Tensor-structured Dictionaries for Hyperspectral Imaging

Cassio Dantas, Jérémy Cohen, Rémi Gribonval

Tensor-structured Dictionaries for Hyperspectral Imaging

Cássio F. Dantas, Jérémy E. Cohen, Rémi Gribonval
Univ Rennes, Inria, CNRS, IRISA
Rennes, France

***Abstract*—Dictionary learning, paired with sparse coding, aims at providing sparse data representations. When dealing with large datasets, the dictionary obtained by applying unstructured dictionary learning methods may be of considerable size, posing both memory and computational complexity issues. We show how a previously proposed structured dictionary learning model, HO-SuKro, can be used to obtain more compact and readily-applicable dictionaries when the targeted data is a collection of multiway arrays. We introduce an alternating optimization learning algorithm and apply it to a hyperspectral image denoising task.**

I. Problem Statement

Let $\mathcal{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3 \times n}$ be a collection of n tensor data $\{\mathcal{Y}_1, \ldots, \mathcal{Y}_n\}$. The Dictionary Learning problem [1] is classically formulated for vector input data. Each $\mathcal{Y}_i$ is thus vectorized, yielding

$$\underset{D \in \mathcal{S}_D, x_i}{\operatorname{argmin}} \sum_{i=1}^{n} \| \operatorname{vec}(\mathcal{Y}_i) - D x_i \|_2^2 + g(x_i) \tag{1}$$

where $D \in \mathbb{R}^{m_1 m_2 m_3 \times d}$ is the dictionary which is often overcomplete ($d \geq m_1 m_2 m_3$) and belongs to a constraint set $\mathcal{S}_D$. Function g is a sparsity-inducing penalty and x_i is the sparse vector of coefficients describing the vectorized tensor $\mathcal{Y}_i$ in the set of atoms D. Most often, $\mathcal{S}_D$ is the set of matrices with unit ℓ_2-norm columns.

When tensor data is considered, two main drawbacks emerge in this formulation: i) it ignores the original multidimensional structure of the data; ii) data sizes m_j may be relatively large (even more so the product $m_1 m_2 m_3$) and both the storage of matrix D and the computation of products such as $D x_i$ may be cumbersome.

One way to tackle the second issue is to restrict the class $\mathcal{S}_D$ of dictionaries that are sought in problem (1). To also tackle the first mentioned issue, we previously proposed the High-Order Sum of Kronecker model (HO-SuKro) [2], [3], a class of tensor-structured dictionaries which are particularly suited to tensorial input data:

$$D = \sum_{q=1}^{r} D_{1,q} \boxtimes D_{2,q} \boxtimes D_{3,q} \tag{2}$$

where $\boxtimes$ is the Kronecker product [4], $D_{j,q}$ are matrices in $\mathbb{R}^{m_j \times d_j}$ and $d_1 d_2 d_3 = d$. To provide some extra intuition, for a fixed q we can see the terms $D_{j,q}$ as linear operators acting independently in the j-th mode of the input data $\mathcal{Y}$.

An obvious remark about HO-SuKro is that the number of parameters to represent D using (2), which is $r(m_1 d_1 + m_2 d_2 + m_3 d_3)$, is much smaller than the number of entries $m_1 m_2 m_3 d_1 d_2 d_3$ as long as r is small. Moreover, matrix products with the dictionary can be computed markedly faster.

II. Proposed algorithm (a glimpse)

As usually done in the literature, problem (1) is tackled by alternating between estimating only D and estimating only the coefficients x_i: respectively the dictionary update and sparse coding steps.

The sparse coding problem remains formally unchanged. Therefore, one may apply the usual greedy heuristics such as Orthogonal Matching Pursuit (OMP) [5] or convex relaxations based algorithms such as FISTA [6]. The Kronecker structure can be exploited to positively impact the complexity and running time of these methods. Indeed, any matrix-vector product with the dictionary becomes a sequence of mode products with the smaller matrices $D_{j,q}$ which involves approximately $r(\sum_j m_j)(\prod_j d_j)$ operations instead of $(\prod_j m_j d_j)$ for a unstructured matrix-vector product (cf. Table I).

Dictionary update, on the other hand, is heavily modified in the HO-SuKro formulation. The partial cost function, with respect to the blocks $\{D_{j,q}\}_{q=1,\ldots,r}^{j=1,\ldots,3}$, may be written as follows:

$$f(\{D_{j,q}\}) = \sum_{i=1}^{n} \| \operatorname{vec}(\mathcal{Y}_i) - \sum_{q=1}^{r} (D_{1,q} \boxtimes D_{2,q} \boxtimes D_{3,q}) \, x_i \|_2^2 \tag{3}$$

In the spirit of Alternating Least Squares algorithms (ALS), it is possible to gather all elementary blocks $D_{j,q}$ for a fixed mode j and alternate between three macro-blocks (one per mode) $\Delta_j = \{D_{j,q}\}_{q \in [1,r]}$, $1 \leq j \leq 3$, obtaining a closed-form solution. For instance, for $j = 1$ (and similarly for other modes)

$$\widehat{\Delta_1} = Y_{[1]} U^{\dagger}, \tag{4}$$

where $Y_{[1]} \in \mathbb{R}^{m_1 \times m_2 m_3 n}$ is the first-mode unfolding [4] of $\mathcal{Y}$, $U = [U_1; \; \ldots \; ; U_r] \in \mathbb{R}^{r d_1 \times m_2 m_3 n}$ with $U_q = X_{[1]} (D_{2,q} \boxtimes D_{3,q} \boxtimes I_n)^T$ and $U^{\dagger} = U^T \left(U U^T\right)^{-1}$ is the right inverse of U.

III. Experiments

Given a hyperspectral image from *San Diego* and *Houston* [8] datasets corrupted with random Gaussian noise uniform over all spectral modes, we consider $n = 10^5$ uniformly-spaced overlapping 3D-patches centralized to zero mean with dimensions $\{m_1, m_2, m_3\}$ from the noisy image to form our training data.

A dictionary is learned from this data with 20 iterations of alternating optimization. Sparse coding is performed by OMP with an error threshold proportional to the noise level. The learned dictionary is then used to reconstruct patches with a one-pixel step. The recovered patches are averaged in the overlapping pixels along with the noisy image itself to form the final denoised image.

Table II shows the denoising performance of the proposed structured dictionaries compared to K-SVD [9] as an unstructured counterpart. Increasing the patch size improves significantly the performance of HO-SuKro, while that of K-SVD doesn't benefit as much and may even deteriorate, indicating the onset of overfitting.

Table III compares HO-SuKro's performance to that of other techniques from the literature. Both HO-SuKro and K-SVD outperform the wavelet-based approaches (2D and 3D [10]), corroborating the interest of learning the sparsifying dictionary from data. HO-SuKro also consistently outperforms FORPDN [11] which exploits the correlation within spectral bands. Naturally, for this task, we don't manage to reach state-of-the-art performance (HyRes [12]) which makes use of a meaningful low-rank prior on the HSI. An example of denoised image is provided in Figure 1.

IV. Conclusion

An alternate minimization algorithm was proposed for learning tensor-structured dictionaries designed to tackle the inherent multidimensional structure of the data. The proposed structure proved to be beneficial in terms of performance in Hyperspectral image denoising experiments compared to an completely unstructured dictionary.

TABLE I: Speedups in matrix-vector products

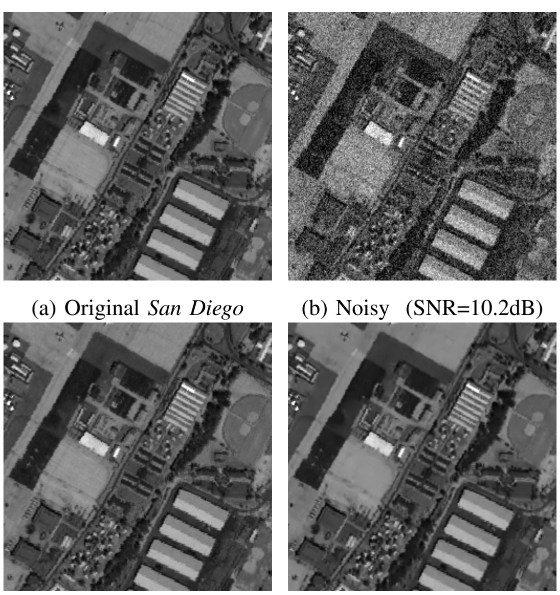
Dimensions		Theoretical speedup | *Empirical speedup*					
m	d	$r=1$		$r=3$		$r=5$	
$[6, 6, 6]$	$[12, 12, 12]$	20.3	*0.8*	6.8	*0.4*	4.9	*0.3*
$[8, 8, 8]$	$[16, 16, 16]$	36.6	*4.5*	12.2	*2.0*	7.3	*1.4*
$[10, 10, 10]$	$[20, 20, 20]$	57.1	*13.2*	19.1	*5.7*	11.4	*3.5*

TABLE II: Output SNR for various patch sizes – San Diego

Algorithm	Patch size		Input SNR			
	m	d	10	15	20	25
	[6, 6, 6]	[12,12,12]	21.77	25.31	29.09	32.68
K-SVD	[8, 8, 8]	[16,16,16]	21.51	25.19	29.2	32.95
	[10,10,10]	[20,20,20]	21.52	25.34	29.42	33.24
	[6, 6,20]	[12,12,20]	22.49	26.06	29.91	33.21
	[6, 6, 6]	[12,12,12]	22.05	25.35	28.9	32.3
HO-SuKro	[8, 8, 8]	[16,16,16]	22.73	25.82	29.22	32.64
($r = 3$)	[10,10,10]	[20,20,20]	23.08	26.35	29.73	**33.27**
	[6, 6,20]	[12,12,20]	**24.10**	**27.07**	**30.09**	33.22

TABLE III: Output SNR [dB] comparison with literature

Image	Algorithm	Input SNR [dB]			
		10	15	20	25
	Wavelet 2D	14.75	18.00	21.70	25.92
	Wavelet 3D	23.11	26.04	28.91	31.68
San Diego	FORPDN	22.23	24.17	26.42	29.00
	HyRes	**25.38**	**28.60**	**31.75**	**34.70**
	HO-SuKro	24.10	27.07	30.09	33.22
	Wavelet 2D	14.22	17.67	21.43	25.80
	Wavelet 3D	22.35	25.54	28.65	31.86
Houston	FORPDN	22.80	25.46	28.09	30.74
	HyRes	**26.00**	**29.35**	**33.24**	**37.05**
	HO-SuKro	23.29	26.63	29.93	33.20



(a) Original *San Diego* (b) Noisy (SNR=10.2dB)

(c) HyRes (SNR=25.4dB) (d) HO-SuKro (SNR=24dB)

Fig. 1: Example of denoised images (100$^{\text{th}}$ spectral band). A closer look reveals that our approach may over-smooth some details.

References

[1] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[2] C. F. Dantas, J. E. Cohen, and R. Gribonval, "Learning fast dictionaries for sparse representations using low-rank tensor decompositions," in *LVA/ICA*, Guildford, UK, 2018. [Online]. Available: https://hal.inria.fr/hal-01709343

[3] C. F. Dantas, M. N. da Costa, and R. R. Lopes, "Learning Dictionaries as a sum of Kronecker products," *IEEE Signal Processing Letters*, 2017.

[4] J. E. Cohen, "About notations in multiway array processing," *arXiv preprint arXiv:1511.01306*, 2015.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.

[6] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: http://dx.doi.org/10.1137/080716542

[7] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[8] "2013 IEEE GRSS Data Fusion Contest," http://www.grss-ieee.org/community/technical-committees/data-fusion/, Accessed: 2019-02-02.

[9] M. Aharon and M. Elad, "Sparse and redundant modeling of image content using an image-signature-dictionary," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 228–247, 2008. [Online]. Available: http://dx.doi.org/10.1137/07070156X

[10] A. A. Basuhail and S. P. Kozaitis, "Wavelet-based noise reduction in multispectral imagery," in *Algorithms for multispectral and hyperspectral imagery IV*, vol. 3372. International Society for Optics and Photonics, 1998, pp. 234–241.

[11] B. Rasti, J. Sveinsson, M. Ulfarsson, and J. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2458–2467, 2014.

[12] B. Rasti, M. O. Ulfarsson, and P. Ghamisi, "Automatic hyperspectral image restoration using sparse and low-rank modeling," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2335–2339, Dec 2017.