COUPLED DICTIONARY LEARNING AND FAST SPARSE CODING ALGORITHMS FOR PIECEWISE SMOOTH SIGNALS IN SEVERE NOISE

Alexandros Gkillas, Dimitris Ampeliotis, Kostas Berberidis

Dept. of Computer Engineering & Informatics, University of Patras & C.T.I RU-2 26500, Patras, Greece. E-mails: {st1003586, ampeliot, berberid}@ceid.upatras.gr

ABSTRACT

Given two datasets, usually with (significantly) different dimensions, that both correspond to the same underlying signal, 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 encoding matrix. In this work, the focus is on a particular form of this problem in which the datasets correspond to a slowly varying (piecewise smooth) signal, and the measurements contain severe noise. A novel coupled dictionary learning technique is developed by properly including a total variation regularization term in the cost function. Furthermore, exploiting the smoothness of the datasets, new fast sparse coding algorithms are derived. The new techniques achieve correct modeling of the smooth signal and significantly alleviate the effects of noise. Finally, simulation results for the problem of spectral super-resolution of hyperspectral images, that demonstrate the effectiveness of the derived theoretical findings, are provided.

***Index Terms*—** Coupled dictionary learning, sparse coding, total variation, smooth signals, 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, and super-resolution, 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 numerous applications and settings [3], [7]–[12], coupled sparse feature spaces arise, as, for example, in low and high-resolution hyperspectral images [7]. In such cases, coupled dictionary learning methods are employed that yield a pair of dictionaries that describe the inherent relation in the data.

The coupled dictionary learning (CDL) model seeks to reveal the fundamental relationship between the two spaces, often referred to as the *observation* and *latent* feature space, 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 [13]. Formally, the CDL problem can be defined as the learning of a pair of dictionaries $\boldsymbol{D}_x \in \mathbb{R}^{PxK}$ and $\boldsymbol{D}_y \in \mathbb{R}^{MxK}$ in such a way that the signals $\boldsymbol{X} \in \mathbb{R}^{PxN}$ in the latent feature space and the signals $\boldsymbol{Y} \in \mathbb{R}^{MxN}$ in the observation space can be approximated through the respective dictionary and a common sparse coding matrix $\boldsymbol{G} \in \mathbb{R}^{KxN}$, as shown by the relations

$$\boldsymbol{X} \approx \boldsymbol{D}_x\boldsymbol{G}, \;\; \boldsymbol{Y} \approx \boldsymbol{D}_y\boldsymbol{G}\,. \tag{1}$$

In this study, we consider the problem of learning coupled overcomplete dictionaries from *locally homogeneous (piecewise smooth) signals,* as for example, hyperspectral images in which neighboring pixels exhibit strong spatial similarity [14]. Furthermore, we assume that the considered data is corrupted by severe noise. Due to the presence of noise, the learning of the dictionaries becomes a more challenging task. As shown later, this difficulty can be overcome by exploiting the underlying homogeneity of the noisy data via proper incorporation of a total-variation (TV) regularizer [15], [16], at the cost functions of the proposed algorithms.

Related prior work and contributions: The combination of the total variation regularizer and the l_1-norm was first introduced in [14], leading to a sparse coding algorithm named SUnSAL-TV, which was applied to the hyperspectral unmixing problem. Leveraging upon that idea, and different from [14], we propose to employ the TV regularizer throughout the (coupled) dictionary learning procedure. This approach is shown to lead to the construction of dictionaries which turn out to be more suitable for the considered smooth, noisy signals. Coupled dictionary learning in the context of hyper-spectral 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 locally homogeneous, noisy signals is developed, as described in Section 3.
- Exploiting the homogeneity 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, that demonstrate the superior performance of the new approaches, are given in Section 5.

2. PROBLEM FORMULATION

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

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{Z} + \boldsymbol{W_x} \\ \boldsymbol{Y} &= \boldsymbol{A} + \boldsymbol{W_y} \end{aligned} \tag{2}$$

where $\boldsymbol{W}_x \in \mathbb{R}^{PxN}, \boldsymbol{W}_y \in \mathbb{R}^{MxN}$ 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, $\boldsymbol{z}^{(i)}$, $\boldsymbol{z}^{(i+1)}$ and $\boldsymbol{a}^{(i)}$, $\boldsymbol{a}^{(i+1)}$, are expected to satisfy some similarity relation, as for example

$$\left\|\boldsymbol{z}^{(i)} - \boldsymbol{z}^{(i+1)}\right\|_1 \leq \boldsymbol{\varepsilon_z} \;\&\; \left\|\boldsymbol{a}^{(i)} - \boldsymbol{a}^{(i+1)}\right\|_1 \leq \boldsymbol{\varepsilon_a} \tag{3}$$

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 for matrix $\boldsymbol{G}$, and a total-variation cost that captures the local homogeneity (smoothness) of the underlying signals. Thus, the proposed CDL problem is formulated as

$$\min_{\boldsymbol{D}_x,\boldsymbol{D}_y,\boldsymbol{G}} \|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\|\boldsymbol{G}\|_1 + \mu TV(\boldsymbol{G}) \tag{4}$$

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

$$TV(\boldsymbol{G}) = \sum_{i=1}^{N-1} \left|\boldsymbol{g}^{(i)} - \boldsymbol{g}^{(i+1)}\right| \tag{5}$$

denotes a vector extension of the non-isotropic TV [14], which promotes smooth variations between subsequent elements of the sparse coding matrix columns $\boldsymbol{g}^{(i)}$ and $\boldsymbol{g}^{(i+1)}$. Equation (4) can be written in a more compact form, as

$$\min_{\boldsymbol{D}_x,\boldsymbol{D}_y,\boldsymbol{G}} \|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\|\boldsymbol{G}\|_1 + \mu\|\boldsymbol{RG}\|_1 \tag{6}$$

where matrix $\boldsymbol{R}$ is the horizontal finite difference operator. Note that, for $\mu = 0$, the above problem becomes equivalent to the classic CDL problem introduced in [3].

It should be highlighted that equation (6) 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 [17], [18], [19]. 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], [20], [21],[22], [23] that is able to treat such issues.

3.2 Optimization via ADMM

Applying ADMM to problem (6), the objective function is equivalent to the following constrained structure

$$\begin{aligned} \min_{\boldsymbol{D}_x,\boldsymbol{D}_y,\boldsymbol{G}} \; & \|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\|\boldsymbol{V}_1\|_1 + \mu\|\boldsymbol{V}_3\|_1 \\ s.t.\; & \boldsymbol{V}_1 - \boldsymbol{G} = 0, \boldsymbol{V}_2 - \boldsymbol{G} = 0, \boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2 = 0, \\ & \|\boldsymbol{D}_x(:,i)\|_2^2 \leq 1, \left\|\boldsymbol{D}_y(:,i)\right\|_2^2 \leq 1 \end{aligned} \tag{7}$$

where $\boldsymbol{D}_x(:,i)$ and $\boldsymbol{D}_y(:,i)$ denote the $i\text{-}th$ atom of the respective dictionary and $\boldsymbol{V}_1, \boldsymbol{V}_2$ and $\boldsymbol{V}_3$ are auxiliary variables.

The augmented Lagrangian function of problem (7) is defined as

$$\begin{aligned} \mathcal{L}\big(\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3, \boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3\big) = \tfrac{1}{2}\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\|_F^2 + \\ \tfrac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda\|\boldsymbol{V}_1\|_1 + \mu\|\boldsymbol{V}_3\|_1 + \tfrac{b_1}{2}\|\boldsymbol{V}_1 - \boldsymbol{G} + \\ \boldsymbol{B}_1\|_F^2 + \tfrac{b_2}{2}\|\boldsymbol{V}_2 - \boldsymbol{G} + \boldsymbol{B}_2\|_F^2 + \tfrac{b_3}{2}\|\boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2 + \boldsymbol{B}_3\|_F^2 \end{aligned} \tag{8}$$

where $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ denote the Lagrange multiplier matrices associated with the constraints, and b_1, b_2, b_3 stand for the penalty parameters. Hence, the following update rules are formed. More specifically:

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

$$\begin{aligned} \nabla_{\boldsymbol{G}}\mathcal{L} = 0 \Rightarrow \boldsymbol{G} = \big(\boldsymbol{D}_x{}^T\boldsymbol{D}_x + \boldsymbol{D}_y{}^T\boldsymbol{D}_y + b_1\boldsymbol{I} + \\ b_2\boldsymbol{I}\big)^{-1}\big(\boldsymbol{D}_x{}^T\boldsymbol{X} + \boldsymbol{D}_y{}^T\boldsymbol{Y} + \boldsymbol{B}_1 + b_1\boldsymbol{V}_1 + \boldsymbol{B}_2 + b_2\boldsymbol{V}_2\big) \end{aligned} \tag{9}$$

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

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

$$\nabla_{\boldsymbol{V}_1}\mathcal{L} = \boldsymbol{0} \Rightarrow \boldsymbol{V_1} = soft(\boldsymbol{G} - \boldsymbol{B_1}/b_1, \lambda/b_1) \tag{10}$$

where the $soft(.,\tau)$ denotes the soft-thresholding function $x = sign(x)max\,(|x| - \tau)$.

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

$$\begin{aligned} \nabla_{\boldsymbol{V}_2}\mathcal{L} = 0 \Rightarrow \boldsymbol{V}_2 = (b_3\boldsymbol{R}^T\boldsymbol{R} + b_2\boldsymbol{I})^{-1}(b_2\boldsymbol{G} - \boldsymbol{B}_2 \\ + b_3\boldsymbol{R}^T\boldsymbol{V}_3 + b_3\boldsymbol{R}^T\boldsymbol{B}_3) \end{aligned} \tag{11}$$

Also, $\boldsymbol{V}_3$ can be updated as

$$\nabla_{\boldsymbol{V}_3}\mathcal{L} = \boldsymbol{0} \Rightarrow \boldsymbol{V_3} = soft(\boldsymbol{RV_2} - \boldsymbol{B_3}/b_3, \mu/b_3) \tag{12}$$

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

$$\nabla_{\boldsymbol{D}_x}(\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\|_F^2) = 0 \;\&\; \nabla_{\boldsymbol{D}_y}\left(\left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2\right) = 0 \tag{13}$$

In order to accelerate this step, we follow the procedure proposed in [7] and [22] by updating the dictionaries column by column. More analytically, the updated scheme becomes

$$\begin{aligned} \boldsymbol{D}_x^{j+1}(:,i) &= \boldsymbol{D}_x^j(:,i) + (\boldsymbol{XG}(i,:)^T/(\zeta_i + \delta)) \\ \boldsymbol{D}_y^{j+1}(:,i) &= \boldsymbol{D}_y^j(:,i) + (\boldsymbol{YG}(i,:)^T/(\zeta_i + \delta)) \end{aligned} \tag{14}$$

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

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

$$\begin{aligned} \boldsymbol{B}_1^{j+1} &= \boldsymbol{B}_1^j + b_1(\boldsymbol{V}_1 - \boldsymbol{G}) \\ \boldsymbol{B}_2^{j+1} &= \boldsymbol{B}_2^j + b_2(\boldsymbol{V}_2 - \boldsymbol{G}) \\ \boldsymbol{B}_3^{j+1} &= \boldsymbol{B}_3^j + b_3(\boldsymbol{V}_3 - \boldsymbol{R}\boldsymbol{V}_2) \end{aligned} \tag{15}$$

The overall algorithm is summarized below.

Algorithm 1 Coupled Dictionary Learning from Smooth and Noisy data

Input: training signals $\boldsymbol{X} \in \mathbb{R}^{PxN}$, $\boldsymbol{Y} \in \mathbb{R}^{MxN}$, number of iterations J, penalty parameters b_1, b_2, b_3
Initialize: $\boldsymbol{D}_x \in \mathbb{R}^{PxK}$, $\boldsymbol{D}_y \in \mathbb{R}^{MxK}$, $\boldsymbol{B}_1 = \boldsymbol{B}_2 = \boldsymbol{B}_3 = 0$
Result: $\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{G}$
1. for $j = 1 \dots J$ do
2. Update $\boldsymbol{G}$ via (9)
3. Update $\boldsymbol{V}_1$ via (10)
4. Update $\boldsymbol{V}_2$ via (11)
5. Update $\boldsymbol{V}_3$ via (12)
7. for $i = 1 \dots K$ do
8. Update dictionaries $\boldsymbol{D}_x, \boldsymbol{D}_y$ by atom via (14)
9. end
10. Normalize the atoms of the dictionaries
11. Update the Lagrange multipliers via (15)
12. end for

4. FAST TV PROMOTING SPARSE CODING

It is important to realize that after the dictionary learning phase, detailed in the previous section, the dictionaries computed are to be used in some real-world application. Since dictionary learning is an offline task, the involved computational complexity may not be a critical issue. However, it is often the case in real-world applications, that only the sparse coding sub-problem is employed, hence time constraints become much more demanding. Thus, in order to deal with such an issue, a fast, sparse coding scheme is derived in this section, as a proper approximation of the orthogonal matching pursuit (OMP) algorithm.

4.1. Fast OMP-based sparse approximation

The OMP [24] is considered to be one of the most prominent and fastest algorithms for tackling the sparse coding problem. In case of smooth signals, however, an approximate solution based on the OMP algorithm could be effectively derived, assuming that a block of signals displaying homogeneity can be represented by the same sparse representation support $\mathcal{S}$, defined by the atoms involved in the representation. Thus, by exploiting the above-mentioned observation, 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.

Defining the set of signals $\boldsymbol{X} \in \mathbb{R}^{PxN}$ consisting of k blocks $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \dots, \boldsymbol{X}_k]$, such that each block $\boldsymbol{X}_i = [\boldsymbol{x}_1^i, \boldsymbol{x}_2^i, \dots \boldsymbol{x}_n^i]$, $i = 1, \dots, k$ contains n homogeneous vectors, the cost function proposed for the computation of the weights for block m becomes

$$\begin{aligned} &\min_{\boldsymbol{G}_\mathcal{S}} \|\boldsymbol{X}_m - \boldsymbol{D}_\mathcal{S}\boldsymbol{G}_\mathcal{S}\|_F^2 + \mu\, TV(\boldsymbol{G}_\mathcal{S}) \sim \\ &\min_{\boldsymbol{G}_\mathcal{S}} \|\boldsymbol{X}_m - \boldsymbol{D}_\mathcal{S}\boldsymbol{G}_\mathcal{S}\|_F^2 + \mu\|\boldsymbol{R}\boldsymbol{G}_\mathcal{S}\|_1 \end{aligned} \tag{17}$$

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 (17) can be solved via ADMM, following a similar procedure as in the previous section. Due to space limitations, we omit the intermediate derivations and give below the complete description of the algorithm.

Algorithm 2 Fast Sparse Coding Promoting Total Variation

Input: data $\boldsymbol{X} \in \mathbb{R}^{PxN}$, dictionary $\boldsymbol{D} \in \mathbb{R}^{PxK}$, sparsity level s, stopping error ε, number of iterations I,
Precompute: $(\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{I})^{-1}$
Result: $\boldsymbol{G} \in \mathbb{R}^{KxN}$
1. for (each block) $m = 1 \dots k$
2. find the corresponding $x_{m,c} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_m(:, i)$
3. Use the OMP to find the support $\mathcal{S}$ of the centroid signal x_c
4. Based on centroid's support compute the sparse representation coefficients for all signals in the block
$$\min_{\boldsymbol{G}_\mathcal{S}} \|\boldsymbol{X}_m - \boldsymbol{D}_\mathcal{S}\boldsymbol{G}_\mathcal{S}\|_F^2 + \mu\|\boldsymbol{R}\boldsymbol{G}_\mathcal{S}\|_1$$
5. if $\mu = 0$ then
6. $\boldsymbol{G}_\mathcal{S} = (\boldsymbol{D}_\mathcal{S}^T\boldsymbol{D}_\mathcal{S})^{-1}\boldsymbol{D}_\mathcal{S}^T\boldsymbol{X}_m$
7. else
8. Precompute $\left({\boldsymbol{D}_\mathcal{S}}^T\boldsymbol{D}_\mathcal{S} + b\boldsymbol{I}\right)^{-1}$, ${\boldsymbol{D}_\mathcal{S}}^T\boldsymbol{X}_m$
9. for $j = 1 \dots J$ do
10. Update $\boldsymbol{G}_\mathcal{S}$ via
$\boldsymbol{G}_\mathcal{S} = (\boldsymbol{D}_\mathcal{S}^T\boldsymbol{D}_\mathcal{S} + b\boldsymbol{I})^{-1}(\boldsymbol{D}_\mathcal{S}^T\boldsymbol{X}_m + \boldsymbol{B}_1 + b\boldsymbol{V}_1)$
11. Update $\boldsymbol{V}_1$ via
$\boldsymbol{V}_1 = (\boldsymbol{R}^T\boldsymbol{R} + \boldsymbol{I})^{-1}(\boldsymbol{G}_\mathcal{S} - \boldsymbol{B}_1/b + \boldsymbol{R}^T\boldsymbol{V}_2 + \boldsymbol{R}^T\boldsymbol{B}_2/b)$
12. Update $\boldsymbol{V}_2$ via
$\boldsymbol{V}_2 = soft(\boldsymbol{R}\boldsymbol{V}_1 - \boldsymbol{B}_2/b, \mu/b)$
13. Update the Lagrange multipliers via
$\boldsymbol{B}_1^{j+1} = \boldsymbol{B}_1^j + b(\boldsymbol{V}_1 - \boldsymbol{G}_\mathcal{S})$
$\boldsymbol{B}_2^{j+1} = \boldsymbol{B}_2^j + b(\boldsymbol{V}_2 - \boldsymbol{R}\boldsymbol{V}_1)$
14. end for
15. end if
16. end for

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, hyper-spectral images from the iCVL [25] dataset were used to generate the high-dimensionality dataset $\boldsymbol{X} \in \mathbb{R}^{31\times N}$, containing data at 31 wavelengths in the $400 - 700nm$ 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 hyper-spectral 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 (hyper-spectral) patches from the training images, so that $\boldsymbol{X}$ was a 31×100.000 matrix. Accordingly, $\boldsymbol{Y}$ was generated by downsampling $\boldsymbol{X}$, leading to a 8×100.000 matrix. Random 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 so-called batch-OMP method [18] is used for sparse coding, for all the examined schemes. Also, the dictionaries employed $K = 1.024$ atoms, while 6 non-zero coefficients were used to encode each data vector.

From Fig. 1, we can deduce that our proposed algorithm (CDL-TV) 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 additive Gaussian noise. In more detail, we use the dictionaries computed by the CDL-TV algorithm 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 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 the 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 correctly exploit the homogeneity of the data.

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

Algorithms	Algorithm 2 μ>0	SUnSAL-TV [14]	Algorithm 2 μ=0	Lasso [26] [27]	Batch-OMP [18]
runtime[sec]	50.62	1835.64	5.35	543.19	47.31

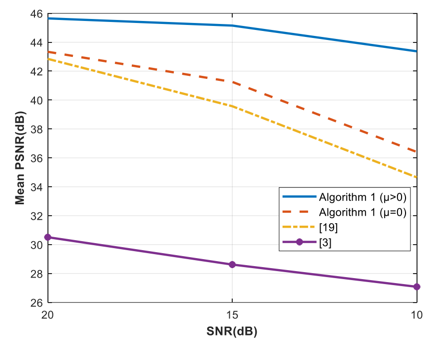

Fig. 1. Average PSNR over 100 images for different levels of SNRs between the proposed Algorithm 1 (μ>0), the Algorithm 1 without the TV regularizer (μ=0), the K-SVD based method in [3] using the OMP at the sparse coding stage, and the method in [19] (equation 15.42, 15.43).

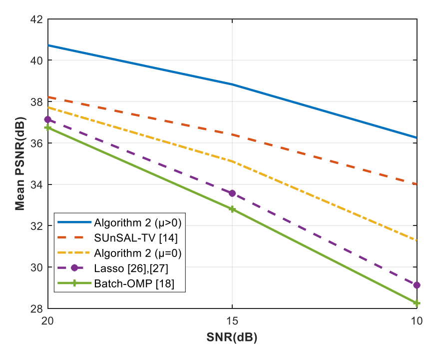

Fig. 2. Average PSNR over 10 images for different levels of SNRs between various sparse coding algorithms.

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 properly employing the total-variation regularizer, 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 Trans. Image Process.*, 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 Trans. Image Process.*, 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 Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[5] K. Engan, S. O. 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)*, 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*, 2017, vol. 2017-January, pp. 2294–2298.

[7] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, “Spectral Super Resolution of Hyperspectral Images via Coupled Dictionary Learning,” *IEEE Trans. Geosci. Remote Sens.*, 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*, 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*, 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 J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 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. 12587–12597, 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,” *Eng. Appl. Artif. Intell.*, vol. 72, pp. 43–53, Jun. 2018.

[13] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, “Coupled dictionary training for image super-resolution,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.

[14] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, “Total variation spatial regularization for sparse hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11 PART1, pp. 4484–4502, 2012.

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

[16] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[17] I. Tošić and P. Frossard, “Dictionary learning,” *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, 2011.

[18] R. Rubinstein, M. Zibulevsky, and M. Elad, “Efficient Implementation of the K-SVD Algorithm and the Batch-OMP Method.”

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

[20] 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.

[21] S. Boyd *et al.*, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Found. Trends R Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

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

[23] 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 Trans. Comput. Imaging*, vol. 3, no. 2, pp. 160–174, Apr. 2017.

[24] 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,” *Proc. 27 TH Annu. ASILOMAR Conf. SIGNALS, Syst. Comput.*, vol. 1, pp. 40--44, 1993.

[25] B. Arad and O. Ben-Shahar, “Sparse recovery of hyperspectral signal from natural RGB images,” in *European Conference on Computer Vision*, pages 19-34, Springer, 2016.

[26] B. Efron *et al.*, “Least angle regression,” *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[27] R. Tibshirani and R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso,” *J. R. Stat. Soc. Ser. B*, vol. 58, pp. 267--288, 1994.