Journal of Mathematical Modeling
Vol. 9, No. 1, 2021, pp. 13–30

JMM

A combined dictionary learning and TV model for image restoration with convergence analysis

Souad Mohaoui*, Abdelilah Hakim, Said Raghay

Department of Mathematics, University of Cadi Ayad, Marrakesh, Morocco

Email(s): email: souad.mohaoui@ced.uca.ma, a.hakim@uca.ma, s.raghay@uca.ma

Abstract. We consider in this paper the l_0-norm based dictionary learning approach combined with total variation regularization for the image restoration problem. It is formulated as a nonconvex nonsmooth optimization problem. Despite that this image restoration model has been proposed in many works, it remains important to ensure that the considered minimization method satisfies the global convergence property, which is the main objective of this work. Therefore, we employ the proximal alternating linearized minimization method whereby we demonstrate the global convergence of the generated sequence to a critical point. The results of several experiments demonstrate the performance of the proposed algorithm for image restoration.

Keywords: Image deblurring, dictionary learning, sparse approximation, total variation, proximal methods, nonconvex optimization.
AMS Subject Classification 2010: 34A34, 65L05.

1 Introduction

Image restoration is one of the earliest and most classical linear problems in imaging. Generally, the information in the image consists of a degraded representation of the original object and one can roughly distinguish two sources of degradation, the process of image formation and the process of image recording. The degradation due to the image formation process is usually called the blur operation, which is a sort of band-limiting of the object. While the degradation introduced by the recording process refers to the noise. Hence, image deblurring aims to recover a clean image from its degraded form. For an observed image y the problem of image restoration can be expressed by the following equation:

$$y = Hu + v,$$

*Corresponding author.
Received: 9 January 2020 / Revised: 19 April 2020 / Accepted: 22 June 2020
DOI: 10.22124/jmm.2020.15408.1369

where H is a linear operator that represents the degradation matrix, u is the original image, and v represents the noise. In general, image restoration problem recovering u from y is known as image deconvolution or image deblurring. When H is the identity operator, estimating u from y is referred to as image denoising. The problem of estimating u from y is called a linear inverse problem which, for most scenarios of practical interest, is ill-posed, either the direct operator does not have an inverse or it is nearly singular with noise sensitivity, or due to the ill-conditioning of H [35,36] . Therefore, in order to stabilize the recovery of u, it is necessary to incorporate prior-enforcing regularization on the solution. For regularization methods, finding and modeling appropriate prior knowledge of images is one of the most important concerns. This is especially true whenever information is incomplete, damaged or degraded. Consequently, variational models for image processing in general consist of two terms. The first term is fidelity or data fitting term and the second one is a regularization term which is formulated as:

$$\arg\min_{u} \ \frac{1}{2}\|Hu - y\|_2^2 + \lambda J(u),$$

where the first term in the objective function represents the fidelity term that measures the discrepancy between Hu and the observation y. Also, $J(u)$ is the regularization term and λ is a positive parameter used to balance the two terms for the minimization problem.

Several regularization models have been studied in the literature. Total variation approach, first introduced for image denoising [13,15], is based on minimizing certain energy function and it has been widely used in the image restoration problem. The success of total variation based methods for image processing is due to their ability to preserve edges in images. Since it was introduced by [29], intensive research has focused on developing efficient methods to solve TV-based minimization problem [9, 15, 22], where one of the main minimization difficulties is the presence of the nonsmooth TV norm in the objective function. Another popular approach that has been successfully used is the dictionary learning. In recent past decades, this unsupervised leaning technique has been emerged as an efficient way to solve various computer vision tasks such as: image recovery, super-resolution problem, image classification and speech separation [18–20, 28]. This success is mainly due to its good reputation in both theoretical research and practical applications in different disciplines. Many methods for learning sparse representation have been developed over the recent years. Among those methods, the KSVD [3] with OMP [27] algorithm, the online dictionary learning and proximal based methods [3,7]. Recent models have proposed a hybrid model where they combined the sparse representation with TV regularization [25]. Concerning image restoration, as long as sparse representation allows well extracting features from data, combined with total variation, the method performs excellent combination of feature extraction and edge preservation.

This paper takes a step in this direction by considering image restoration model that contains three terms, the nonconvex sparse representation prior, the total variation regularization and the fidelity term. Regarding the resolution of this model, most classical strategy tended to focus on using the alternating minimization method. This technique provides several sub-problems to solve, where each sub-problem is solved separately by using some existing algorithms. The most considered scenario is the use of KSVD algorithm to learn the dictionary matrix, the orthogonal matching pursuit (OMP) to get the sparse approximation, while for total variation, splitting variable techniques [1, 8, 16] are used to decouple the problem. To the best of our knowledge, no one has studied its global convergence. Therefore, the main aim of this work is the study of convergence analysis by considering an appropriate minimization method with multiple variables. Thereby, we demonstrate that the whole sequence generated by this method globally converges to a critical point.

2 Problem statement

2.1 Dictionary learning framework

The dictionary learning model has led to a large success in image denoising before being extended to numerous applications in computer vision. Without making any a priori assumption about the data except a parsimony principle, the method is able to produce an unsupervised dictionary learning devoted to noise reduction. Since dictionary learning is often considered as a computationally demanding process especially when the size of the data is large as the images, it is often learned over image patches. Given an image in its vectorized form $u \in \mathbb{R}^N$, we note $u_l \in \mathbb{R}^n$ the vector representation of the square 2D image patch of size $\sqrt{n} \times \sqrt{n}$ pixels. The image patch u_l is obtained by multiplication of u by a matrix R_l of size $n \times N$ whose columns are indexed by the image pixels. The l_0-quasi norm is used to encode the sparsity of the patch representation, T is the required sparsity level, and z is used to denote the set of sparse representations $\{z_l\}_l$ of all patches. Such an approach follows a methodology that allows the use of different schemes for unsupervised learning dictionary and the sparse coding steps as well as it is valid for the other reconstruction tasks. Formally, the dictionary learning model is given by the minimization problem:

$$\min_{D \in \mathcal{D}, z \in \mathbb{R}^{K,n}} \quad \frac{1}{2} \sum_{l}^{L} \|R_l x - D z_l\|_2^2 \; + \; \mu \|z_l\|_0, \tag{1}$$

where $\|.\|_0$ is the sparsity measure defined as the number of nonzeros entries in the input, D is referred to the dictionary, a matrix of values which gets us a sparse from z of the image patches. The usual strategy to solve problem (1) is by using an alternating optimization procedure which basically consists in alternatively learning the sparse approximation when the dictionary is considered fixed and then in updating the dictionary with the current sparse coefficients. The cardinality constraint is used to enforce the sparsity. However, such sparsity constraint is nonconvex and the resulting problem is known to be NP-hard. The first and most obvious approach to getting “pretty good” solutions to NP-Hard problems is to devise greedy algorithms. Greedy algorithms has been developed to solve the l_0-norm minimization in its current form [26, 27]. This algorithms attempt to solve the exact problem by different linear algebraic tools. The greedy algorithms known to be easy to implement and computationally can be extremely fast. The main principle is to detect or estimate the underlying support set of a given sparse vector followed by evaluating the associated signal values. Among the mainly existing greedy algorithms, the most used one is the OMP algorithm [27]. Differently, faced with the challenge of the nonconvexity and the associated NP-hardness, a traditional workaround in literature has been to modify the problem formulation itself for which existing tools can be applied. This is often done by relaxing the problem so that it becomes a convex optimization problem. Therefore, the l_0-norm formulation can be relaxed and solved optimally with the l_1-norm minimization, this problem is known as the convex sparse coding problem. However, l_0-regularization problem still has some advantages over l_1-regularization. The l_1-norm may fail to recover sparse solutions for some ill-posed inverse problems, and when dealing with large vectors, these methods are computationally very expensive [33]. Hence, compared with the l_1-norm, l_0-norm can directly recover sparser solutions. Fortunately, in recent years, a new process has permeated the fields of machine learning and image processing, consists of directly solve the nonconvex problems instead of relaxing them. In this vein, the so-called nonconvex optimization methods have been consider. At first glance, however, this approach seems doomed to fail, given the NP-hardness results. However in numerous deep and illuminating results, it has been revealed that if the problem possesses some structure, such as sparsity or low rankness, then nonconvex optimization methods are able to avoid NP-hardness as well as provide probably optimal solution (see for more insight [34]). Furthermore, the recent development of proximal methods led to significant advances in the design and analysis of algorithms, such as the iterative hard thresholding algorithms [32] that operate directly on the l_0 regularized problem.

2.2 Image restoration by dictionary learning and TV model

Natural images and images encountered in real applications are structured and often have many repetitive local patterns, in particular edges, smooth regions, and textures. The so-called nonlocal self-similarity is among the most successful priors for image restoration properties. Evidently, due to such properties, methods incorporating sparse representation and adaptive patch based models have exhibited very good results. Therefore, we consider the patch based dictionary learning model for image restoration model. Moreover, to overcome the artifacts sometimes caused by the patch based priors, the TV regularization is added in the minimization model. Hence, the restoration model is a sum of three terms, the fidelity term, the total variation regularization, and the nonconvex sparse representation prior. The minimization problem is given as follows:

$$\arg\min_{z,D,u} \ \frac{1}{2}\|Hu-y\|_2^2 \ + \ \mu_1\|\nabla u\|_1 \ + \ \frac{\mu_2}{2}\sum_l \|R_l u - Dz_l\|_2^2 \ + \ \mu_3\|z\|_0, \tag{2}$$

In the expression (2), the first term is the fidelity term that controls the global proximity of u to the blurred image y. The second term is the total variation prior defined by $\|\nabla u\|_1 = |\nabla_x u|_1 + |\nabla_y u|_1$ with $\nabla = [\nabla_x, \nabla_y]$ is the gradient operator in the discrete setting, whose two components at each pixel (i,j) are defined as follows

$$(\nabla_x u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j}, & \text{for } i < m, \\ 0, & \text{for } i = m, \end{cases}$$
$$(\nabla_y u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j}, & \text{for } j < n, \\ 0, & \text{for } j = n. \end{cases}$$

For the last two terms in equation (2), one controls the proximity of the patch $R_l u$ of the reconstruction to the deblurred patch Dz_l for each l, and the other one controls the sparsity of the representation of the patches.

Indeed such a model is rather difficult to solve, because it is neither convex nor smooth. The recent development of proximal methods led to significant advances in the design and analysis of algorithms for solving problems involving sum of nonsmooth convex or nonconvex functions [12, 23, 24]. The recent paper [12] introduced a proximal alternating linearized minimization algorithm denoted by PALM algorithm for solving a broad class of nonconvex and nonsmooth minimization problems. The fundamental advantages of the presented PALM method is that it only requires the objective function to be smooth at the term that uses all variables and that this smooth part has component-wise Lipschitz continuous gradient for convergence. Moreover, it has been proven that under certain conditions, each sequence generated by the PALM scheme globally converges to a critical point.

3 Notation and preliminaries

For an extended real valued function $f : \mathbb{R}^d \longrightarrow (-\infty, +\infty]$, we say that f is proper if it is never $-\infty$ and its domain defined by $Dom(f) = \{x \in \mathbb{R}^d : f(x) < +\infty\}$ is nonempty. For a closed $\mathcal{C} \in \mathbb{R}^n$, its indicator function $I_{\mathcal{C}}$ is defined by:

$$I_{\mathcal{C}}(s) = \begin{cases} 0, & \text{if} \quad s \in \mathcal{C}, \\ \infty, & \text{otherwise}, \end{cases}$$

Let $f : \mathbb{R}^d \longrightarrow (-\infty, +\infty]$ be a proper and lower semi-continuous function:
(1) For a given $x \in Dom(f)$, the Frechet sub-differential of f at x written $\hat{\partial} f(x)$ is the set of all vectors $u \in \mathbb{R}^d$ which satisfy:

$$\lim_{y \neq x; y \longrightarrow x} \frac{f(y) - f(x) - < u, y - x >}{\|y - x\|} \geq 0,$$

when $x \notin Dom(f)$ then $\hat{\partial} f(x) = \emptyset$.
(2) The limiting sub-differential of f at x is defined as:

$$\partial f(x) = \{u \in \mathbb{R}^n, \exists\, x^k \longrightarrow x, f(x^k) \longrightarrow f(x) \text{ and } u^k \in \hat{\partial} f(x^k) \longrightarrow u;\ k \longrightarrow \infty\}. \tag{3}$$

The sub-differential (3) reduces to the derivative of f (denoted by∇f) if f is continuously differentiable.

A necessary (but not sufficient) condition for $x \in \mathbb{R}^d$ to be a minimizer of f is $0 \in \partial f(x)$ and a point that satisfies this condition is called limiting-critical or simply critical.

(3) For any subset $S \subset \mathbb{R}^d$ and any point $x \in \mathbb{R}^d$ the distance from x to S is defined and denoted by :

$$dist(x, S) = \inf\{\|y - x\|; y \in S\}.$$

Note that when S is empty, we have $dist(x, S) = +\infty$ for all $x \in \mathbb{R}^d$. Note also that for any real extended valued function f on $\mathbb{R}^d$ and any $x \in \mathbb{R}^d$:

$$dist(x, \partial f(x)) = \inf\{\|u\|; u \in \partial f(x)\}.$$

(4) Given $x \in \mathbb{R}^d$ and $\lambda > 0$, the *proximal operator* associated to f is defined by:

$$\mathrm{Prox}_{\lambda}^{f}(x) = \arg\min \left\{ f(u) \quad + \frac{\lambda}{2}\|u-x\|_2^2, \;\; u \in \mathbb{R}^d \right\}.$$

When the function f is the indicator function of a nonempty and closed set $\mathcal{X}$ i.e for the function $\delta_{\mathcal{X}} : \mathbb{R}^d \longrightarrow (-\infty, +\infty]$, the proximal map reduces to the projection operator onto $\mathcal{X}$ defined by :

$$\mathrm{P}_{\mathcal{X}}(x) = \arg\min \;\; \left\{ \|u-x\|_2^2, \;\; u \in \mathcal{X} \right\}.$$

4 Image restoration via a proximal-based algorithm

4.1 Proximal alternating linearized minimization method

For many years, convex optimization which consists of minimizing a sum of convex or smooth functions has been received great interest. Several works on such convex problems have provided a sound theoretical foundation. Both, the theoretic and computing advantages created many benefits to practical use. However, what deserves the special consideration is the fact that the convex or smooth models are often approximations of nonconvex. A very well-known example is the l_0 minimization norm in sparse recovery problems which is commonly relaxed as convex l_1-norm. Despite the numerical complexities arising with nonconvex regularization, a variety applications have shown their importance in many disciplines. As a result, increasing alternating methods have been paid to a broad nonconvex optimization problem. Among those methods, the PALM scheme build on the Kurdyka-Lojasiewicz (KL) property allows to analyze various classes of nonconvex nonsmooth problems as well as establishing the global convergence of the sequence generated by PALM scheme. The KL property played a fundamental role on the convergence analysis of PALM method. For our purpose, we review the method for the problems of the form:

$$\min_{x,y,z} \;\; B(x,y,z) = H(x,y,z) + F(x) + R(y) + G(z), \tag{4}$$

where $F : \mathbb{R}^n \longrightarrow (-\infty, +\infty]$, $R : \mathbb{R}^m \longrightarrow (-\infty, +\infty]$ and $G : \mathbb{R}^d \longrightarrow (-\infty, +\infty]$ are extended valued, proper and lower semi-continuous functions. And $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \longrightarrow \mathbb{R}$ is C^1: smooth function with Lipschitz gradient on any bounded set and convex with respect to either x,y or z.

At the k^{th} iteration the PALM scheme updates the estimate of (x,y,z) alternatively by solving the following linearized proximal problems:

$$\begin{aligned}
x^{k+1} &= \arg\min_{x} \;\; < \nabla_x H(x^k, y^k, z^k), x - x^k > \;+\; \frac{t_x^k}{2}\|x - x^k\|_2^2 \;+\; F(x),\\
y^{k+1} &= \arg\min_{y} \;\; < \nabla_y H(x^{k+1}, y^k, z^k), y - y^k > \;+\; \frac{t_y^k}{2}\|y - y^k\|_2^2 \;+\; R(y),\\
z^{k+1} &= \arg\min_{z} \;\; < \nabla_z H(x^{k+1}, y^{k+1}, z^k), x - z^k > \;+\; \frac{t_z^k}{2}\|z - z^k\|_2^2 \;+\; G(z),
\end{aligned}$$

where t_x^k, t_y^k and t_z^k are appropriately chosen step sizes. Using the *the Proximal operator*, the minimization problems are equivalent to the following proximal problems:

$$
\begin{aligned}
x^{k+1} &\in \mathrm{Prox}_{t_x^k}^{F}\,(x^k - \frac{1}{t_x^k}\nabla_x H(x^k, y^k, z^k)),\\
y^{k+1} &\in \mathrm{Prox}_{t_y^k}^{R}\,(y^k - \frac{1}{t_y^k}\nabla_y H(x^{k+1}, y^k, z^k)),\\
z^{k+1} &\in \mathrm{Prox}_{t_z^k}^{G}\,(z^k - \frac{1}{t_z^k}\nabla_z H(x^{k+1}, y^{k+1}, z^k)).
\end{aligned}
$$

4.2 PALM algorithm for the proposed restoration model

In this subsection, we describe an adaptation of the PALM algorithm to the suggested deblurring model (2). In doing so, we first define the following sets:

$$
\begin{aligned}
\chi &= \left\{u \in \mathbb{R}^{n\times m} \quad \text{such that} \quad l \le u_{ij} \le h\right\},\\
\mathcal{C} &= \left\{z \in \mathbb{R}^{p} \quad \text{such that} \quad \|z_k\|_\infty \le M\right\},\\
\mathcal{D} &= \left\{D \in \mathbb{R}^{N\times p} \quad \text{such that} \quad \|d_j\|_2 \le 1 \quad j = 1,\dots,p\right\}.
\end{aligned}
$$

In order to express the image deblurring model as in the form of (4) we set:

$$
\begin{cases}
H(u, D, z) = \frac{1}{2}\|Hu - y\|_2^2 + \frac{\mu_2}{2}\sum_l \|R_l u - Dz_l\|_2^2,\\
F(u) = \mu_1\|\nabla u\|_1 + I_\chi(u),\\
R(D) = I_{\mathcal{D}}(D),\\
G(z) = \mu_3\|z\|_0 + I_{\mathcal{C}}(z),
\end{cases} \tag{5}
$$

where $I_\chi(u)$, $I_{\mathcal{D}}(D)$ and $I_{\mathcal{C}}(z)$ denotes the indicator functions of the sets χ, $\mathcal{D}$ and $\mathcal{C}$. The bound constraints on u model the situation where all pixel values have lower and upper bounds, for example when u is an image with pixel values in the range $[0, 256]$ or $[0, 1]$. While for z_k, the bound constraint is to prevent some elements of sparse coefficients to have unusual large values. To avoid the scenario that dictionary columns have arbitrary large norm which would lead to small values of the sparse approximations, we restrict its column to have an l_2-norm less or equal than 1. Then to use PALM algorithm for solving (2), let $D^{(0)}$ be the initial dictionary and $u^{(0)} = y$ then for $k = 0, 1, ..$

$$
z^{k+1} \in \mathrm{Prox}_{t_z^k}^{G}\,(z^k - \frac{1}{t_z^k}\nabla_z H(\hat{z}^k, D^k, u^k)), \tag{6}
$$

$$
D^{k+1} \in \mathrm{Prox}_{t_d^k}^{R}\,(D^k - \frac{1}{t_d^k}\nabla_D H(z^{k+1}, \hat{D}^k, u^k)), \tag{7}
$$

$$
u^{k+1} \in \mathrm{Prox}_{t_u^k}^{F}\,(u^k - \frac{1}{t_u^k}\nabla_u H(z^{k+1}, D^{k+1}, \hat{u}^k)), \tag{8}
$$

and $\hat{z}^k$, $\hat{D}^k$, $\hat{u}^k$ are the extrapolated points given by: $\hat{z}^k = z^{k-1} + w_z^k(z^{k-1} - z^{k-2})$, $\hat{D}^k = D^{k-1} + w_d^k(D^{k-1} - D^{k-2})$ and $\hat{u}^k = u^{k-1} + w_u^k(u^{k-1} - u^{k-2})$ respectively, with $w_z^k\,; w_d^k\,; w_u^k \le 1$. Hence, the update (6), (7) and (8) can be explicitly obtained by direct minimization.

Sparse Coefficients: Given the dictionary matrix D^k, and the image u^k, the task is to get the sparse coefficient z^k that satisfies (6), in which each iteration requires solving a minimization problem. We define

$$e^k = z^k - \frac{1}{t_z^k}\nabla_z H(\hat{z}^k, D^k, u^k) = z^k - \frac{\mu_2}{t_z^k}\sum_l D^T(D^k\hat{z}_l^k - R_l u^k),$$

and solve the minimization problem

$$z^{k+1} = \arg\min_{z\in\mathcal{C}} \ \frac{1}{2}\|z - e^k\|_2^2 \ + \ \mu_3\|z\|_0 = \arg\min_{z\in\mathcal{C}} \ \frac{1}{2}\sum_{i=1}^{p}(z_i - e_i^k)^2 \ + \ \mu_3\mathbb{I}_{\{z_i\neq 0\}}. \tag{9}$$

Basically, one could solve the scalar problem due to the component wise property. The term $\Gamma(z_i) = \mu_3\mathbb{I}_{\{z_i\neq 0\}}$ is the weighted l_0 penalty defined by

$$\Gamma(z_i) = \begin{cases} 0, & \text{if} \quad z_i = 0, \\ \mu_3, & \text{if} \quad z_i \neq 0. \end{cases}$$

Its proximity operator is a set valued mapping defined by:

$$\mathrm{Prox}_\Gamma(z_i) = \begin{cases} \{0\}, & \text{if} \quad |z_i| < \sqrt{2\mu_3}, \\ \{0, \sqrt{2\mu_3}\}, & \text{if} \quad z_i = \sqrt{2\mu_3}, \\ \{-\sqrt{2\mu_3}, 0\}, & \text{if} \quad z_i = -\sqrt{2\mu_3}, \\ \{z_i\}, & \text{if} \quad |z_i| > \sqrt{2\mu_3}. \end{cases}$$

By choosing $\sqrt{2\mu_3}$ as the value of $z_i = \pm\sqrt{2\mu_3}$, we get the Hard thresholding operator [30, 32] with $\alpha = \sqrt{2\mu_3}$

$$(\mathcal{T}_\alpha(z))_i = \begin{cases} 0, & \text{if} \quad |z_i| < \alpha, \\ z_i & \text{if} \quad |z_i| \geq \alpha. \end{cases}$$

Hence, it is easy to see that the solution of (9) is

$$z^{k+1} = \mathrm{P}_\mathcal{C}(\mathcal{T}_\alpha(z^k - \frac{1}{t_z^k}\nabla_z H(\hat{z}^k, D^k, u^k)) = \min(\mathcal{T}_\alpha(z^k - \frac{1}{t_z^k}\nabla_z H(\hat{z}^k, D^k, u^k), M)).$$

Dictionary matrix: Given the obtained sparse approximation z^{k+1} and the image u^k, we update the dictionary matrix by computing the proximal operator which corresponds to solving a minimization problem. Since the proximity mapping over the indicator function is the projection operator, thus the dictionary learning is obtained by

$$D^{k+1} = \mathrm{P}_\mathcal{D}(D^k - \frac{1}{t_d^k}\nabla_D H(z^{k+1}, \hat{D}^k, u^k)),$$

where $\mathrm{P}_\mathcal{D}$ denotes the Euclidean projection to $\mathcal{D}$ defined for any D as:

$$(\mathrm{P}_\mathcal{D})_i = \frac{d_i}{\max(1, \|d_i\|_2)} \quad \forall i = 1, \ldots, p.$$

Image Reconstruction: The reconstruction step is to solve the denoising problem based total variation:

$$\begin{aligned} s^k &= u^k - \frac{1}{t_u^k}\nabla_u H(z^{k+1}, D^{k+1}, \hat{u}^k), \\ u^{k+1} &= \arg\min_v \;\; \mu_1\|\nabla v\|_1 + I_\chi(v) + \frac{\lambda}{2}\|v - s^k\|_2^2, \\ u^{k+1} &= \arg\min_{v\in\chi} \;\; \mu_1\|\nabla v\|_1 + \frac{\lambda}{2}\|v - s^k\|_2^2. \end{aligned} \tag{10}$$

Problem (10) is a very known denoising problem in the area of image processing. A variety of methods and algorithms that cover the minimization of this problem efficiently were developed. The classical method used for the minimization of total variation denoising problem is [15], where the author proposed to consider a dual approach and a gradient based algorithm for solving the resulting dual problem. In the same direction, the forward backward splitting (FBS) method and its accelerated versions also used this dual approach. Following the same strategy, we first form a dual formulation of the problem (10):

$$\min_{v\in\chi} \mu_1\|\nabla v\|_1 \;\; + \frac{\lambda}{2}\|v - s^k\|_2^2 = \max_{\|q\|_\infty\leq 1}\min_{v\in\chi} \mu_1\langle q, \nabla v\rangle + \frac{\lambda}{2}\|v - s^k\|_2^2. \tag{11}$$

Hence, for a given q, the minimal value of v satisfies: $\mathrm{P}_\chi(s^k + \mu_1/\lambda\nabla^T q)$, where ∇^T is the discrete divergence (which is the negative adjoint of the gradient operator). Plugging this expression for v back into the second part of (11), we get the dual problem:

$$\min_{\|q\|_\infty\leq 1} \;\; \{\|(s^k + \mu_1/\lambda\nabla^T q)\|_2^2 - \|\mathrm{P}_\chi(s^k + \mu_1/\lambda\nabla^T q) - (s^k + \mu_1/\lambda\nabla^T q)\|_2^2\}. \tag{12}$$

Then, we simply minimize a quadratic function with an infinity-norm constraint using a recent variation of FBS algorithm [22]. To do so, we only need to determine the gradient of the objective function which is continuously differentiable. Thus, define $\Lambda_\chi(x) = \|x - \mathrm{P}_\chi(x)\|_2^2$. Its gradient is $\nabla\Lambda_\chi(x) = 2(x - \mathrm{P}_\chi(x))$. Hence, the dual problem is reformulated as

$$\min_{h(q) = \|q\|_\infty\leq 1} \;\; \{\|(s^k + \mu_1/\lambda\nabla^T q)\|_2^2 - \Lambda_\chi(s^k + \mu_1/\lambda\nabla^T q)\}.$$

The gradient of the dual objective function is given by

$$\nabla h(q) \; = \; \nabla_q(\|(s^k + \mu_1/\lambda\nabla^T q)\|_2^2 - \Lambda_\chi(s^k + \mu_1/\lambda\nabla^T q)) = \frac{2\mu_1}{\lambda}\nabla(\mathrm{P}_\chi(s^k + \mu_1/\lambda\nabla^T q)).$$

The solution of the dual problem is obtained using a variant of FBS algorithm, in which resulting scheme alternately performs gradient descent step $q = q - t\nabla h(q)$, and then re-projects the result back into the infinity-norm ball using the formula $q^*_{ij} \leftarrow q_{ij}/\max(1, q_{ij})$. Once we obtain q^* the solution of (12), the optimal (denoised) image is then given by

$$u^* \; = \; \mathrm{P}_\chi(s^k + \mu_1/\lambda\nabla^T q^*).$$

To make the whole objective non-increasing, at the k^{th} iteration we set $w_z^k = w_d^k = w_u^k = 0$ (i.e no extrapolation) if $B(u^k, D^k, z^k) > B(u^{k-1}, D^{k-1}, z^{k-1})$.

Algorithm 1: PALM algorithm for the combined deblurring model

---

Initialisation: Set $D^{(0)}$, $z^{(0)}$ and $u^{(0)} = y$

for $k = 1, 2, 3, \ldots, Maxiter$ **do**

1. Sparse approximation step:

$$z^{k+1} = \min(\mathcal{T}_\alpha(z^k - \frac{1}{t_z^k}\nabla_z H(\hat{z}^k, D^k, u^k), M).$$

2. Dictionary update step:

$$D^{k+1} = \mathcal{P}_{\mathcal{D}}(D^k - \frac{1}{t_k^d}\nabla_D H(z^{k+1}, \hat{D}^k, u^k)).$$

3. Reconstruction step:

$$u^{k+1} = \mathcal{P}_\chi(s^k + \mu_1/\lambda\nabla^T p^*).$$

if $B(u^{k+1}, D^{k+1}, z^{k+1}) > B(u^k, D^k, z^k)$ **then**

Re-update $u^{k+1}, D^{k+1}, z^{k+1}$ with $\hat{z}^{k+1} = z^k$, $\hat{D}^{k+1} = D^k$, $\hat{u}^{k+1} = u^k$

end

end

---

5 Convergence Analysis

Here, first we give some important definitions and results used in convergence analysis.

The Kurdyka Lojasiewicz property: This property plays a basic role in the convergence analysis.

Let $\eta \in (0, +\infty]$, we denote by Ψ_η the class of all concave and continuous functions $\varphi : (0, +\infty] \longrightarrow \mathbb{R}_+$ which satisfy the following conditions

1. $\varphi(0) = 0$,
2. φ is C^1 on $(0, \eta]$ and continuous at 0,
3. for all $(0, \eta) : \varphi'(s) > 0$.

Definition 1 (Kurdyka-Lojasiewcz property). *Let $f : \mathbb{R}^d \longrightarrow (-\infty, +\infty]$ be a proper and lower semi-continuous function.*

(i) *The function f is said to have the Kurdyka-Lojasiewcz property at $x \in Dom(\partial f)$ where $Dom(\partial f) = \{y \in \mathbb{R}^d : \partial f(y) \neq \emptyset\}$ if there exists $\eta \in (0, +\infty]$, a neighborhood U of x and a function $\varphi \in \Psi_\eta$ such that the following inequality holds for all $y \in U \cup [f(x) < f(y) < f(x) + \eta]$*

$$\varphi'(f(y) - f(x))dist(0, \partial f(y)) \geq 1.$$

(ii) *If f satisfies the KL property at each point of $Dom(\partial f)$, then f is called a KL function.*

Remark 1. *A proper lower semi-continuous function $F : \mathbb{R}^n \longrightarrow (-\infty, +\infty]$ has the Kurdyka Lajasiewicz property at any non critical point.*

Definition 2 (Semi-algebraic sets and functions)**.** *A subset S of $\mathbb{R}^n$ is called semi-algebraic if there exists a finite number of real polynomial functions P_{ij} and Q_{ij} such that*

$$S = \cup_j \cap_i \{x \in \mathbb{R}^n \;\; P_{ij}(x) = 0 \;\text{ and }\; Q_{ij}(x) < 0\}.$$

A function is called semi-algebraic if its graph

$$Graph(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \;\text{ such that }\; t = f(x)\},$$

is a semi-algebraic set.

Semi-algebraic functions satisfy the Kurdyka-Lojasiewicz property (see [10], [11]) with $\varphi(s) = cs^{1-\theta}$ for θ belong to $[0,1)$ and some positive real number c. We have the following results.

Theorem 1 (Convergence result)**.** *The sequence $s^k = (z^k, D^k, u^k)$ generated by Algorithm 1 converge to the critical point of problem (2) if the following conditions hold:*

(1) *The functions F, G and R are lower semi-continuous;*

(2) *$H(z,D,u)$ is a C^1 function;*

(3) *$B(z,D,u)$ is a KL function;*

(4) *The partials gradient $\nabla_z H$, $\nabla_D H$ and $\nabla_u H$ are globally Lipschitz with moduli L_z, L_d and L_u respectively;*

(5) *$\{s^k\}_{k\in\mathbb{N}}$ is a bounded sequence and the steps size must be such that $t_z^k > L_z$, $t_d^k > L_d$, $t_u^k > L_u$;*

(6) *$\nabla H(z,D,u)$ is Lipschitz continuous on bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d$.*

Remark 1. Note that the KL property is an assumption on the class of function to be minimized and does not depend on the structure of the considered algorithm. And as the goal is to prove that the whole sequence converges to a critical point of B, the role of KL property is to demonstrate that the generated sequence $\{s^k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. (see [12] for the proof).

The convergence rate of the method is given through the following estimations. Thereby, Algorithm 1 has at least sub-linear convergence rate.
If the desingularizing function φ of Ψ is of the form $\varphi(s) = cs^{1-\theta}$, where c is positive real number and θ belong to $[0,1)$, then the following estimations hold (for more detail see [5]).

(i) If $\theta = 0$ then the sequence z^k converges in a finite number of steps.

(ii) If $\theta \in [0, 1/2]$ then there exist $\omega > 0$ and $\tau \in [0.1)$ such that $\|s^k - s^*\| \leq \theta\tau^k$.

(iii) If $\theta \in [1/2, 0)$ then there exist $\omega > 0$ such that $\|s^k - s^*\| \leq k^{-\frac{1-\theta}{2\theta-1}}$.

Proof of Theorem 1. We show that Algorithm 1 satisfies the conditions of Theorem 1.

(1), (2) Observe that χ, $\mathcal{C}$ and $\mathcal{D}$ are closed sets and hence the indicator functions $I_\chi(u)$, $I_\mathcal{D}$ and $I_\mathcal{C}$ are lower semi-continuous, and indeed the $\|.\|_0$ is also lower semi-continuous. Hence, the functions F, R and G are lower semi-continuous while the function H is a C^1 function.

(3) As we mentioned above, to prove the first condition, that is, the objective function satisfies the Kurdyka-Lojasiwicz property in its effective domain, it is sufficient to prove that the objective function is a semi-algebraic function, i.e., to prove that each term in the function (4) defined via (5) is a semi-algebraic function.

(i) $H(u, D, z) = \frac{1}{2}\|Hu - y\|_2^2 + \frac{\mu_2}{2}\sum_l \|R_l u - Dz_l\|_2^2$ is a real polynomial functions and thus by definition, $H(u, D, z)$ is a semi-algebraic function.

(ii) $F(u) = \mu_1\|\nabla u\|_1 + I_\chi(u)$, where

$$\begin{aligned}\chi &= \cap_{k=l}^{nm}\{u \in \mathbb{R}^{nm} \quad \text{such that} \quad l \leq u_k \leq h\}\\ &= \cap_{k=l}^{nm}\cup_{s=l}^{h}\{u \in \mathbb{R}^{nm} \quad \text{such that} \quad u_k = s\}\end{aligned}$$

which is a semi-algebraic set and the indicator function of a semi-algebraic set is a semi-algebraic function.

For $u \longrightarrow \|u\|_1$, we set $\varphi : s > 0 \longrightarrow s^{\frac{n_1}{n_2}}$ where n_1, n_2 are positive numbers. The graph of φ is given by

$$\begin{aligned}Graph(\varphi) &= \left\{(s,t) \in \mathbb{R}_+^2, \quad \varphi(s) = t\right\} = \left\{(s,t) \in \mathbb{R}_+^2, \quad s^{\frac{n_1}{n_2}} = t\right\}\\ &= \left\{(s,t) \in \mathbb{R}^2, \quad s^{n_1} - t^{n_2} = 0\right\} \cap \mathbb{R}_+^2.\end{aligned}$$

Therefore $Graph(\varphi)$ is a semi-algebraic set, and hence φ is semi-algebraic function. Consequently, the function $\varphi : \ x \longrightarrow |x_i|^p \quad \forall i$ with fractional integers $p = \frac{n_1}{n_2}$ is a semi-algebraic function. Particularly, the function $\varphi : x \longrightarrow \|x\|^p$ is a semi-algebraic function for fractional integers p, since the sum of semi-algebraic functions is a semi-algebraic function. In our case $p = 1$. Moreover, $\nabla : \mathbb{R}^{n\times m} \longrightarrow \mathbb{R}^{2n\times m}$ is a linear operator and thus the composite function $\varphi \circ \nabla : u \longrightarrow \|\nabla u\|_1$ is semi-algebraic. Therefore, the function $F(u)$ is semi-algebraic.

(iii) For $R(D) = I_\mathcal{D}(D)$, we have

$$\begin{aligned}\mathcal{D} &= \left\{D \in \mathbb{R}^{N\times p} \quad \text{such that} \quad \|d_j\|_2 \leq 1 \quad \forall j = 1, \ldots, p\right\}\\ &= \cap_{j=1}^{p}\left\{D \in \mathbb{R}^{N\times p} \quad \text{such that} \quad \sum_{k=1}^{N} d_{jk}^2 \leq 1 \quad \right\},\end{aligned}$$

which is a semi-algebraic set and then the indicator function of $\mathcal{D}$ is also semi-algebraic.

(iv) $G(z) = \mu_3\|z\|_0 + I_\mathcal{C}(z)$, where $\|.\|_0$ is the sparsity term that counts the number of nonzero elements of a vector $z \in \mathbb{R}^p$. The graph of $\|.\|_0$ is given by

$$Graph(\|.\|_0) = \{(x,t) \in \mathbb{R}^p \times \mathbb{R}_+, \ \|x\|_0 = t\} = \left\{(x,t) \in \mathbb{R}^p \times \mathbb{R}_+, \ \sum_{i=1}^{p} \mathbb{I}_{\{x_i \neq 0\}} = t\right\},$$

where $\mathbb{I}_{\{x_i \neq 0\}} = \{1 \text{ if } i \in \{x_i \neq 0\} \text{ and } 0 \text{ otherwise}\}$. We set $J_t = \{I \subseteq \{1, \dots, p\} \;\; s.t \;\; |I| = t\}$, then we have

$$\begin{aligned} Graph(\|.\|_0) &= \{(x,t) \in \mathbb{R}^p \times \mathbb{R}_+, \;\; |\{x_i \neq 0, \forall i = 1, \dots, p\}| = t\} \\ &= \cup_{t=0}^{p} \cup_{I \in J_t} \{(x,t) \quad s.t. \;\; x_i \neq 0 \;\text{ for } i \in I, \; x_i = 0 \;\text{ for } i \in I^c\}. \end{aligned}$$

Thus $Graph(\|.\|_0)$ is semi-algebraic. While for $I_{\mathcal{C}}(z)$ we observe that

$$\mathcal{C} = \{z \in \mathbb{R}^p \text{ such that } \;\; \|z_k\|_\infty \leq M\} = \cup_{i=1}^{M} \{z : \;\; \|z_k\|_\infty = i\},$$

which is a semi-algebraic set, hence the indicator function is semi-algebraic. Therefore each term of the objective function is semi-algebraic function then it is KL function.

(4) (i) $\|\nabla_z H(z_1, D^k, u^k) - \nabla_z H(z_2, D^k, u^k)\| \leq \|(D^k)^T D^k\|_2 \|z_1 - z_2\| = L(D^k)\|_2 \|z_1 - z_2\|$.
(ii) $\|\nabla_D H(z^k, D_1, u^k) - \nabla_D H(z^k, D_2, u^k)\|_2 \leq \|z^k (z^k)^T\| \|z_1 - z_2\| = L(z^k)\|z_1 - z_2\|$
(iii) $\|\nabla_u H(z^k, D^k, u_1) - \nabla_z H(z^k, D^k, u_2)\| \leq (\|H^T H\|_2 + \mu_2 \sum_l \|R_l^T R_l\|_2)\|z_1 - z_2\|$

$$= L(H,R)\|z_1 - z_2\|,$$

where $H^T H$ is a circulant square matrix and $R_l^T R_l$ is a diagonal matrix its entries correspond to image pixel locations and their values are equal to the number of overlapping patches contributing at those pixel locations.

(5) The $s^k = (z^k, D^k, u^k)$ is a bounded sequence since $z^k \in \mathcal{C}$, $D^k \in \mathcal{D}$ and $u^k \in \chi$ for any $k = 1, 2, \dots$. In addition, the step sizes can be chosen as
$t_z^k = \max(\gamma_1 L(D^k), l_1^+)$ where $\gamma_1 > 1$, $l_1^+ > 0$,
$t_d^k = \max(\gamma_2 L(z^k), l_2^+)$ where $\gamma_2 > 1$, $l_2^+ > 0$,
$t_u^k = \max(\gamma_3 L(H,R), l_3^+)$ where $\gamma_3 > 1$, $l_3^+ > 0$.

(6) For the last condition in Theorem 1, as we have mentioned above, the function

$$H(z,D,u) = \frac{1}{2}\|Hu - y\|_2^2 + \frac{\mu_2}{2}\sum_l \|R_l u - D z_l\|_2^2,$$

is a smooth function. Moreover, the gradient of the objective function is given by

$$\nabla H(z,D,u) = (H^T(Hu - y) + \mu_2 \sum_l R_l^T(R_l u - D z_l), \mu_2 \sum_l (D z_l - R_l u) z_l^T, \mu_2 \sum_l D^T(D z_l - R_l u)).$$

which has a Lipschitz constant on any bounded set, i.e., for all bounded sets S there exists a constant $C > 0$ such that $\{(z_1, D_1, u_1), (z_2, D_2, u_2)\} \subseteq S$, we have

$$\|\nabla H(z_1, D_1, u_1) - \nabla H(z_2, D_2, u_2)\| \leq C\|(z_1, D_1, u_1) - (z_2, D_2, u_2)\|.$$

So the proof is completed. □

6 Results and discussion

In this section, we present the experimental results of the proposed approach for image deblurring. For the experimental setting, first it should be noted that all implementations were carried out using MATLAB R2015 in a desktop pc equipped with an intel core i3. The Algorithm 1 is evaluated on classical gray-level images: Barbara, House, Boats and Peppers. Through all the experiments we use the Gaussian blur generated by the MATLAB function fspecial with two standard deviation 1.6 and 2 and noise= $\sqrt{2}$. For the comparison, some image restoration approaches including TVMM [9] and BM3DDEB [17] have been implemented. The quality of recovered images was evaluated by *Peak signal to noise ration* PSNR defined by

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}},$$

where

$$\text{MSE} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[u_0(i,j) - u(i,j)\right]^2.$$

The u_0 and u are the original and recovered images respectively. On the other hand, the visual comparison of reconstructed images obtained by the three approaches is presented in Figures 1, 2 and 3. It can be seen visually that Algorithm 1 present remarkable results with respect to other algorithms in several images. Table 1 represents the obtained PSNR values of the two cases fspecial$(Gaussian, 25, 1.6)$ and fspecial$(Gaussian, 25, 2)$.

Table 1: The reconstruction PSNR values of the tested images using Gaussian blur.

Image	House	Barbara	Boats	Pepper
Gaussian blur, $\sigma = \sqrt{2}$ and $s.v = 1.6$				
TVMM	33.05	24.62	30.33	26.63
BM3DDEB	32.87	**28.19**	30.63	27.48
Ours	**33.41**	26.07	**30.99**	**27.72**
Gaussian blur, $\sigma = \sqrt{2}$ and $s.v = 2$				
TVMM	31.95	24.15	28.80	25.75
BM3DDEB	31.10	24.35	29.05	26.14
Ours	**32.33**	**24.36**	**29.52**	**26.60**

7 Conclusion

In this paper, we interested in proximal methods for the minimization of image restoration problem based dictionary learning and the total variation model. Therefore, an adaptation of the proximal alternating linearized method for this nonconvex nonsmooth minimization problem was presented. Thereby, we demonstrate that the generated sequence by the proposed algorithm globally converges to a critical point at least in sub-linear rate. Numerical performance of the proposed algorithm is evaluated by some experiments results of image deblurring problem compared with two popular deblurring methods.

Figure 1: Visual comparison of image deblurring on gray image Boats (256 × 256). From left to right and to bottom: original image, noisy and blurred image by Gaussian blur with the standard deviation of 1.6, the deblurred image by TVMM, BM3DDEB, and our proposed approach.

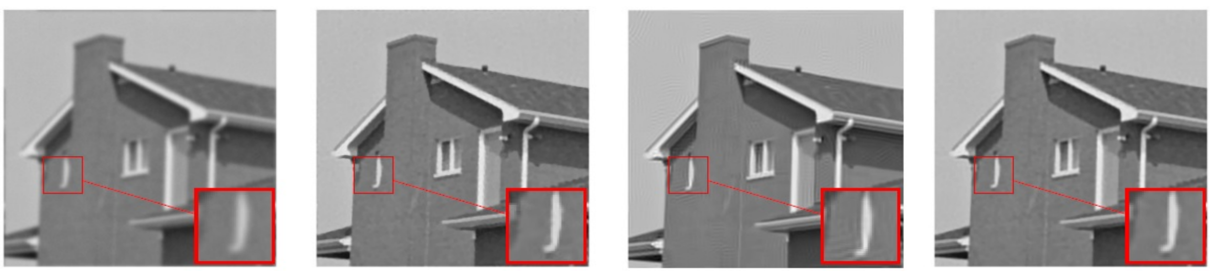

Figure 2: Visual comparison of image deblurring on gray image House (256 × 256). From left to right and to bottom: original image, noisy and blurred image by Gaussian blur with the standard deviation of 1.6, the deblurred image by TVMM, BM3DDEB, and our proposed approach.

Figure 3: Visual comparison of image deblurring on gray image Peppers (256 × 256). From left to right and to bottom: original image, noisy and blurred image by Gaussian blur with the standard deviation of 1.6, the deblurred image by TVMM, BM3DDEB, and our proposed approach.

Acknowledgements

Authors thank the editor and reviewers for the corrections and useful suggestions that have improved this work.

References

[1] M. Afonso, V.M. Jos D. Bioucas-Dias, M.T. Figueiredo, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process. **19** (2010) 2345-2356.

[2] A. Asaei, M. Golbabaee, H. Bourlard, V. Cevher, *Structured sparsity models for reverberant speech separation*, IEEE/ACM Trans. Au. Sp. Lang. Proc. **22** (2014) 620–633.

[3] M. Aharon, M. Elad, A. Bruckstein, *K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Trans. Signal Process. **54** (2006) 4311-4322.

[4] Z. Alexander, R. Tichatschke, *P*roximal point methods and nonconvex optimization, J. Global Optim. **13** (1998) 389-406.

[5] H. Attouch, J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program. **116** (2009) 5-16.

[6] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality,* Math. Oper. Res. **35**(2) (2010) 438-457.

[7] C. Bao, H. Ji, Y. Quan, Z. Shen, *Dictionary learning for sparse coding: Algorithms and convergence analysis*, IEEE Trans. Pattern Anal. Mach. Intell **38**(7) (2015) 1356-1369.

[8] J.M. Bioucas-Dias *A variable splitting augmented Lagrangian approach to linear spectral unmixing*, First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009.

[9] J.M. Bioucas-Dias, M.A.T. Figueiredo, J.P. Oliveira, *Total variation-based image deconvolution: a majorization-minimization approach*, IEEE Int. Conf. ICASSP, 2006.

[10] J. Bolte, A. Daniilidis, A. Lewis, *The kojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim. **17**(4) (2007) 1205-1223.

[11] J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim. **18** (2007) 556–572 .

[12] S. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization or nonconvex and nonsmooth problems*, Math. Program. **146** (1-2) (2014) 459-494.

[13] T. Chan, A. Marquina, P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput. **22**(2) (2000) 503-516.

[14] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision **20** (2004) 89-97.

[15] A. Chambolle, P. Lions, *Image recovery via total variation minimization and related problems,* Numer. Math. **76** (2) (1997) 167-188.

[16] Y. Chen, W. W. Hager, M. Yashtini, X. Ye, H. Zhang, *Bregman operator splitting with variable stepsize for total variation image reconstruction,* Comput. Optim. Appl. **54** (2) (2013) 317-342.

[17] K. Dabov, S. Foi, V. Katkovnik, K. Egiazarian, *Image restoration by sparse 3D transform-domain collaborative filtering,* IPAS VI. **6812** (2008) 681207 .

[18] D.L. Donoho, M. Elad, V.N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise,* IEEE Trans. Inform. Theory **52** (2006) 6–18.

[19] M. Elad, M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process. **15** (2006) 3736–3745.

[20] S. Esedoglu, S.J. Osher. *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model,* Comm. Pure Appl. Math. **57** (2004) 1609–1626.

[21] T. Goldstein, S. Osher, *The split Bregman method for L1-regularized problems,* SIAM J. Imaging Sci. **2** (2009) 323–343.

[22] T. Goldstein, C. Studer, R. Baraniuk, *A field guide to forward-backward splitting with a FASTA implementation,* Available: http://arxiv.org/abs/1411.3406, 2014.

[23] Z. Huan, Z. Lin, *Accelerated proximal gradient methods for nonconvex programming,* Adv. Neural. Inf. Process. Syst. (2015) 379-387.

[24] A. Kaplan, W. Tichatschke, *Proximal point methods and nonconvex optimization,* J. Global Optim. **13** (4) (1998) 389-406.

[25] H. Lu, J. Wei, Q. Liu, Y. Wang, X. Deng, *A dictionary learning method with total generalized variation for MRI reconstruction,* Int. J. Biomed. Imaging., Volume 2016, Article ID 7512471, 13 pages.

[26] L. Rebollo-Neira, D. Lowe. *Optimized orthogonal matching pursuit approach*, IEEE Signal Process. Lett. **9**(4) (2002) 137-140.

[27] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition*, Proc. 27th Asilomar Conf. Signals, Syst.. and Comput., Pacific Grove, CA, USA, **1** (1993) 40–44.

[28] I. Ramirez, P. Sprechmann, G. Sapiro, *Classification and clustering via dictionary learning with structured incoherence and shared features,* Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., (2011) 35013508

[29] L.I. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D. **60** (1992) 259–268.

[30] R. Gribonval, M. Nikolova, *A characterization of proximity operators*, J. Math. Imaging Vision (2020) 1-17.

[31] D. Strong, T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization,* Inverse Probl. **19** (2006) S165.

[32] T. Blumensath, M.E. Davies, *Iterative hard thresholding for compressed sensing,* Appl. Comput. Harmon. Anal. **27**(3) (2009) 265-274.

[33] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, D.L. Donoho, *Morphological component analysis*, Proc. SPIE 5914, Wavelets XI, 59140Q, 17 September 2005.

[34] P. Jain, P. Kar, *Non-convex optimization for machine learning*, Found. Trends Mach. Learn. **10** (3-4) (2017) 142-336.

[35] T. Poggio, V. Torre, *Ill-posed problems and regularization analysis in early vision,* MIT AI Lab, Tech. Rep. AI Memo. (1984) 773.

[36] M. Bertero, P. Boccacci, *Introduction to inverse problems in imaging,* Inst. of Physics, Bristol, U.K, 1998.