EFFICIENT COUPLED DICTIONARY LEARNING AND SPARSE CODING FOR NOISY PIECEWISE-SMOOTH SIGNALS: APPLICATION TO HYPERSPECTRAL IMAGING

Alexandros Gkillas, Dimitris Ampeliotis, Kostas Berberidis

Dept. of Computer Engineering & Informatics, University of Patras 26500, Patras, Greece.
E-mails: {st1003586, ampeliot berberid}@ceid.upatras.gr

ABSTRACT

Given two datasets that belong to different feature spaces and both correspond to the same underlying phenomenon, the scope of *coupled dictionary learning* is to compute two dictionaries, one for each dataset, so that each dataset is approximated using the respective dictionary but the same sparse coding matrix. In this work, the focus is on a particular, yet widespread, form of this problem in which the datasets correspond to slowly varying (piece-wise smooth) signals, and the measurements contain severe noise. A novel coupled dictionary learning technique is developed by including a suitable total-variation-based regularization term in the cost function. Furthermore, exploiting the smoothness of the datasets, new fast sparse coding algorithms are derived. The new techniques achieve effective modeling of the smooth signal and significantly alleviate the effects of noise. Finally, extensive simulation results for the problem of spectral super-resolution of hyperspectral images are provided, demonstrating the performance improvements offered by the derived techniques.

***Index Terms*—** Coupled dictionary learning, domain adaptation, sparse coding, total variation, hyperspectral imaging

1. INTRODUCTION

Dictionary learning and the associated sparse representation theory have been particularly ground-breaking in the field of signal processing, achieving remarkable results in a variety of applications, including image denoising, image inpainting, super-resolution and classification, among others [1], [2], [3]. So far, research efforts have mainly focused on dictionary learning in a single sparse feature space, in centralized [4], [5] and distributed scenarios [6]. However, in several applications and settings [3], [7–12], coupled sparse feature spaces arise, as, for example, in low and high-resolution hyperspectral images [7]. Coupled dictionary learning as a domain adaptation procedure aims at transferring knowledge across different but related feature spaces (domains) [13] .

This work was supported partly by the University of Patras, the RPF, Cyprus, under the project INFRASTRUCTURES/1216/0017 (IRIDA), and the GSRT, Greece, under the project FP7 ERANET COM-MED.

The coupled dictionary learning (CDL) model seeks to reveal the fundamental relationship between the two spaces, often referred to as the *observation* and *latent*, so that the sparse representation of the signals in the *observation* space can be effectively used to describe the corresponding signals in the *latent* space [14]. Formally, the CDL problem can be defined as the learning of a pair of dictionaries $\boldsymbol{D}_x \in \mathbb{R}^{P\times K}$ and $\boldsymbol{D}_y \in \mathbb{R}^{M\times K}$ in such a way that the signals $\boldsymbol{X} \in \mathbb{R}^{P\times N}$ in the latent feature space and the signals $\boldsymbol{Y} \in \mathbb{R}^{M\times N}$ in the observation space can be approximated through the respective dictionary and a common sparse coding matrix $\boldsymbol{G} \in \mathbb{R}^{K\times N}$, as shown by the relations $X \approx \boldsymbol{D}_x\mathbf{G}$, $Y \approx \boldsymbol{D}_y\boldsymbol{G}$.

In this study, we consider the problem of learning coupled overcomplete dictionaries from locally homogeneous (piece-wise smooth) signals, as for example, hyperspectral images in which neighboring pixels exhibit strong spatial and spectral similarities [15]. Also, we assume that the considered data is corrupted by severe noise as, for example, is the case in remote sensing applications where the hyperspectral images are affected by various factors such as atmospheric degradations and sensor imperfections [16]. Due to the presence of noise, the learning of the dictionaries becomes a more challenging task. This difficulty can be overcome by exploiting the underlying homogeneity of the noisy data via incorporation of a proper total-variation (TV) regularizer [17, 18] at the cost functions of the proposed algorithms. Thus, the derived methods are suitable for several hyperspectral applications such as spatial and spectral super-resolution and unmixing, without the necessity for a costly denoising pre-processing step.

The combination of the total-variation regularizer and the l_1-norm was first introduced in [15], leading to a sparse coding algorithm named SUnSAL-TV, which was applied to the hyperspectral unmixing problem. In this paper we investigate the employment of a TV regularizer in the learning procedure as well. Note that this different from [15] where the TV term was confined to the sparse coding step. The new approach is shown to lead to the construction of coupled dictionaries which turn out to be more suitable for the considered smooth, noisy signals. Coupled dictionary learning in the context of hyperspectral images is also considered in [7], however, no smoothness priors are considered. To sum up, the key contributions of this paper are the following:

- A novel coupled dictionary learning algorithm, suitable for piecewise smooth and noisy signals is developed, as described in Section 3.
- Exploiting the piecewise smoothness of the considered datasets, two variants of a fast, sparse coding algorithm are derived, as described in Section 4.

Finally, simulation results, for the problem of spectral super-resolution are given in Section 5.

2. PROBLEM FORMULATION

Consider the signals $\boldsymbol{X}$ and $\boldsymbol{Y}$ in the latent and observation space respectively. Assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are modelled as

$$\boldsymbol{X} = \boldsymbol{Z} + \boldsymbol{W}_x \quad \& \quad \boldsymbol{Y} = \boldsymbol{A} + \boldsymbol{W}_y \tag{1}$$

where $\boldsymbol{W}_x \in \mathbb{R}^{P \times N}$, $\boldsymbol{W}_y \in \mathbb{R}^{M \times N}$ denote zero-mean noise terms, whereas $\boldsymbol{Z}$ and $\boldsymbol{A}$ stand for locally homogeneous (piecewise smooth) signals, in the sense that neighboring vectors in $\boldsymbol{Z}$ and $\boldsymbol{A}$, say, $\mathbf{z^i}$, $\mathbf{z^{i+1}}$ and $\mathbf{a^i}$, $\mathbf{a^{i+1}}$, are expected to satisfy some similarity relation, as for example

$$\left\|\mathbf{z^i} - \mathbf{z^{i+1}}\right\|_1 \leq \varepsilon_z, \quad \left\|\mathbf{a^i} - \mathbf{a^{i+1}}\right\|_1 \leq \varepsilon_a, \tag{2}$$

where ε_z and ε_a denote some small, positive constants.

Given the noisy signals $\boldsymbol{X}$ and $\boldsymbol{Y}$, our goal is to learn two coupled dictionaries $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$, based on the signals $\boldsymbol{X}$ and $\boldsymbol{Y}$, in such a way that the original smooth signals $\boldsymbol{Z}$ and $\boldsymbol{A}$ are accurately encoded by the same sparse coding matrix $\boldsymbol{G}$.

3. COUPLED DICTIONARY LEARNING

3.1. A new cost function for CDL

Taking into consideration the underlying structure of the noisy data $\boldsymbol{X}$ and $\boldsymbol{Y}$, we propose a cost function that includes the required data fidelity Frobenius norm terms, a sparsity promoting l_1-norm, and a total-variation cost that captures the local homogeneity (smoothness) of the underlying signals. Thus, the proposed CDL problem is formulated as

$$\underset{\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}}{\arg\min} \left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\left\|\boldsymbol{G}\right\|_1 + \mu TV(\boldsymbol{G}), \tag{3}$$

where λ and μ are positive scalar constants controlling the relative importance of the sparsity level and the smoothness, respectively. Also,

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

denotes a vector extension of the non-isotropic TV [15], which promotes smooth variations between subsequent elements of the sparse coding matrix columns $\mathbf{g}^i$ and $\mathbf{g}^{i+1}$. Problem (3) can be written in a more compact form, as

$$\underset{\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}}{\arg\min} \left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\left\|\boldsymbol{G}\right\|_1 + \mu\left\|\boldsymbol{R}\boldsymbol{G}\right\|_1, \tag{5}$$

where matrix $\boldsymbol{R}$ is the horizontal finite difference operator.

It should be highlighted that relation (5) constitutes a non-convex problem. To overcome this difficulty, we employ an alternating optimization (AO) scheme, splitting the dictionary learning problem into two sub-problems, namely, dictionary update and sparse coding [19–21]. In our case, the sparse coding sub-problem, although convex, requires special treatment, due to the non-smooth l_1 and TV terms. In light of this, we follow the ADMM strategy [7], [22–25] that is able to treat such issues.

3.2. Optimization via ADMM

Following the ADMM optimization methodology, we consider an equivalent constrained version of (5) given by

$$\begin{aligned} \min_{\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}} & \left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\left\|\boldsymbol{V}_1\right\|_1 \\ & + \mu\left\|\boldsymbol{V}_3\right\|_1 \\ s.t. \quad & \boldsymbol{V}_1 - \boldsymbol{G} = 0,\, \boldsymbol{V}_2 - \boldsymbol{G} = 0,\, \boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2 = 0, \\ & \left\|\boldsymbol{D}_x(:,i)\right\|_2^2 \leq 1, \left\|\boldsymbol{D}_y(:,i)\right\|_2^2 \leq 1, \end{aligned} \tag{6}$$

where $\boldsymbol{D}_x(:,i)$ and $\boldsymbol{D}_y(:,i)$ denote the i-th atom of the respective dictionary and $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ and $\boldsymbol{V}_3$ are auxiliary variables. Note that in (6) we have followed a similar procedure as in [15] by introducing an additional dummy variable $\boldsymbol{V}2$, an approach that makes the optimization procedure significantly more tractable.

The augmented Lagrangian function of problem (6) is

$$\begin{aligned} &\mathcal{L}(\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3, \boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3) = \tfrac{1}{2}\left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2 + \\ &\lambda\left\|\boldsymbol{V}_1\right\|_1 + \tfrac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \tfrac{b_1}{2}\left\|\boldsymbol{V}_1 - \boldsymbol{G} + \boldsymbol{B}_1\right\|_F^2 + \\ &\mu\left\|\boldsymbol{V}_3\right\|_1 + \tfrac{b_2}{2}\left\|\boldsymbol{V}_2 - \boldsymbol{G} + \boldsymbol{B}_2\right\|_F^2 + \tfrac{b_3}{2}\left\|\boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2 + \boldsymbol{B}_3\right\|_F^2, \end{aligned} \tag{7}$$

where $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ denote the Lagrange multiplier matrices associated with the constraints, following the exact procedure as in [15], [7], [24], and $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$ stand for the penalty parameters. Hence, the following update rules are formed:

The sub-problem for $\boldsymbol{G}$ is solved via the relation

$$\begin{aligned} \nabla_{\boldsymbol{G}}\mathcal{L} = 0 \Rightarrow \boldsymbol{G} = & (\boldsymbol{D}_x^T\boldsymbol{D}_x + \boldsymbol{D}_y^T\boldsymbol{D}_y + b_1\boldsymbol{I} + b_2\boldsymbol{I})^{-1} \\ & (\boldsymbol{D}_x^T\boldsymbol{X} + \boldsymbol{D}_y^T\boldsymbol{Y} + \boldsymbol{B}_1 + b_1\boldsymbol{V}_1 + \boldsymbol{B}_2 + b_2\boldsymbol{V}_2), \end{aligned} \tag{8}$$

where $\boldsymbol{I}$ stands for the identity matrix.

The solution of the sub-problem for $\boldsymbol{V}_1$ derives from

$$\nabla_{\boldsymbol{V}_1}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_1 = soft(\boldsymbol{G} - \boldsymbol{B}_1/b_1, \lambda/b_1), \tag{9}$$

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

The closed form solution for $\boldsymbol{V}_2$ is given by

$$\begin{aligned} \nabla_{\boldsymbol{V}_2}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_2 = & (b_3\boldsymbol{R}^T\boldsymbol{R} + b_2\boldsymbol{I})^{-1} \\ & (b_2\boldsymbol{G} - \boldsymbol{B}_2 + b_3\boldsymbol{R}^T\boldsymbol{V}_3 + b_3\boldsymbol{R}^T\boldsymbol{B}_3). \end{aligned} \tag{10}$$

The variable $\boldsymbol{V}_3$ can be updated as

$$\nabla_{\boldsymbol{V}_3}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_3 = soft(\boldsymbol{R}\boldsymbol{V}_2 - \boldsymbol{B}_3/b_3, \mu/b_3). \tag{11}$$

The update rule for the coupled dictionaries derives from solving the following equation

$$\nabla_{\boldsymbol{D}_x}(\left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2) \quad \& \quad \nabla_{\boldsymbol{D}_y}(\left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2). \tag{12}$$

In order to accelerate this step, we follow the procedure proposed in [7], [24] by updating the dictionaries column by column. More analytically, the updated scheme becomes

$$\begin{aligned} \boldsymbol{D}_x^{j+1}(:,i) &= \boldsymbol{D}_x^j(:,i) + (\boldsymbol{X}\boldsymbol{G}(i,:)^T/(\zeta_i + \delta)), \\ \boldsymbol{D}_y^{j+1}(:,i) &= \boldsymbol{D}_y^j(:,i) + (\boldsymbol{Y}\boldsymbol{G}(i,:)^T/(\zeta_i + \delta)), \end{aligned} \tag{13}$$

where j stands for the number of iterations, δ is a small regularization value, and $\zeta_i = \boldsymbol{G}(i,:)\boldsymbol{G}(i,:)^T$.

Algorithm 1: CDL from Noisy and Smooth data

Input: training signals $\boldsymbol{X} \in \mathbb{R}^{P \times N}$,$\boldsymbol{Y} \in \mathbb{R}^{M \times N}$, number of iterations $\boldsymbol{J}$, penalty parameter b_1, b_2, b_3
Output: $\boldsymbol{D}_x \in \mathbb{R}^{P \times K}$, $\boldsymbol{D}_y \in \mathbb{R}^{M \times K}$, $\boldsymbol{G} \in \mathbb{R}^{K \times N}$

Precompute $\boldsymbol{D}_x$, $\boldsymbol{D}_y$, $\boldsymbol{B}_1 = \boldsymbol{B}_2 = \boldsymbol{B}_3 = 0$
for $j = 1$ **to** J **do**
  Update $\boldsymbol{G}$ via (8)
  Update $\boldsymbol{V}_1$ via (9)
  Update $\boldsymbol{V}_2$ via (10)
  Update $\boldsymbol{V}_3$ via (11)
  for $i = 1$ **to** K **do**
    Update dictionaries $\boldsymbol{D}_x$, $\boldsymbol{D}_y$ by atoms via (13)
  end for
  Normalize the atoms of the dictionaries
  Update the Lagrange multipliers via (14)
end for

Finally, the update rules of the Lagrangian multiplier matrices are given by

$$\begin{aligned} \boldsymbol{B}_1^{j+1} &= \boldsymbol{B}_1^j + b_1(\boldsymbol{V}_1 - \boldsymbol{G}), \\ \boldsymbol{B}_2^{j+1} &= \boldsymbol{B}_2^j + b_2(\boldsymbol{V}_2 - \boldsymbol{G}), \\ \boldsymbol{B}_3^{j+1} &= \boldsymbol{B}_3^j + b_1(\boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2). \end{aligned} \tag{14}$$

The overall algorithm is summarized in Algorithm 1.

4. FAST TV PROMOTING SPARSE CODING

Among the most widely used algorithms for sparse coding of hyperspectral images is the so-called SUnSAL-TV algorithm [15]. However, a major drawback of this approach is its high computational complexity, making it inappropriate for online real-world applications. Thus, in order to deal with such a critical issue, in this section we propose a novel procedure which effectively surmounts this drawback. The novel approach follows a block-processing methodology, where the assumption of piecewise smooth signals is efficiently employed to yield a remarkably reduced computational complexity. The proposed sparse coding scheme can be used in several hyperspectral imaging applications such as spatial and/or spectral super-resolution, denoising and unmixing.

4.1. Fast OMP-based sparse approximation

The OMP [26] is considered to be one of the most prominent algorithms for tackling the sparse coding problem. In the case of smooth signals, however, an approximate solution based on the OMP could be derived, assuming that a block of signals e.g., neighboring pixels, displaying homogeneity can be represented by the same sparse representation support $\boldsymbol{S}$, defined by the atoms involved in the representation. Thus, instead of building the support of each signal separately, we propose to calculate the support of their corresponding centroid signal using the OMP, considering that it can efficiently be used for all signals in the block. After finding the support, a simpler optimization problem can be employed to compute the optimal weights, for each of the signals in the block. We propose two schemes, where the first one consists in solving a linear least squares problem, and the second one utilizes a TV regularized linear least squares cost function, which is optimized using the ADMM method.

Algorithm 2: Fast Sparse Coding Promoting Total Variation

Input: Data matrix $\boldsymbol{X} \in \mathbb{R}^{P \times N}$, dictionary $\boldsymbol{D} \in \mathbb{R}^{P \times K}$, sparsity level s, number of iterations $\boldsymbol{J}$,
Output: Sparse coding matrix $\boldsymbol{G} \in \mathbb{R}^{K \times N}$

Precompute $(\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{I})^{-1}$
for $m = 1$ **to** k **do**
  Find the centroid signal of block m, $x_{m,c} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_m(:, i)$
  Use OMP to find the support $\boldsymbol{S}$ of the centroid $x_{m,c}$
  if $\mu = 0$ **then**
    $\boldsymbol{G}_{\mathcal{S}} = (\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{D}_{\mathcal{S}})^{-1}\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{X}_m$
  else if $\mu > 0$ **then**
    Precompute $(\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{D}_{\mathcal{S}} + b\boldsymbol{I})^{-1}$, $\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{X}_m$
    for $j = 1$ **to** J **do**
      Update $\boldsymbol{G}_{\mathcal{S}}$ via
      $\boldsymbol{G}_{\mathcal{S}} = (\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{D}_{\mathcal{S}} + b\boldsymbol{I})^{-1}(\boldsymbol{D}_{\mathcal{S}}^T\boldsymbol{X}_m + \boldsymbol{B}_1 + b\boldsymbol{V}_1)$
      Update $\boldsymbol{V}_1$ via
      $\boldsymbol{V}_1 = (\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{I})^{-1}(\boldsymbol{G}_{\mathcal{S}} - \boldsymbol{B}_1/b + \boldsymbol{R}^T\boldsymbol{V}_2 + \boldsymbol{R}^T\boldsymbol{B}_2/b)$
      Update $\boldsymbol{V}_2$ via
      $\boldsymbol{V}_2 = soft(\boldsymbol{R}\boldsymbol{V}_1 - \boldsymbol{B}_2/b, \mu/b)$
      Update the Lagrange multipliers via
      $\boldsymbol{B}_1^{j+1} = \boldsymbol{B}_1^j + b(\boldsymbol{V}_1 - \boldsymbol{G}_{\mathcal{S}})$
      $\boldsymbol{B}_2^{j+1} = \boldsymbol{B}_2^j + b(\boldsymbol{V}_2 - \boldsymbol{R}\boldsymbol{V}_1)$
    end for
  end if
end for

Defining the set of signals $\boldsymbol{X} \in \mathbb{R}^{P \times N}$ consisting of k blocks $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k]$, so that each block $\boldsymbol{X}_m = [x_1^m, x_2^m, \ldots, x_n^m]$, $m = 1, \ldots, k$ contains n homogeneous vectors such as neighboring pixels, the cost function proposed for the computation of the weights for block m becomes

$$\begin{aligned} &\min_{\boldsymbol{G}_{\mathcal{S}}} \|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{G}_{\mathcal{S}}\|_F^2 + \mu TV(\boldsymbol{G}_{\mathcal{S}}) \sim \\ &\min_{\boldsymbol{G}_{\mathcal{S}}} \|\boldsymbol{X}_m - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{G}_{\mathcal{S}}\|_F^2 + \mu \|\boldsymbol{R}\boldsymbol{G}_{\mathcal{S}}\|_1, \end{aligned} \tag{15}$$

where $\boldsymbol{D}_{\mathcal{S}}, \boldsymbol{G}_{\mathcal{S}}$ and $\boldsymbol{R}$ denote the selected atoms from the dictionary, the corresponding representation coefficients and the horizontal finite difference operator, respectively. Note that the first scheme corresponds to the case where $\mu = 0$, which can be solved in closed form.

When $\mu > 0$ optimization problem in (15) can be solved via ADMM, following a similar procedure as in the previous section. Due to space limitations, we omit the derivation steps and give below the complete description of the algorithm. The size n of the blocks employed can be selected through experimentation to verify that the involved signals are sufficiently homogeneous. Future work will focus on methodologies for the dynamic selection of this parameter. Algorithm 2 summarizes the proposed sparse coding algorithms.

5. NUMERICAL RESULTS

To demonstrate the efficacy and applicability of the proposed schemes some appropriate experimental tests were performed, in the context of the spectral super-resolution problem [7]. In more detail, hyperspectral images from the iCVL [27] dataset were used to generate the high-dimensionality dataset $\boldsymbol{X} \in \mathbb{R}^{31 \times N}$, containing data at 31 wavelengths in the $400 - 700$ nm spectrum for each "hyper-pixel", while the low-dimensionality dataset $\boldsymbol{Y} \in \mathbb{R}^{8 \times N}$ was generated by downsampling $\boldsymbol{X}$ along the spectral dimension.

Two sets of experimental results are given, where the first set focuses on coupled dictionary learning from noisy data, while the second set focuses on the sparse coding problem given that the coupled dictionaries are available.

5.1. Coupled dictionary learning from noisy data

The first set of experiments quantifies the performance of the proposed scheme, as compared to other approaches, for the problem of coupled dictionary learning from noisy data. We used 100 hyperspectral images to generate the training datasets, while another 100 images were used for testing. In particular, dataset $\boldsymbol{X}$ was generated by randomly selecting 1000 different 10×10 (hyperspectral) patches from the training images, so that $\boldsymbol{X}$ was a 31×100000 matrix. Accordingly, $\boldsymbol{Y}$ was generated by downsampling $\boldsymbol{X}$, leading to a 8×100000 matrix. Additive white Gaussian noise was added to both datasets corresponding to three different Signal to Noise Ratios (SNRs), namely 20, 15, and 10 dB. Various algorithms for coupled dictionary learning were used to compute the dictionaries $\boldsymbol{D}_x$ and $\boldsymbol{D}_y$. Given these dictionaries, the performance of each algorithm is measured in terms of the quality of super-resolution as follows. For each of the 100 testing images, $\boldsymbol{D}_y$ is used along with a sparse coding algorithm to compute $\boldsymbol{G}$. Then, $\boldsymbol{D}_x\boldsymbol{G}$ is computed as an estimate of the high spectral resolution image, and the Peak SNR is computed. At the testing phase, the batch-OMP method [20] is used for sparse coding, for all the examined schemes. Also, the dictionaries employed $K = 1024$ atoms, while the sparsity level was 6.

From Fig.1, we can deduce that our proposed algorithm namely, Algorithm 1 for $\mu > 0$ notably outperforms the other methods. In particular, for high levels of noise, the total variation term becomes more significant, allowing our algorithm to maintain high PSNR values in all cases.

5.2. Fast sparse coding of locally homogeneous data

In this scenario, having a pair of dictionaries fixed, derived via the previous procedure, we consider again the problem of spectral super-resolution in the case where the low spectral resolution images are corrupted by AWGN noise. In more detail, we use the dictionaries computed by the Algorithm 1 and examine various approaches for sparse coding.

According to Fig.2, the proposed Algorithm 2, for $\mu > 0$, exhibits superior performance as compared to the other approaches. Although, SUnSAL-TV [15] demonstrates good results, its high computational complexity is its basic drawback, rendering it slow in comparison with our fast Algorithm 2 for $\mu > 0$. It is notable that Algorithm 2, for $\mu = 0$, outperforms Batch-OMP, although it was derived as an approximation to the OMP. This may be explained by considering that the centroid computed for each block is, in essence, a denoised, average vector that represents all noisy signals in the block. Finally, Table 1 gives the average time, required for constructing one hyperspectral image. The simulations were performed in a Matlab (2018a) implementation running on an Intel i7-2700, CPU at 3,40GHz with 16 GB RAM. It is evident that the proposed sparse coding algorithms achieve significantly smaller computation times, without sacrificing performance. Thus, we conclude that the proposed algorithms exploit effectively the homogeneity of the data.

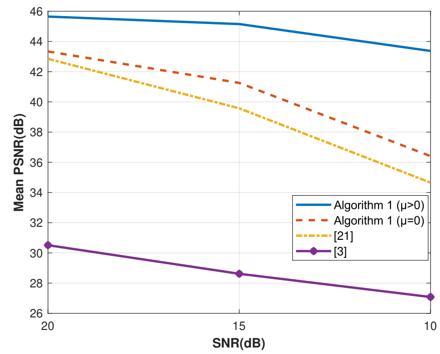

Fig. 1: Average PSNR over 100 images for different levels of SNRs between the proposed Algorithm 1 ($\mu > 0$), the Algorithm 1 without the TV regularizer ($\mu = 0$), the K-SVD based method in [3] using the OMP at the sparse coding stage, and the method in [21] (equations 15.42, 15.43).

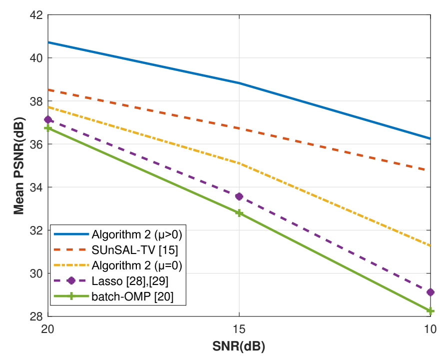

Fig. 2: Average PSNR over 10 images for different levels of SNRs between various sparse coding algorithms.

Table 1: Average runtime for Sparse Coding Algorithms to reconstruct a hyper-spectral image of size $1000 \times 1000 \times 31$.

Method	Algorithm 2 ($\mu > 0$)	SUnSAL-TV [15]	Algorithm 2 ($\mu = 0$)	Lasso [28, 29]	batch-OMP [20]
time[sec]	50.62	1835.64	5.35	543.19	47.31

6. CONCLUSIONS

In this work, the problem of coupled dictionary learning and the problem of fast sparse coding were investigated, for the case of locally homogeneous (smooth) data. For both problems, efficient algorithms were derived by employing proper total-variation regularizers, and solving the resulting problems by using the ADMM optimization algorithm. Simulation results for the problem of spectral super-resolution of hyperspectral images were conducted using real data and confirmed the effectiveness of the derived techniques.

7. REFERENCES

[1] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[2] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3501–3508.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov.2010.

[4] 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.

[5] K. Engan, S. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*. IEEE, 1999, pp. 2443–2446 vol.5.

[6] D. Ampeliotis, C. Mavrokefalidis, and K. Berberidis, "Distributed dictionary learning via projections onto convex sets," in *25th European Signal Processing Conference, EUSIPCO 2017*, vol. 2017-January. Institute of Electrical and Electronics Engineers Inc., Oct 2017, pp. 2294–2298.

[7] 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.

[8] D. A. Huang and Y. C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proceedings of the IEEE International Conference on Computer Vision*. Institute of Electrical and Electronics Engineers Inc., 2013, pp. 2496–2503.

[9] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, Sep 2014, pp. 3550–3557.

[10] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1284–1294, 2014.

[11] L. Sun, Z. Jiang, H. Song, Q. Lu, and A. Men, "Semi-Coupled Dictionary Learning with Relaxation Label Space Transformation for Video-Based Person Re-Identification," *IEEE Access*, vol. 6, pp. 12 587–12 597, Feb. 2018.

[12] Y. Zhang, E. Zhang, W. Chen, Y. Chen, and J. Duan, "Sparsity-based inverse halftoning via semi-coupled multi-dictionary learning and structural clustering," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 43–53, Jun. 2018.

[13] L. Zhang, W. Zuo, and D. Zhang, "LSDT: Latent Sparse Domain Transfer Learning for Visual Adaptation," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1177–1191, Mar. 2016.

[14] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

[15] 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.

[16] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sensing*, vol. 10, no. 3, Mar. 2018.

[17] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *J. of Math. Imaging and Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

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

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

[21] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*. Springer New York, 2010.

[22] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.

[23] 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 R in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

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

[25] 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.

[26] Y. C. Pati, Y. C. Pati, R. Rezaiifar, Y. C. P. R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition," *Conference Record of the Asilomar Conference on Signals, Systems & Computers*, vol. 1, pp. 40–44, 1993.

[27] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *European Conference on Computer Vision*, vol. 9911 LNCS. Springer, Cham, 2016, pp. 19–34.

[28] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, H. Ishwaran, K. Knight, J. M. Loubes, P. Massart, D. Madigan, G. Ridgeway, S. Rosset, J. I. Zhu, R. A. Stine, B. A. Turlach, S. Weisberg, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[29] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. STAT SOC. SER. B*, vol. 58, pp. 267–288, 1994.