Inertial nonconvex alternating minimizations for the image deblurring

Tao Sun, Roberto Barrio, Marcos Rodríguez, and Hao Jiang

Abstract—In image processing, *Total Variation (TV) regularization models* are commonly used to recover blurred images. One of the most efficient and popular methods to solve the convex TV problem is the *Alternating Direction Method of Multipliers* (ADMM) algorithm, recently extended using the inertial proximal point method. Although all the classical studies focus on only a convex formulation, recent articles are paying increasing attention to the nonconvex methodology due to its good numerical performance and properties. In this paper, we propose to extend the classical formulation with a novel *nonconvex Alternating Direction Method of Multipliers with the Inertial technique* (IADMM). Under certain assumptions on the parameters, we prove the convergence of the algorithm with the help of the Kurdyka-Łojasiewicz property. We also present numerical simulations on classical TV image reconstruction problems to illustrate the efficiency of the new algorithm and its behavior compared with the well established ADMM method.

Index Terms—Inertial algorithms, nonconvex method, Kurdyka-Łojasiewicz property, image deblurring, ADMM, inertial proximal ADMM.

I. INTRODUCTION

Denoising and deblurring have numerous applications in communications, control, machine learning, and many other fields of engineering and science. Restoration of distorted images is, from the theoretical, as well as from the practical point of view, one of the most interesting and important problems of image processing. One special case is the blurring, due, for instance, to incorrect focus and/or blurring due to movement, or added Gaussian noise (a Gaussian blur).

A mathematical model for the process of blurred images can be expressed as follows. Let $\Omega \equiv \mathbb{R}^{N^2}$ be a two-dimensional index set representing the image domain, $\tilde{u} \in \Omega$ be the original

Tao Sun is with the College of Computer, National University of Defense Technology, Changsha, 410073, Hunan, China (e-mail: nudtsuntao@163.com).

Roberto Barrio is with Departamento de Matemática Aplicada, IUMA and CoDy group, Universidad de Zaragoza, Pedro Cerbuna, 12, Zaragoza, 50009, Spain. (e-mail: rbarrio@unizar.es).

Marcos Rodríguez is with CoDy group, Zaragoza, Universidad de Zaragoza, Pedro Cerbuna, 12, Zaragoza, 50009, Spain. (e-mail: imark.rodriguez@gmail.com).

Hao Jiang is with the College of Computer, National University of Defense Technology, Changsha, 410073, Hunan, China (e-mail: haojiang@nudt.edu.cn).

image, $\tilde{f}$ be the observed image, and $\tilde{K}$ be a linear blurring operator. Then, the blurred image can be written [1] as

$$\tilde{f} = \tilde{K}(\tilde{u}) + e, \qquad (1)$$

where e is an unknown additive noise vector. In this paper, the blurring operator $\tilde{K}$ is assumed to be known, otherwise, one will deal with the blind image deblurring problem [2], in which $\tilde{K}$ also needs to be solved.

In image processing one typically aims at recovering an image from noisy data while still keeping edges in the image, and this goal is the main reason of the tremendous success of the *Total Variation (TV) regularization* [3] for solving the deblurring problem (although other methods are also used). The TV method can be presented as

$$\min_{\tilde{u}} \left(\frac{1}{2} \|\tilde{K}\tilde{u} - \tilde{f}\|_{L^2(\Omega)}^2 + \sigma \|\nabla \tilde{u}\|_{L^1(\Omega^*)} \right), \qquad (2)$$

being $\sigma > 0$ a parameter and where ∇ is the gradient operator, $\Omega^* := \nabla(\Omega)$, and the norms $\|\tilde{v}\|_{L^1(\Omega^*)} := \sum_{i,j \in \Omega^*} |\tilde{v}_{i,j}|$, $\|\tilde{u}\|_{L^2(\Omega)} := \sqrt{\sum_{i,j \in \Omega} |\tilde{u}_{i,j}|^2}$.

In most situations, rather than directly minimizing the support of the image, one is interested in minimizing the support of the gradient of the recovered image. In most references, the convex methodology is considered [4], [5], [6], but in recent years, some nonconvex methods have been developed [7], [8], [9]. The use of a suitable nonconvex and nondifferentiable function allows possibly a smaller number of measurements than the convex one in compressed sensing [7]. In [10] the authors showed that nonconvex regularization terms in total variation-based image restoration yields even better edge preservation when compared to the convex-type regularization. Moreover, they showed that it seems to be also more robust with respect to noise. Nonconvex regularization in image restoration poses significant challenges on the existence of solutions of associated minimization problems and on the development of efficient solution algorithms.

The main difference between the convex and nonconvex methods is replacing the l_1-norm of the variational term by the nonconvex and nondifferentiable function $\|v\|_{\varphi(\Omega^*)} := \sum_{i,j \in \Omega^*} \varphi(v_{i,j})$ that uses the nonconvex regulation function φ, and that we refer to as the ℓ^q semi-norm ($0 < q < 1$). Therefore, the general nonconvex deblurring model is presented as

$$\min_{\tilde{u}} \left(\frac{1}{2} \|\tilde{K}\tilde{u} - \tilde{f}\|_{L^2(\Omega)}^2 + \sigma \|\nabla \tilde{u}\|_{\varphi(\Omega^*)} \right). \qquad (3)$$

Many efficient numerical algorithms have been developed for solving the TV regularization problem. One of the most

efficient methods for the convex problem (2) is the Alternating Direction Method of Multipliers (ADMM) algorithm [11], [12], [13]. In the general case (convex and nonconvex cases depend of the function φ), the method is constructed by introducing an auxiliary variable $\tilde{v}$, which actually represents $\nabla\tilde{u}$, to reformulate (3) into a composite optimization problem with linear constraints. The augmented Lagrange dual function is then

$$\tilde{\mathcal{L}}_\delta^\varphi(\tilde{u},\tilde{v},\tilde{p}) := \frac{1}{2}\|\tilde{K}\tilde{u} - \tilde{f}\|_{L^2(\Omega)}^2 + \sigma\|\tilde{v}\|_{\varphi(\Omega^*)}$$
$$- \langle\tilde{p},\nabla\tilde{u} - \tilde{v}\rangle + \frac{\delta}{2}\|\nabla\tilde{u} - \tilde{v}\|_{L^2(\Omega^*)}^2, \quad (4)$$

where $\delta > 0$ is a parameter, and the norm $\|X\|_{L^2(\Omega^*)} := \sqrt{\sum_{i,j\in\Omega^*}|X_{i,j}|^2}$. If $\varphi(\cdot) = |\cdot|_1$, we use the notation $\tilde{\mathcal{L}}_\delta^1(\tilde{u},\tilde{v},\tilde{p})$ for representing $\tilde{\mathcal{L}}_\delta^\varphi(\tilde{u},\tilde{v},\tilde{p})$. Now, the standard convex ADMM method ($\varphi(\cdot) = |\cdot|_1$) for the deblurring problem can be presented as

$$\begin{cases} \tilde{v}^{k+1} &= \arg\min_{\tilde{v}} \tilde{\mathcal{L}}_\delta^1(\tilde{u}^k,\tilde{v},\tilde{p}^k) \\ \tilde{u}^{k+1} &= \arg\min_{\tilde{u}} \tilde{\mathcal{L}}_\delta^1(\tilde{u},\tilde{v}^{k+1},\tilde{p}^k) \\ \tilde{p}^{k+1} &= \tilde{p}^k - \delta(\nabla\tilde{u}^{k+1} - \tilde{v}^{k+1}) \end{cases} \quad (5)$$

The earlier analyses of convergence and performance of the ADMM algorithms directly depended on the existing results of ADMM framework [14], [15], [16], [17], [18]. More recently, motivated by acceleration techniques proposed in [19], inertial algorithms have been proposed for many areas such as (distributed) optimization and imaging sciences in references [20], [21], [22], [23], [24]. The ideas of the inertial strategy have been also applied to ADMM in [4], [25]; and under several assumptions in convex case, some convergence results are proved in those articles. As the nonconvex penalty functions perform more efficiently in some applications, as above commented, nonconvex ADMM has been also developed and studied [26], [27], [28], [29], [30], [31], [32], [33], [34]. The main goal of this paper is to propose a new algorithm that combines the nonconvex methods and the inertial strategy organically.

In this paper, when φ is nonconvex, we consider a new inertial scheme for the image deblurring model (3). One of the main differences (and new difficulties) with the convex ADMM, is that in order to properly define the nonconvex ADMM some extra assumptions are needed to prove the convergence. First, at least one of the objective functions has to be smooth. And more, matrix corresponding to the smooth function is required to be injective, i.e., reversible. Thus, a direct application of the ADMM scheme to the image deblurring model cannot guarantee the convergence because the operator ∇ fails to be injective (although the numerical performance may be good in some cases). Considering this, we first modify the model (3), and then we develop the new nonconvex inertial ADMM. By using the Kurdyka-Łojasiewicz property, we prove the convergence of the new algorithm under several requirements on the parameters. In opposite to the convex case, selecting a suitable parameter δ is crucial to obtain the convergence of the new algorithm. In order to make the method more useful, we provide a probabilistic strategy for selecting a suitable δ.

$\|\cdot\|$ stands for $\|\cdot\|_{L^2(\Omega)}$ or $\|\cdot\|_{L^2(\Omega^*)}$ (L_2 norm)	
$\text{dist}(x,C) := \min_{y\in C}\|x - y\|$	$\|A\|_2 := \max_{\|x\|=1}\|Ax\|$
$\|\cdot\|_\varphi := \|\cdot\|_{\varphi(\Omega^*)}$	$\otimes$ stands for the Kronecker product
$\mathcal{C}^1$ stands for the function class whose derivatives are continuous	
for a matrix A, rank(A) stands for its rank, $\text{Null}(A) := \{x \mid Ax = \mathbf{0}\}$	
$\lambda_{\min}(A)$ stands for the minimum eigenvalue of A	

TABLE I: Basic notations, where x, y stand for points, C stands for a set, A stands for matrix

The rest of the paper is organized as follows. In Section II we collect some mathematical preliminaries needed for the convergence analysis. Section III presents the details for the new algorithm (inertial alternating minimization algorithm, IADMM) including the schemes and parameters. In Section IV, we prove the convergence of the new algorithm. Section V reports the numerical results and compares the algorithm with convex and nonconvex ADMM. Section VI gives some conclusions. Finally, we provide in the Appendixes all the detailed proofs of the proposed results.

II. MATHEMATICAL TOOLS

In this section we present the definitions and basic properties of the subdifferentials and the Kurdyka-Łojasiewicz functions used later in the convergence analysis. The basic notations used in this paper are detailed in Table I.

A. Subdifferentials

We collect several definitions as well as some useful properties in variational and convex analysis (see the monographes [35], [36], [37], [38]). For any matrix A, we define A^* to be the adjoint of A.

Definition 1 (subdifferentials [35], [37]): Let $J : \mathbb{R}^N \to (-\infty, +\infty]$ be a proper and lower semicontinuous function.

1) For a given $x \in \text{dom}(J)$, the Fréchet subdifferential of J at x, written as $\widehat{\partial}J(x)$, is the set of all vectors $u \in \mathbb{R}^N$ which satisfy

$$\liminf_{y\neq x\ y\to x} \frac{J(y) - J(x) - \langle u, y - x\rangle}{\|y - x\|_2} \geq 0.$$

When $x \notin \text{dom}(J)$, we set $\widehat{\partial}J(x) = \emptyset$.

2) The (limiting) subdifferential, or simply the subdifferential, of J at $x \in \mathbb{R}^N$, written as $\partial J(x)$, is defined through the following closure process

$$\partial J(x) := \{u \in \mathbb{R}^N : \exists x^k \to x, J(x^k) \to J(x)$$
$$\text{and } u^k \in \widehat{\partial}J(x^k) \to u \text{ as } k \to \infty\}. \quad (6)$$

It is easy to verify that the Fréchet subdifferential is convex and closed while the subdifferential is closed. When J is convex, the definition agrees with the subgradient in convex analysis [38] as

$$\partial J(x) := \{v : J(y) \geq J(x) + \langle v, y - x\rangle \text{ for any } y \in \mathbb{R}^N\}.$$

We call J is *strongly convex* with constant $a \in \mathbb{R}^+$ if for any $x, y \in \text{dom}(J)$ and any $v \in \partial J(x)$, it holds that

$$J(y) \geq J(x) + \langle v, y - x\rangle + \frac{a}{2}\|y - x\|_2^2.$$

And J is called as *L-gradient continuous* (Lipschitz) with constant $L > 0$ if

$$\|\nabla J(x) - \nabla J(y)\|_2 \le L\|x - y\|_2.$$

Noting the closedness of the subdifferential, we have the following simple proposition.

Proposition 1: If $v^k \in \partial J(x^k)$, $\lim_k v^k = v$ and $\lim_k x^k = x$, then we have

$$v \in \partial J(x). \tag{7}$$

Definition 2: A necessary condition for $x \in \mathbb{R}^N$ to be a minimizer of $J(x)$ is

$$\mathbf{0} \in \partial J(x), \tag{8}$$

which is also sufficient when J is convex. A point that satisfies (8) is called *(limiting) critical point*. The set of critical points of $J(x)$ is denoted by $\mathrm{crit}(J)$.

With these basics, we can easily obtain the following proposition.

Proposition 2: If $(\bar{u}, \bar{v}, \bar{p})$ is a critical point of $\mathcal{L}_\delta^\varphi$ whose definition given in (18), it must hold that

$$\begin{cases} -\bar{p} & \in & \sigma\partial\|\bar{v}\|_\varphi, \\ \nabla^*\bar{p} & = & K^*(K\bar{u} - f), \\ \nabla\bar{u} & = & \bar{v}. \end{cases} \tag{9}$$

Finally, the *proximal map* of J is defined as

$$\mathcal{S}_J(x) \in \arg\min_y \left\{ J(y) + \frac{1}{2}\|x - y\|_2^2 \right\}. \tag{10}$$

Note that J can be nonconvex. If J is convex, $\mathcal{S}_J$ is a point-to-point operator; otherwise, it may be point-to-set.

B. Kurdyka-Łojasiewicz property

In this paper the convergence analysis is based on the Kurdyka-Łojasiewicz functions, originated in the seminal works of Łojasiewicz [39] and Kurdyka [40]. This kind of functions has played a key role in several recent convergence results on nonconvex minimization problems and they are ubiquitous in applications.

Definition 3 ([41], [42]): (a) The function $J : \mathbb{R}^N \to (-\infty, +\infty]$ is said to have the Kurdyka-Łojasiewicz property at $\widehat{x} \in \mathrm{dom}(\partial J)$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\widehat{x}$ and a continuous concave function $\rho : [0, \eta) \to \mathbb{R}^+$ such that

1) $\rho(0) = 0$.
2) ρ is $\mathcal{C}^1$ on $(0, \eta)$.
3) For all $s \in (0, \eta)$, $\rho'(s) > 0$.
4) For all x in $U \bigcap \{x | J(\widehat{x}) < J(x) < J(\widehat{x}) + \eta\}$, the Kurdyka-Łojasiewicz inequality holds

$$\rho'(J(x) - J(\widehat{x})) \cdot \mathrm{dist}(\mathbf{0}, \partial J(x)) \ge 1. \tag{11}$$

(b) Proper lower semicontinuous functions which satisfy the Kurdyka-Łojasiewicz inequality at each point of $\mathrm{dom}(\partial J)$ are called *KŁ functions*.

Remark 1: There are large sets of functions that are KŁ functions [41].

Lemma 1 ([42]): Let $J : \mathbb{R}^N \to \mathbb{R}$ be a proper lower semi-continuous function and Π be a compact set. If J is a constant on Π and J satisfies the KŁ property at each point on Π, then there exists a concave function ρ satisfying the four properties given in Definition 3, and constants $\eta, \varepsilon > 0$ such that for any $\widehat{x} \in \Pi$ and any x satisfying that $\mathrm{dist}(x, \Pi) < \varepsilon$ and $f(\widehat{x}) < f(x) < f(\widehat{x}) + \eta$, it holds that

$$\rho'(J(x) - J(\widehat{x})) \cdot \mathrm{dist}(\mathbf{0}, \partial J(x)) \ge 1. \tag{12}$$

III. NONCONVEX IADMM ALGORITHM

In this section we introduce the new extended Inertial Alternating Direction Method of Multipliers (ADMM) algorithm for nonconvex functions.

In this paper, we consider $\Omega = [1, 2, \ldots, N] \times [1, 2, \ldots, N]$ (equivalent to space $\mathbb{R}^{N^2}$) as the two-dimensional index set representing the image domain. In this case, the image variable constrained on Ω is actually a $N \times N$ matrix. We use the symbol $\tilde{u}$ to present its vectorization (a vector of all the columns of the image variable). And then the original total variation operator then enjoys the following form

$$\nabla(\tilde{u}) = \left(\begin{array}{c} I_N \otimes D \\ D \otimes I_N \end{array} \right) \cdot \tilde{u}, \tag{13}$$

where I_N the identity matrix and D the banded matrix

$$D = \left(\begin{array}{ccccc} 1 & -1 & & & \\ & 1 & \ddots & & \\ & & \ddots & -1 & \\ & & & 1 & -1 \end{array} \right)_{(N-1) \times N}.$$

If we directly apply the inertial ADMM, the convergence is hard to be proved as ∇ fails to be injective. Therefore, we need to modify the image deblurring model (3). To that goal, we define

$$\mathcal{T} := \left(\begin{array}{cc} I_N \otimes D & \\ & D \otimes I_N \end{array} \right), \quad u := \left(\begin{array}{c} u_1 \in \mathbb{R}^{N^2} \\ u_2 \in \mathbb{R}^{N^2} \end{array} \right).$$

Obviously, we have $\mathcal{T} \in \mathbb{R}^{2N(N-1) \times 2N^2}$. Noting that

$$\begin{aligned} \mathrm{rank}(\mathcal{T}) &= \mathrm{rank}(I_N \otimes D) + \mathrm{rank}(D \otimes I_N) \\ &= \mathrm{rank}(I_N) \cdot \mathrm{rank}(D) + \mathrm{rank}(D) \cdot \mathrm{rank}(I_N) \\ &= 2N(N-1), \end{aligned}$$

and thus, $\mathcal{T}$ is injective. The following technical lemma focuses on giving a lower bound for the operator $\mathcal{T}^*$.

Lemma 2: For any $x \in \mathbb{R}^{2N(N-1)}$, it holds that

$$\|\mathcal{T}^*x\| \ge \frac{\|x\|}{\theta}, \tag{14}$$

where $\theta = 1/(2\sin(\frac{\pi}{2N}))$.

Then, the image deblurring model (3) is equivalent to

$$\min_{u_1 = u_2} \left(\frac{1}{2}\|\tilde{K}u_1 - \tilde{f}\|^2 + \sigma\|\mathcal{T}u\|_\varphi \right). \tag{15}$$

Instead, we consider its extended penalty form

$$\min_u \left(\frac{1}{2}\|\tilde{K}u_1 - \tilde{f}\|^2 + \sigma\|\mathcal{T}u\|_\varphi + \frac{\beta^2}{2}\|u_1 - u_2\|^2 \right), \tag{16}$$

where $\beta > 0$ is a large weight parameter. Therefore, we apply the nonconvex inertial ADMM to

$$\min_{u \in \mathbb{R}^{2N^2}} \left(\frac{1}{2} \|Ku - f\|^2 + \sigma \|\mathcal{T}u\|_\varphi \right), \qquad (17)$$

where $K = \begin{pmatrix} \tilde{K} & \mathbf{0} \\ \beta\mathbb{I} & -\beta\mathbb{I} \end{pmatrix}$ and $f = \begin{pmatrix} \tilde{f} \\ \mathbf{0} \end{pmatrix}$. This leads us to define the function

$$\mathcal{L}_\delta^\varphi(u, v, p) := \frac{1}{2}\|Ku - f\|^2$$
$$+ \sigma\|v\|_\varphi - \langle p, \mathcal{T}u - v \rangle + \frac{\delta}{2}\|\mathcal{T}u - v\|^2. \qquad (18)$$

Inertial methods have witnessed great success in convex ADMM and nonconvex first-order algorithms. In the nonconvex optimization community, the inertial style ADMM has never been proposed and analyzed. The convex inertial ADMM has been proposed in [4], in which one first uses the "inertial method" to refresh the current sequence with last iteration, and then performs the ADMM scheme with the updated variables. However, the direct extension of convex ADMM is not allowed in the nonconvex settings. This is because without convexity, several descents are heavily dependent on the continuities of the functions, which φ may fail to obey. And the difference of function values at two different points is hard to estimate, which leads to troubles in the convergence proof. Thus, in the updating of v^{k+1}, we used u^k rather than the updated one. And the nonconvex IADMM scheme proposed in this paper is defined as follows

$$\begin{cases} (\hat{u}^k, \hat{v}^k, \hat{p}^k) = (u^k, v^k, p^k) \\ \qquad\qquad + \alpha(u^k - u^{k-1}, v^k - v^{k-1}, p^k - p^{k-1}), \\ v^{k+1} = \arg\min_v \mathcal{L}_\delta^\varphi(u^k, v, \hat{p}^k), \\ u^{k+1} = \arg\min_u \mathcal{L}_\delta^\varphi(u, v^{k+1}, \hat{p}^k), \\ p^{k+1} = \hat{p}^k - \delta(\mathcal{T}u^{k+1} - v^{k+1}), \end{cases} \qquad (19)$$

where $\alpha > 0$ is a free parameter chosen by the user. Actually, if $\alpha = 0$, the algorithm then will reduce to basic ADMM.

Now we can focus on rewriting the inertial scheme for the image deblurring model (19). First, we rearrange the minimization for v^{k+1},

$$v^{k+1} = \mathcal{S}_{\frac{\sigma}{\delta}\|\cdot\|_{\varphi(\Omega^*)}} \left(\mathcal{T}u^k - \frac{\hat{p}^k}{\delta} \right), \qquad (20)$$

being S_J the proximal map of J (10). For a matrix v, and indices $(i, j) \in \Omega^*$,

$$[\mathcal{S}_{\frac{\sigma}{\delta}\|\cdot\|_\varphi}(v)]_{i,j} = \mathcal{S}_{\frac{\sigma}{\delta}\varphi}(v_{i,j}). \qquad (21)$$

The scheme for updating u^{k+1} can be rewritten as

$$K^*(Ku^{k+1} - f) - \mathcal{T}^*\hat{p}^k + \delta\mathcal{T}^*(\mathcal{T}u^{k+1} - v^{k+1}) = \mathbf{0}. \qquad (22)$$

That is also

$$u^{k+1} = (K^*K + \delta\mathcal{T}^*\mathcal{T})^{-1}(K^*f + \delta\mathcal{T}^*v^{k+1} + \mathcal{T}^*\hat{p}^k). \qquad (23)$$

Taking into account (19), (20) and (23) we propose a nonconvex inertial version of the IADMM algorithm (Algorithm 1).

Assumption 1: We assume that $\text{Null}(K_0) \bigcap \text{Null}(\mathcal{T}) = \{\mathbf{0}\}$, where $K_0 := \begin{pmatrix} \tilde{K} & \mathbf{0} \\ \mathbb{I} & -\mathbb{I} \end{pmatrix}$. And the minimum single

---

Algorithm 1 Nonconvex Inertial Alternating Minimization (IADMM) for Image Deblurring
***
Require: parameters $\alpha > 0$, $\delta > 0$,
  Initialization: $u^0 = u^1$, $v^0 = v^1$, $p^0 = p^1$
  for $k = 1, 2, \ldots$
    $(\hat{u}^k, \hat{v}^k, \hat{p}^k) = (u^k, v^k, p^k)$
        $+ \alpha(u^k - u^{k-1}, v^k - v^{k-1}, p^k - p^{k-1})$
    $v^{k+1} = \mathcal{S}_{\frac{\sigma}{\delta}\|\cdot\|_\varphi} \left(\mathcal{T}u^k - \dfrac{\hat{p}^k}{\delta} \right)$
    $u^{k+1} = (K^*K + \delta\mathcal{T}^*\mathcal{T})^{-1}(K^*f + \delta\mathcal{T}^*v^{k+1} + \mathcal{T}^*\hat{p}^k)$
    $p^{k+1} = \hat{p}^k - \delta(\mathcal{T}u^{k+1} - v^{k+1})$
  end for
***

value is given as ν.

This hypothesis also indicates that the matrix $\begin{pmatrix} \mathcal{T} \\ K_0 \end{pmatrix}$ is reversible. Note that the rank of $\mathcal{T}$ is $2N(N - 1)$. Then, the assumed hypothesis is easy to be satisfied.

We remark that u^{k+1} is the minimizer of $\mathcal{L}_\delta^\varphi(u, v^{k+1}, \hat{p}^k)$, which is strongly convex with $\lambda_{\min}(K^*K + \delta\mathcal{T}^*\mathcal{T})$. If we set $\delta, \rho > 1$, then we have

$$\mathcal{L}_\delta^\varphi(u^k, v^{k+1}, \hat{p}^k) - \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \hat{p}^k)$$
$$\geq \frac{\lambda_{\min}(K^*K + \delta\mathcal{T}^*\mathcal{T})}{2}\|u^{k+1} - u^k\|^2 \geq \frac{\nu}{2}\|u^{k+1} - u^k\|^2, \quad (24)$$

where we used the fact $\lambda_{\min}(K^*K + \delta\mathcal{T}^*\mathcal{T}) \geq \lambda_{\min}(K_0^*K_0 + \mathcal{T}^*\mathcal{T})$ when $\delta, \beta \geq 1$.

The following problem is what ν exactly is. In a real situation, the dimensions of $\mathcal{T}$ are large, and so, a direct calculation leads to a large computational cost. Therefore, we provide a probabilistic method to estimate a suitable value of θ. If $K^*K + \mathcal{T}^*\mathcal{T}$ is reversible, it is easy to see that $1/\nu = \|(K^*K + \mathcal{T}^*\mathcal{T})^{-1}\|_2$. Then, if we obtain a bound $\|(K^*K + \mathcal{T}^*\mathcal{T})^{-1}\|_2 \leq c$, we then have $\nu \geq 1/c$. To that goal we employ a Lemma proposed in [43]:

Lemma 3 (Lemma 4.1, [43]): For a fixed positive integer M and a real number $b > 1$, and given an independent family $\{w_i\}_{i=1,2,\ldots,M}$ of standard Gaussian vectors, we have that

$$\|(K^*K + \mathcal{T}^*\mathcal{T})^{-1}\|_2$$
$$\leq b\sqrt{\frac{2}{\pi}} \max_{i=1,2,\ldots,M} \|(K^*K + \mathcal{T}^*\mathcal{T})^{-1}w_i\|_2 \qquad (25)$$

with probability at least $1 - b^{-M}$.

Note that for computing $(K^*K + \mathcal{T}^*\mathcal{T})^{-1}w_i$ we just need several FFT and inverse FFT. Therefore, its computational cost is low ($\mathcal{O}(N \log N)$), and the estimation of ν is very fast.

IV. CONVERGENCE ANALYSIS

This section consists of two parts and provides a complete analysis of the convergence problem of the nonconvex IADMM algorithm. The first subsection contains the main convergence results, the proof sketch, the difficulties in the proof and theoretical contributions. While the second subsection introduces the necessary technical lemmas. Assumption 1 holds through this section.

A. Main results

Theorem 1 (Stationary point convergence): Assume that the free parameter δ satisfies the condition

$$\delta > \max\left\{1, \frac{6\theta^2\|K\|_2^4}{\nu} + \frac{7\alpha^2\theta^2\|K\|_2^4}{\nu}\right\} \quad (26)$$

with $\theta = 1/(2\sin(\frac{\pi}{2N}))$, then any cluster point (u^*, v^*, p^*) is also a critical point of $\mathcal{L}_\delta^\varphi$.

Theorem 1 describes the stationary point convergence result for the IADMM method, which is free of using the KŁ property of the functions. If the KŁ property is further assumed, the sequence convergence can be proved giving us the Theorem 2.

Theorem 2 (Sequence convergence): Let condition (36) hold, and the auxiliary function F (30) be a KŁ function. Then the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$, generated by Algorithm 1, converges to a critical point of $\mathcal{L}_\delta^\varphi$.

The proof can be divided into two parts, and in order to help the reader we first give a *brief sketch of the proof*:

I. In the first part we introduce an auxiliary sequence $\{w^k\}_{k=0,1,2,...}$, where w^k are composite points from $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$. An auxiliary function F is also proposed. In Lemma 5, we prove a "sufficient descent condition" of the values of F at $\{w^k\}_{k=0,1,2,...}$, i.e.,

$$F(w^k) - F(w^{k+1}) \geq \widehat{h}\|u^{k+1} - u^k\|^2, \quad (27)$$

where $\widehat{h} > 0$ is a positive constant.

II. We prove a "relative error condition" of $\{w^k\}_{k=0,1,2,...}$, i.e., there exists $v^{k+1} \in \partial F(w^{k+1})$ such that

$$\|v^{k+1}\| \leq \gamma(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|), \quad (28)$$

where $\gamma > 0$ is a positive constant. Note that this condition is different from the "real" relative error condition proposed in [41].

The major difficulty in deriving these two conditions is the use of inertial terms, with which the descent values are lower bounded by $\|v^{k+1} - \widehat{v}^k\|^2$ and $\|p^{k+1} - \widehat{p}^k\|^2$ rather than $\|v^{k+1} - v^k\|^2$ and $\|p^{k+1} - p^k\|^2$. Similarly, the relative error is bounded by $\|v^{k+1} - \widehat{v}^k\|$ and $\|p^{k+1} - \widehat{p}^k\|$. The relative error can be expanded by triangle inequalities, which is relatively proved. However, for the sufficient descent, the use of the triangle inequalities is much more difficult and technical for the lower boundedness.

The theoretical contributions in this paper are two-fold. The first one is, of course, dealing with the difference caused by the inertial terms. This part also includes how to design the scheme of the algorithm, whose details have been presented in previous section. The second one is to determine the parameters for IADMM applied to the image deblurring.

The main results can be proved with the following lemmas.

B. Technical lemmas

Lemma 4: Let the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$ be generated by Algorithm 1 to solve problem (3), then

$$\|p^k - p^{k+1}\| \leq \theta\|K\|_2^2 \|u^{k+1} - u^k\|, \quad (29)$$

where θ is given in Lemma 2 and $\|K\|_2$ denotes the spectral radius of K.

Now we provide the main technical lemma that states the descent condition for a suitable function of the sequences of the Algorithm 1.

Lemma 5: Let the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$ be generated by Algorithm 1 and conditions of Theorem 1 hold. By defining the auxiliary function F

$$F(u, v, p, x) := \mathcal{L}_\delta^\varphi(u, v, p) + \frac{7\alpha^2\theta^2\|K\|_2^4}{2\delta}\|u - x\|^2, \quad (30)$$

and

$$w^k := (u^k, v^k, p^k, u^{k-1}), \quad (31)$$

we have that

$$F(w^k) - F(w^{k+1}) \geq \widehat{h}\|u^{k+1} - u^k\|^2, \quad (32)$$

where

$$\widehat{h} := \frac{\nu}{2} - \left(\frac{3\theta^2\|K\|_2^4}{\delta} + \frac{7\alpha^2\theta^2\|K\|_2^4}{2\delta}\right) > 0. \quad (33)$$

If the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$ is bounded, then, it holds that

$$\lim_k \|u^{k+1} - u^k\| = \lim_k \|v^{k+1} - v^k\| = \lim_k \|p^{k+1} - p^k\| = 0. \quad (34)$$

Remark 2: Based on Lemma 5, it is important to guarantee that the condition (26) can be satisfied. This fact can be reached if δ is large enough. Fortunately, in the Algorithm 1 the parameter δ can be fixed by the user. Thus, the parameter δ shall be chosen large enough to guarantee the convergence considering condition (26).

Lemma 6: If the nonconvex regulation function φ is coercive and

$$\delta > \theta^2\|K\|_2^4, \quad (35)$$

the sequence $\{w^k\}_{k=0,1,2,...}$ is bounded.

Remark 3: By combining the conditions (26) and (35), we just need that

$$\delta > \max\left\{1, \frac{6\theta^2\|K\|_2^4}{\nu} + \frac{7\alpha^2\theta^2\|K\|_2^4}{\nu}, \theta^2\|K\|_2^4\right\}. \quad (36)$$

Lemma 7: Let the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$ be generated by Algorithm 1. Then, for any k, there exists $\gamma > 0$ and $s^{k+1} \in \partial F(w^{k+1})$ such that

$$\|s^{k+1}\| \leq \gamma(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|). \quad (37)$$

Now, we recall a definition about the limit point set $\mathcal{M}$ introduced in [42], which denotes the set of all the stationary points generated by the nonconvex IADMM. The specific mathematical definition of $\mathcal{M}$ is given as follows.

Definition 4: Let $\{w^k\}_{k=0,1,2,...}$ be generated by the nonconvex IADMM. We define the set $\mathcal{M}$ by

$$\mathcal{M} := \big\{x \in \mathbb{R}^N : \exists \text{ an increasing sequence of integers}$$
$$\{k_j\}_{j\in\mathbb{N}} \text{ such that } x^{k_j} \to x \text{ as } j \to \infty\big\}. \quad (38)$$

Lemma 8: Let the sequence $\{(u^k, v^k, p^k)\}_{k=0,1,2,...}$ be generated by Algorithm 1, F the auxiliary function defined in

(30) and suppose that condition (36) holds. Then, we have the following results.

(1) $\mathcal{M}$ is nonempty and $\mathcal{M} \subseteq \mathrm{crit}(F)$.
(2) $\lim_k \mathrm{dist}(w^k, \mathcal{M}) = 0$.
(3) The objective function F is finite and constant on $\mathcal{M}$.

V. NUMERICS

In this section, we illustrate the effectiveness of the proposed algorithm on different numerical blurred images with Gaussian blur.

All the programs have been written entirely in `C++`, and all the experiments are implemented under Linux running on a desktop computer with an Intel Core i5-2400S CPU (2.5 GHz) and 4 GB Memory. The FFT subroutines used in the algorithms are taken from the `fftw-3`[1] library. As test problems we have selected twelve images (see Figure 1), which include seven images from USC-SIPI image database[2], two classical test images (Lena and cameraman), one text image from "El Quijote" book and two medical images. In order to obtain the blurred images, we use, as it is common in literature, the blurring operator generated using a convolution with Gaussian kernel (`KernelSize` $= 17 \times 17$, `KernelMu` $= 0$, `KernelSigma` $= 7$) and circular mapping on the edges of the image.

The proposed algorithm IADMM (Algorithm 1) is compared with the widely used augmented Lagrangian methods (ADMM [22]) for image deblurring. We mainly consider two models, i.e., $\varphi(\cdot) = |\cdot|$ and $\varphi(\cdot) = |\cdot|^{\frac{1}{2}}$. We call them as TV1 and TV(1/2) methods, respectively. Note that TV1 is a convex method, while TV(1/2) is nonconvex. In the tests we have considered (unless so indicated) for all the methods the value of the penalty parameter $\delta = 0.001$ (a small one) and/or $\delta = 10$ (a large one) just to see the behaviour of the IADMM algorithm.

The performance of the deblurring algorithms is quantitatively measured by means of the *objective function* (Equations 2 or 3), the *Real Error* as $\|\cdot\|_2$ of the difference between the original and deblurred images, the *signal-to-noise ratio (SNR)* [22]

$$\mathrm{SNR}(u, u^*) = 10 \times \log_{10}\left(\frac{\|u - \bar{u}\|^2}{\|u - u^*\|^2}\right), \quad (39)$$

where u and u^* denote the original image and the restored image, respectively, and $\bar{u}$ represents the mean of the original image u, and the *residual* ($\mathrm{res}(k, k+1)$ and $\mathrm{res}_i(k, k+1)$ in the standard and inertial versions, respectively) as described in [22]:

$$\begin{aligned}
\mathrm{res}(k, k+1) &:= \frac{\|(u^{k+1}, p^{k+1}) - (u^k, p^k)\|}{1 + \|(u^k, p^k)\|}, \\
\mathrm{res}_i(k, k+1) &:= \frac{\|(u^{k+1}, p^{k+1}) - (\widehat{u}^k, \widehat{p}^k)\|}{1 + \|(\widehat{u}^k, \widehat{p}^k)\|}.
\end{aligned} \quad (40)$$

In the tests we do not provide CPU tests as all the algorithms have a very similar computation cost per iteration (mainly from the FFT routines). Therefore, there is almost no difference

[1] http://www.fftw.org/
[2] http://sipi.usc.edu/database/

between pictures showing iterations or CPU cost, and the consumed CPU is basically proportional to the respective number of iterations.

On our first test we use the brain M0 image with a low value of $\delta = 0.001$ and we show in Figure 2 that the performances of the ADMM TV1 and TV(1/2) are quite similar in error and SNR. Therefore, in the rest of comparisons we will just consider the TV1 method. For the IADMM method the inertial parameter α in Algorithm 1 is investigated firstly by using two different values ($\alpha = 0.2$ and 0.5). The main difference observed in these tests is that the nonconvex method is more unstable once reached the maximum precision (at this point the ADMM TV1 convex method seems to be the more stable with a quite smooth behaviour). The fastest convergence is observed using the IADMM (with the largest *stepsize* value $\alpha = 0.5$) method, but when the maximum precision is obtained unstable behaviour appears. Therefore, using mainly the information provided by the `residual` (Eq. (40)), we provide a *stop control* criterion (in the same spirit as [22]) that stops the iterative process at the black points of Figure 2 (d) in the tests. That is, we stop when the following condition is hold

$$\mathrm{res}(k, k+1) < \varepsilon \quad \text{or} \quad \mathrm{res}(k-1, k) < \mathrm{res}(k, k+1). \quad (41)$$

In the first case the convergence till the desired tolerance error is obtained, while in the second case the algorithm has reached its stability limit and the residual grows, behaving later in an unstable way. Note that the residual is used in the stop criterion, as it uses known data from the iterations (and which does not depend on the original image that is unknown). We remark that the use of stop control techniques avoids the use of unnecessary iterations, and also to stop at the limit accuracy of the method. Also, from the pictures we observe that the IADMM algorithm provides enough precision in a lower number of iterations. The larger α means faster method, but at the price of a more unstable method as it can be seen on Fig. 2(d). On that picture we observe that when the residual begins to behave chaotically, with sudden increases, it means that it is advisable to stop the iterative process as considered in the stop criterion (41) (black dot points on Fig. 2(d)). With that criterion, the IADMM method seems to be an interesting option for fast deblurring problems.

To observe more clearly the influence of the parameter α in Algorithm 1 we perform several tests on the brain M0 image on Figure 3 for values $\alpha = 0.1, 0.2, 0.5, 0.8, 1.0$ and 2.0. Note that this parameter plays a role similar to the *stepsize* (as it also occurs to the parameter δ), as it controls the perturbation at each step. A large value will provide, when the method works, a quite fast method, but on the other hand it makes the method more unstable. In fact, from the plot 3(b) we observe that in this case it would be optimal to use the parameter value $\alpha = 0.8$ in combination with the stop criterion, giving the maximum precision in just 28 iterations. Besides, it is shown that after the values selected by the stop criterion the residual begins to oscillate among values that provides similar error but that generates an unstable behaviour giving rise to an increment of the error in subsequent iterations (this instability is delayed when the parameter α decreases, what
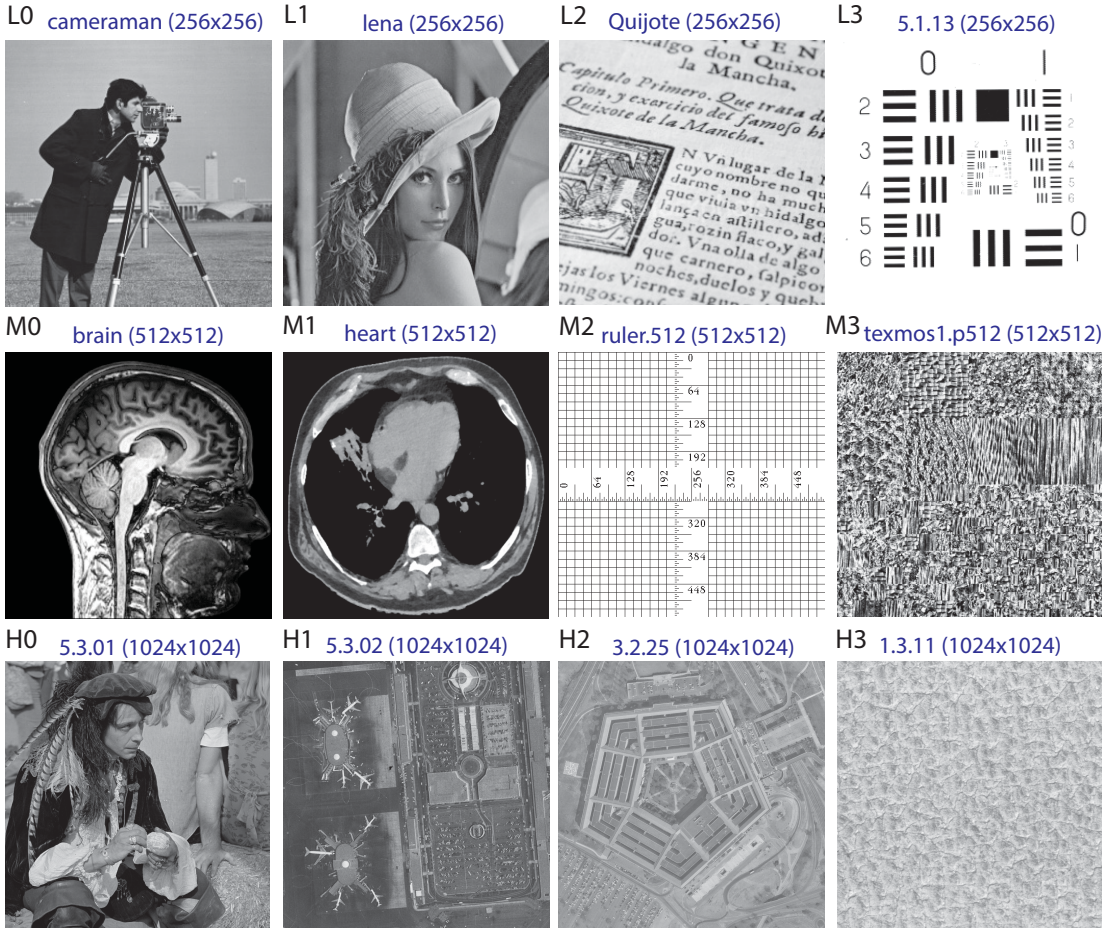
Fig. 1: Original images (some of them from USC-SIPI image database) in low (L), medium (M) and hight-resolution (H). (L0) Cameraman (256×256); (L1) Lena (256×256); (L2) "El Quijote" (256×256); (L3) 5.1.09 (256×256); (M0) Brain (512×512); (M1) Heart (512×512); (M2) ruler.512 (512×512); (M3) texmos1.p512 (512×512); (H0) 5.3.01 (1024×1024); (H1) 5.3.02 (1024×1024); (H2) 3.2.25 (1024×1024); (H3) 1.3.11 (1024×1024).

is expected because the increment is smaller, as the vertical lines connecting the error and residual plots show).

The influence of the penalty parameter δ is also quite relevant, but a detailed analysis is out of the scope of this paper. On the Figure 4 we show the evolution of the residual using two values of $\alpha = 0.2, 0.5$ and several values of the parameter $\delta = 0.001, 0.01, 0.1, 1$ and 10. We observe that low values of the penalty parameter δ gives a lower residual, but the error is lower for large δ providing a faster convergence, and it has a big effect on the empirical performance of the methods as shown in [33], but it remains to study optimal combinations of the parameters δ and α and suitable criteria for an automatic selection (this will be part of the next steps in our study of these methods).

On Figure 5, we present the original medium resolution brain M0 image, the blurred one using, as indicated, a con-volution with Gaussian kernel, and the results of the IADMM deblurring images using two error tolerances ($\varepsilon = 5 \times 10^3$, 5×10^4) in the stop criterion (41). We can see that in both cases the quality of the recovered image is visually good.

A more detailed analysis is shown on Table II, where we present for the IADMM ($\alpha = 0.5$ and 0.2) and the ADMM

(TV1) methods (we do not show results for the TV(1/2) as they are quite similar to those of the TV1 ones), the iteration number, real error (ERROR), SNR, the residual (RES) and the *efficiency rates* of the IADMM ($\alpha = 0.5$) method over the two other ones, I2/I1 and I3/I1, for different values of the tolerance ε and using a small value of $\delta = 0.001$. In all the methods we have used the same values of the parameters and a similar stop criterion (41) with the respective definition of the residual (Eqn. (40)). From the tests, the inertial IADMM nonconvex methods are the fastest, as expected, but also with a more unstable behaviour, also as expected due to the nonconvex version of the methods. Note that from Figure 2 the ADMM TV1 and TV(1/2) perform similarly for low δ, but if we make simulations small differences appear in medium-high resolution, obtaining for the M1 image the TV1 method 113 iterations and a ratio 2.26, but the TV(1/2) 99 iterations and a ratio 1.98; and for the H0 image the TV1 uses 110 iterations and a ratio 2.75, but the TV(1/2) 81 iterations and a ratio 2.03. That is, the nonconvex TV(1/2) may perform better in some circumstances, but the instability may also appear. In case of using larger values of the parameter δ, as the convergence theorems suggest, all the methods are much faster, giving a

deblurring results for $\delta = 0.001$, $\varepsilon = 0.005$

	IADMM ($\alpha=0.5$)				IADMM ($\alpha=0.2$)					ADMM(TV1)				
IMG	I1	ERROR	SNR	RES	I2	ERROR	SNR	RES	I2/I1	I3	ERROR	SNR	RES	I3/I1
L0	4	17.4	11.1	4.59e-03	6	17.4	11.1	4.62e-03	1.50	7	17.6	11.0	4.80e-03	1.75
L1	5	16.7	10.0	4.67e-03	8	16.5	10.1	4.60e-03	1.60	9	16.8	9.9	4.95e-03	1.80
L2	2	19.4	8.7	4.36e-03	2	19.9	8.5	4.69e-03	1.00	2	20.2	8.3	4.88e-03	1.00
L3	2	16.6	13.2	3.01e-03	2	17.1	13.0	3.25e-03	1.00	2	17.9	12.6	3.67e-03	1.00
M0	8	30.6	13.8	4.50e-03	13	30.4	13.8	4.67e-03	1.62	16	30.6	13.8	4.88e-03	2.00
M1	9	28.5	13.7	4.33e-03	14	28.7	13.6	4.86e-03	1.56	18	28.5	13.7	4.78e-03	2.00
M2	6	36.6	12.8	3.48e-03	7	40.1	12.0	4.82e-03	1.17	9	39.9	12.1	4.86e-03	1.50
M3	7	35.5	12.5	4.34e-03	11	35.6	12.4	4.50e-03	1.57	13	36.3	12.3	4.97e-03	1.86
H0	8	71.4	10.3	4.50e-03	13	71.0	10.3	4.65e-03	1.62	16	71.4	10.3	4.84e-03	2.00
H1	9	69.2	6.3	4.47e-03	14	69.7	6.2	4.96e-03	1.56	18	69.2	6.3	4.89e-03	2.00
H2	7	74.2	3.3	4.16e-03	10	76.2	3.0	4.87e-03	1.43	13	75.5	3.1	4.77e-03	1.86
H3	6	77.4	1.5	4.08e-03	8	80.8	1.1	4.87e-03	1.33	11	78.9	1.3	4.48e-03	1.83

deblurring results for $\delta = 0.001$, $\varepsilon = 0.001$

	IADMM ($\alpha=0.5$)				IADMM ($\alpha=0.2$)					ADMM(TV1)				
IMG	I1	ERROR	SNR	RES	I2	ERROR	SNR	RES	I2/I1	I3	ERROR	SNR	RES	I3/I1
L0	19	12.3	14.1	9.96e-04	32	12.2	14.2	9.75e-04	1.68	40	12.2	14.2	9.82e-04	2.11
L1	22	12.0	12.9	9.92e-04	37	11.8	13.0	9.69e-04	1.68	46	11.8	13.0	9.82e-04	2.09
L2	14	13.2	12.0	9.86e-04	23	13.1	12.1	1.00e-03	1.64	29	13.1	12.1	9.99e-04	2.07
L3	11	12.2	15.9	9.10e-04	17	12.2	15.9	9.76e-04	1.55	21	12.3	15.8	9.96e-04	1.91
M0	26	21.7	16.8	9.84e-04	43	21.5	16.8	9.78e-04	1.65	54	21.4	16.9	9.79e-04	2.08
M1	29	20.3	16.6	9.91e-04	48	20.1	16.7	9.80e-04	1.66	60	20.1	16.7	9.86e-04	2.07
M2	16	27.2	15.4	9.52e-04	26	27.1	15.4	9.84e-04	1.62	33	27.0	15.5	9.79e-04	2.06
M3	24	25.1	15.5	9.66e-04	39	25.0	15.5	9.80e-04	1.62	48	25.1	15.5	1.00e-03	2.00
H0	28	50.7	13.2	9.92e-04	46	50.3	13.3	9.89e-04	1.64	58	50.2	13.3	9.85e-04	2.07
H1	33	48.5	9.4	9.82e-04	53	48.5	9.4	9.99e-04	1.61	67	48.3	9.4	9.91e-04	2.03
H2	22	54.3	6.0	9.88e-04	36	53.9	6.0	9.98e-04	1.64	46	53.6	6.1	9.80e-04	2.09
H3	18	57.3	4.1	9.43e-04	29	57.2	4.1	9.81e-04	1.61	36	57.4	4.1	9.99e-04	2.00

deblurring results for $\delta = 0.001$, $\varepsilon = 0.0005$

	IADMM ($\alpha=0.5$)				IADMM ($\alpha=0.2$)					ADMM(TV1)				
IMG	I1	ERROR	SNR	RES	I2	ERROR	SNR	RES	I2/I1	I3	ERROR	SNR	RES	I3/I1
L0	36	10.8	15.2	4.94e-04	59	10.8	15.3	4.92e-04	1.64	73	10.8	15.3	4.99e-04	2.03
L1	42	10.5	14.0	4.96e-04	68	10.5	14.1	4.99e-04	1.62	86	10.5	14.1	4.95e-04	2.05
L2	27	11.5	13.2	4.91e-04	44	11.4	13.3	4.95e-04	1.63	55	11.4	13.3	4.97e-04	2.04
L3	21	10.6	17.2	4.80e-04	34	10.5	17.2	4.88e-04	1.62	42	10.6	17.1	4.97e-04	2.00
M0	49	18.2	18.3	4.99e-04	80	18.1	18.3	4.96e-04	1.63	100	18.1	18.3	4.97e-04	2.04
M1	50	17.4	17.9	5.50e-04	79	17.5	17.9	5.64e-04	1.58	113	16.8	18.2	4.96e-04	2.26
M2	34	22.1	17.2	5.83e-04	54	22.1	17.2	5.90e-04	1.59	158	17.8	19.1	5.00e-04	4.65
M3	41	22.0	16.6	6.65e-04	65	22.0	16.6	6.74e-04	1.59	401	16.3	19.2	5.00e-04	9.78
H0	40	45.8	14.1	6.73e-04	64	45.8	14.1	6.85e-04	1.60	110	41.8	14.9	4.96e-04	2.75
H1	68	39.5	11.1	4.99e-04	108	39.6	11.1	5.05e-04	1.59	138	39.3	11.2	4.97e-04	2.03
H2	42	45.3	7.6	4.90e-04	67	45.3	7.6	4.99e-04	1.60	84	45.3	7.6	4.99e-04	2.00
H3	32	48.8	5.5	4.91e-04	52	48.6	5.5	4.94e-04	1.62	65	48.7	5.5	4.96e-04	2.03

TABLE II: Deblurring results for a small value of δ ($\delta = 0.001$) using different values of the tolerance ε. *Iteration number* (I1, I2, I3), *Real error* (ERROR), *SNR*, *Residual* (RES) and *efficiency rates* I2/I1 and I3/I1 using the IADMM ($\alpha = 0.5$ and 0.2) and the ADMM (TV1) methods applied to all the 12 test images (IMG).

better performance as shown on Table III for $\delta = 10$. Finally, we present on Table IV some results obtained by changing the value of the parameter δ and with the fixed value of the tolerance $\varepsilon = 0.01$. Again, the IADMM ($\alpha = 0.5$) method presents the best performance. Therefore, globally, the inertial IADMM version seems to be an interesting option, being the fastest one.

As above commented, how to select optimal combinations of the parameters δ and α, the use on other types of blur, development of suitable criteria for automatic selection of all the parameters and so on, remains a research issue, but these goals are out of the scope of this article.

VI. CONCLUSION

In this paper, a more efficient nonconvex inertial alternating minimization algorithm (IADMM) is developed for solving the Total Variation model for image deblurring. The proposed scheme is based on using the inertial proximal strategy on nonconvex ADMM methods. By the using the Kurdyka-Łojasiewicz property, we prove the convergence of the algorithm under several reasonable assumptions. Numerical experiments demonstrate that the proposed algorithm overall

deblurring results for $\delta = 10$, $\varepsilon = 0.01$

IMG	IADMM ($\alpha = 0.5$)				IADMM ($\alpha = 0.2$)					ADMM (TV1)				
	I1	ERROR	SNR	RES	I2	ERROR	SNR	RES	I2/I1	I3	ERROR	SNR	RES	I3/I1
L0	5	3.0	26.5	4.72e-01	5	2.8	27.1	4.31e-01	1.00	10	6.2	20.0	9.56e-02	2.00
L1	3	3.3	24.1	5.38e-01	5	3.2	24.3	4.62e-01	1.67	10	5.8	19.2	9.48e-02	3.33
L2	1	3.8	22.9	8.22e-01	3	3.7	23.0	5.13e-01	3.00	10	9.3	15.0	9.68e-02	10.00
L3	6	1.4	34.6	2.41e-01	7	1.3	35.5	2.00e-01	1.17	10	7.0	20.7	9.86e-02	1.67
M0	3	6.2	27.6	5.79e-01	5	6.0	28.0	4.98e-01	1.67	10	7.0	26.5	9.74e-02	3.33
M1	7	4.3	30.2	4.40e-01	7	3.9	31.0	4.12e-01	1.00	10	5.4	28.1	9.78e-02	1.43
M2	4	3.0	34.5	1.50e-01	4	3.0	34.7	1.00e-01	1.00	10	20.6	17.8	9.90e-02	2.50
M3	1	14.6	20.2	8.94e-01	1	14.5	20.2	8.68e-01	1.00	10	31.7	13.4	9.64e-02	10.00
H0	5	13.8	24.5	4.96e-01	5	12.6	25.3	4.44e-01	1.00	10	15.2	23.7	9.52e-02	2.00
H1	3	16.6	18.7	5.73e-01	5	16.2	18.9	4.98e-01	1.67	10	20.4	16.9	9.47e-02	3.33
H2	5	15.7	16.8	5.11e-01	5	14.6	17.4	4.57e-01	1.00	10	20.7	14.4	9.29e-02	2.00

TABLE III: Deblurring results for a large value of δ ($\delta = 10$) using the tolerance $\varepsilon = 0.01$. *Iteration number* (I1, I2, I3), *Real error* (ERROR), *SNR*, *Residual* (RES) and *efficiency rates* I2/I1 and I3/I1 using the IADMM ($\alpha = 0.5$ and 0.2) and the ADMM (TV1) methods applied to all the 12 test images (IMG).

deblurring results for $\varepsilon = 0.01$ on M0 image

δ	IADMM ($\alpha = 0.5$)				IADMM ($\alpha = 0.2$)					ADMM (TV1)				
	I1	ERROR	SNR	RES	I2	ERROR	SNR	RES	I2/I1	I3	ERROR	SNR	RES	I3/I1
0.001	6	33.5	13.0	7.40e-03	8	35.3	12.5	8.70e-03	1.33	9	34.0	12.8	7.50e-03	1.50
0.01	4	19.1	17.9	8.88e-03	4	21.2	16.9	9.82e-03	1.00	5	20.0	17.4	7.68e-03	1.25
0.10	3	11.2	22.5	2.39e-02	4	11.4	22.3	2.31e-02	1.33	5	10.8	22.8	2.32e-02	1.67
1.00	1	8.0	25.4	2.25e-01	2	7.5	26.0	2.03e-01	2.00	3	7.3	26.3	2.20e-01	3.00
10.00	3	6.2	27.6	5.79e-01	5	6.0	28.0	4.98e-01	1.67	10	7.0	26.5	9.74e-02	3.33

TABLE IV: Deblurring results for several values of parameter δ using the tolerance $\varepsilon = 0.01$ on M0 image. *Iteration number* (I1, I2, I3), *Real error* (ERROR), *SNR*, *Residual* (RES) and *efficiency rates* I2/I1 and I3/I1 using the IADMM ($\alpha = 0.5$ and 0.2) and the ADMM (TV1) methods.

outperforms the widely used augmented Lagrangian ADMM methods, being a fast option, although it can be unstable for high precision. This instability is easily controlled via a suitable stop control criteria.

APPENDIX A: PROOF OF LEMMA 2

Direct computation yields

$$\mathcal{T}^*\mathcal{T} = \begin{pmatrix} (I_N \otimes D)(I_N \otimes D)^* & \\ & (D \otimes I_N)(D \otimes I_N)^* \end{pmatrix}. \quad (42)$$

Then, we have

$$\lambda_{\min}(\mathcal{T}^*)^2 = \lambda_{\min}(\mathcal{T}\mathcal{T}^*) = \quad (43)$$
$$\min\left\{\lambda_{\min}\big((I_N \otimes D)(I_N \otimes D)^*\big), \lambda_{\min}\big((D \otimes I_N)(D \otimes I_N)^*\big)\right\}.$$

With Lemma 7.2 in [44], we obtain

$$\lambda_{\min}\big((I_N \otimes D)(I_N \otimes D)^*\big)$$
$$= \lambda_{\min}\big((D \otimes I_N)(D \otimes I_N)^*\big) = \lambda_{\min}(DD^*). \quad (44)$$

Noting DD^* is a symmetric tridiagonal matrix

$$DD^* = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}, \quad (45)$$

and using the result in [45],

$$\lambda_{\min}(DD^*) = 2 + 2\cos\left(\pi - \frac{\pi}{N}\right)$$
$$= 4\cos^2\left(\frac{\pi}{2} - \frac{\pi}{2N}\right) = 4\sin^2\left(\frac{\pi}{2N}\right), \quad (46)$$

we obtain $\lambda_{\min}(\mathcal{T}^*) \geq 2\sin\left(\frac{\pi}{2N}\right)$. Now, taking into account that $\mathcal{T}^*$ is full row-rank,

$$\|\mathcal{T}^*x\| \geq \lambda_{\min}(\mathcal{T}^*)\|x\| \geq 2\sin\left(\frac{\pi}{2N}\right)\|x\|. \quad (47)$$

APPENDIX B: PROOF OF LEMMA 4

As the point u^{k+1} is the minimization of $\mathcal{L}_\delta^\varphi(u, v^{k+1}, \widehat{p}^k)$, then the optimization condition yields

$$K^*(Ku^{k+1} - f) - \mathcal{T}^*\widehat{p}^k + \delta\mathcal{T}^*(\mathcal{T}u^{k+1} - v^{k+1}) = \mathbf{0}. \quad (48)$$

Therefore,

$$K^*(Ku^{k+1} - f) - \mathcal{T}^*p^{k+1} = \mathbf{0}, \quad (49)$$

and replacing $k + 1$ by k

$$K^*(Ku^k - f) - \mathcal{T}^*p^k = \mathbf{0}. \quad (50)$$

Subtracting (49) and (50), and using Lemma 2,

$$\|p^k - p^{k+1}\| \leq \theta\|\mathcal{T}^*(p^k - p^{k+1})\|$$
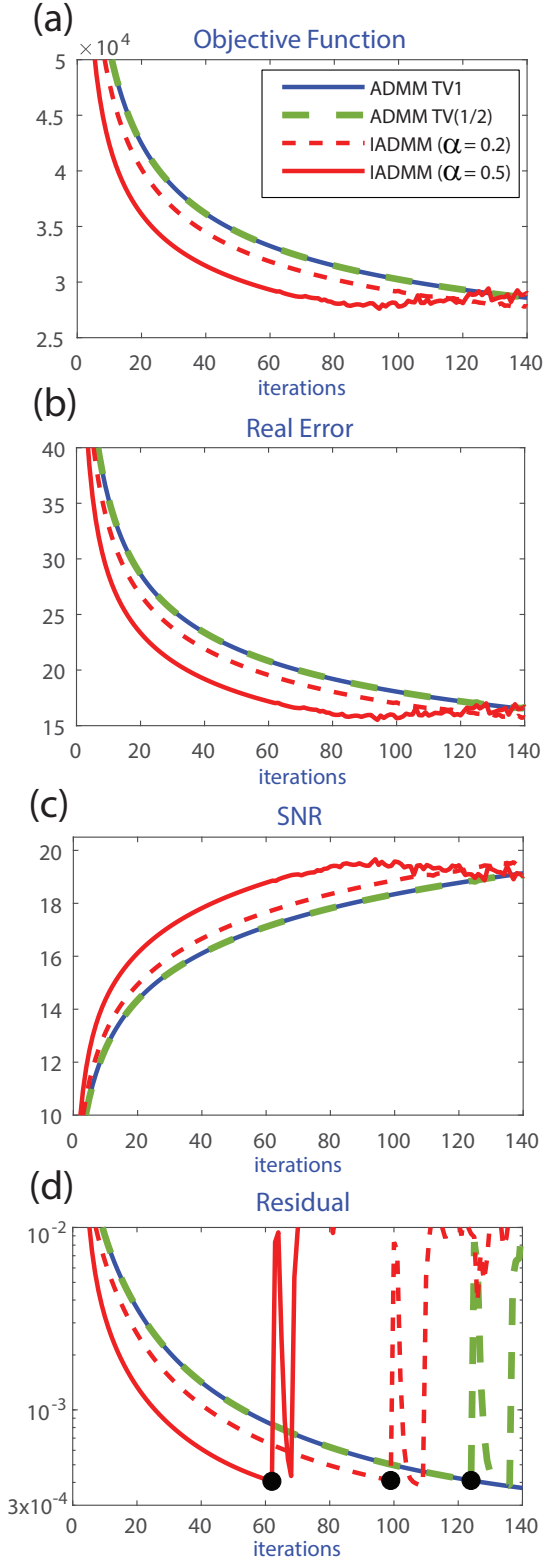$$= \theta\|K^*K(u^{k+1} - u^k)\| \leq \theta\|K\|_2^2 \cdot \|u^{k+1} - u^k\|. \quad (51)$$

Fig. 2: The evolution curves of the (a) *objective function*, (b) the *Real Error*, (c) the *signal-to-noise ratio (SNR)* and (d) the *Residual*, for the brain M0 image versus the iteration number. All the simulations have been done using the ADMM TV1 and TV(1/2) algorithms and the IADMM algorithm using $\alpha = 0.2$ and 0.5, and $\delta = 0.001$.
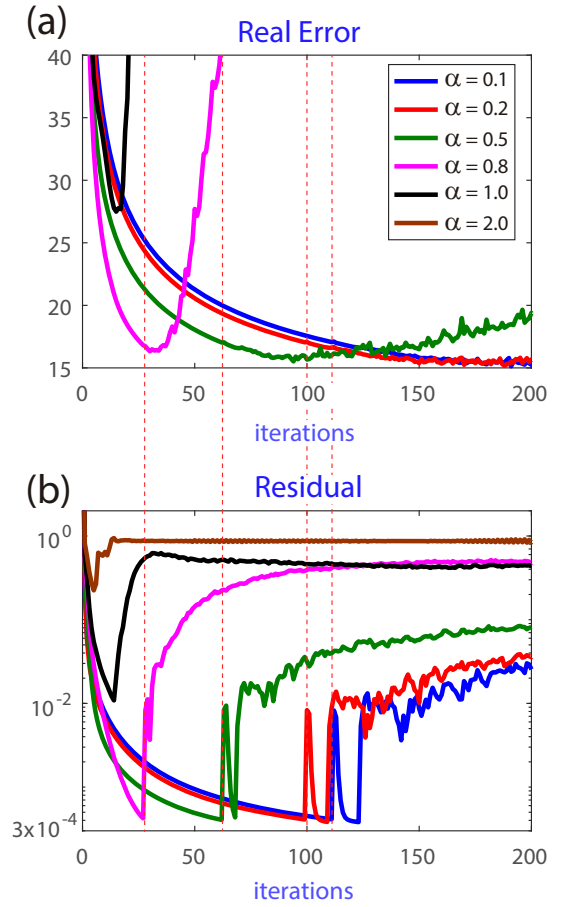

Fig. 3: The evolution curves of (a) the *Real Error* and (b) the *Residual*, for the brain M0 image versus the iteration number. All the simulations have been done using the IADMM algorithm with several values of the parameter $\alpha = 0.1, 0.2, 0.5, 0.8, 1.0$ and 2.0 and $\delta = 0.001$.

APPENDIX C: PROOF OF LEMMA 5

Note that v^{k+1} is the minimizer of $\mathcal{L}_\delta^\varphi(u^k, v, \widehat{p}^k)$ with respect to the variable v. Then we have:

$$\mathcal{L}_\delta^\varphi(u^k, v^{k+1}, \widehat{p}^k) \leq \mathcal{L}_\delta^\varphi(u^k, v^k, \widehat{p}^k). \tag{52}$$

In (24), we have obtained

$$\mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \frac{\nu}{2}\|u^{k+1} - u^k\|^2 \leq \mathcal{L}_\delta^\varphi(u^k, v^{k+1}, \widehat{p}^k). \tag{53}$$

By direct calculations, we obtain

$$
\begin{aligned}
&\mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, p^{k+1}) \\
&= \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \langle \widehat{p}^k - p^{k+1}, \mathcal{T}u^{k+1} - v^{k+1} \rangle \\
&= \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \frac{1}{\delta}\|\widehat{p}^k - p^{k+1}\|_2^2 \\
&\leq \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \frac{2}{\delta}\|\widehat{p}^k - p^k\|_2^2 + \frac{2}{\delta}\|p^k - p^{k+1}\|_2^2 \\
&\leq \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \frac{2\alpha^2}{\delta}\|p^k - p^{k-1}\|_2^2 + \frac{2}{\delta}\|p^k - p^{k+1}\|_2^2 \\
&\leq \mathcal{L}_\delta^\varphi(u^{k+1}, v^{k+1}, \widehat{p}^k) + \frac{2\theta^2\|K\|_2^4}{\delta}\|u^{k+1} - u^k\|^2 \\
&\quad + \frac{2\alpha^2\theta^2\|K\|_2^4}{\delta}\|u^k - u^{k-1}\|^2. \tag{54}
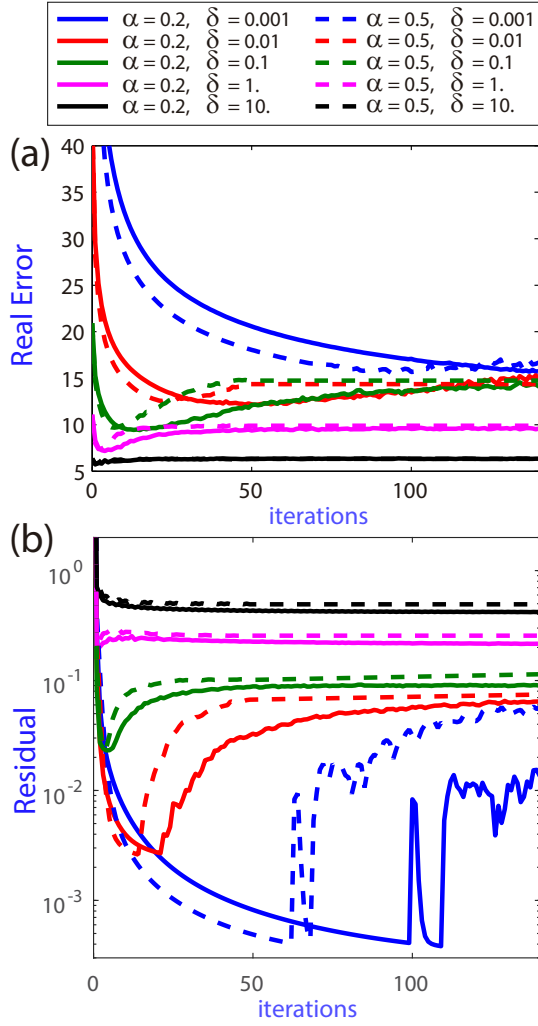\end{aligned}
$$

Fig. 4: The evolution curves of (a) the *Real Error* and (b) the *Residual*, for the brain M0 image versus the iteration number. All the simulations have been done using the IADMM algorithm with several values of the parameters $\alpha = 0.2, 0.5$ and $\delta = 0.001, 0.01, 0.1, 1$ and 10.

and

$$
\begin{aligned}
& \mathcal{L}_\delta^\varphi(u^k, v^k, \widehat{p}^k) \\
& = \mathcal{L}_\delta^\varphi(u^k, v^k, p^k) + \langle p^k - \widehat{p}^k, \mathcal{T}u^k - v^k \rangle \\
& = \mathcal{L}_\delta^\varphi(u^k, v^k, p^k) + \frac{1}{\delta} \langle p^k - \widehat{p}^k, p^{k+1} - \widehat{p}^k \rangle \\
& \leq \mathcal{L}_\delta^\varphi(u^k, v^k, p^k) + \frac{\|p^k - \widehat{p}^k\|^2}{2\delta} + \frac{\|p^{k+1} - \widehat{p}^k\|^2}{2\delta} \\
& \leq \mathcal{L}_\delta^\varphi(u^k, v^k, p^k) + \frac{3\alpha^2 \|p^k - p^{k-1}\|^2}{2\delta} + \frac{\|p^{k+1} - p^k\|^2}{\delta} \\
& \leq \mathcal{L}_\delta^\varphi(u^k, v^k, p^k) + \frac{\theta^2 \|K\|_2^4}{\delta} \|u^{k+1} - u^k\|^2 \\
& \quad + \frac{3\alpha^2 \theta^2 \|K\|_2^4}{2\delta} \|u^k - u^{k-1}\|^2.
\end{aligned}
\tag{55}
$$

(a) ORIGINAL IMAGE: brain

(b) BLURRED IMAGE: brain

(c) IADMM ($\varepsilon = 0.005$)
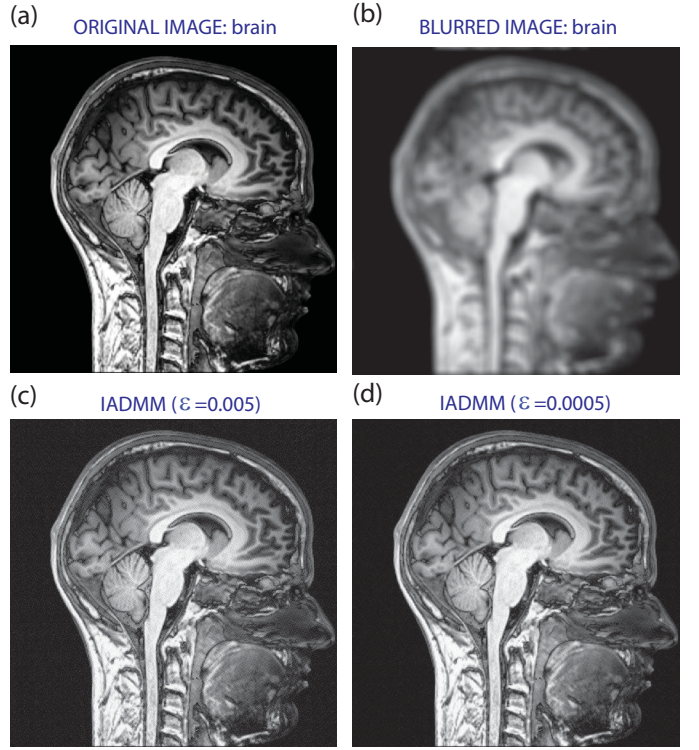
(d) IADMM ($\varepsilon = 0.0005$)

Fig. 5: Deblurred images at different stages of the IADMM method for brain M0 image. (a) Original image; (b) Blurred image; (c) Recovery by IADMM using error tolerance $\varepsilon = 5 \times 10^{-3}$; (d) Recovery by IADMM using error tolerance $\varepsilon = 5 \times 10^{-4}$.

Combining the above equations, we obtain

$$
\begin{aligned}
& \mathcal{L}_\delta^\varphi(x^{k+1}, y^{k+1}, p^{k+1}) + \frac{\nu}{2} \|u^{k+1} - u^k\|^2 \\
& \leq \mathcal{L}_\delta^\varphi(x^k, y^k, p^k) + \frac{3\theta^2 \|K\|_2^4}{\delta} \|u^{k+1} - u^k\|^2 \\
& \quad + \frac{7\alpha^2 \theta^2 \|K\|_2^4}{2\delta} \|u^k - u^{k-1}\|^2.
\end{aligned}
$$

Thus, taking into account the definition of F (30), we have

$$
\begin{aligned}
& F(w^k) - F(w^{k+1}) \tag{56} \\
& \geq \left[\frac{\nu}{2} - \left(\frac{3\theta^2 \|K\|_2^4}{\delta} + \frac{7\alpha^2 \theta^2 \|K\|_2^4}{2\delta} \right) \right] \|u^{k+1} - u^k\|^2.
\end{aligned}
$$

Now, using the condition (26) and the definition (33) of $\widehat{h}$, we obtain the descent result given in Eq. (32).

The boundness of $\{(u^k, v^k, p^k)\}_{k=0,1,2,\ldots}$ implies the boundness of $\{w^k\}_{k=0,1,2,\ldots}$, and the continuity of F implies that $\{F(w^k)\}_{k=0,1,2,\ldots}$ is bounded. From the above proven result Eq. (32), $F(w^k)$ is decreasing. Thus, the sequence $\{F(w^k)\}_{k=0,1,2,\ldots}$ is convergent, i.e., $\lim_k [F(w^k) - F(w^{k+1})] = 0$. Therefore, from (32), we have

$$
\lim_k \|u^{k+1} - u^k\|_2 \leq \lim_k \sqrt{\frac{F(w^k) - F(w^{k+1})}{\widehat{h}}} = 0, \tag{57}
$$

which indicates $\lim_k \|p^{k+1} - p^k\| = 0$ due to (29). The scheme of updating p^{k+1} gives us $\lim_k \|v^{k+1} - v^k\| = 0$.

APPENDIX D: PROOF OF LEMMA 6

From (50), we derive

$$\|p^k\|^2 \le \theta^2\|K\|_2^4 \cdot \|Ku^k - f\|^2. \tag{58}$$

By direct calculations, we obtain

$$
\begin{aligned}
\mathcal{L}_\delta^\varphi(u^k, v^k, p^k) &= \frac{1}{2}\|Ku^k - f\|^2 + \sigma\|v^k\|_\varphi \\
&\quad - \langle p^k, \mathcal{T}u^k - v^k\rangle + \frac{\delta}{2}\|\mathcal{T}u^k - v^k\|^2 \\
&= \frac{1}{2}\|Ku^k - f\|^2 + \sigma\|v^k\|_\varphi \\
&\quad + \frac{\delta}{2}\left\|\mathcal{T}u^k - v^k - \frac{p^k}{\delta}\right\|^2 - \frac{\|p^k\|^2}{2\delta} \\
&\ge \frac{1}{2}\|Ku^k - f\|^2 + \sigma\|v^k\|_\varphi + \frac{\delta}{2}\left\|\mathcal{T}u^k - v^k - \frac{p^k}{\delta}\right\|^2 \\
&\quad - \frac{\theta^2\|K\|_2^4}{2\delta}\|Ku^k - f\|^2 \\
&= \left(\frac{1}{2} - \frac{\theta^2\|K\|_2^4}{2\delta}\right)\|Ku^k - f\|^2 \\
&\quad + \frac{\delta}{2}\left\|\mathcal{T}u^k - v^k - \frac{p^k}{\delta}\right\|^2 + \sigma\|v^k\|_\varphi.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
F(w^k) \ge &\left(\frac{1}{2} - \frac{\theta^2\|K\|_2^4}{2\delta}\right)\|Ku^k - f\|^2 \\
&+ \frac{\delta}{2}\left\|\mathcal{T}u^k - v^k - \frac{p^k}{\delta}\right\|^2 + \sigma\|v^k\|_\varphi.
\end{aligned}
$$

From Lemma 5, $\{F(w^k)\}_{k=0,1,2,\dots}$ is bounded. This means, using the definition of $F(w^k)$, the boundedness of sequences $\{\|v^k\|_\varphi\}_{k=0,1,2,\dots}$, $\{\mathcal{T}u^k - v^k - \frac{p^k}{\delta}\}_{k=0,1,2,\dots}$ and $\{Ku^k - f\}_{k=0,1,2,\dots}$. Taking into account the coercivity of φ, $\{v^k\}_{k=0,1,2,\dots}$ is bounded. Now, taking into account (58), $\{p^k\}_{k=0,1,2,\dots}$ is bounded. Then, $\left\{\begin{pmatrix} \mathcal{T} \\ K \end{pmatrix}u^k\right\}_{k=0,1,2,\dots}$ is bounded, and using the reversibility of $\begin{pmatrix} \mathcal{T} \\ K \end{pmatrix}$, we obtain that $\{u^k\}_{k=0,1,2,\dots}$ is bounded. Therefore, all the components of the sequence $\{w^k\}_{k=0,1,2,\dots}$ are bounded, and so the result is proved.

APPENDIX E: PROOF OF LEMMA 7

For the v^{k+1} component, we have

$$-\widehat{p}^k + \delta(\mathcal{T}u^k - v^{k+1}) \in \partial\sigma\|v^{k+1}\|_\varphi. \tag{59}$$

Easy computations on the new defined term s_v^{k+1} give

$$s_v^{k+1} := (1+\delta)(p^{k+1} - \widehat{p}^k) \in \partial_v F(w^{k+1}). \tag{60}$$

Direct calculations hold

$$
\begin{aligned}
\|s_v^{k+1}\| &\le (1+\delta)\|p^{k+1} - p^k\| + (1+\delta)\|p^k - \widehat{p}^k\| \\
&= (1+\delta)\|p^{k+1} - p^k\| + \alpha(1+\delta)\|p^k - p^{k-1}\| \\
&\le (1+\delta)\theta\|K\|_2^2\|u^{k+1} - u^k\| \\
&\quad + \alpha(1+\delta)\theta\|K\|_2^2\|u^k - u^{k-1}\|.
\end{aligned}
$$

Thus, we obtain that

$$\|s_v^{k+1}\| \le \gamma_v(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|), \tag{61}$$

where the parameter γ_v is given by

$$\gamma_v = \max\left\{(1+\delta)\theta\|K\|_2^2,\ \alpha(1+\delta)\theta\|K\|_2^2\right\}.$$

For the u^{k+1} component, we have

$$\mathcal{T}^*\widehat{p}^k - \delta\mathcal{T}^*(\mathcal{T}u^{k+1} - v^{k+1}) = K^*(Ku^{k+1} - f). \tag{62}$$

With (62), we have the new term s_u^{k+1}

$$
\begin{aligned}
s_u^{k+1} :=\ &\mathcal{T}^*\widehat{p}^k - \mathcal{T}^*p^{k+1} + \frac{7\alpha^2\theta^2\|K\|_2^4}{2\delta}(u^{k+1} - u^k) \\
&\in \partial_u F(w^{k+1}).
\end{aligned}
\tag{63}
$$

It is easy to obtain that

$$\|s_u^{k+1}\| \le \gamma_u(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|), \tag{64}$$

where

$$\gamma_u = \max\left\{\theta\|\mathcal{T}\|_2\|K\|_2^2 + \frac{7\alpha^2\theta^2\|K\|_2^4}{2\delta},\ \alpha\theta\|\mathcal{T}\|_2\|K\|_2^2\right\}.$$

Obviously, it holds

$$s_p^{k+1} := -\mathcal{T}u^{k+1} + v^{k+1} \in \partial_p F(w^{k+1}). \tag{65}$$

From the scheme, we can easily see

$$
\begin{aligned}
\|s_p^{k+1}\| &= \frac{1}{\delta}\|p^{k+1} - \widehat{p}^k\| \le \frac{\|p^{k+1} - p^k\|}{\delta} + \frac{\alpha\|p^k - p^{k-1}\|}{\delta} \\
&\le \gamma_p(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|),
\end{aligned}
$$

where $\gamma_p = \max\left\{\frac{\theta\|K\|_2^2}{\delta},\ \frac{\alpha\theta\|K\|_2^2}{\delta}\right\}$. Therefore, we have

$$\|\nabla_x F(w^{k+1})\| \le \frac{7\alpha^2\theta^2\|K\|_2^4}{\delta}(\|d^{k+1} - d^k\| + \|u^k - u^{k-1}\|). \tag{66}$$

Letting $s^{k+1} := (s_v^{k+1}, s_u^{k+1}, s_p^{k+1}, \nabla_x F(w^{k+1}))$, we have $s^{k+1} \in \partial F(w^{k+1})$. With all the above equations (61), (64), (65), and (66), we obtain the global bound

$$
\begin{aligned}
\|s^{k+1}\| \le\ &\left(\gamma_u + \gamma_v + \gamma_p + \frac{7\alpha^2\theta^2\|K\|_2^4}{\delta}\right) \\
&\times (\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|).
\end{aligned}
\tag{67}
$$

Denoting $\gamma = \gamma_u + \gamma_v + \gamma_p + \frac{7\alpha^2\theta^2\|K\|_2^4}{\delta}$, and then, we finish the proof.

APPENDIX F: PROOF OF THEOREM 1

For any cluster point (u^*, v^*, p^*), there exists $\{k_j\}_{j=0,1,2,\dots}$ such that $\lim_j(u^{k_j}, v^{k_j}, p^{k_j}) = (u^*, v^*, p^*)$. Then, from Lemma 5, we have that

$$\lim_j(u^{k_j+1}, v^{k_j+1}, p^{k_j+1}) = \lim_j(\widehat{u}^{k_j}, \widehat{v}^{k_j}, \widehat{p}^{k_j}) = (u^*, v^*, p^*).$$

From the scheme of Algorithm 1, we have the following conditions

$$
\begin{aligned}
-\widehat{p}^{k_j} + \delta(\mathcal{T}u^{k_j} - v^{k_j+1}) &\in \partial\sigma\|v^{k_j+1}\|_\varphi, \\
\mathcal{T}^*\widehat{p}^{k_j} - \delta\mathcal{T}^*(\mathcal{T}u^{k_j+1} - v^{k_j+1}) &= K^*(Ku^{k_j+1} - f), \\
p^{k_j+1} &= \widehat{p}^{k_j} - \delta(\mathcal{T}u^{k_j+1} - \widehat{v}^{k_j+1}).
\end{aligned}
$$

Letting $j \to +\infty$, with Proposition 1, we have

$$
\begin{aligned}
-p^* &\in \partial\sigma\|v^*\|_\varphi, \\
\mathcal{T}^*p^* &= K^*(Ku^* - f), \\
\mathcal{T}u^* &= v^*.
\end{aligned}
$$

Finally, from Proposition 2, we obtain that (u^*, v^*, p^*) is a critical point of $\mathcal{L}_\delta^\varphi$.

APPENDIX G: PROOF OF LEMMA 8

(1) From Lemma 6 the sequences $\{(u^k, v^k, p^k)\}_{k=0,1,2,\dots}$ and $\{w^k\}_{k=0,1,2,\dots}$ are bounded. Thus, $\mathcal{M}$ is nonempty. Assume that $w^* \in \mathcal{M}$, then from the definition there exists a subsequence $w^{k_i} = (u^{k_i}, v^{k_i}, p^{k_i}, u^{k_i-1}) \to w^* = (u^*, v^*, p^*, u^\bullet)$. In that case, from Lemma 5, we also have $u^{k_i-1} \to u^*$. That is also $u^\bullet = u^*$, i.e., $w^* = (u^*, v^*, p^*, u^*)$. Besides, from Lemma 7 and Lemma 5, we have $s^{k_i} \in \partial F(w^{k_i})$ and $s^{k_i} \to \mathbf{0}$. Finally, Proposition 1 indicates that $\mathbf{0} \in \partial F(w^*)$, i.e. $w^* \in \mathrm{crit}(F)$.

(2) This item follows as a consequence of the definition of the limit point set (Definition 4).

(3) In Eq. (53), by taking limits on i, we obtain $\limsup_i \|v^{k_i}\|_\varphi \le \|v^*\|_\varphi$. And with the closedness of φ, we have the other bound $\liminf_i \|v^{k_i}\|_\varphi \ge \|v^*\|_\varphi$. That means $\lim_i \|v^{k_i}\|_\varphi = \|v^*\|_\varphi$. Using the continuity of the rest of functions that defines F, we have $\lim_i F(w^{k_i}) = F(w^*)$. On the other hand, Lemma 5 indicates $\{F(w^k)\}_{k=0,1,2,\dots}$ is decreasing. Noting the boundedness of $\{F(w^k)\}_{k=0,1,2,\dots}$, this sequence has a lower bound. Thus, $\{F(w^k)\}_{k=0,1,2,\dots}$ is convergent. And then, we have $\lim_k F(w^k) = \lim_i F(w^{k_i}) = F(w^*)$.

APPENDIX H: PROOF OF THEOREM 2

From Lemma 8, F is constant on $\mathcal{M}$. Let $w^* = (u^*, v^*, p^*, x^*)$ be a stationary point of $\{w^k\}_{k=0,1,2,\dots}$. Then, from the definition of w^k we have $u^* = x^*$. Also from Lemma 8, we have $\mathrm{dist}(w^k, \mathcal{M}) < \varepsilon$ and $F(w^k) < F(w^*) + \eta$ for any $k > K$, for some K. Hence, from Lemma 1, we have

$$
\mathrm{dist}(\mathbf{0}, \partial F(w^k))\rho'(F(w^k) - F(w^*)) \ge 1, \qquad (68)
$$

that together with Lemma 7 give that

$$
\begin{aligned}
\frac{1}{\rho'(F(w^k) - F(w^*))} &\le \mathrm{dist}(\mathbf{0}, \partial F(w^k)) \\
&\le \gamma(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|). \qquad (69)
\end{aligned}
$$

Then, the concavity of ρ yields that

$$
\begin{aligned}
&F(w^k) - F(w^{k+1}) \\
&= F(w^k) - F(w^*) - [F(w^{k+1}) - F(w^*)] \\
&\le \frac{\rho[F(w^k) - F(w^*)] - \rho[F(w^{k+1}) - F(w^*)]}{\rho'[F(w^k) - F(w^*)]} \\
&\le \gamma(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|) \\
&\quad \times \{\rho[F(w^k) - F(w^*)] - \rho[F(w^{k+1}) - F(w^*)]\}. \qquad (70)
\end{aligned}
$$

Using Lemma 7, we have

$$
\begin{aligned}
\widehat{h}\|u^{k+1} - u^k\|^2 &\le \gamma(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|) \\
&\times \{\rho[F(w^k) - F(w^*)] - \rho[F(w^{k+1}) - F(w^*)]\}, \qquad (71)
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
\frac{\widehat{h}}{\gamma}\|u^{k+1} - u^k\| &\le 2 \cdot \frac{1}{2}\sqrt{\frac{\widehat{h}}{\gamma}}\sqrt{\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|} \\
&\times \sqrt{\rho[F(w^k) - F(w^*)] - \rho[F(w^{k+1}) - F(w^*)]}. \qquad (72)
\end{aligned}
$$

Using the Schwartz's inequality, we then derive that

$$
\begin{aligned}
\frac{\widehat{h}}{\gamma}\|u^{k+1} - u^k\| &\le \frac{\widehat{h}}{4\gamma}(\|u^{k+1} - u^k\| + \|u^k - u^{k-1}\|) \\
&+ \{\rho[F(w^k) - F(w^*)] - \rho[F(w^{k+1}) - F(w^*)]\}. \qquad (73)
\end{aligned}
$$

Summing (73) from K to $K + j$ yields that

$$
\begin{aligned}
\frac{\widehat{h}}{2\gamma} &\sum_{k=K}^{K+j-2} \|u^{k+1} - u^k\| + \frac{3\widehat{h}}{4\gamma}\|u^{K+j+1} - u^{K+j}\| \\
&\le \rho[F(w^K) - F(w^*)] - \varphi[F(w^{K+j+1}) - F(w^*)] \\
&+ \frac{\widehat{h}}{4\gamma}\|u^{K-1} - u^{K-2}\|.
\end{aligned}
$$

Letting $j \to +\infty$, and applying Lemma 5, we have

$$
\frac{\widehat{h}}{2\gamma}\sum_{k=K}^{+\infty}\|u^{k+1} - u^k\| < +\infty. \qquad (74)
$$

Then, $\{u^k\}_{k=0,1,2,\dots}$ is convergent. And noting that u^* is a stationary point of $\{u^k\}_{k=0,1,2,\dots}$ we also have that $\{u^k\}_{k=0,1,2,\dots}$ converges to u^*. With (29), the convergence of $\{u^k\}_{k=0,1,2,\dots}$ indicates the convergence of $\{p^k\}_{k=0,1,2,\dots}$. And then the sequence $\{\widehat{p}^k\}_{k=0,1,2,\dots}$ is also convergent. With the identity $p^{k+1} = \widehat{p}^k - \delta(\mathcal{T}u^{k+1} - v^{k+1})$, $\{v^k\}_{k=0,1,2,\dots}$ is convergent. That means (u^k, v^k, p^k) converges to some point (u^*, v^*, p^*). Finally, from Theorem 1, we have that (u^*, v^*, p^*) is a critical point of $\mathcal{L}_\delta^\varphi$.

REFERENCES

[1] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2006.

[2] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, p. 745, 1974.

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

[4] C. Chen, R. H. Chan, S. Ma, and J. Yang, "Inertial proximal ADMM for linearly constrained separable convex optimization," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2239–2267, 2015.

[5] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[6] F. Chen, H. Zhou, C. Grecos, and P. Ren, "Segmenting oil spills from blurry images based on alternating direction method of multipliers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 6, pp. 1858–1873, 2018.

[7] M. Hintermüller and T. Wu, "Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1385–1415, 2013.

[8] M. Hintermüller and T. Wu, "A smoothing descent method for nonconvex TV^q-models," in *Efficient Algorithms for Global Optimization Methods in Computer Vision*, pp. 119–133, Springer, 2014.

[9] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l_0 gradient minimization," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, p. 174, 2011.

[10] M. Nikolova, M. K. Ng, S. Zhang, and W.-K. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 2–25, 2008.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[12] J. Eckstein and M. Fukushima, "Some reformulations and applications of the alternating direction method of multipliers," in *Large scale optimization*, pp. 115–134, Springer, 1994.

[13] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[14] D. Boley, "Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2183–2207, 2013.

[15] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *Journal of Scientific Computing*, pp. 1–28, 2012.

[16] B. He and X. Yuan, "On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers," *Numerische Mathematik*, vol. 130, no. 3, pp. 567–577, 2012.

[17] B. He and X. Yuan, "On the o(1/n) convergence rate of the Douglas-Rachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.

[18] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, pp. 165–199, Mar 2017.

[19] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[20] S. Banert and R. I. Boţ, "Backward penalty schemes for monotone inclusion problems," *Journal of Optimization Theory and Applications*, vol. 166, no. 3, pp. 930–948, 2015.

[21] R. I. Bot and E. R. Csetnek, "Penalty schemes with inertial effects for monotone inclusion problems," *Optimization*, vol. 66, no. 6, pp. 965–982, 2017.

[22] C. Chen, S. Ma, and J. Yang, "A general inertial proximal point algorithm for mixed variational inequality problem," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2120–2142, 2015.

[23] P. Ochs, Y. Chen, T. Brox, and T. Pock, "iPiano: Inertial proximal algorithm for nonconvex optimization," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1388–1419, 2014.

[24] T. Sun, P. Yin, D. Li, C. Huang, L. Guan, and H. Jiang, "Non-ergodic convergence analysis of heavy-ball algorithms," in *AAAI Conference on Artificial Intelligence*, 2019.

[25] R. I. Bot and E. R. Csetnek, "An inertial alternating direction method of multipliers," *arXiv:1404.4582*, 2014.

[26] G. Li and T. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.

[27] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[28] T. Sun, H. Jiang, and L. Cheng, "Iteratively linearized reweighted alternating direction method of multipliers for a class of nonconvex problems," *arXiv:1709.00483*, 2017.

[29] T. Sun, P. Yin, L. Cheng, and H. Jiang, "Alternating direction method of multipliers with difference of convex functions," *Advances in Computational Mathematics*, pp. 1–22, 2017.

[30] F. Wang, Z. Xu, and H.-K. Xu, "Convergence of alternating direction method with multipliers for non-convex composite problems," *arXiv:1410.8625*, 2014.

[31] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.

[32] T. Sun, H. Jiang, L. Cheng, and W. Zhu, "A convergence frame for inexact nonconvex and nonsmooth algorithms and its applications to several iterations," *arXiv preprint arXiv:1709.04072*, 2017.

[33] Z. Xu, S. De, M. Figueiredo, C. Studer, and T. Goldstein, "An empirical study of admm for nonconvex problems," *arXiv preprint arXiv:1612.03349*, 2016.

[34] G. Taylor, R. Burmeister, Z. Xu, B. Singh, A. Patel, and T. Goldstein, "Training neural networks without gradients: A scalable ADMM approach," in *International Conference on Machine Learning*, pp. 2722–2731, 2016.

[35] B. S. Mordukhovich, *Variational analysis and generalized differentiation I: Basic theory*, vol. 330. Springer Science & Business Media, 2006.

[36] Y. Nesterov, *Introductory lectures on convex optimization*, vol. 87. Springer Science & Business Media, 2004.

[37] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317. Springer Science & Business Media, 2009.

[38] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[39] S. Łojasiewicz, *Les Équations aux Dérivées Partielles*, ch. Une propriété topologique des sous-ensembles analytiques réels, pp. 8–89. Paris: Éditions du Centre National de la Recherche Scientifique, 1963.

[40] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de l'Institut Fourier*, vol. 48, pp. 769–783, 1998.

[41] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Mathematical Programming*, vol. 137, no. 1-2, pp. 91–129, 2013.

[42] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[43] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.

[44] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, "Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results," *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 746–768, 1998.

[45] T. I. Lakoba, "The eigenvalues of tridiagonal matrices," *Technical Note of the Norwegian University of Science and Technology, Department of Mathematical Sciences, NoteTMA4205*, 2009.