ℓ_0 and ℓ_1 approaches to sparse coding of octonion valued signals

Srđan Lazendić*, Hendrik De Bie* and Aleksandra Pižurica†

*Department of Electronics and Information Systems, Clifford Research Group,

†Department of Telecommunications and Information Processing, Group for Artificial Intelligence and Sparse Modelling,

Faculty of Engineering and Architecture, Ghent University, Belgium

{Srdan.Lazendic; Hendrik.DeBie; Aleksandra.Pizurica}@UGent.be

***Abstract*— Octonion based signal representations attract the attention for multichannel signal processing due to their property of treating multiple channels in a holistic manner. This leads to better preservation of the correlation among different channels of a given signal. The aim of this article is to compare two different sparse coding techniques for octonion valued signals by using a unified octonion algebra approach.**

1 Introduction

Sparse representation, i.e. representation of a signal as a linear combination of only few elements from a given dictionary of prototype functions called atoms [1, 2], has been underlying most of the recent developments in signal and image processing [3]. Most methods usually treat multidimensional signals as long vectors, by concatenating all channels into a single column vector, regardless of the number of channels that the signal has. In this way also multispectral and hyperspectral images can be treated although the interrelationship between the spectral channels is neglected [4, 5].

The interchannel dependencies can be exploited through a tensor representation model [6, 7]. Another idea of treating all channels simultaneously lies in the combination of the sparse representation model with a hypercomplex algebra representation of multichannel data. Since the quaternion model is not applicable to signals and images with more than three spectral bands, the octonion model has been introduced [4, 8]. The main goal of the hypercomplex algebra approach is to better preserve the interrelationship and channel dependencies [8]. The key idea is to represent multichannel signals and images as octonion vectors and matrices by assigning different channels to different imaginary units of the octonion algebra (Fig. 1). In this way different channels can be treated simultaneously and the interchannel dependencies can be better preserved.

In this paper we address the sparse coding problem for octonion valued signals. First, we consider the ℓ_0-minimization problem, where the exact number of nonzero elements of the code vector is prescribed and we show that it can be solved by using the octonion version of the OMP algorithm. Second, we show that the ℓ_1 minimization problem over octonions can be represented as a real Second-Order Cone Programming (SOCP) problem [9]. Both problems will be represented in a unified way by introducing suitable matrix operators.

2 Octonion sparse representation

Two main representatives of the hypercomplex algebras are Clifford algebras and Cayley-Dickson algebras [10]. The latter ones are obtained by doubling a smaller algebra and adding

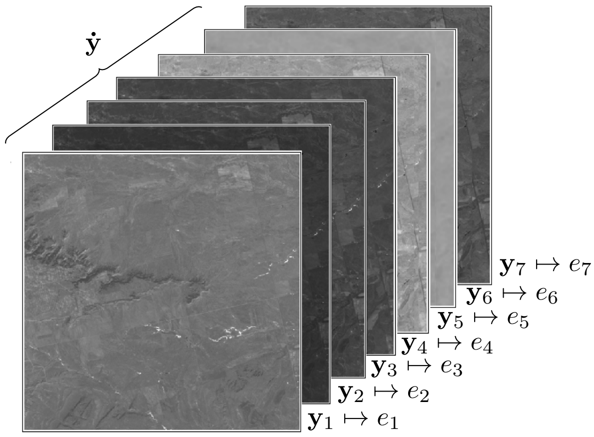

Figure 1: Landsat 7 image patch can be represented as an octonion vector $\dot{\mathbf{y}}$, where each spectral channel $\mathbf{y}_1, \ldots, \mathbf{y}_7$, is assigned to one of the imaginary units $e_1, \ldots, e_7$ of the octonion algebra $\mathbb{O}$.

an additional imaginary unit. Starting with the real numbers $\mathbb{R}$, higher dimensional algebras are constructed as:

$$\mathbb{C} = \mathbb{R} \oplus i\mathbb{R} \quad \to \quad \mathbb{H} = \mathbb{C} \oplus j\mathbb{C} \quad \to \quad \mathbb{O} = \mathbb{H} \oplus \ell\mathbb{H},$$

where i, j, ℓ are imaginary units. Note that $\mathbb{C}$ and $\mathbb{H}$ are special cases of the octonion algebra.

Let $\dot{\mathbf{y}}$ be an m-dimensional multichannel signal with no more than seven spectral channels as in Fig. 1. Then it can be written in the form of an octonion vector $\dot{\mathbf{y}} \in \mathbb{O}^{m\times 1}$ as

$$\dot{\mathbf{y}} = 0 + \mathbf{y}_1 e_1 + \cdots + \mathbf{y}_7 e_7, \quad \mathbf{y}_i \in \mathbb{R}^{m\times 1}. \tag{1}$$

2.1 Octonion sparse coding problem

The octonion sparse coding aims to represent a given signal $\dot{\mathbf{y}} \in \mathbb{O}^{m\times 1}$ as a linear combination of only a few elements from a redundant set $\dot{\mathbf{D}} = \{\dot{\mathbf{d}}_k\}_{k=1}^{m} \in \mathbb{O}^{m\times n}$ which can be represented as $\dot{\mathbf{D}} = \mathbf{D}_0 + \mathbf{D}_1 e_1 + \cdots + \mathbf{D}_7 e_7$. We will refer to $\dot{\mathbf{D}}$ as a *dictionary* whose columns $\dot{\mathbf{d}}_k$ are unit norm vectors known as *atoms*. Thus if the dictionary $\dot{\mathbf{D}}$ is given, the aim is to find the sparse vector of coefficients $\dot{\mathbf{x}} \in \mathbb{O}^{n\times 1}$ known as the *sparse code* such that $\|\dot{\mathbf{x}}\|_0 \leq L$ and $\dot{\mathbf{y}} \approx \dot{\mathbf{D}}\dot{\mathbf{x}}$. $L = const.$ is the prescribed number of non-zero octonion elements in the octonion vector $\dot{\mathbf{x}}$, also known as the *sparsity level*. The pseudo-norm ℓ_0 counts the number of non-zero components of the given octonion vector. Formally the sparse coding problem can be expressed as

$$\hat{\dot{x}} = \underset{\dot{\mathbf{x}} \in \mathbb{O}^{n\times 1}}{\arg\min} \|\dot{\mathbf{y}} - \dot{\mathbf{D}}\dot{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad \|\dot{\mathbf{x}}\|_0 \leq L. \tag{2}$$

An alternative approach is to convexify the original problem (2) by replacing the ℓ_0 norm with the ℓ_1 norm, which leads to the LASSO problem [11], whose equivalent Lagrangian formulation known as Basis Pursuit Denoising (BPDN) [12] over octonions is given by

$$\hat{\dot{x}} = \underset{\dot{\mathbf{x}} \in \mathbb{O}^{n\times 1}}{\arg\min} \|\dot{\mathbf{y}} - \dot{\mathbf{D}}\dot{\mathbf{x}}\|_2^2 + \lambda \|\dot{\mathbf{x}}\|_1. \tag{3}$$

This problem was first studied in [13]. The approach presented there leads to a solution which gives nearly perfect reconstruction of the signal, but at the expense of a very high computational complexity compared to sparse coding of real-valued signals. We explore thus a unified framework for formulating octonion sparse coding models in terms of the corresponding problems over real numbers, using thereby the efficient available solvers, and we compare the performance of ℓ_0 and ℓ_1 formulations under this setup.

3 Sparse coding techniques

Here we introduce a unified octonion algebra sparse coding approach that will allow us to formulate diverse minimization problems over the octonion algebra as the equivalent minimization problems over the real numbers. We shall use the vectorization operator $\nu : \mathbb{O}^{n\times 1} \to \mathbb{R}^{8n\times 1}$ such that $\|\dot{\mathbf{x}}\|_2 = \|\nu(\dot{\mathbf{x}})\|_2$. Note that in this way the real vector space isomorphism $\nu : \mathbb{O}^{n\times 1} \to \mathbb{R}^{8n\times 1}$ is obtained. Another useful operator is a matrix operator $\chi : \mathbb{O}^{n\times m} \to \mathbb{R}^{8n\times 8m}$, with the property $\nu(\dot{\mathbf{D}}\dot{\mathbf{x}}) = \chi(\dot{\mathbf{D}})\nu(\dot{\mathbf{x}})$. The explicit forms of the operators ν and χ together with the proofs of many useful properties can be found in [14, 15, 4].

3.1 ℓ_0 minimization problem

We will show here that the idea of the Orthogonal Matching Pursuit (OMP) [16] can be adapted to the octonion setting. In each step k the algorithm selects the atom $\dot{\mathbf{d}}_k$ that produces the strongest decrease in the residual $\|\dot{\mathbf{r}}^k\|_2^2$ where $\dot{\mathbf{r}}^k = \dot{\mathbf{r}}^{k-1} - \dot{\mathbf{d}}_k\dot{\mathbf{x}}_k$, and $\dot{\mathbf{r}}^0 = \dot{\mathbf{y}}$. In the octonion setting, this is equivalent with selecting the atom that is most correlated with the residual vector i.e. $\langle \dot{\mathbf{r}}^k, \dot{\mathbf{d}}_k\rangle$. After choosing the atom that produces the maximum absolute value in the inner product with the residual, the active dictionary is formed $\dot{\mathbf{D}}^k = [\dot{\mathbf{D}}^{k-1}, \dot{\mathbf{d}}_k]$ and the coding coefficients $\dot{\mathbf{x}}^k$ are selected so that the norm $\|\dot{\mathbf{y}} - \dot{\mathbf{D}}^k\dot{\mathbf{x}}\|_2^2$ is minimized. Solving linear least square problems in the octonion setting is difficult. Therefore, we transform this minimization problem into a real vector minimization problem. Observe that

$$\left\|\dot{\mathbf{y}} - \dot{\mathbf{D}}^k\dot{\mathbf{x}}\right\|_2^2 = \left\|\nu\left(\dot{\mathbf{y}} - \dot{\mathbf{D}}^k\dot{\mathbf{x}}\right)\right\|_2^2 = \left\|\nu\left(\dot{\mathbf{y}}\right) - \chi\left(\dot{\mathbf{D}}^k\right)\nu\left(\dot{\mathbf{x}}\right)\right\|_2^2.$$

By minimizing the last expression on the right we are solving the real minimization problem. The octonion coefficient vector $\dot{\mathbf{x}}^k$ is then obtained by using the inverse map ν^{-1}. Note that the inverse ν^{-1} exists since ν is bijective. Finally, the solution is given by $\dot{\mathbf{x}}^k = \nu^{-1}(\chi(\dot{\mathbf{D}}^k)^\dagger \nu(\dot{\mathbf{y}}))$, where $(\cdot)^\dagger$ is the Moore-Penrose (pseudo)inverse of a real matrix $\chi(\dot{\mathbf{D}}^k)$.

3.2 ℓ_1 minimization problem

Now we will show that the ℓ_1 minimization problem that was earlier addressed in [13] can be compactly written and implemented by using the previously introduced operators ν and χ. This allows us to represent both approaches in a unified way. The problem (3) can be equivalently written as

$$\min_{t\in\mathbb{R}^+} t \ \text{ s.t. } \ \dot{\mathbf{y}} = \dot{\mathbf{D}}\dot{\mathbf{x}}, \ \ \|\dot{\mathbf{x}}\|_1 \le t. \tag{4}$$

By decomposing $t = \sum_{i=1}^{n} t_i$, $t_i \in \mathbb{R}^+$ we can write the last constraint as

$$\|\dot{\mathbf{x}}\|_1 = \sum_{i=1}^{n} \|\dot{\mathbf{x}}_i\|_2 = \sum_{i=1}^{n} \|\nu(\dot{\mathbf{x}}_i)\|_2 \le \mathbf{1}^T\mathbf{t} = \mathbf{1}^T \cdot [t_1, \dots, t_n] = t.$$

The minimization problem in (4) then becomes

$$\min_{t\in\mathbb{R}^+} \mathbf{1}^T\mathbf{t} \ \text{ s.t. } \ \dot{\mathbf{y}} = \dot{\mathbf{D}}\dot{\mathbf{x}}, \ \ \|\nu(\dot{\mathbf{x}}_i)\|_2 \le t_i, \tag{5}$$

for every $i = 1, \dots, n$. By denoting

$$\begin{aligned}
\hat{\mathbf{x}} &= [t_1, \nu(\dot{\mathbf{x}}_1), \dots, t_n, \nu(\dot{\mathbf{x}}_n)]^T \in \mathbb{R}^{9n\times 1},\\
\hat{\mathbf{c}} &= [c_j]_j = \begin{cases} 1, & \text{if } j = 9i - 8 \\ 0, & \text{otherwise,} \end{cases},\\
\hat{\mathbf{y}} &= \nu(\dot{\mathbf{y}}) \in \mathbb{R}^{8m\times 1},\\
\hat{\mathbf{D}} &= \left[\mathbf{0}, \chi\left(\dot{\mathbf{d}}_1\right), \dots, \mathbf{0}, \chi\left(\dot{\mathbf{d}}_n\right)\right] \in \mathbb{R}^{8m\times 9n}
\end{aligned}$$

for every $i = 1, \dots, n$, (5) can be equivalently written in the form of the classical second-order cone optimization problem [9] as:

$$\min_{\hat{\mathbf{x}}\in\mathbb{R}^{9n}} \hat{\mathbf{c}}^T\hat{\mathbf{x}} \ \text{ s.t. } \ \hat{\mathbf{y}} = \hat{\mathbf{D}}\hat{\mathbf{x}}, \ \ \|\nu(\dot{\mathbf{x}}_i)\|_2 \le t_i. \tag{6}$$

This problem can be solved by SeDuMi software [17, 18, 19].

It can be shown that ℓ_1 minimization problem over octonions can also be represented as a convex problem suited for application of the Alternating Direction Method of Multipliers (ADMM). In order to do so, apart from the operators ν and χ, we need another operator $\mathcal{R} : \mathbb{O}^{n\times 1} \to \mathbb{R}^{n\times 8}$ such that $\|\dot{\mathbf{x}}\|_1 = \|\mathcal{R}(\dot{\mathbf{x}})\|_{2,1}$. A detailed derivation is in [20].

	O-OMP	ℓ_1 with ADMM	ℓ_1 as SOCP
PSNR	39.28 dB	63.34 dB	98.06 dB
Time	290 s	14 min	101 min

Table 1: Average reconstruction values with the computation times for 1D octonion signals which are produced by extracting random patches from Landsat 7 images. Sparsity level for O-OMP has been set to 5.

4 Results and discussions

Table 1 reports average reconstruction errors for 1D octonion signals extracted from Landsat 7 images. For the ℓ_1 problem we give the results using two solvers SOCP [9] and ADMM [21]. It can be observed that the SOCP model gives nearly perfect reconstruction of the signal, but it is more time consuming in comparison with the other two methods. The results obtained by ADMM based model showed improvements in terms of the mean squared error values compared to the O-OMP model. Moreover, this model is soundly fast compared to the SOCP model. The indicative computation times in Table 1 were obtained in MATLAB on Intel Core-i7 CPU with 16GB memory. Most importantly, we showed that both ℓ_0 and ℓ_1 minimization problems for octonion signals can be equivalently formulated as real minimization problems by introducing suitable matrix operators.

Acknowledgment

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

References

[1] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[2] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[3] P. Song, X. Deng, J. F. Mota, N. Deligiannis, P.-L. Dragotti, and M. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," *IEEE Transactions on Computational Imaging*, 2019.

[4] S. Lazendić, H. De Bie, and A. Pižurica, "Octonion sparse representation for color and multispectral image processing," in *2018 26TH European Signal Processing Conference (EUSIPCO)*. EURASIP, 2018, pp. 608–612.

[5] C. Zou, K. I. Kou, and Y. Wang, "Quaternion collaborative and sparse representation with application to color face recognition," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3287–3302, 2016.

[6] N. Qi, Y. Shi, X. Sun, and B. Yin, "TenSR: Multi-dimensional tensor sparse representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5916–5925.

[7] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2949–2956.

[8] S. Lazendić, A. Pižurica, and H. De Bie, "Hypercomplex algebras for dictionary learning," in *Early Proceedings of the AGACSE 2018 Conference*. Unicamp/IMECC, Campinas, Brazil, 2018, pp. 57–64.

[9] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[10] T. Dray and C. A. Manogue, *The geometry of the octonions*. World Scientific, 2015.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[13] R. Wang, G. Xiang, and F. Zhang, "L_1-norm minimization for octonion signals," in *2016 International Conference on Audio, Language and Image Processing (ICALIP)*. IEEE, 2016, pp. 552–556.

[14] L. Rodman, "Hermitian octonion matrices and numerical ranges," *Electronic Journal of Linear Algebra*, vol. 27, no. 1, p. 33, 2014.

[15] Y. Tian, "Matrix representations of octonions and their applications," *Advances in Applied Clifford Algebras*, vol. 10, no. 1, pp. 61–90, 2000.

[16] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[17] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[18] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[19] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.

[20] S. Lazendić, H. De Bie, and A. Pižurica, "Sparse coding for octonion valued signals: performance comparison and analysis," 2020, in preparation.

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