on matrices with real quaternion elements," Canadian Journal of Mathematics, vol. 7, pp. 191-201, 1955.
[19] F. Zhang, "Quaternions and matrices of quaternions," Linear Algebra and Its Applications, vol. 251, pp. 21-57, 1997.
[20] T. Dray and C. A. Manogue, "The octonionic eigenvalue problem," Advances in Applied Clifford Algebras, vol. 8, no. 2, pp. 341-364, 1998.
[21] T. Mizoguchi and I. Yamada, "An algebraic translation of CayleyDickson linear systems and its applications to online learning," IEEE Trans. Signal Process., vol. 62, no. 6, pp. 1438-1453, Mar. 2014.
[22] - , "Hypercomplex tensor completion with Cayley-Dickson singular value decomposition," in Proc. IEEE ICASSP, Calgary, Alberta, Canada, 2018.
[23] -, "Hypercomplex tensor completion via convex optimization," Submitted to IEEE Trans. Signal Process., 2018.
[24] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" Journal of the ACM, vol. 58, no. 3, pp. 1-37, May 2011.
[25] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in Proc. IEEE ICASSP, 2012, pp. 57-60.
[26] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," IEEE Trans. Pattern Anal. Mach. Intell., pp. 2233-2246, Jan. 2012.
[27] T.-S. T. Chan and Y.-H. Yang, "Complex and quaternionic principal component pursuit and its application to audio separation," IEEE Signal Processing Letters, vol. 23, no. 2, pp. 287-291, Feb. 2016.
[28] 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.
[29] 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.
[30] G. W. Stewart, "On the early history of the singular value decomposition," SIAM Review, vol. 35, no. 4, pp. 551-566, 1993.
[31] E. J. Candès and B. Recht, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," Foundations of Computational Mathematics, vol. 9, no. 6, pp. 717-772, Dec. 2008.
[32] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low- $n$-rank tensor recovery via convex optimization," Inverse Problems, vol. 27, no. 2, 2011.
[33] D. Wipf and B. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," IEEE Trans. Signal Process., vol. 55, no. 7, pp. 3704-3716, Jun. 2007.
[34] F. R. Deutsch, Best Approximation in Inner Product Spaces. SpringerVerlag, New York, 2001.
[35] 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.
[36] 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.
[37] 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.
[38] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted lowrank matrix," UIUC Technical Report UILU-ENG-09-2214, Tech. Rep., Jul. 2009.
[39] 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.
[40] 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.
[41] 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.
[42] 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.

