

HYPERCOMPLEX ALGEBRAS FOR DICTIONARY LEARNING

Srdan Lazendić^a, Aleksandra Pižurica^b and Hendrik De Bie^a

^a Department of Mathematical Analysis, Clifford Research Group

^b Department of Telecommunications and Information Processing, UGent-IPI-imec

Faculty of Engineering and Architecture, Ghent University, Belgium

Srdan.Lazendic@UGent.be [presenter, corresponding]

Aleksandra.Pizurica@UGent.be

Hendrik.DeBie@UGent.be

ABSTRACT. This paper presents an application of hypercomplex algebras combined with dictionary learning for sparse representation of multichannel images. Two main representatives of hypercomplex algebras, Clifford algebras and algebras generated by the Cayley-Dickson procedure are considered. Related works reported quaternion methods (for color images) and octonion methods, which are applicable to images with up to 7 channels. We show that the current constructions cannot be generalized to dimensions above eight.

1. INTRODUCTION

The complex (\mathbb{C}) and quaternion (\mathbb{H}) algebras are special cases of more general hypercomplex algebras [18]. Rooted in Hamilton's seminal paper about quaternions published in 1843, the theory of hypercomplex algebras evolved, finding applications in many disciplines, ranging from theoretical physics [6] to robotics [17] and signal processing [2, 3, 7, 16]. In general, a *hypercomplex algebra* is defined as unital, distributive algebra, not necessarily associative, over the field of real or complex numbers with n generators (e_1, \dots, e_n) . Usually $e_k^2 \in \{-1, 0, 1\}$ and different multiplication rules between the basis elements generate different algebras [18]. Basis elements are referred to as the *imaginary units*. We address here a recent application of hypercomplex algebras in sparse image representation, with some new insights, regarding applicability to general multichannel images.

Dictionary learning techniques provide the most succinct representation of signals and images. Initiated by the classical work of Olsahusen and Field on sparse coding [22], many dictionary learning methods emerged, K-SVD [1] being among the best known ones, especially in image and video processing. An overview of different models for color image processing by using dictionary learning techniques has been given in [5]. State-of-the-art methods for sparse representation of color images typically concatenate pixel values from collocated image patches in the three channels and treat them then by the standard K-SVD or by a variant thereof [20]. Recently reported quaternion-based K-SVD known as K-QSVD [28, 29], introduced quaternions into dictionary construction, where the three color channels are assigned to the three imaginary units.

The quaternion model demonstrated improvements compared to the classical K-SVD model, especially in the terms of color fidelity as reported in [28, 30]. A limitation of the model is that it cannot treat more than three spectral channels. Recently, in [19] a new model was introduced, called octonion dictionary learning (ODL), which is a generalization of the K-QSVD, in the sense that it can handle up to 7 spectral channels. This is of interest e.g., for multispectral imaging. To the best of our knowledge, the quaternion and the octonion algebras are the only examples of hypercomplex algebra that have been used for multichannel image processing in the combination with dictionary learning techniques. From a different perspective, the approach

This research has been partially supported by the institutional grant no. BOF15/24J/078 of Ghent University.

of [27] treated reconstruction of octonion signals from incomplete and noisy data, by solving a convex ℓ_1 minimization problem.

The rest of the paper is organized as follows. In Section 2 we review briefly the dictionary learning approach and the particular techniques for solving the involved sub-problems. The theory of hypercomplex algebras, and in particular Clifford algebras and algebras generated by the Cayley-Dickson procedure, is reviewed in Section 3. In Section 4, we address sparse representation of images in a hypercomplex algebra framework with two particular instances: quaternion- and octonion-based approach. We explore further generalization in Section 5 and we prove that the current constructions do not admit generalizations beyond \mathbb{C} , \mathbb{H} and \mathbb{O} . Section 6 concludes the paper.

2. DICTIONARY LEARNING APPROACH

2.1. Sparse Coding. The goal of sparse coding is to represent a given signal $\mathbf{y} \in \mathbb{R}^n$ by a linear combination of only a few elements from a redundant (overcomplete) set of basis vectors $\mathbf{D} = \{\mathbf{d}_k\}_{k=1}^m \in \mathbb{R}^{n \times m}$ ($m \gg n$), which is called a *dictionary* and whose elements \mathbf{d}_k are unit norm vectors known as *atoms*. Thus, if the dictionary \mathbf{D} is given, the aim is to find the L -sparse vector of coefficients $\mathbf{x} \in \mathbb{R}^m$, $\|\mathbf{x}\|_0 \leq L$, such that

$$(1) \quad \mathbf{y} \approx \mathbf{D}\mathbf{x} = \sum_{k=1}^m \mathbf{d}_k x_k$$

where $L = \text{const.}$ is a prescribed number of non-zero elements, i.e. the sparsity level. The linear expansion given in (1) is known as *sparse approximation*. The pseudo-norm ℓ_0 counts the number of non-zero elements of the given vector. The *sparse coding problem*

$$(2) \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq L$$

is NP-hard, and thus its approximate solutions are sought instead, typically using greedy algorithms such as Matching Pursuit (MP) [21] and Orthogonal Matching Pursuit (OMP) [23]. An alternative solution is to relax, i.e., convexify the original problem (2) by replacing the ℓ_0 problem with the ℓ_1 norm, which leads to the LASSO problem [26]. Its equivalent non-constrained (Lagrangian) formulation

$$(3) \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

is known as Basis Pursuit Denoising (BPDN) [9] and is solved by convex optimization algorithms [8]. The sparse coding problem was generalized to the quaternion [4, 28] and the octonion [19, 27] settings, where instead of working with real vectors and matrices, elements in \mathbb{H} and \mathbb{O} are used.

2.2. Dictionary Learning. For a set of training samples $\mathbf{Y} = \{\mathbf{y}_k\}_{k=1}^p \in \mathbb{R}^{n \times p}$, where each \mathbf{y}_k is a signal vector, the *dictionary learning problem* consists of finding the dictionary $\mathbf{D} = \{\mathbf{d}_k\}_{k=1}^m \in \mathbb{R}^{n \times m}$ ($m \gg n$), which best adapts to the given training set \mathbf{Y} , and the sparse code $\mathbf{X} = \{\mathbf{x}_k\}_{k=1}^p \in \mathbb{R}^{m \times p}$ such that $\mathbf{Y} \approx \mathbf{D}\mathbf{X}$. Formally, the dictionary learning problem can be expressed as the following minimization problem

$$(4) \quad \{\hat{\mathbf{D}}, \hat{\mathbf{X}}\} = \arg \min_{\mathbf{D}, \mathbf{X}} \{\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2\} \\ \text{s.t.} \quad \|\mathbf{x}_k\|_0 \leq L, \quad k = 1, \dots, p$$

where F denotes the Frobenius norm. The dictionary learning algorithms typically alternate between two steps: sparse coding step (fix \mathbf{D} and find a sparse code \mathbf{X}) and dictionary update step. Different dictionary learning algorithms differ mostly in this second, dictionary update step, the most notable ones being Maximum Likelihood Method [22], Method of Optimal Directions (MOD) [15], K-SVD [1] and Approximate K-SVD [25]. Since the quaternion algebra

is well suited for color image processing, several recent works, including [28, 30] addressed the dictionary learning problem in the quaternion framework. The existing quaternion model is limited to color images with three channels. A generalization in the octonion setting which can handle up to seven channels was reported in [19].

3. HYPERCOMPLEX ALGEBRAS

Here we review two main representatives of the hypercomplex algebras, viz. Clifford algebras and algebras generated by the Cayley-Dickson procedure, that we shall later discuss in the framework of dictionary learning.

3.1. Clifford Algebras. By virtue of its rich structure that includes both geometrical and algebraic properties of Euclidean space, Clifford algebras found their application in computer graphics, robotics, signal and image processing [3, 11, 14, 24].

The real Clifford algebra with generators (e_1, \dots, e_n) is a real associative algebra with identity 1, containing \mathbb{R} and \mathbb{R}^n as subspaces, where the generators satisfy the conditions

$$(5) \quad e_i e_j + e_j e_i = -2\delta_{ij}, \quad i, j = 1, \dots, n,$$

with δ_{ij} being Kronecker delta function which takes value 1 if $i = j$, and 0 otherwise. A basis for the real Clifford algebra $\mathbb{R}_n := \text{Alg}_{\mathbb{R}}(e_1, \dots, e_n)$ is given by the elements

$$(6) \quad e_A = e_{\alpha_1} \dots e_{\alpha_k}, \quad A = (\alpha_1, \dots, \alpha_k), \\ 1 \leq \alpha_1 < \dots < \alpha_k \leq n, \quad k \leq n \quad \text{and} \quad e_{\emptyset} := 1.$$

Hence the Clifford algebra \mathbb{R}_n is 2^n -dimensional.

Examples

- (1) The complex algebra is a special case of a Clifford algebra. Namely, the Clifford algebra \mathbb{R}_1 generated by (e_1) with the basis $(1, e_1)$ is isomorphic with the algebra of complex numbers \mathbb{C} , where we identify $e_1 = i$.
- (2) The Clifford algebra \mathbb{R}_2 has the basis $(1, e_1, e_2, e_{12})$ and is isomorphic with the algebra of quaternions, i.e. $\mathbb{H} \cong \mathbb{R}_2$, where we make the identification $e_1 = i, e_2 = j, e_{12} = k$.
- (3) It can be shown that $\mathbb{R}_3 \cong {}^2\mathbb{R}_2 = \mathbb{R}_2 \oplus \mathbb{R}_2$ (see [10]), so we can conclude that $\mathbb{R}_3 \cong \mathbb{H} \oplus \mathbb{H}$. This is an 8-dimensional associative algebra.

3.2. Cayley-Dickson Algebras. Complex numbers and quaternions are also special cases of the Cayley-Dickson algebra. Higher-dimensional algebras in the Cayley-Dickson process are obtained by doubling a smaller algebra and adding an additional imaginary unit [12]. More precisely, starting with the real numbers \mathbb{R} , higher dimensional Cayley-Dickson algebras are constructed as:

$$\begin{aligned} \mathbb{C} &= \mathbb{R} \oplus i\mathbb{R}, \\ \mathbb{H} &= \mathbb{C} \oplus j\mathbb{C}, \\ \mathbb{O} &= \mathbb{H} \oplus \ell\mathbb{H}, \end{aligned}$$

where i, j, ℓ are imaginary units. Note that although \mathbb{C} and \mathbb{H} are special cases of both the Clifford algebras and the Cayley-Dickson algebras, already the 8-dimensional representatives are different (see Example (3) in Subsection 3.1). Although all Clifford algebras are associative, with Cayley-Dickson algebras already the octonion algebra \mathbb{O} is non-associative. Since every higher-dimensional Cayley-Dickson algebra is obtained by doubling a smaller algebra, this means that also all the algebras of the dimension higher than 8 are not associative.

Remark Note that in every step Cayley-Dickson algebras loose some "nice" properties. For example, the quaternions are not commutative and the octonions are neither commutative nor associative. For the sedenions and higher dimensional algebras, the property $|ab| = |a||b|$ doesn't

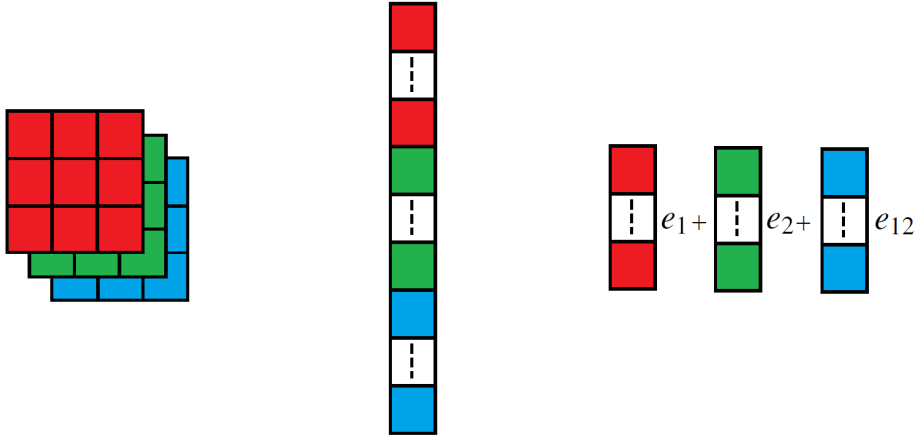


FIGURE 1. RGB color channels (left) represented in the standard concatenated model (middle) and in the quaternion model (right).

hold, which means that they are not composition algebras. Moreover, the Hurwitz theorem tells us that $\mathbb{R}, \mathbb{C}, \mathbb{H}$ and \mathbb{O} are the only real composition algebras with positive-definite norm and thus without zero divisors [12].

4. HYPERCOMPLEX MODELS FOR MULTICHANNEL IMAGES

For monochrome (single-channel) images, the training samples \mathbf{Y} are represented as vectors obtained by stacking the pixel values within a small (typically square) window in some raster-scanning fashion. For n -channel images (including color images), the traditional dictionary learning methods are again stacking the pixel values from collocated windows in the n -channels into one long vector. In this approach the learning process is restricted to relatively low-dimensional signals, i.e., after concatenation the length of the obtained vectors should not exceed 1000 pixel values. An attempt to go beyond this limit, introduces certain problems [13], so we will rather keep the number of channels n small as well as the size of collocated windows. A recently introduced quaternion-based dictionary learning approach [28, 30] for color images assigns the three color channels to three imaginary units. Figure 1 highlights the differences between the traditional stacking and the quaternion-based approach for dictionary learning on color images.

Here, we explain first the main concepts of quaternion-based dictionary learning, and then we address extensions to higher dimensions, with a more in-depth analysis in the following section. A quaternion vector $\mathbf{x} \in \mathbb{H}^n$ is defined as

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1 i + \mathbf{x}_2 j + \mathbf{x}_3 k, \text{ where } \mathbf{x}_i \in \mathbb{R}^n;$$

and a quaternion matrix $\mathbf{M} \in \mathbb{H}^{m \times n}$ equivalently as

$$\mathbf{M} = \mathbf{M}_0 + \mathbf{M}_1 i + \mathbf{M}_2 j + \mathbf{M}_3 k, \text{ where } \mathbf{M}_i \in \mathbb{R}^{m \times n}.$$

A color image can be seen as a quaternion matrix \mathbf{M} with the scalar part \mathbf{M}_0 equal to zero matrix and each of \mathbf{M}_i , $i \in \{1, 2, 3\}$ corresponding to one of the three color channels. Since we are dealing with patch based image processing, an extracted color image patch \mathbf{y} with its three color components \mathbf{y}_r , \mathbf{y}_g and \mathbf{y}_b can be represented as a pure quaternion vector

$$\mathbf{y} = 0 + \mathbf{y}_r i + \mathbf{y}_g j + \mathbf{y}_b k.$$

A quaternion-based sparse representation codes a quaternion image patch \mathbf{y} sparsely as

$$(7) \quad \mathbf{y} \approx \mathbf{D}\mathbf{x} \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq L$$

where \mathbf{D} is a quaternion dictionary given in the general form of a quaternion matrix

$$\mathbf{D} = \mathbf{D}_s + \mathbf{D}_r i + \mathbf{D}_g j + \mathbf{D}_b k.$$

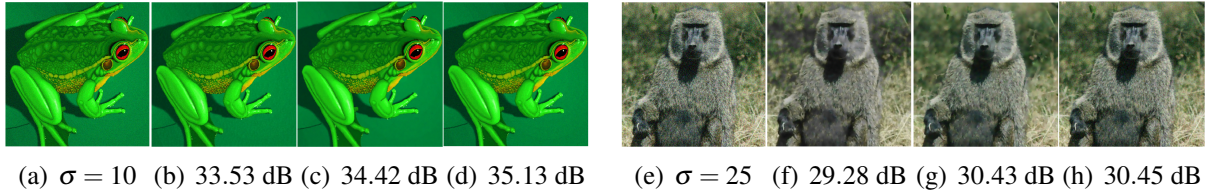


FIGURE 2. Color image denoising. Noisy images with Gaussian noise with standard deviation σ in (a) and (e) reconstructed by K-SVD (in (b) and (f)), Q-KSVD (in (c) and (g)) and ODL (in (d) and (h)) method. PSNR values are below the images.

The expression $\mathbf{y} = \mathbf{D}\mathbf{x}$ can be written in the expanded form as

$$(8) \quad 0 = \mathbf{D}_s\mathbf{x}_0 - \mathbf{D}_r\mathbf{x}_1 - \mathbf{D}_g\mathbf{x}_2 - \mathbf{D}_b\mathbf{x}_3$$

$$(9) \quad \mathbf{y}_r = \mathbf{D}_s\mathbf{x}_1 + \mathbf{D}_r\mathbf{x}_0 + \mathbf{D}_g\mathbf{x}_3 - \mathbf{D}_b\mathbf{x}_2$$

$$(10) \quad \mathbf{y}_g = \mathbf{D}_s\mathbf{x}_2 - \mathbf{D}_r\mathbf{x}_3 + \mathbf{D}_g\mathbf{x}_0 + \mathbf{D}_b\mathbf{x}_1$$

$$(11) \quad \mathbf{y}_b = \mathbf{D}_s\mathbf{x}_3 + \mathbf{D}_r\mathbf{x}_2 - \mathbf{D}_g\mathbf{x}_1 + \mathbf{D}_b\mathbf{x}_0.$$

In this way, the linear correlation given in the equation (8) as well as the orthogonality property between the column vectors of the coefficients remains valid, which can be seen in (8)-(11). This has shown to be useful in the preservation of spectral fidelity in the reconstructed image [28]. The same property holds also for larger number of channels, which can be seen in [19], where the quaternion model K-QSVD was generalized to the octonion setting, and a new ODL model was introduced. The idea in the approach of the ODL model was initiated by the promising results obtained by the K-SVD model, although certain problems occurred. Indeed, since the eigenvalue problem in the octonion setting is not solved for the general dimension of the octonion matrix, straightforward generalization of the K-SVD model from the complex and the quaternion case was not possible. These difficulties were overcome by using the approximate model for dictionary learning introduced in [25]. Detailed explanations with a more in depth analysis of the ODL model can be found in [19]. In Figure 2 we compare the previously mentioned models for color image denoising with different levels of Gaussian additive white noise. As it was stressed in [19], K-QSVD and ODL models outperform classical K-SVD model for lower levels of noise. For higher levels of noise, ODL and K-QSVD are showing more or less similar results.

Apart from the application in color image processing, the ODL model can be applied for processing of multispectral images with up to 7 spectral channels. Landsat 7 satellite images are a notable example of image data with 7 spectral bands (three visible and four infrared bands). Figure 3 shows examples of image reconstruction from incomplete data with traditional K-SVD and ODL. A more extensive evaluation of ODL is in [19]. In particular the method was validated in image denoising of additive white Gaussian noise with standard deviation σ .

5. HIGHERDIMENSIONAL GENERALIZATION

Motivated by the encouraging results of quaternion- and octonion-based dictionary learning [19, 28, 30], we want to explore whether the same idea can be generalized to a hypercomplex algebra of arbitrary dimension.

Consider a multichannel image \mathbf{Y} with $k \in \mathbb{N}$ channels. Let \mathcal{A} be an algebra of dimension 2^n with $n \in \mathbb{N}$ being the first natural number such that $2^n > k$ (see Subsection 3.1). We know that this algebra has $2^n - 1$ imaginary units, let us denote them by e_1, \dots, e_{2^n-1} . The idea is to assign k spectral channels to k out of $2^n - 1$ imaginary units, e_1, \dots, e_k , of a given hypercomplex

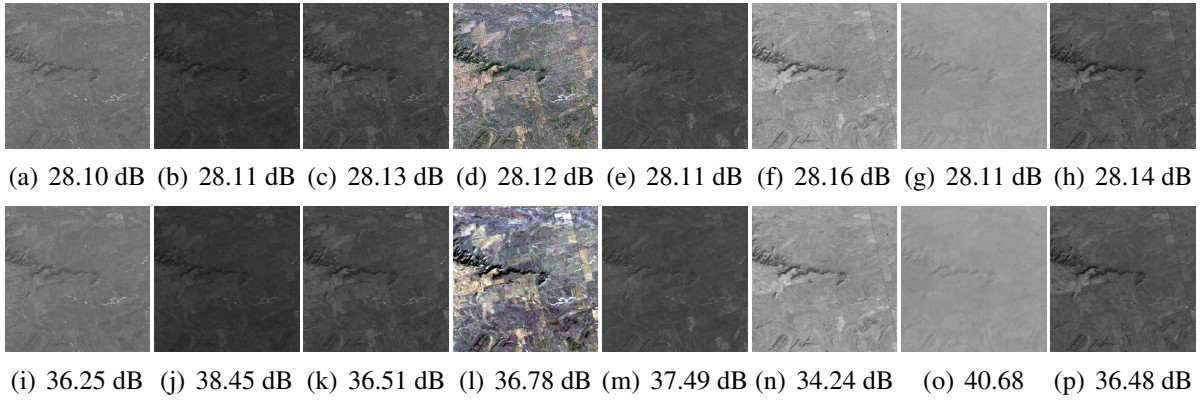


FIGURE 3. Landsat 7 image denoising. First row noisy image channels, second row image reconstructed by the ODL model - PSNR values are shown below.

algebra \mathcal{A} . The sample $\mathbf{y} \in \mathcal{A}^{m \times 1}$ is then represented as

$$\mathbf{y} = \sum_{i=1}^k \mathbf{y}_i e_i$$

where we set $\mathbf{y}_0 = 0$ and $\mathbf{y}_i = 0$, for $i \in \{k+1, \dots, 2^n - 1\}$. The goal is to find a sparse code $\mathbf{x} \in \mathcal{A}^{p \times 1}$ and a dictionary $\mathbf{D} \in \mathcal{A}^{m \times p}$ such that (7) holds. The main motivation is the preservation of spectral fidelity (i.e., inter-channel dependencies) as a generalization of the color fidelity to k -channel images. This implies ensuring the orthogonality between the column vectors of the coefficient matrix, and linear correlation among the spectral channels. Indeed, in the expanded form of multiplication for quaternion and octonion signals we have that the orthogonality property of the coefficient matrix, as well as linear correlation among the spectral channels hold [19, 28].

Keeping the same notation as in [19], we want to define operators \mathbf{v}, χ, ψ that act on elements from \mathcal{A} as well as on vectors and matrices with values in \mathcal{A} , transforming (7) into a real matrix problem. The operator \mathbf{v} is a vectorization operator and transforms elements from \mathcal{A} as well as vectors with values in \mathcal{A} into real vectors. The operators χ and ψ are acting on elements from \mathcal{A} , and can be extended to matrices with values in \mathcal{A} , transforming them into real vectors and real matrices, such that the following relations

$$\mathbf{v}(\mathbf{y}) = \mathbf{v}(\mathbf{D}\mathbf{x}) = \chi(\mathbf{D})\mathbf{v}(\mathbf{x})$$

hold (for more detailed explanations see [19, 30]).

In the expanded form (e.g., in the quaternion case we have (8)-(11)) we can extract all the coefficients in a matrix form, which gives us the real coefficient matrix equal to $\chi(\mathbf{x})$ (or $\psi(\mathbf{x})$). This is just a special case, when instead of a matrix $\mathbf{D} \in \mathcal{A}^{m \times p}$, the map χ (or ψ) acts on a coefficient vector $\mathbf{x} \in \mathcal{A}^{p \times 1}$. The obtained coefficient matrix should have the orthogonality property, i.e., the columns of the obtained matrix should be mutually orthogonal real vectors (like the column coefficient vectors from the coefficient matrix extracted from (8)-(11)). Depending on our preference we can work interchangeably with the right ($\psi(\mathbf{x})$) or left ($\chi(\mathbf{x})$) representation map.

It turns out that this orthogonality property does not hold in higher dimensional Clifford algebras as we show next. For two vectors \mathbf{a}, \mathbf{b} with values in \mathcal{A} , we define

$$(12) \quad \langle \mathbf{a}, \mathbf{b} \rangle = \Re \left(\sum_{i=1}^n \bar{a}_i b_i \right) = \Re(\mathbf{a}^* \mathbf{b}),$$

with $*$ being the conjugate transpose operator and \Re being the real part of the expression. Since the operator \mathbf{v} assigns real vectors $\mathbf{v}(\mathbf{a}), \mathbf{v}(\mathbf{b})$ to \mathbf{a} and \mathbf{b} , it can be shown that the standard real

inner product $\langle \mathbf{v}(\mathbf{a}), \mathbf{v}(\mathbf{b}) \rangle$ equals the one defined in (12). For a vector $\mathbf{x} = [x_i]_{i=1}^p \in \mathcal{A}^{p \times 1}$, we can show that the obtained coefficient matrix $\chi(\mathbf{x})$ is not orthogonal in general, because

$$\begin{aligned} \langle \chi(\mathbf{x})\mathbf{v}(\mathbf{a}), \chi(\mathbf{x})\mathbf{v}(\mathbf{b}) \rangle &= \langle \mathbf{v}(\mathbf{x}\mathbf{a}), \mathbf{v}(\mathbf{x}\mathbf{b}) \rangle = \langle \mathbf{x}\mathbf{a}, \mathbf{x}\mathbf{b} \rangle \\ &= \Re(\mathbf{a}^* \mathbf{x}^* \mathbf{x} \mathbf{b}) = \Re\left(\sum_{i=1}^n \bar{a}_i \bar{x}_i x_i b_i\right) \neq \mathbf{x}^* \mathbf{x} \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{x}^* \mathbf{x} \langle \mathbf{v}(\mathbf{a}), \mathbf{v}(\mathbf{b}) \rangle \end{aligned}$$

since $\bar{x}_i x_i$ is not necessarily a real number. Indeed, for an element $x = e_{12} + e_{34} \in \mathcal{A}$, we have that

$$\bar{x}x = -2e_{1234} + 2$$

which is not a real number, so the matrix $\chi(\mathbf{x})$, in general, is not orthogonal. Here we decided to present the previous proof, although we could also have shown by tedious calculations that \mathbb{R}_3 does not have the orthogonality property. This then implies that all the algebras of dimension larger than 8 do not retain this property as well, since they contain algebras of smaller dimension as subalgebras.

Similarly, in Cayley-Dickson algebras we do not have the desired orthogonality property either for dimensions greater than eight. This follows from the fact that we do not have the associativity property, so we cannot use that

$$\Re\left(\sum_{i=1}^n (\bar{a}_i \bar{x}_i) (x_i b_i)\right) = \Re\left(\sum_{i=1}^n \bar{a}_i (\bar{x}_i x_i) b_i\right).$$

Although the octonion algebra is not associative, it retains the orthogonality property as presented in [19]. Unfortunately, already sedenions do not satisfy this property, which follows from direct calculations but also the fact that in general Clifford and Cayley-Dickson algebras possess zero divisors, could be used to argue the presented results.

The analysis above shows that the only examples of Clifford and Cayley-Dickson algebras with the orthogonality property of the coefficient matrix are \mathbb{C} , \mathbb{H} and \mathbb{O} .

6. CONCLUSION

In this paper we gave an overview of quaternion and octonion-based methods in dictionary learning and we studied their possible extensions with hypercomplex algebras of arbitrary dimension. Such generalizations would be of much interest for processing general multichannel images, such as arbitrary multispectral and hyperspectral images. Our analysis shows that the current approaches to dictionary learning with Clifford and Cayley-Dickson algebras do not admit generalizations to dimensions higher than eight because the orthogonality property of the coefficient matrix is not retained.

REFERENCES

- [1] AHARON, M., ELAD, M., AND BRUCKSTEIN, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing* 54, 11 (2006), 4311–4322.
- [2] ALFSMANN, D. On families of 2^N -dimensional hypercomplex algebras suitable for digital signal processing. In *14th European Signal Processing Conference (EUSIPCO)* (2006), IEEE, pp. 1–4.
- [3] ALFSMANN, D., GÖCKLER, H. G., SANGWINE, S. J., AND ELL, T. A. Hypercomplex algebras in digital signal processing: Benefits and drawbacks. In *15th European Signal Processing Conference (EUSIPCO)* (2007), IEEE, pp. 1322–1326.
- [4] BARTHÉLEMY, Q., LARUE, A., AND MARS, J. I. Sparse approximations for quaternionic signals. *Advances in Applied Clifford Algebras* 24, 2 (2014), 383–402.
- [5] BARTHÉLEMY, Q., LARUE, A., AND MARS, J. I. Color sparse representations for image processing: review, models, and prospects. *IEEE Transactions on Image Processing* 24, 11 (2015), 3978–3989.
- [6] BAYLIS, W. E. Applications of Clifford algebras in physics. In *Lectures on Clifford (geometric) algebras and applications*. Springer, 2004, pp. 91–133.

- [7] BAYRO-CORROCHANO, E. The theory and use of the quaternion wavelet transform. *Journal of Mathematical Imaging and Vision* 24, 1 (2006), 19–35.
- [8] BOYD, S., AND VANDENBERGHE, L. *Convex optimization*. Cambridge university press, 2004.
- [9] CHEN, S. S., DONOHO, D. L., AND SAUNDERS, M. A. Atomic decomposition by basis pursuit. *SIAM review* 43, 1 (2001), 129–159.
- [10] DELANGHE, R., SOMMEN, F., AND SOUČEK, V. *Clifford algebra and spinor-valued functions*, vol. 53 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1992.
- [11] DORST, L., FONTIJNE, D., AND MANN, S. *Geometric algebra for computer science: an object-oriented approach to geometry*. Morgan Kaufmann Publishers Inc., 2009.
- [12] DRAY, T., AND MANOGUE, C. A. *The geometry of the octonions*. World Scientific, 2015.
- [13] ELAD, M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
- [14] ELL, T. A., LE BIHAN, N., AND SANGWINE, S. J. *Quaternion Fourier transforms for signal and image processing*. John Wiley & Sons, 2014.
- [15] ENGAN, K., AASE, S. O., AND HUSOY, J. H. Method of optimal directions for frame design. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1999), vol. 5, pp. 2443–2446.
- [16] HAHN, S. L., AND SNOPEK, K. M. *Complex and Hypercomplex Analytic Signals: Theory and Applications*. Artech House, 2016.
- [17] HITZER, E., NITTA, T., AND KUROE, Y. Applications of Clifford’s geometric algebra. *Advances in Applied Clifford Algebras* 23, 2 (2013), 377–404.
- [18] I.L. KANTOR, A. S. *Hypercomplex Numbers: An Elementary Introduction to Algebras*. Springer-Verlag, New York, 1989.
- [19] LAZENDIĆ, S., DE BIE, H., AND PIŽURICA, A. Octonion sparse representation for color and multispectral image processing. *Submitted to 26th European Signal Processing Conference (EUSIPCO), IEEE*, 2018.
- [20] MAIRAL, J., ELAD, M., AND SAPIRO, G. Sparse representation for color image restoration. *IEEE Transactions on image processing* 17, 1 (2008), 53–69.
- [21] MALLAT, S. G., AND ZHANG, Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing* 41, 12 (1993), 3397–3415.
- [22] OLSHAUSEN, B. A., AND FIELD, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research* 37, 23 (1997), 3311–3325.
- [23] PATI, Y. C., REZAIIFAR, R., AND KRISHNAPRASAD, P. S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers* (1993), IEEE, pp. 40–44.
- [24] PERWASS, C., EDELSBRUNNER, H., KOBELT, L., AND POLTHIER, K. *Geometric algebra with applications in engineering*, vol. 4. Springer, 2009.
- [25] RUBINSTEIN, R., ZIBULEVSKY, M., AND ELAD, M. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Technion* 40, 8 (2008), 1–15.
- [26] TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.
- [27] WANG, R., XIANG, G., AND ZHANG, F. L_1 -norm minimization for octonion signals. In *IEEE International Conference on Audio, Language and Image Processing (ICALIP)* (2016), pp. 552–556.
- [28] XU, Y., YU, L., XU, H., ZHANG, H., AND NGUYEN, T. Vector sparse representation of color image using quaternion matrix analysis. *IEEE Transactions on Image Processing* 24, 4 (2015), 1315–1329.
- [29] YU, L., XU, Y., XU, H., AND ZHANG, H. Quaternion-based sparse representation of color image. In *IEEE International Conference on Multimedia and Expo (ICME)* (2013), pp. 1–7.
- [30] ZOU, C., KOU, K. I., AND WANG, Y. Quaternion collaborative and sparse representation with application to color face recognition. *IEEE Transactions on Image Processing* 25, 7 (2016), 3287–3302.