Fast Sparse Coding Algorithms for Piece-wise Smooth Signals

A. Gkillas, D. Ampeliotis, K. Berberidis
Department of Computer Engineering and Informatics, University of Patras
Rio - Patras, Greece
E-mails: {st1003586, ampeliot, berberid}@ceid.upatras.gr

***Abstract*—The problem of computing a proper sparse representation matrix for a signal matrix that obeys some local smoothness property, given an over-complete dictionary, is considered. The focus is on piece-wise smooth signals, defined as signals that comprise a number of blocks that each fulfills the considered smoothness property. A computationally efficient sparse coding algorithm is derived by limiting the number of times that a new support set of dictionary atoms is computed, exploiting the smoothness of the signal. Furthermore, a new, total-variation regularized problem is proposed for computing the required sparse coding coefficients, exploiting further the smoothness priors of the signals. The considered problem is solved using the alternating direction method of multipliers. Finally, numerical results considering hyperspectral images are provided, that demonstrate the applicability and complexity - denoising performance benefits of the novel algorithms.**

***Index Terms*—Sparse coding, piece-wise smooth signals, total variation, hyperspectral imaging, Alternating Direction Method of Multipliers**

I. Introduction

The sparse representation theory has proven to be a remarkably powerful analysis and modelling tool with application to various fields such as signal processing, image processing and machine learning. In particular, it has offered many techniques that achieve state-of-the-art results in a variety of problems, ranging from image denoising, inpainting and spectral/spatial super-resolution to image classification and segmentation [1]. Based on its theoretical foundations and ground-breaking results [2], the sparse representation theory has evolved into a universal mathematical model, aiming to reveal the intrinsic sparsity that possibly exists in natural signals. More precisely, the sparse representation model seeks to describe each signal as a linear combination of a small number of columns, called atoms, from a given overcomplete matrix, called dictionary, leading to a corresponding sparse representation [2], [3].

However, the procedure of selecting the optimal set of dictionary atoms for the representation of a signal constitutes an NP-hard problem, and hence only sub-optimal solutions can be derived in polynomial time [4]. In this context, existing algorithms alleviate this inherent difficulty by employing either greedy approaches, such as orthogonal matching pursuit (OMP) [5], batch-OMP [6] and compressive sampling matching pursuit (CoSaMP) [7], or convex relaxation based techniques which replace the l_0 pseudo-norm with the l_1 norm, such as, basis pursuit [8] and least absolute shrinkage and selection operator (Lasso) [9]. In terms of computationally complexity, the greedy approaches outperform significantly the convex relaxation based techniques [1]. However, when problems with high dimensionality are to be tackled, even greedy approaches may become a computational burden.

In this study, we consider the problem of sparse coding, focusing on a widespread category of signals, that is, those that can be characterized as piece-wise smooth (locally homogeneous). Smooth signals are often encountered in a wide range of engineering fields, such as in image processing, control systems and environment monitoring [10]. A typical example of such signals is hyperspectral images, a main property of which is that spatially neighboring pixels demonstrate strong spectral similarity [11], [12]. In light of this, we propose a novel approach which employs a block-processing strategy, where the structure of piece-wise smooth signals is effectively exploited to offer a notably reduced computational complexity, hence resulting in an efficient sparse coding scheme.

Apart from the block-processing strategy employed by the previous scheme, the smoothness of the signals considered is further taken into account by introducing a proper total-variation (TV) regularizer [13] at the cost function used for computing the representation weights, thus leading to a modified technique that promotes small variations in these coefficients. The use of the TV regularizer, which is well known for its denoising properties [14], renders this second scheme resilient to noisy input signals. It should be stated that a TV regularizer has also been proposed in the so-called SUnSAL-TV sparse coding algorithm [11], combined with an l_1 norm term, achieving state-of-the-art results in the hyperspectral unmixing problem. However, this algorithm is excessively expensive in computational terms [11], [15], inducing its limited applicability in real-time and/or high-dimensional applications with large data-sets. By contrast, the proposed schemes enjoy a remarkably lower computational complexity, without sacrificing accuracy, and turns out to be ideal for several applications in which the involved signals are piece-wise smooth, such as in hyperspectral imaging, especially real-time, e.g., spatial/spectral super resolution, unmixing, denoising and classification [16], [11], [15], [17]. In our previous work [18] we have incorporated the above-mentioned

This work was supported partly by the University of Patras, the RPF, Cyprus, under the project INFRASTRUCTURES/1216/0017 (IRIDA), and the EPAnEK, GSRT, Greece, under the project SOilless culture UPgrade - SOUP code:T1EDK-04171.

idea during a coupled dictionary learning procedure.

The remainder of the paper is organized as follow. Section II formulates the problem under study. Section III describes the proposed algorithms. Section IV presents some proper experiments on hyperspectral images, that demonstrate the efficacy of the new algorithms. Section V concludes the paper.

II. Problem Formulation

Consider an over-complete dictionary $\boldsymbol{D} \in \mathbb{R}^{M\times K}$ and a data matrix $\boldsymbol{X}_m \in \mathbb{R}^{M\times N}$, which consists of vectors $\mathbf{x}_m^{(i)}$ ($i = 1, \ldots, N$) that are expected to obey some smoothness property. The smoothness property of interest in this work is a relation of the form

$$\left\|\mathbf{x}_m^{(i)} - \mathbf{x}_m^{(i+1)}\right\|_1 \leq \varepsilon, \quad i = 1, 2 \ldots, N-1 \;, \tag{1}$$

for some suitable small positive constant ε, that informs us that adjacent vectors in $\boldsymbol{X}_m$ are expected to have similar values. As an example, consider that $\boldsymbol{X}_m$ may be a sufficiently small block of neighboring hyper-pixels of some hyperspectral image. In the following, we use the term piece-wise smooth signal to refer to signals that can be decomposed into a number of such smooth blocks $\boldsymbol{X}_m$, $m = 1, \ldots, P$. *Note that since, in general, a given dataset may contain blocks that violate (1), a mechanism to detect and handle such non-smooth blocks must also be considered, as explained further in Section III-C.*

The scope of this work is to compute a so-called sparse approximation matrix $\boldsymbol{G}_m \in \mathbb{R}^{K\times N}$, so that

$$\boldsymbol{X}_m \approx \boldsymbol{D}\,\boldsymbol{G}_m \;. \tag{2}$$

Different from general sparse coding algorithms that exist in literature, the focus of this work is on exploiting the smoothness relation (1) in order to derive low complexity solutions, without sacrificing performance. As it will be demonstrated in the sequel, the inherent structure of the data can be effectively used as a-priori information so as to design sparse coding algorithms with remarkably reduced computational cost as compared to standard sparse coding algorithms that ignore the existing relations among homogeneous signals.

III. Efficient Sparse Coding

In this section, we derive two novel schemes that can be used for the efficient sparse coding of piece-wise smooth signals.

A. The proposed schemes

In general, the computational complexity of sparse coding algorithms is mainly due to the problem of determining a proper support set, defined as the set of dictionary atoms employed in the representation, for the signal of interest. In particular, an optimal solution would require a combinatorial optimization algorithm that examines all possible support sets, an approach which is of-course not feasible in practise. Furthermore, the determination of the support set has dominant complexity also in various practical algorithms, such as the OMP, which constitutes a greedy approach to approximate the solution of exhaustive search.

In the case of interest in this work, where a matrix $\boldsymbol{X}_m$ that contains "similar" signals as described by (1) is to be sparsely coded, significant computational gains would be offered by the assumption that the same support set can be effectively used for all vectors $\mathbf{x}_m^{(i)}$. In particular, a proper support set, denoted as $\mathcal{S}_m$, could be computed by executing some suitable algorithm only once for the whole block of signals, considering some suitable "representative" signal for the block.

Thus, instead of building the support for each data-vector independently, it is proposed here to calculate the support of their centroid signal, using some suitable sparse coding algorithm (e.g. OMP), considering that it can accurately represent all signals in the block. Consequently, considering the part of the dictionary that comprises of the respective atoms identified by the support $\mathcal{S}_m$ as $\boldsymbol{D}_{\mathcal{S}_m}$, a new regularized optimization problem emerges to compute efficiently the optimal weights for each data-vector in $\boldsymbol{X}_m$, exploiting further the smoothness priors. More analytically, by incorporating a suitable total-variation (TV) regularizer, the following optimization problem is proposed for the identification of the optimal weights $\boldsymbol{G}_{\mathcal{S}_m}^{(o)}$ for the block of signals $\boldsymbol{X}_m$,

$$\boldsymbol{G}_{\mathcal{S}_m}^{(o)} = \arg\min_{\boldsymbol{G}_{\mathcal{S}_m}} \|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}_m}\boldsymbol{G}_{\mathcal{S}_m}\|_F^2 + \lambda TV(\boldsymbol{G}_{\mathcal{S}_m}) \;, \tag{3}$$

where $\boldsymbol{G}_{\mathcal{S}_m}$ and λ denote the corresponding representation coefficients matrix and a positive constant that controls the relative importance of the smoothness, respectively. Note that since the support set has been identified, $\boldsymbol{G}_{\mathcal{S}_m}$ is not a sparse matrix, as it contains only the non-zero elements of the respective sparse matrix $\boldsymbol{G}_m$. Moreover,

$$TV(\boldsymbol{G}_{\mathcal{S}_m}) = \sum_{i=1}^{N-1}\left\|\mathbf{g}_{\mathcal{S}_m}^{(i)} - \mathbf{g}_{\mathcal{S}_m}^{(i+1)}\right\|_1 \tag{4}$$

which denotes a vector extension of the non-isotropic TV [11], indicating that the sparse coding coefficients of signals that satisfy relation (1) should not alter significantly.

Furthermore, the optimization problem in (3) can be given in a more compact form, as

$$\boldsymbol{G}_{\mathcal{S}_m}^{(o)} = \arg\min_{\boldsymbol{G}_{\mathcal{S}_m}} \|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}_m}\boldsymbol{G}_{\mathcal{S}_m}\|_F^2 + \lambda\,\|\boldsymbol{R}\boldsymbol{G}_{\mathcal{S}_m}\|_1 \tag{5}$$

where matrix $\boldsymbol{R}$ is a linear operator calculating the horizontal finite differences of the coding matrix $\boldsymbol{G}_{\mathcal{S}_m}$.

When $\lambda = 0$, the optimization problem in (5) reduces to a simple linear least-squares problem that has a closed-form solution. In the following, we denote as *Scheme 1* this particular case.

Accordingly, when $\lambda > 0$, the second proposed scheme arises, denoted as *Scheme 2*. By incorporating the above mentioned TV regularizer, Scheme 2 takes into account the smoothness attribute of the signals. As expected, this scheme enjoys significant noise removal properties. It should be mentioned that when $\lambda > 0$, relation (5) constitutes a challenging problem to be solved due to the non-smooth TV term. In view of this, the extended alternating direction method of multipliers

(ADMM) strategy [19]–[22] is proposed to surmount this difficulty. The details of this approach are given in the following sub-section.

B. Optimization via ADMM

Employing the ADMM methodology to (5), we arrive at the following constrained optimization problem,

$$\min_{\boldsymbol{G}_{\mathcal{S}_m}} \|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}_m}\boldsymbol{G}_{\mathcal{S}_m}\|_F^2 + \lambda\|\boldsymbol{V}_2\|_1$$
$$s.t.\, \boldsymbol{V}_1 - \boldsymbol{G}_{\mathcal{S}_m} = 0, \boldsymbol{V}_2 - \boldsymbol{R}\boldsymbol{V}_1 = 0\ , \tag{6}$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ denote auxiliary variables. It is noteworthy that in (6) we introduced two auxiliary variables, namely $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$.

The corresponding augmented Lagrangian function is given by

$$\mathcal{L}(\boldsymbol{G}_{\mathcal{S}_m}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{B}_1, \boldsymbol{B}_2) = \frac{1}{2}\|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}_m}\boldsymbol{G}_{\mathcal{S}_m}\|_F^2$$
$$+\lambda\|\boldsymbol{V}_2\|_1 + \frac{b}{2}\|\boldsymbol{V}_1 - \boldsymbol{G}_{\mathcal{S}_m} + \boldsymbol{B}_1/b\|_F^2$$
$$+\frac{b}{2}\|\boldsymbol{V}_2 - \boldsymbol{R}\boldsymbol{V}_1 + \boldsymbol{B}_2/b\|_F^2\ , \tag{7}$$

where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ denote the Lagrange multiplier matrices associated with the constraints, [11], [21], and $b > 0$ stands for the penalty parameter. Thus, a sequence of individual sub-problems emerges.

The closed form solution of the sub-problem for $\boldsymbol{G}_{\mathcal{S}_m}$ is

$$\nabla_{\boldsymbol{G}_{\mathcal{S}_m}}\mathcal{L} = 0 \Rightarrow \boldsymbol{G}_{\mathcal{S}_m} = (\boldsymbol{D}_{\mathcal{S}_m}^T\boldsymbol{D}_{\mathcal{S}_m} + b\boldsymbol{I})^{-1}$$
$$+ (\boldsymbol{D}_{\mathcal{S}_m}^T\boldsymbol{X}_m + \boldsymbol{B}_1 + b\boldsymbol{V}_1)\,, \tag{8}$$

where $\boldsymbol{I}$ denotes the identity matrix. Note that the inverse matrix $(\boldsymbol{D}_{\mathcal{S}_m}^T\boldsymbol{D}_{\mathcal{S}_m} + b\boldsymbol{I})^{-1}$ and the product $\boldsymbol{D}_{\mathcal{S}_m}^T\boldsymbol{X}_m$ can be calculated only once, as they remain fixed at each iteration.

The closed form solution of sub-problem $\boldsymbol{V}_1$ is

$$\nabla_{\boldsymbol{V}_1}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_1 = (\boldsymbol{R}^T\boldsymbol{R} + b\boldsymbol{I})^{-1}$$
$$+(\boldsymbol{G}_{\mathcal{S}_m} - \boldsymbol{B}_1/b + \boldsymbol{R}^T\boldsymbol{V}_2 + \boldsymbol{R}^T\boldsymbol{B}_2/b)\,. \tag{9}$$

Again, the inverse matrix $(\boldsymbol{R}^T\boldsymbol{R} + b\boldsymbol{I})^{-1}$ can be calculated only once at the beginning of the proposed algorithm.

Additionally, the solution of sub-problem $\boldsymbol{V}_2$ is given by

$$\nabla_{\boldsymbol{V}_2}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_2 = soft(\boldsymbol{R}\boldsymbol{V}_1 - \boldsymbol{B}_2/b, \lambda/b)\,, \tag{10}$$

where the $soft(., \tau)$ stands for the soft-thresholding function $x = sign(x)max(\mid x\mid - \tau)$.

Finally, the update rules for the Lagrangian multiplier matrices are

$$\boldsymbol{B}_1^{k+1} = \boldsymbol{B}_1^k + b(\boldsymbol{V}_1 - \boldsymbol{G}_{\mathcal{S}})$$
$$\boldsymbol{B}_2^{k+1} = \boldsymbol{B}_2^k + b(\boldsymbol{V}_2 - \boldsymbol{R}\boldsymbol{V}_1)\,. \tag{11}$$

C. Detection and handling of non-smooth blocks

In this sub-section, we summarise the proposed sparse coding techniques. To complete the derivation of an effective block-based sparse coding algorithm, as developed in the previous paragraphs, it is very important to furnish it with a mechanism, which will detect and handle the case where $\boldsymbol{X}_m$ contains vectors that cannot be accurately encoded by the same support set. In our approach, we employ a simple condition that examines if the representation error for each vector $\mathbf{x}_m^{(i)}$ is sufficiently small. In more detail, we examine if

$$\left\|\mathbf{x}_m^{(i)} - \boldsymbol{D}_{\boldsymbol{S}_m}\mathbf{g}_{\boldsymbol{S}_m}^{(i)}\right\|_2 < T\ , \tag{12}$$

where T is some properly defined threshold and $\mathbf{g}_{\boldsymbol{S}_m}^{(i)}$ denotes the i-th column of matrix $\boldsymbol{G}_{\boldsymbol{S}_m}^{(o)}$. All vectors $\mathbf{x}_m^{(i)}$ for which relation (12) is violated, participate in a new step of the algorithm where their centroid is computed, the corresponding support is identified and the respective encoding coefficients are computed, as detailed in the previous paragraph. This procedure could be employed recursively, until the number of signals that violate the condition becomes very small, in which case the OMP algorithm could be employed for each one. In the numerical experiments conducted, however, it was found out that such a recursive procedure is rarely required for more than a few recursions.

Given a piece-wise smooth input signal $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_P]$, where $\boldsymbol{X}_m$ may be non-overlapping patches derived from a hyperspectral image, a proper value for the threshold T, required in relation (12), can be estimated in a more insightful way by the following procedure,

1) Compute the centroids $x_{m,c}$, for each block $\boldsymbol{X}_m$, $m = 1, 2, \ldots, P$.
2) Compute the support sets $\mathcal{S}_m$, $m = 1, 2, \ldots, P$, using a sparse coding algorithm such as OMP.
3) Compute the sparse approximation coefficients for each of the centroids, and the respective representation errors, say e_m, $m = 1, 2, \ldots, P$.
4) Compute the mean $\boldsymbol{\mu}$ and variance σ^2 of e_m.
5) Set $T = \boldsymbol{\mu} + c \cdot \sigma$, where the positive constant c controls the approximation error that is acceptable, offering a balance between the complexity and accuracy of the algorithm.

Thus, the overall algorithm consists of two main steps, where the first step employs the above mentioned procedure for the estimation of the threshold T, while the second step implements the sparse coding method analyzed in Section III-A. *It should be stated that the proposed algorithm is not limited to use the OMP algorithm for determining the support of the centroid. Rather, any sparse coding algorithm could be also employed.* Table I summarizes the proposed scheme.

IV. Numerical Results

To demonstrate the effectiveness of the proposed algorithm, some proper numerical experiments were conducted, in the context of computing sparse representations of hyperspectral images. To validate the accuracy of the sparse coding algorithms we compared the reconstructed hyperspectral images with the corresponding original images. In particular, hyperspectral images from the iCVL dataset [23] were used to evaluate the proposed schemes. The provided images are $1300 \times 1300 \times 31$ dimensional cubes, with $d = 31$ spectral wavelengths, pertaining to a variety of natural scenes.

TABLE I
THE PROPOSED SPARSE CODING ALGORITHM

Input: Data matrix $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_P]$, dictionary $\boldsymbol{D} \in \mathbb{R}^{M \times K}$, sparsity level, number of iterations $\boldsymbol{J}$, penalty parameter λ
Output: Sparse coding matrix $\boldsymbol{G} \in \mathbb{R}^{K \times P}$

Precompute $(\boldsymbol{R}^T \boldsymbol{R} + \boldsymbol{I})^{-1}$
// Stage A: Threshold estimation procedure
Find the support $\mathcal{S}_m$ for each block ($m = 1, 2, \ldots, P$) and the threshold error T, following the procedure in Section III-C
// Stage B: Sparse coding procedure
for $m = 1$ **to** P **do**
  if $\lambda = 0$ **then**
    $\boldsymbol{G}_{\mathcal{S}} = (\boldsymbol{D}_{\mathcal{S}_m}^T \boldsymbol{D}_{\mathcal{S}_m})^{-1} \boldsymbol{D}_{\mathcal{S}_m}^T \boldsymbol{X}_m$
    Identify the signals $\mathbf{x}_m^{(i)}$ which violate condition (12) and repeat the process until the condition is met for all signals in the block $\boldsymbol{X}_m$.
  else if $\lambda > 0$ **then**
    Precompute $(\boldsymbol{D}_{\mathcal{S}_m}^T \boldsymbol{D}_{\mathcal{S}_m} + b\boldsymbol{I})^{-1}$, $\boldsymbol{D}_{\mathcal{S}_m}^T \boldsymbol{X}_m$
    for $j = 1$ **to** J **do**
      Update $\boldsymbol{G}_{\mathcal{S}_m}$ via (8)
      Update $\boldsymbol{V}_1$ via (9)
      Update $\boldsymbol{V}_1$ via (10)
      Update the Lagrange multipliers via (11)
    end for
    Identify the signals $\mathbf{x}_m^{(i)}$ which violate condition (12) and repeat the process until the condition is met for all signals in the block $\boldsymbol{X}_m$.
  end if
end for

Furthermore, a second dataset which we used was the well-known AVIRIS Cuprite hyperspectral image [24], which is a $2776 \times 754 \times 224$ dimensional cube with $d = 224$ spectral bands. The images were processed into small non-overlapping patches with size $(n \times n \times d)$. Thus, each patch forms a block $\boldsymbol{X}_m \in \mathbb{R}^{d \times N}$, where $N = n^2$. Note that, such blocks were described in Section II. After some experimentation, $n = 10$ was found to be a proper choice for the cases considered. Also, in all experiments, a fixed dictionary was used which was learned by employing the K-SVD [25] and OMP algorithms, on a suitable image dataset that did not include the images considered for sparse encoding. For all cases, the dictionaries employed $K = 1.024$ atoms. For the proposed schemes the OMP was used for determining the support for each block.

To this end, two sets of experimental results are given, the first of which focuses on sparse coding using the original images of the datasets used, while the second one focuses on the problem of sparse coding of images corrupted by severe noise. The simulations were performed using Matlab (2019a), running on a personal computer with an Intel i7-8700 CPU at 3.40 GHz with 16 GB of RAM.

A. Sparse coding of locally homogeneous data

In our first experiment, hyperspectral images from the iCVL and the AVIRIS datasets were employed, while measurements of the execution time as well as the peak-to-noise ratio (PSNR) between the original and the reconstructed images were collected. We examine the first scheme of the proposed algorithm, corresponding to the case where $\lambda = 0$, and compare its performance to the batch-OMP algorithm [6].

TABLE II
AVERAGE PSNR AND EXECUTION TIME OVER 40 HYPERSPECTRAL OF SIZE $1300 \times 1300 \times 31$ FROM ICVL [23], BETWEEN THE PROPOSED SCHEME 1 FOR $\lambda = 0$ AND THE BATCH-OMP [6].

Method	Sparsity level = 9		Sparsity level = 12	
	Time [sec]	PSNR [dB]	Time [sec]	PSNR [dB]
batch-OMP	57.90	51.78	76.70	54.39
Scheme 1 ($\lambda = 0$)	4.20	51.30	4.84	53.90

TABLE III
PSNR AND EXECUTION TIME OVER THE AVIRIS CUPRITE HYPERSPECTRAL IMAGE [24] OF SIZE $2776 \times 754 \times 224$, BETWEEN THE PROPOSED SCHEME 1 FOR $\lambda = 0$ AND THE BATCH-OMP [6].

Method	Sparsity level = 20		Sparsity level = 23	
	Time [sec]	PSNR [dB]	Time [sec]	PSNR [dB]
batch-OMP	246.80	50.1623	314.66	53.13
Scheme 1 ($\lambda = 0$)	16.67	49.6970	17.39	52.75

Table II summarizes the results for the iCVL dataset while Table III summarizes the results for the AVIRIS Cuprite image [24]. It is evident that Scheme 1 ($\lambda = 0$) significantly outperforms the batch-OMP algorithm in terms of execution time. Furthermore, the proposed scheme has a very small performance degradation as compared to the more computationally demanding batch-OMP algorithm. Also, it is worth noting that this degradation can be alleviated by a small increase of the sparsity level, while the execution time still remains significantly lower than that of the batch-OMP algorithm.

B. Sparse coding of noisy locally homogeneous data

In this scenario, we consider the problem of computing the sparse representations of 20 hyperspectral images from the iCVL dataset, after they have been corrupted by severe noise. In greater detail, the images were degraded by additive white Gaussian noise, corresponding to three different levels of Signal to Noise Ratios (SNRs), namely 20, 15, 10 dB. Various algorithms for sparse coding were examined in order to quantify the merits of the proposed schemes.

Figure 1 demonstrates the average PSNRs achieved by various algorithms for the SNR values considered. It is clear, that the proposed Scheme 2 ($\lambda > 0$), notably outperforms the other methods. Although, SUnSAL-TV [11] exhibits similarly good performance, its high execution time is its major bottleneck, rendering it significantly slow in comparison with our fast Scheme 2 ($\lambda > 0$). Furthermore, it is noteworthy that even Scheme 1 ($\lambda = 0$), outperforms the batch-OMP algorithm. Although the proposed Scheme 1 does not employ the noise-resilient TV regularizer, it still enjoys some denoising properties. This can be explained by considering that the centroid computed for each block is, in essence, a denoised, average vector that represents all noisy signals in this block.

Overall, it is evident that the proposed sparse coding schemes not only accomplish remarkably lower computation times than the other considered sparse coding techniques, but also exhibit superior performance in terms of accuracy.

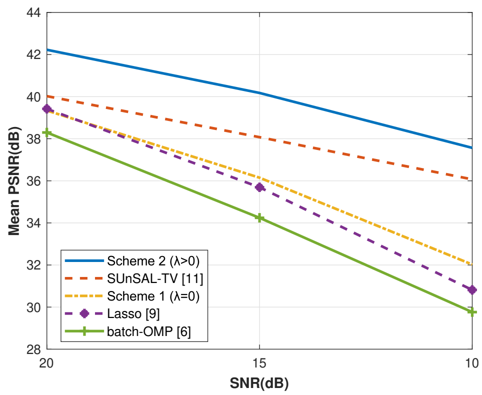

Fig. 1. Average PSNR over 20 images for different levels of SNRs between the proposed Scheme 2 ($\lambda > 0$),the Scheme 1 ($\lambda = 0$), the SUnSAL-TV [11], the Lasso [9] and the batch-OMP [6].

TABLE IV
AVERAGE RUNTIME FOR SPARSE CODING ALGORITHMS TO RECONSTRUCT A HYPERSPECTRAL IMAGE OF SIZE $1300 \times 1300 \times 31$.

Method	Scheme 2 ($\lambda > 0$)	SUnSAL-TV [11]	Scheme 1 ($\lambda = 0$)	Lasso [9]	batch-OMP [6]
time[sec]	54.61	2617	4.20	720.81	57.90

V. CONCLUSIONS

In this work, the problem of sparse coding of piece-wise smooth (locally homogeneous) signals was considered. The smoothness property of the input signals was exploited so as to derive fast schemes that alleviate the need for computing the support of dictionary atoms for each input signal separately. Furthermore, a total-variation regularized cost function was proposed for the problem of computing the required sparse representation coefficients. An alternating direction method of multipliers based method was employed for optimizing the proposed cost function. The novel schemes were shown to offer significantly lower computational complexity as compared to other, state-of-the-art algorithms, without lacking in terms of accuracy. Moreover, the proposed schemes offer state-of-the-art denoising performance, at a fraction of the computational cost.

REFERENCES

[1] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[2] V. Papyan, Y. Romano, J. Sulam, and M. Elad, "Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 72–89, Jul. 2018.

[3] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[4] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary Learning for Sparse Coding: Algorithms and Convergence Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1356–1369, Jul. 2016.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of the Asilomar Conference on Signals, Systems & Computers*, vol. 1. Publ by IEEE, 1993, pp. 40–44.

[6] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit," *CS Technion*, vol. 40, 2008.

[7] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, Jan. 1998.

[9] R. Tibshirani and R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 58, pp. 267—-288, 1994. [Online]. Available: http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7574

[10] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Representations of piecewise smooth signals on graphs," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 6370–6374.

[11] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11 PART1, pp. 4484–4502, 2012.

[12] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5096–5110, 2013.

[13] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.

[14] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. on Image Processing*, vol. 18, no. 11, pp. 2419–2434, Nov 2009.

[15] X. Song, L. Wu, H. Hao, and W. Xu, "Hyperspectral Image Denoising Based on Spectral Dictionary Learning and Sparse Coding," *Electronics*, vol. 8, no. 1, p. 86, Jan. 2019. [Online]. Available: http://www.mdpi.com/2079-9292/8/1/86

[16] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral Super Resolution of Hyperspectral Images via Coupled Dictionary Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2777–2797, May. 2019.

[17] T. Dundar and T. Ince, "Sparse Representation-Based Hyperspectral Image Classification Using Multiscale Superpixels and Guided Filter," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 246–250, Feb. 2019.

[18] A. Gkillas, D. Ampeliotis, and K. Berberidis, "Efficient coupled dictionary learning and sparse coding for noisy piecewise-smooth signals: Application to hyperspectral imaging," in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2020.

[19] 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, no. 1-3, pp. 293–318, apr 1992.

[20] S. Boyd, N. Parikh, E. Chu, J. Eckstein, 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. 1, pp. 1–122, 2010.

[21] Y. Li, X. Xie, and Z. Yang, "Alternating Direction Method of Multipliers for Solving Dictionary Learning Models," *Communications in Mathematics and Statistics*, vol. 3, no. 1, pp. 37–55, Mar. 2015.

[22] C. G. Tsinos, A. A. Rontogiannis, and K. Berberidis, "Distributed Blind Hyperspectral Unmixing via Joint Sparsity and Low-Rank Constrained Non-Negative Matrix Factorization," *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 160–174, Apr. 2017.

[23] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*. Springer, 2016, pp. 19–34.

[24] "AVIRIS - Airborne Visible / Infrared Imaging Spectrometer - Data." [Online]. Available: https://aviris.jpl.nasa.gov/data/free_data.html

[25] 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, Nov. 2006.