This is a peer-reviewed, accepted author manuscript of the following research article: Pang, Z-F, Meng, G, Li, H & Chen, K 2020, 'Image restoration via the adaptive TVp regularization', Computers and Mathematics with Applications, vol. 80, no. 5, pp. 569-587. https://doi.org/10.1016/j.camwa.2020.04.030

Half-quadratic adaptive TV^p to the image restoration problem

Zhi-Feng Pang, Ge Meng, Hui Li and Ke Chen

***Abstract*—To keep structures in the restoration problem is very important via coupling the local information of the image with the proposed model. In this paper we propose a local self-adaptive ℓ^p-regularization model for $p \in (0, 2)$ based on the total variation scheme, where the choice of p depends on the local structures described by the eigenvalues of the structure tensor. Since the proposed model as the classic ℓ^p problem unifies two classes of optimization problems such as the nonconvex and nonsmooth problem when $p \in (0, 1)$, and the convex and smooth problem when $p \in (1, 2)$, it is generally challenging to find a ready algorithmic framework to solve it. Here we propose a new and robust numerical method via coupling with the half-quadratic scheme and the alternating direction method of multipliers(ADMM). The convergence of the proposed algorithm is established and the numerical experiments illustrate the possible advantages of the proposed model and numerical methods over some existing variational-based models and methods.**

***Index Terms*—Image restoration, Proximal point scheme, Half-quadratic scheme, Alternating direction method of multipliers(ADMM), Structure tensor.**

I. Introduction

IMAGE processing technique plays an important part in various applied areas such as medical and astronomical imaging, image and video coding, computer vision and many others [28], [32]. One of the most challenging tasks in image processing is the restoration problem, where one aims at providing an acceptable estimate of the original image from degraded observations. These degradations such as noise or blurring contamination may arise due to various phenomena which are often unavoidable in practical situations. Many restoration techniques such as the original filter method or recent spectral analysis and variational partial differential equations (VPDE) have all been successfully applied by employing some fast and stable numerical methods [3], [13], [48].

In this work we mainly focus on the VPDE-based restoration models, which usually are formed as an energy minimization problem

$$\min_u \mathcal{E}(\mathcal{A}u, f) := \frac{\lambda}{2}\mathcal{F}(\mathcal{A}u, f) + \mathcal{R}(u) \tag{1}$$

to deal with the ill-posedness due to the lacks of some prior information. Here $f : \Omega \to \mathbb{R}$ is the degraded image, $\mathcal{A}$ denotes some linear operator such as the blurring operator, Ω denotes two dimensional $M \times N$ image domain. The main principle behind the model (1) is that the data fitting term $\mathcal{F}(\mathcal{A}u, f)$ keeps some prior information of the image, the regularized term $\mathcal{R}(u)$ penalizes some structure features such as edges and corners, the regularization parameter $\lambda > 0$ balances the weight between two terms. Among the models as the form (1), a very popular method for a wide variety of image restoration problems is the pioneering work of Rudin et al [42] (called the ROF model)

$$\min_u \frac{\lambda}{2}\|\mathcal{A}u - f\|_2^2 + \|\nabla u\|_1. \tag{2}$$

Note that the original ROF model sets $\mathcal{A}$ as an identity operator. Minimization of the ROF model (2) corresponds to finding the steady solution of

$$\frac{\partial u}{\partial t} = \lambda\mathcal{A}^*(\mathcal{A}u - f) - \operatorname{div}\left(\frac{\nabla u}{\|\nabla u\|_1}\right) \tag{3}$$

with the suitable boundary condition by using the time marching method [42]. If we here define $\mathcal{N}(x) = \frac{\nabla u}{\|\nabla u\|_1}$ and $\mathcal{T}(x)$ orthogonal to $\mathcal{N}(x)$, the term $\operatorname{div}\left(\|\nabla u\|^{-1}\nabla u\right)$ in (3) can be rewritten as

$$\operatorname{div}\left(\frac{\nabla u}{\|\nabla u\|_1}\right) = \frac{1}{\|\nabla u\|_1}\mathcal{T}(x)^T\begin{bmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy}\end{bmatrix}\mathcal{T}(x) = \frac{1}{\|\nabla u\|_1}u_{\mathcal{T}\mathcal{T}}$$

in the sense of local coordinate. Thus the field flow (3) is an anisotropic smoothing diffusion in the tangent direction $\mathcal{T}(x)$ to the isophote line of u at points where the gradient does not vanish, which implies that the TV-based model does not smooth the gradient direction. Then the edges can be preserved while smoothing the homogeneous regions. However, the TV-based model has still some limitations that restrict its performance. Specifically, the penalization to the ℓ^1 norm of the gradient encourages the recovery of images with sparse gradients, thus resulting in reconstructed images with patchy or painting-like staircase artifacts. To overcome this problem, several functionals involving total generalized variation-based models [7], [49], higher order TV-based models [12], [43], nonlocal-based models [8], [25] and fractional order models [4], [58] were introduced, mainly in the context of image restoration. These schemes are reported to give better restored performance than the standard TV regularizer in some regions of image. However, these functionals may not be ideally suited for the regularizer of ill-posed inverse problems. Overall, a single regularizer is insufficient for all regions in an image.

Zhi-Feng Pang, Ge Meng and Hui Li are with the Department of Mathematics and statistics, and Laboratory of Data Analysis Technology, Henan University, Kaifeng, 475004, China; E-mails: zhifengpang@163.com; mengge1113@163.com; huili_2017@163.com.

Ke Chen is with the Centre for Mathematical Imaging Techniques and Department of Mathematical Sciences, University of Liverpool, United Kingdom E-mail: k.chen@liv.ac.uk

A. Related Works

The $\ell^{p(x)}$ regularization for choosing a suitable $p(x) \in (0,2)$ has advantages over smooth, convex regularization for restoring image with near edges, sparse signal reconstruction and variable selection. Recently, motivated by the effect of the compressive sensing (CS), TV^p-based models [52], [38], [33] for $p \in (0,1)$ have received extensive attention for the image restoration problem. According to the CS theory, most signals are sparse or approximately sparse in some transform domain. So the regularization methods can be applied to effectively solve the sparsity-constrained optimization problems. However, the TV^p regularizer is based on the gradient information, it tends to a constant or approximate constant image when choosing it as the transform basis. Then for the approximately smooth image it is more suitable for choosing $p \in (1,2)$ to TV^p-based models [33], [45], [5], [15]. These observations can be traced back to early works for overcoming the staircase effect in the smoothing regions for the ROF model, where p is usually chosen as an adaptive parameter. Specifically, p tends to 1 near edges and it behaves exactly like the ROF model; p to 2 away from the edges and it may behave more like the Dirichlet energy. This leads to much smoother restorations in regions of moderate gradient and thus prevents staircase effect. These observations imply that more robust restoration models depend on choosing an adaptive function $p(x)$ to describe local structures in the image.

In order to efficiently describe the local feature, the authors in [15] set $p(x)$ to be an edge indicator function which can apply less smoothing near significant edges, while in [16] the authors proposed $p(x)$ to use the difference curvature information to effectively distinguish between edges and ramps. Different from the methods for only focusing on the adaption of $p(x)$, some schemes considered to replace $|\nabla u|_1$ by $|\nabla u|_{1,g}$, here $g(x)$ denotes the weighted function such as the spatially adapted and edge indicator schemes [20], [36]. In [21], the authors used another weighted function based on the structure tensor. Comparing to the weighted function via directly using the gradient information, its importance stems from its eigenvalue decomposition, which summarizes the orientation of the image gradient in a neighborhood of a point. Formally, its principal eigenvector indicates the direction of the largest contrast. So the geometric structures can be efficiently described by using its eigenvalue information [54]. The aforementioned facts imply that it is more suitable to combine $p(x)$ as exponent weigh in the proposed TV-based model, so we consider an adaptive weighted model as

$$\min_u \frac{\lambda}{2}\|\mathcal{A}u - f\|_2^2 + \|\nabla u\|^{p(x)}, \tag{4}$$

where $\|\nabla u\|^{p(x)} := \sum_{i=1}^{M}\sum_{j=1}^{N} \|\nabla u_{i,j}\|^{p_{i,j}(x)}$.

The proposed model (4) is a $\ell^2 - \ell^{p(x)}$ problem, which is convex $p(x) \in [1,2)$ and nonconvex $p(x) \in (0,1)$. To the specific cases by setting $p(x) = 1$, the numerical solution can be efficiently obtained by employing fast numerical methods since the objective functional owns better structures and also convex. For example, we can use the alternating direction method of multipliers (ADMM) methods [56], [27], primal dual methods [31], [9] or forward-backward methods [17], [50]. However, to the general $p(x)$, it is not trivial to find the solution since we can not obtain a closed expression to the original problem or some related subproblems. Some efforts have been committed to overcome these drawbacks. For $p(x) \in (1,2)$, it is a smoothing convex problem and then time marching method has been proposed [45], [15]. This method requires the time step to be small enough and the solution also tends to the unique minimizer as time increases. Actually, it is slow due to the Courant-Friedrichs-Lewy (CFL) stability constraints [18], which put a very tight bound on the time step when the solution develops flat regions. Hence we only obtain a low-accuracy solutions. For $p(x) \in (0,1)$, it is a nonconvex, nonsmooth and non-Lipschitz optimization problem, thus we only expect to find a local minimizer. There are two popular schemes to deal with it, i.e. the iteratively reweighted norm method and the thresholding method. The iteratively reweighted norm method is originally proposed in [47] for minimizing the so called generalized total variation functional. The key of this method is in transforming the $\ell^{p(x)}$ regularizer into the ℓ^1-norm form by employing a suitable weighted function and then some efficient methods to deal with the ℓ^1-norm can be used. This method was also extended to deal with the $\ell^{p(x)}$ problem for $p(x) \in (0,2)$ in [33]. The thresholding method is a simple iterative process followed by a thresholding operator, which mainly follows from the motivation of the well-known soft-thresholding for the ℓ^1 regularization and the hard-thresholding for the ℓ_0. Then the efficient restoring image with sparse structure characteristics can be obtained by using the derived analytic thresholding representation [19], [55], [57].

B. Contributions

Here we propose a new and novel numerical method to solve the adaptive model (4) via coupling with the half-quadratic scheme and the variable splitting scheme [11], [34], [41], [24]. The main purpose of the half-quadratic scheme is introduction of a sequence of quadratic convex minimization problems to approximate the $\ell^{p(x)}$-(quasi)norm. So we can obtain an approximation solution by using the alternating minimization scheme. Based on this transformation, we then use the ADMM scheme and also efficiently related subproblems. Our main contribution can be summarized as

- The proposed model (4) is an adaptive variational-based model by employing the total variation penalty exponent $p(x) \in (0,2)$. The choice of the function $p(x)$ depends on the eigenvalues of the local structure tensor, which can efficiently distinguish different image structures such as edges and corners. So we can expect to obtain a better restoration image comparing to choosing $p(x)$ as a constant.
- We present a novel and unified numerical framework via coupling with the half-quadratic scheme and the ADMM to solve the $\ell^{p(x)}$–(quasi)norm problem (4) for $p(x) \in (0,2)$. In order to efficiently solve this problem, we propose a fixed point iteration method based on the half-quadratic scheme and also use the continuation method

[10] to the smoothing $\ell^{p(x)}$ problem. Simultaneously we prove that the proposed method converges to the global minimizer for $p(x) \in [1, 2)$ and to the local minimizer for $p(x) \in (0, 1)$.
- Numerical implementations show that the proposed schemes give superior results to the state-of-art variation-based models and numerical methods.

The rest of this paper is organized as follows. Section II describes the general framework of the proposed numerical method including the half-quadratic scheme to approximate the ℓ^p-(quasi)norm, the ADMM to solve the proposed model and some convergence analysis for our proposed numerical method. We then discuss how to choose a suitable exponent $p(x)$ based on the structure tensor strategy. Section III mainly considers the numerical details via comparing different variation-based models and numerical methods to show the effectiveness of our proposed models and numerical algorithm. We conclude our paper in Section IV.

C. Notations

Let us describe the notations used throughout this paper. For simplification, we use $\mathrm{X} = \mathbb{R}^n \times \mathbb{R}^n$ and $\mathrm{Y} = \mathrm{X} \times \mathrm{X}$. $(\cdot, \cdot)$ denotes the usual scalar product on X with $(\bar{u}_{i,j}, \bar{v}_{i,j}) = \sum_{i=1}^{M}\sum_{j=1}^{N} \bar{u}_{i,j}\bar{v}_{i,j}$ and the ℓ^2-norm in X is defined as $\|\bar{u}\|_2^2 := \sum_{i=1}^{M}\sum_{j=1}^{N} \bar{u}_{i,j}^2$. $\langle \cdot, \cdot \rangle$ denotes the usual scalar product on Y with $\langle \bar{g}_{i,j,s}, \bar{h}_{i,j,s} \rangle = \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{s=1}^{2} \bar{g}_{i,j,s}\bar{h}_{i,j,s}$. For $\hat{\mathbf{g}} \in \mathrm{Y}$, the ℓ^p-(quasi)norm in Y is defined as $\|\hat{\mathbf{g}}\|^p := \sum_{i=1}^{M}\sum_{j=1}^{N} \left(\sqrt{\hat{g}_{i,j,1}^2 + \hat{g}_{i,j,2}^2}\right)^p$ for $p \in (0, \infty)$. For a matrix operator A, we also define $\|\mathbf{z}\|_A^2 = \mathbf{z}^T A\mathbf{z}$ for $\mathbf{z} \in \mathbb{R}^n$ and its norm as $\|A\|_2 := \max_{\mathbf{z}} \{\|A\mathbf{z}\|_2 : \ \mathbf{z} \in \mathbb{R}^n \text{ with } \|\mathbf{z}\| \le 1\}$.

For a proper function $\phi : \mathbb{R}^n \to \bar{\mathbb{R}} := (-\infty, \infty]$, its subdifferential at $x \in \mathrm{dom}\phi(x)$ is defined by

$$\partial\phi(x) := \Bigg\{ y \in \mathbb{R}^n : \ \exists f(x^k) \to f(x) \text{ and } z^k \to z \text{ as } k \to \infty \text{ with } \liminf_{k\to\infty} \frac{f(y) - f(x^k) - (v^k, y - x^k)}{\|y - x^k\|} \ge 0 \Bigg\}$$

Especially, for the convex function ϕ it reduces to the classical subdifferential in the convex analysis [46] as

$$\partial\phi(x) := \{y \in \mathbb{R}^n : (y, z - x) \le \phi(z) - \phi(x)\}.$$

Moreover, the above subdifferential reduces to the derivative denoted by $\nabla\phi$ when ϕ is a continuously differentiable function.

II. General framework to the proposed model

In this section, we start with a brief introduction the proximal point scheme and the half-quadratic scheme. Then we discuss our proposed numerical method to solve our proposed TV^p-based model (4) via coupling with the half-quadratic scheme and the ADMM. We also give some convergence analysis and discuss how to choose suitable $p(x)$ based on the structure tensor in our proposed model. Note in this section we omit the variable $x \in p(x)$ for the simplification unless explicitly explained.

A. Proximal Point Scheme

The proximal point scheme as the iteration method first arisen in the works of Moreau [39] to cope with the convex optimization problems. Recently, this scheme has been used in the image processing problem [56], [31], [9], [27], where one of subproblems including ℓ^1-norm proximal point problem can be solved relatively by using the soft-threshold operator without any iteration. However, to the general ℓ^p-(quasi)norm proximal point problem, this convexity is not always satisfied when $p \in (0, 1)$. Especially, it is also the nonsmoothing and non-Lipschitz problem in this case. So we need to extend the definition of the general ℓ^p-(quasi)norm proximal point problem as follows.

Definition 1. The proximal point of ℓ^p-(quasi)norm at $\mathbf{y} \in \mathrm{Y}$ is defined by

$$\mathrm{Prox}_{\ell^p}^{\tau}(\mathbf{y}) := \underset{\mathbf{x}\in\mathrm{Y}}{\mathrm{argmin}} \left\{ \|\mathbf{x}\|^p + \frac{\tau}{2}\|\mathbf{x} - \mathbf{y}\|^2 \right\}, \tag{5}$$

where $\tau > 0$ is a prox-parameter and $p \in (0, 2)$.

In this definition, τ and $\mathbf{y}$ are called the prox-parameter and prox-center. The proximal envelope as a function actually assigns each $\mathbf{y}$ to be the optimal objective value. In general, the proximal point $\mathbf{x}$ can be efficiently solved by using the iteratively re-weighted least squares method (IRLSM) [47], where it is described as the infimum over a family of quadratic functions, eventually tracked by the tools of numerical linear algebra. However, the IRLSM does not outperform in general well-established first methods such as the fast iteration soft-thresholding algorithm (FISTA) [13]. Furthermore, the weighted parameter as a smoothing parameter in the IRLSM is fixed, which implies that we only obtain the approximate solution to the problem (5). So we propose a half-quadratic scheme [11], [41], [24] with the continuation method [10] to overcome above approximation, where authors proposed to couple the Newton method with the smoothing reduction parameter along with the iteration number.

Lemma 2. *The problem $|t|$ for $t \in \mathbb{R} \setminus \{0\}$ can be written as*

$$|t| = \min_{v>0} \left\{ \bar{w}t^2 + \frac{1}{v} \right\}$$

and the minimization value is reached at $\bar{w} = 1/|t|$.

This Lemma 2 can be easily proved by using the triangular inequality due to the convexity when $\bar{w} > 0$. In the following we extend it to the vector space Y for ℓ^p-(quasi)norm, which has been used in the paper [11].

Theorem 3. *For any $p \in (0, 2)$ and $\boldsymbol{t} \in \mathrm{Y} \setminus \{(0, 0)^T\}$, then*

$$\|\boldsymbol{t}\|^p = \min_{w>0} \left\{ w\|\boldsymbol{t}\|^2 + \frac{1}{\delta w^{\gamma}} \right\},$$

where

$$\gamma = \frac{p}{2-p} \text{ and } \delta = \frac{2^{\frac{2}{2-p}}}{(2-p)p^{\frac{p}{2-p}}},$$

and the minimizer is reached at

$$w^* = \frac{p}{2}\|\boldsymbol{t}\|^{p-2}.$$

In the problem (5), the variable $\mathbf{x}$ is possible equal to zero vector. In order to use Theorem 3, we modify it by adding a smoothing parameter $\varepsilon > 0$ to avoid this special case. Specifically, we transform this problem into the following optimization problem

$$\min_{\mathbf{x}} \underbrace{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\sqrt{\|\mathbf{x}_{i,j}\|^2+\varepsilon}\right)^p}_{:=\|\mathbf{x}\|_\varepsilon^p} + \frac{\tau}{2}\|\mathbf{x}-\mathbf{y}\|^2 := \mathcal{F}(\mathbf{x}). \quad (6)$$

Furthermore we have

$$\min_{\mathbf{x},w>0} w\|\mathbf{x}\|_\varepsilon^2 + \frac{1}{\delta w^\gamma} + \frac{\tau}{2}\|\mathbf{x}-\mathbf{y}\|^2 := \mathcal{G}(w,\varepsilon,\mathbf{x}), \quad (7)$$

where γ and δ are defined in Theorem 3. This problem is a multi-variable optimization problem and its variables is separable, so we employ the alternating direction method (ADM) to solve it summarized in Algorithm 1.

Algorithm 1: ADM to solve the problem (7)

1.Initialize: $\tau > 0$ and choose original values of $\mathbf{x}^0$ and ε^0;
2.For k=1,2,$\cdots$, obtain $\left(w^{k+1},\varepsilon^{k+1},\mathbf{x}^{k+1}\right)$ by

$$\begin{cases} w^{k+1} := \underset{w>0}{\operatorname{argmin}}\, \mathcal{G}(w,\varepsilon^k,\mathbf{x}^k), & (8a)\\ \varepsilon^{k+1} := \rho\varepsilon^k, & (8b)\\ \mathbf{x}^{k+1} := \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{G}(w^{k+1},\varepsilon^{k+1},\mathbf{x}), & (8c)\end{cases}$$

where $\rho \in (0,1)$ as a contractive constant;
3.End for until stopping rule meets;
4.Set $\mathbf{x} := \mathbf{x}^{k+1}$.

Remark 1. In Algorithm 1, if $\rho = 1$, it essentially uses the fixed method to solve the problem (7). In this case, the objective function can be regarded as the smoothing of the objective function (5). Furthermore, it is also Lipschitz continuous. So general iterative methods can be used to solve this subproblem. However we may expect a more accurate approximate solution to (5) with the dynamic decrease for ε in this algorithm. In fact, the solution of the subproblems (8a) and (8c) can be explicitly obtained by

$$\begin{cases} w^{k+1} = \frac{p}{2}\|\mathbf{x}^k\|_{\varepsilon^k}^{p-2}, \\ \mathbf{x}^{k+1} = \dfrac{\tau\mathbf{y}}{2w^{k+1}+\tau}. \end{cases}$$

In addition, using (8a) and (8c), we also have

$$\begin{aligned}&\mathcal{G}(w^{k+1},\varepsilon,\mathbf{x}^{k+1}) - \mathcal{G}(w^k,\varepsilon,\mathbf{x}^k) = [\mathcal{G}(w^{k+1},\varepsilon,\mathbf{x}^{k+1})\\ &-\mathcal{G}(w^{k+1},\varepsilon,\mathbf{x}^k)] + \left[\mathcal{G}(w^{k+1},\varepsilon,\mathbf{x}^k) - \mathcal{G}(w^k,\varepsilon,\mathbf{x}^k)\right] < 0.\end{aligned}$$

So we can deduce that

$$\mathcal{G}(w^{k+1},\varepsilon^{k+1},\mathbf{x}^{k+1}) < \mathcal{G}(w^{k+1},\varepsilon^k,\mathbf{x}^{k+1}) < \mathcal{G}(w^k,\varepsilon^k,\mathbf{x}^k).$$

It implies the convergence of the sequence $\left\{\mathcal{G}(w^{k+1},\varepsilon^{k+1},\mathbf{x}^{k+1})\right\}$ due to the bound below by zero of $\{\mathcal{G}(w,\varepsilon,\mathbf{x})\}$.

Theorem 4. *Assume that there exists a positive constants such that* $\left\|\boldsymbol{x}^k\right\|_{\varepsilon^k}^2 < c$ *for* $k = 1,2,\cdots$, *in Algorithm 1. Then the sequence* $\left\{\boldsymbol{x}^k\right\}$ *converges to one of the minimizer of the problem (5). Furthermore, the minimizer is unique when* $p \in [1,2)$.

Proof. In order to prove the convergence, we first set the objective function of (8a) as

$$g(w) := w\left\|\mathbf{x}^k\right\|_{\varepsilon^k}^2 + \frac{1}{\delta w^\gamma}.$$

Then we have $w^{k+1} = \frac{p}{2}\left\|\mathbf{x}^k\right\|_{\varepsilon^k}^{p-2}$ and

$$g'(w^{k+1}) = 0 \text{ and } g''(w) = \frac{\gamma(\gamma+1)}{\delta w^{\gamma+2}}$$

Using the Taylor expansion to $\mathcal{G}\left(w,\varepsilon^k,\mathbf{x}^k\right)$ at w^{k+1}, we can deduce that

$$\begin{aligned}&\mathcal{G}\left(w^k,\varepsilon^k,\mathbf{x}^k\right) - \mathcal{G}\left(w^{k+1},\varepsilon^k,\mathbf{x}^k\right)\\ &= g'\left(w^{k+1}\right)\left(w^k - w^{k+1}\right) + \frac{1}{2}g''(\breve{w})\left\|w^k - w^{k+1}\right\|_2^2\end{aligned}$$

for some $\breve{w}$ between w^k and w^{k+1}. Based on Remark 1, we then have

$$\lim_{k\to\infty} g''(\breve{w})\left\|w^k - w^{k+1}\right\|_2^2 = 0 \quad (9)$$

If $\lim_{k\to\infty}\left\|w^k - w^{k+1}\right\|_2^2 \neq 0$, it implies that $\lim_{k\to\infty} g''(\breve{w}) \to 0$ in (9). In the following, we prove that it does not hold for this choice. Setting $\breve{w} = w^{k+1} + \bar{\tau}\left(w^k - w^{k+1}\right)$ if $w^{k+1} > w^k$ with $\bar{\tau} \in (0,1)$, we deduce that

$$\breve{w} = \frac{p(1-\bar{\tau})}{2}\|\mathbf{x}^k\|_{\varepsilon^k}^{p-2} + \frac{p\tau}{2}\|\mathbf{x}^{k-1}\|_{\varepsilon^{k-1}}^{p-2} \to \infty \quad (10)$$

as $k \to \infty$. However, from Algorithm 1, we obtain that

$$\left\|\mathbf{x}^{k+1}\right\|^2 = \left\|(2w^{k+1}+\tau)^{-1}\tau\mathbf{y}\right\|^2 \leq \|\mathbf{y}\|^2 \quad (11)$$

for all of k. Then there is a contradiction due to the unboundedness of (10) and the boundedness (11) when $\varepsilon^k \to 0$ as $k \to \infty$. To the case of $\breve{w} = w^k + \bar{\tau}\left(w^{k+1} - w^k\right)$ if $w^k > w^{k+1}$, we can get the same contradiction. So we have

$$\lim_{k\to\infty}\left\|w^k - w^{k+1}\right\|_2^2 = 0 \quad (12)$$

Now the gradient of (6) at $\mathbf{x}^{k+1}$ based on the problem (8c) can be expressed as

$$\begin{aligned}&\nabla\mathcal{F}\left(\mathbf{x}^{k+1}\left(\varepsilon^{k+1}\right)\right) = 2w^{k+2}\mathbf{x}^{k+1} + \tau\left(\mathbf{x}^{k+1} - \mathbf{y}\right)\\ &= 2\left(w^{k+2} - w^{k+1}\right)\mathbf{x}^{k+1} + \left(2w^{k+1}+\tau\right)\mathbf{x}^{k+1} - \tau\mathbf{y}\\ &= 2\left(w^{k+2} - w^{k+1}\right)\mathbf{x}^{k+1},\end{aligned}$$

where $\mathbf{x}(\varepsilon)$ depends on ε. Using the fact (12), we have

$$\lim_{k\to\infty}\nabla\mathcal{F}\left(\mathbf{x}^k\left(\varepsilon^k\right)\right) = 0, \quad (13)$$

which implies that the sequence $\left\{\mathbf{x}^k\right\}$ converges to one of minimizer of the problem (5) since $\varepsilon^k \to 0$ as $k \to \infty$. Furthermore, the problem (5) is convex while $p \in [1,2)$, so

the sequence $\left\{\mathbf{x}^k\right\}$ converges to the unique minimizer in this case. □

Remark 2. The optimization problem (5) is the nonsmooth problem if choosing $p = 1$. In this case, we need to use the subdifferential operator $\partial(\cdot)$ to replace the gradient operator $\nabla(\cdot)$. Furthermore, if $\mathbf{x}^*$ is the (global/local) minimizer of the problem (5), then the optimization condition can be written as

$$0 = \begin{cases} p\left\|\mathbf{x}^*\right\|^{p-1}\mathbf{x}^* + \tau\left(\mathbf{x}^* - \mathbf{y}\right), & \text{if } p \neq 1, \\ \boldsymbol{\varrho}^* + \tau\left(\mathbf{x}^* - \mathbf{y}\right), & \text{if } p = 1, \end{cases}$$

where $\boldsymbol{\varrho}^* \in \partial\|\mathbf{x}^*\|_1$. So the limit in (13) will tend to a range value $\boldsymbol{\varrho}^*$ while $p = 1$ in Theorem 4.

B. ADMM

The alternating direction method of multipliers (ADMM) was originally introduced in early 1970s [26], [22], and has been studied extensively in the field of machine learning, computer vision, image and signal processing, networking. The basic motivation of ADMM is to first split the original nonsmooth minimization problem into several subproblems by introducing some auxiliary variables, and then solve each subproblem separately by employing some efficient numerical methods. This method is closely related many other methods such as dual decomposition, the method of multipliers, Douglas-Rachford splitting, Spingarn's method of partial inverses, Dykstra's alternating projections, see for instance of the works in [6].

In this subsection, to describe the ADMM for finding the global or local solution of the problem (4), we first reformulate it as

$$\begin{cases} \min\limits_{u,\mathbf{v}} \dfrac{\lambda}{2}\|\mathcal{A}u - f\|_2^2 + \|\mathbf{v}\|^p \\ \text{s.t. } \mathbf{v} = \nabla u \end{cases} \tag{14}$$

to decouple the linear map and ℓ^p-(quasi)norm. The idea of solving the problem (14) is to introduce some Lagrange multipliers for the constraints, and write the global problem as a saddle-point optimization problem

$$\min_{\mathbf{v},u}\max_{\boldsymbol{\alpha}} \mathcal{L}(\mathbf{v}, u, \boldsymbol{\alpha}) = \|\mathbf{v}\|^p + \frac{\lambda}{2}\|\mathcal{A}u - f\|_2^2 + \langle\boldsymbol{\alpha}, \mathbf{v} - \nabla u\rangle + \frac{\beta}{2}\|\mathbf{v} - \nabla u\|^2, \tag{15}$$

where $\mathcal{L}(\mathbf{v}, u, \boldsymbol{\alpha})$ denotes the augmented Lagrangian function and $\beta > 0$ is the penalty parameter. Then the well-known ADMM for solving (14) can be expressed as the following framework.

Algorithm 2: ADMM to solve the problem (14)

1.Intialize: $\lambda > 0$ and choose original value of u^0 and $\boldsymbol{\alpha}^0$;
2.For t=1,2,$\cdots$, obtain $\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^{t+1}\right)$ by

$$\begin{cases} \mathbf{v}^{t+1} := \underset{\mathbf{v}}{\operatorname{argmin}}\, \mathcal{L}\left(\mathbf{v}, u^t, \boldsymbol{\alpha}^t\right), & \text{(16a)} \\ u^{t+1} := \underset{u}{\operatorname{argmin}}\, \mathcal{L}\left(\mathbf{v}^{t+1}, u, \boldsymbol{\alpha}^t\right), & \text{(16b)} \\ \boldsymbol{\alpha}^{t+1} := \boldsymbol{\alpha}^t + \beta\left(\mathbf{v}^{t+1} - \nabla u^{t+1}\right). & \text{(16c)} \end{cases}$$

3.End for until stopping rule meets;
4.Set $u := u^{t+1}$ as the restored image.

Based on the Theorem 4, the optimization condition of the iteration scheme (16a)-(16c) in Algorithm 2 can be written as

$$\begin{cases} 0 \in \mathbf{v}^{t+1} - \operatorname{prox}_{\ell p}^{\beta}\left(\nabla u^{t+1} - \boldsymbol{\alpha}^t/\beta\right), & \text{(17a)} \\ \lambda\mathcal{A}^T\left(f - \mathcal{A}u^{t+1}\right) = \operatorname{div}\left(\boldsymbol{\alpha}^t + \beta\left(\mathbf{v}^{t+1} - \nabla u^{t+1}\right)\right), & \text{(17b)} \\ \boldsymbol{\alpha}^{t+1} := \boldsymbol{\alpha}^t + \beta\left(\mathbf{v}^{t+1} - \nabla u^{t+1}\right). & \text{(17c)} \end{cases}$$

Remark 3. For the subproblem (17a), we use the Algorithm 1 to find the numerical solution $\mathbf{v} := \mathbf{v}^{t+1}$. In the subproblem (17b), we can use the fast fourier transform $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ as

$$u^{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\lambda\mathcal{A}^T f - \operatorname{div}\boldsymbol{\alpha}^t - \beta\operatorname{div}\mathbf{v}^{t+1}\right)}{\left(\lambda\mathcal{A}^T\mathcal{A} - \beta\triangle\right)}\right)$$

to find the solution if the matrix operators $\mathcal{A}$ and ∇ are circulant. For the general stations, we can employ the preconditioned conjugate gradient method to find the numerical solution if the coefficient matrix $(\lambda\mathcal{A}^T\mathcal{A} - \beta\triangle)$ is positive definite. Simultaneously, we can write the subproblem (17b) as

$$\lambda\mathcal{A}^T\left(f - \mathcal{A}u^{t+1}\right) = \operatorname{div}\boldsymbol{\alpha}^{t+1} \tag{18}$$

by using the scheme (17c).

Theorem 5. *If we respectively denote the biggest and smallest singular values to the matrix operator $\mathcal{A}$ and the gradient operator ∇ as $\bar{s}_1$, $\bar{s}_2$ and $\underline{s}_1$, $\underline{s}_2$ and let $\bar{\kappa} := \lambda\bar{s}_1^4\bar{s}_2^2 - (\beta\lambda\underline{s}_1^2 + \beta^2\underline{s}_2^2)/2 < 0$, then the sequence $\{(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t)\}$ generated by the iteration scheme (16a)-(16c) satisfies that*

$$\mathcal{L}\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^{t+1}\right) \leq \mathcal{L}\left(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t\right) + \frac{\bar{\kappa}}{\beta}\left\|u^{t+1} - u^t\right\|_2^2.$$

Proof. Following from the Lagrangian function (15) and using the formulation (17c) and (18), we have

$$\begin{aligned} \mathcal{L}_1 &:= \mathcal{L}\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^{t+1}\right) - \mathcal{L}\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^t\right) \\ &= \langle\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t, \mathbf{v}^{t+1} - \nabla u^{t+1}\rangle \\ &= \frac{1}{\beta}\left\|\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t\right\|^2 \leq \frac{\lambda^2\bar{s}_1^4\bar{s}_2^2}{\beta}\|u^{t+1} - u^t\|_2^2. \end{aligned} \tag{19}$$

Based on the Taylor expansion at the point u^{t+1} and using (17c) and (18) again, we can get

$$\begin{aligned}
\mathcal{L}_2 &:= \mathcal{L}\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^t\right)\left(-\mathcal{L}\ \mathbf{v}^{t+1}, u^t, \boldsymbol{\alpha}^t\right) \\
&= \frac{\lambda}{2}\left\|\mathcal{A}u^{t+1} - f\right\|_2^2 - \frac{\lambda}{2}\left\|\mathcal{A}u^t - f\right\|_2^2 + \frac{\beta}{2}\left\|\mathbf{v}^{t+1} - \nabla u^{t+1}\right\|^2 \\
&\quad - \frac{\beta}{2}\left\|\mathbf{v}^{t+1} - \nabla u^t\right\|^2 + \left\langle \boldsymbol{\alpha}^t, \nabla u^t - \nabla u^{t+1}\right\rangle \\
&= \frac{\lambda}{2}\left(\mathcal{A}(u^{t+1} - u^t) - 2f, \mathcal{A}\left(u^{t+1} - u^t\right)\right) - \left(\mathrm{div}\boldsymbol{\alpha}^t, u^t - u^{t+1}\right) \\
&\quad + \frac{\beta}{2}\left(2v^{t+1} - \nabla u^{t+1} - \nabla u^t, \nabla u^t - \nabla u^{t+1}\right) \\
&= -\frac{\lambda}{2}\left\|\mathcal{A}u^{t+1} - \mathcal{A}u^t\right\|_2^2 - \frac{\beta}{2}\left\|\nabla u^{t+1} - \nabla u^t\right\|^2 \\
&\leq -\frac{\left(\lambda \underline{s}_1^2 + \beta \underline{s}_2^2\right)}{2}\left\|u^{t+1} - u^t\right\|_2^2.
\end{aligned}$$

On the other hand, the minimization subproblem (16a) implies that

$$\mathcal{L}_3 := \mathcal{L}\left(\mathbf{v}^{t+1}, u^t, \boldsymbol{\alpha}^k\right) - \mathcal{L}\left(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t\right) < 0.$$

So we obtain

$$\mathcal{L}\left(\mathbf{v}^{t+1}, u^{t+1}, \boldsymbol{\alpha}^{t+1}\right) \leq \mathcal{L}\left(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t\right) + \frac{\bar{\kappa}}{\beta}\left\|u^{t+1} - u^t\right\|_2^2$$

if $\bar{\kappa} := \lambda \bar{s}_1^4 \bar{s}_2^2 - (\beta\lambda \underline{s}_1^2 + \beta^2 \underline{s}_2^2)/2 < 0$. □

Remark 4. There is a similar result to Theorem 5 in [29], where the authors gave some prior assumptions to keep the monotonicity of the iteration sequence. A key question is then used to determine whether $\bar{\kappa} := \lambda \bar{s}_1^4 \bar{s}_2^2 - (\beta\lambda \underline{s}_1^2 + \beta^2 \underline{s}_2^2)/2 < 0$ is held. Actually, we usually choose a bigger penalized parameter $\beta > 0$ when using the augmented Lagrangian method in (15), so the assumption can be easily satisfied since other parameters are constant.

Theorem 6. *If assumptions in Theorem 5 still hold and* $1 - 2\lambda \underline{s}_1^2 \underline{s}_2^2 \geq 0$*, the iteration sequence* $\{(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t)\}$ *generated by the iteration scheme (16a)-(16c) has a cluster point. Furthermore, this cluster point is also a saddle point of the problem (15).*

Proof. Based on Theorem 5 and the equation (18), we have

$$\begin{aligned}
\mathcal{L}\left(\mathbf{v}^0, u^0, \boldsymbol{\alpha}^0\right) \geq \mathcal{L}\left(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t\right) &= \left\|\mathbf{v}^t\right\|^p + \frac{\lambda}{2}\left\|\mathcal{A}u^t - f\right\|_2^2 \\
&\quad + \frac{\beta}{2}\left\|\mathbf{v}^t - \left(\nabla u^t - \frac{\boldsymbol{\alpha}^t}{\beta}\right)\right\|^2 - \frac{\beta}{2}\left\|\boldsymbol{\alpha}^t\right\|^2 \\
&\geq \left\|\mathbf{v}^t\right\|_g^p + \frac{\lambda}{2}\left(1 - 2\lambda \underline{s}_1^2 \underline{s}_2^2\right)\left\|\mathcal{A}u^t - f\right\|_2^2 \\
&\quad + \frac{\beta}{2}\left\|\mathbf{v}^t - \left(\nabla u^t - \frac{\boldsymbol{\alpha}^t}{\beta}\right)\right\|^2.
\end{aligned}$$

Then we can deduce that the sequence as

$$\left\{\left\|\mathbf{v}^t\right\|_g^p\right\},\ \left\{\left\|\mathcal{A}u^t - f\right\|_2^2\right\} \text{ and } \left\{\left\|\mathbf{v}^t - \left(\nabla u^t - \frac{\boldsymbol{\alpha}^t}{\beta}\right)\right\|^2\right\}$$

are bounded if we assume that $1 - 2\lambda \underline{s}_1^2 \underline{s}_2^2 \geq 0$. So we have the sequence $\{(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t)\}$ is bounded and there exists a cluster point.

Suppose that $\{(\mathbf{v}^*, u^*, \boldsymbol{\alpha}^*)\}$ is a cluster point of the sequence $\{(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t)\}$ and let $\{(\mathbf{v}^{t_j}, u^{t_j}, \boldsymbol{\alpha}^{t_j})\}$ be a convergent subsequence such that

$$\lim_{j\to\infty}\left(\mathbf{v}^{t_j}, u^{t_j}, \boldsymbol{\alpha}^{t_j}\right) = \left(\mathbf{v}^*, u^*, \boldsymbol{\alpha}^*\right).$$

Simultaneously, from Theorem 5 we have

$$\begin{aligned}
-\sum_{t=0}^{t_j}\frac{\bar{\kappa}}{\beta}\left\|u^{t+1} - u^t\right\|_2^2 &\leq \mathcal{L}\left(\mathbf{v}^0, u^0, \boldsymbol{\alpha}^0\right) \\
&\quad -\mathcal{L}\left(\mathbf{v}^{t_j}, u^{t_j}, \boldsymbol{\alpha}^{t_j}\right) < \infty.
\end{aligned}$$

Taking the limit on the side of above inequation, we have

$$\lim_{t\to\infty}\left\|u^{t+1} - u^t\right\|_2^2 = 0. \tag{20}$$

By using the relationship of (18), we have

$$\left\|\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t\right\|_2^2 = \left\|\nabla\mathcal{A}^T\mathcal{A}\left(u^{t+1} - u^t\right)\right\|_2^2 \leq \bar{s}_1^4 \bar{s}_2^2 \|u^{t+1} - u^t\|_2^2.$$

So we can deduce that

$$\lim_{t\to\infty}\left\|\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t\right\|_2^2 = 0. \tag{21}$$

Furthermore, using the updating (17c), we have

$$\begin{aligned}
\left\|\mathbf{v}^{t+1} - \mathbf{v}^t\right\|_2^2 \leq \bar{s}_2^2\left\|\mathbf{v}^{t+1} - \mathbf{v}^t\right\|_2^2 &+ \frac{1}{\beta}\left\|\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t\right\|_2^2 \\
&+ \frac{1}{\beta}\left\|\boldsymbol{\alpha}^t - \boldsymbol{\alpha}^{t-1}\right\|_2^2
\end{aligned}$$

Thus we also obtain that

$$\lim_{t\to\infty}\left\|\boldsymbol{\alpha}^{t+1} - \boldsymbol{\alpha}^t\right\|_2^2 = 0 \tag{22}$$

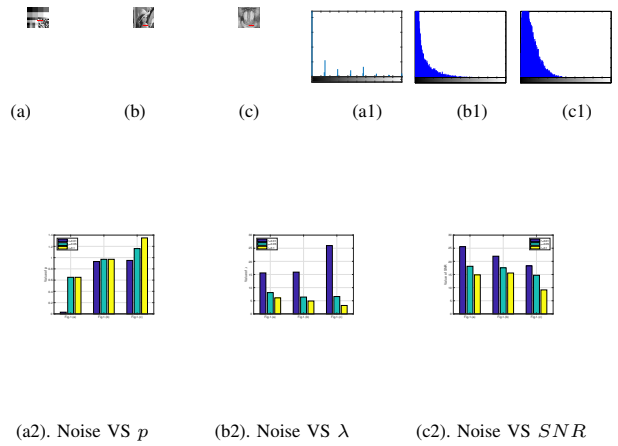
based on the relationships of (20) and (21).

We next show that the cluster point of the sequence $\{(\mathbf{v}^t, u^t, \boldsymbol{\alpha}^t)\}$ is a saddle point of the Lagrangian function (15). For a convergent subsequence $\{(\mathbf{v}^{t_j}, u^{t_j}, \boldsymbol{\alpha}^{t_j})\}$, the facts (20)-(22) imply that the sequences $\{(\mathbf{v}^{t_j}, u^{t_j}, \boldsymbol{\alpha}^{t_j})\}$ and $\{(\mathbf{v}^{t_j+1}, u^{t_j+1}, \boldsymbol{\alpha}^{t_j+1})\}$ have the same limit point $\{(\mathbf{v}^*, u^*, \boldsymbol{\alpha}^*)\}$. Then passing to the limit in the optimization condition (17a) along the subsequence index t_j yields that

$$\begin{cases}
0 \in \mathbf{v}^* - \mathrm{prox}_{\ell^p}^{\frac{\beta}{g}}\left(\nabla u^* - \boldsymbol{\alpha}^*/\beta\right), & \text{(23a)} \\
0 = \lambda\mathcal{A}^T\left(f - \mathcal{A}u^*\right) - \mathrm{div}\boldsymbol{\alpha}^*, & \text{(23b)} \\
0 = \beta\left(\mathbf{v}^* - \nabla u^*\right). & \text{(23c)}
\end{cases}$$

Then the above assertion is held. □

Remark 5. Using the property of the saddle point, we can deduce that the sequence $\{(\mathbf{v}^*, u^*, \boldsymbol{\alpha}^*)\}$ is also the stationary point of the optimization problem (14) by substituting (23c) into the Lagrangian function (15). However we can not expect to obtain the global convergence for the general quasi-morm ℓ^p when $p \in (0, 1)$. In other words, this case does not satisfy the Kurdyka-Łojasoewicz (KL) property when p is a irrational number [29], [1], [2].



(a) (b) (c) (a1) (b1) (c1)

(a2). Noise VS p (b2). Noise VS λ (c2). Noise VS SNR

Fig. 1. Top Row: (a). Piecewise Constant; (b). Lena; (c). Mandrill; (a2)-(c2): Related histograms of the gradient magnitudes for images (a)-(c). (a1)-(c1): Related data for parameters and SNR for restoring different noise images.

C. Choice of $p(x)$ in the model (4)

In the proposed model (4), the choice of the self-adaptive $p \in (0, 2)$ is important and critical to be defined for efficiently describing all kinds of image structures. As mentioned before, a larger value p than 1 should be set to suppress noise and simultaneously blur some details such as edge and texture in the image. So it is suitable to the smoothing regions. In contrast, for edge-area pixels, a smaller p nearing to 1 should be set to preserve the detailed information. To the approximate constant regions, we expect p to be smaller than 1 in order to keep $\|\nabla u\|^p$ to penalize sparsity. These facts can be also found from Fig 1, where three testing images include different structure features. Fig 1(a) includes more constant regions and also a few edge information. Fig 1(c) includes more texture information and the structure of Fig 1(b) is between in Fig 1(a) and 1(c) as observing in the gradient histograms for Fig 1(a1)-(c1).

To efficiently find the relationship between the parameter p and noise level while restoring noisy images based on Fig 1(a)-(c), where we use the Matlab function $imnoise(\cdot)$ to add to different white Gaussian noise with the deviation as $\sigma = 0.01$, 0.05 and 0.1. We can find that the choice of parameter p in Fig 1(a2) depends on image structures and noise levels σ at the same time. Specifically, the parameter p is expected to be low than 1 in the (approximation) piecewise constant regions for the white Gaussian noise with low level and piecewise constant regions. However, we need a bigger p to penalize $\|\nabla u\|^p$ for the texture regions and the white Gaussian noise with high level.

For the dependence of p to the image structure, we need the local activity indicator to determine station of the pixel. Here we employs a structure tensor indicator, which has been proven that this indicator can efficiently distinguish the different structures such as the smooth region, edges, corner and isolated noise via using the gradient information. One of classic structure tensors identifies gradient with the outer product as

$$J_{\hat{\sigma}} = G_{\hat{\sigma}} * (\nabla f \otimes \nabla f),$$

where $G_{\hat{\sigma}}$ is a Gaussian function with the standard derivation $\hat{\sigma}$ and $\otimes$ denotes the convolute operator. This formulation has been used in many image problems such as the image restoration and segmentation [21], [54], [45], [51], [37], [53]. However, the choice of $\hat{\sigma}$ heavily effect image details. Larger $\hat{\sigma}$ values tend to blur edges and corners while smaller scale values can miss certain edge structures. Some works proposed to choose a local adaptive $\hat{\sigma}$ based on the optimization scheme [53], [35], but it increases the calculation cost. In order to simplify, we use the Wiener filter $W(\cdot)$ instead of the Gaussian convolution filter in the tensor structure as

$$\mathcal{W} := W(\nabla f \otimes \nabla f) := \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}.$$

Denoting the eigenvector as e_1 and e_2 with $e_1 /\!\!/ \nabla f$ and $e_2 \perp \nabla f$ for the tensor matrix $\mathcal{W}$, then the corresponding eigenvalues d_1 and d_2[1] are give by

$$d_1 = \frac{1}{2}\left|W_{11} + W_{22} + \sqrt{(W_{11} - W_{22})^2 + 4W_{12}^2}\right|,$$
$$d_2 = \frac{1}{2}\left|W_{11} + W_{22} - \sqrt{(W_{11} - W_{22})^2 + 4W_{12}^2}\right|.$$

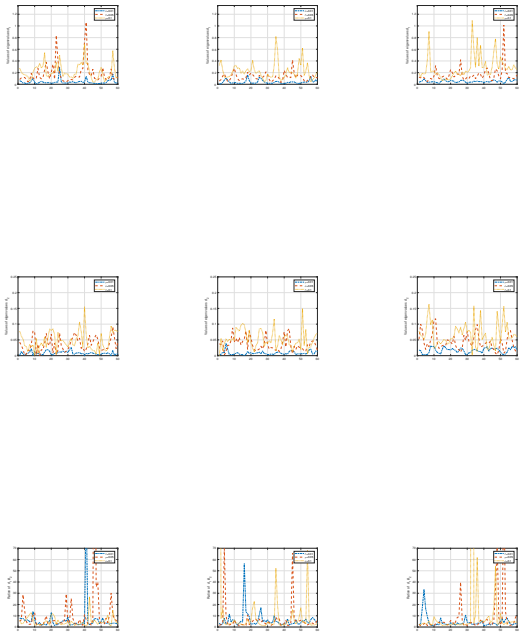
Therefore d_1 and d_2 respectively describe the pixel value fluctuations and preferred local direction of smoothing and also offer a discriminative description of the local structure. For smooth areas with small variations in the degraded image, we have $d_1 \approx d_2 \approx 0$. Under the condition that $d_1 \gg d_2 \approx 0$, the strong variations are only concentrated on a dominant direction. Therefore, the current point is close to the straight edge. As to the points close to corners, both d_1 and d_2 are large, i.e., $d_1 \geq d_2 \gg 0$. These facts can be also observed from Fig 2. In addition, we should notice that the level of noise also effects eigenvalues d_1 and d_2, i.e., the high noise can efficiently destroy the simple structure such as the smoothing region, so it increases the value of d_1 and d_2. In summary, eigenvalues d_1 and d_2 simultaneously include the structure and contamination information. Based on above observations, we first propose to couple d_1 with d_2 by

$$c_i = \frac{d_1 d_2}{k_i + d_1 + d_2} + \frac{|d_1 - d_2|^2}{(k_i + d_1 + d_2)^2}$$

to indicate difference structures for $i = 1, 2$. Specially, $c_1, c_2 \approx 0$ indicate smooth areas and $c_1, c_2 \gg 0$ indicates some significant areas such as edges, flow linear structure area and T-shaped characteristics areas. So c_1 and c_2 are designed as the local activity indicator and then we set the regularization index p as

$$p = \begin{cases} 1, & \text{if } d_1 > \varrho_1 d_2, \\ 1 + \kappa_1 \exp(-c_1), & \text{if } d_1 \in [\varrho_2 d_2, \varrho_1 d_2], \\ 1 - \kappa_2 \exp(-c_2), & \text{if } d_1 < \varrho_1 d_2, \end{cases} \tag{24}$$

[1]We normalize the original image to be in [0,1] before degrading it. So the eigenvalue d_1 shown in Fig. 2 is small.



(a). Piecewise Constant (b). Lena (c). Mandrill

Fig. 2. The variation of eigenvalues d_1 in the top and d_2 in the bottom with different noise deviations as $\sigma = 0.01$, 0.05 and 0.1. The station with 60 pixel width are shown as the red lines in Fig 1(a)-(c).

where ϱ_i is the truncation parameter and $\kappa_i \in (0,1)$ is the weighted parameter for $i = 1$ or 2. In general, how to adaptively choose them is another more difficult issue. Following from the observations in Fig 2, we experimentally use a compromised parameter as $\varrho_1 = 10$ and $\varrho_2 = 4$. By using this setting, the adaptive norm can efficiently preserve image structures by setting $p = 1$ to keep edges, $p < 1$ to keep sparsity of approximation constant regions and $p > 1$ to keep smoothing regions. Although these choices have might be improved further, the proposed method still has advantages over the other regularized schemes as shown the numerical implementations.

III. Numerical Implementations

In this section, some experimental results are conducted to evaluate the performance of the proposed model and numerical method. From various classic and recent state-of-the-art image restoration approaches such as the gradient-based methods, learning-based methods and wavelet-based methods, etc., we only focus on the comparisons between our proposed model (Our) and other the gradient-based models such as the ROF model (TV) [42], the adaptive total variation model (ATV) [5], the high-order total variation model (HOTV) [56], the total generalized variation model (TGV) [7] and non-local total variation model (NLTV) [25]. All experiments are performed using MATLAB (R2017a) on a Windows10(64bit) desktop computer with an Intel Core i7 2.40 GHz processor and 8.0 GB of RAM. Furthermore, all restored images can be obtained when the relative difference between two successive iteration satisfies

$$\frac{\|u^k - u^{k-1}\|_2}{\|u^{k-1}\|_2} \leq 10^{-5}$$

or after a maximum number of 500 iterations. Furthermore, testing images are normalized in the range of [0,1] before degrading them and then use Matlab's function '$imnoise$' to add noise or to add blur to them. Here we set the Greek alphabet σ as the standard deviation of the noise. The restored quality is measured by SNR (signal-noise ratio) and $SSIM$ (structural similarity index). It was demonstrated that the higher value of $SSIM$ and SNR implies a better restoration corresponds to subjective quality of visual perception. Here we still use Matlab's functions snr and $ssim$ with the standard parameter to compute the corresponding values. In addition, we can find the exact solution $\boldsymbol{v}^{k+1}$ of the subproblem (16a) from the theory (See Theorem 4) when iteration t tends to the infinity. However it is unprocurable from numerical implementation. In all numerical implementations, we set $\varepsilon = 0.01$ with the inner iteration $t = 5$. For this choice, we still obtain the satisfactory numerical results.

A. Selection of regularization parameter λ in the model (4)

Except for the parameter p in the (quasi)norm term, the restoration effectiveness also depends on the regularization parameter λ that controls how much filtering is introduced by the regularization. Often the key issue in connection with these methods is to find a regularization parameter that gives a good balance, filtering out enough noise without losing too much information in the computed solution. In order to show this fact, we first fix the regularization parameter λ and then show the variation of the energy to the fitting term and the

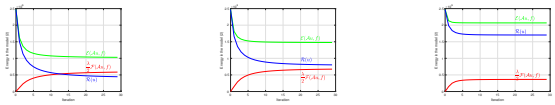


Fig. 3. As an example, we use a fixed model (2) as the ROF model [42] to illustrate the variation for the energy functionals with three different regularization parameters λ. Left: Smaller λ; Middle: Suitable λ; Right: Larger λ.

regularization term in Fig 3 (a). It is obvious that the data fitting energy always increases but the regularization energy always reduces when iterating the proposed numerical method and choosing the original degraded image f as the original value. Over estimating the regularized parameter λ may lead to the domination for the regularization term $\mathcal{R}(u)$. So it will over-penalize image details and obtain over smoothed blurry result with a low energy $\mathcal{E}(\mathcal{A}u, f)$ as showing in Fig 3(a). On

the contrary, while underestimating it may leave the noise in the image unfiltered. So we still obtain a noisy image and a high energy $\mathcal{E}(\mathcal{A}u, f)$ as showing in Fig 3(c). So it is very important to find a suitable λ based on these relation. Some classic methods to choose the suitable λ can be employed such as the L-curve method [30], the generalized cross validation (GCV) [23], the discrepancy principle [40], or the variational Bayes approach [44].

In our experiments, the main aim is focused on the advantages of the proposed adaptive norms model and especially the proposed numerical method. If above methods for selecting regularization parameter are used, it will bring great uncertainties for the comparison between the proposed and other traditional methods. Therefore, for a fair comparison between different models, the regularization parameter was manually determined by attempting a series of values and selecting the one with the highest SNR (in simulated experiments) or the best visual effect (in real experiments). Specifically, we set the related parameters into a bigger range as $[a, b]$ and then find a suitable sunset as $[c, d] \subset [a, b]$. In the next we find a more suitable parameter in $[c, d]$. When the difference between the successive SNR is below 0.001, we set this parameter as the best value to the regularization parameter λ.

B. LRLS Scheme VS Algorithm 2

Except our proposed Algorithm 2, Chartrand [14] extended the lagged diffusivity fixed point method to solve the model (4) as an approximated scheme

$$\left(\lambda \mathcal{A}^T \mathcal{A} - \frac{p}{2}\text{div}\left(|\nabla u^k|_\varepsilon^{p-2}\right)\nabla\right) u^{k+1} = \lambda \mathcal{A}^T f. \qquad (25)$$

for the fixed parameter $p \in (0, 1)$, where the approximated factor $|\nabla u^k|_\varepsilon = \sqrt{(u_y^k)^2 + (u_y^k)^2 + \varepsilon^2}$. For convenience we have ignored factors of $p/2$ since it can be consolidated into the parameter λ. In fact it is easy to observe that the scheme (25) is equivalent to the iteratively reweighted least square (IRLS) scheme if we rewrite its corresponding optimization problem as

$$u^{k+1} := \underset{k}{\text{argmin}} \frac{\lambda}{2}\|\mathcal{A}u - f\|_2^2 + \frac{1}{2}\left\|\left(W^k\right)^{0.5}\nabla u\right\|_2^2, \quad (26)$$

where W^k is the weighting operator defined by $W^k := \frac{p}{2}\left(\left(u_y^k\right)^2 + \left(u_y^k\right)^2 + \varepsilon^2\right)^{\frac{p-2}{2}}$. So the convergence analysis can be found in the work [47]. However, this convergence is only used for the fixed parameter p in the model (4). Furthermore, we also found that the numerical method was not stable when choosing an adaptive parameter p for the IRLS scheme. In order to keep fairness, we compare the computational performance between the LRLS and Algorithm 2 for the fixed parameter p. It is noteworthy that we possibly choose $p \in (0, 1)$ when restoring the approximated piecewise constant image as restoring the image in Fig 1 (a). In this case the model (4) is nonconvex, so we can not expect to find the global solution and then the numerical solution also heavily desponds on the original value.

Let us mention that using these two schemes to solve the model (4) does not mean that we need to choose the same regularization parameter λ and p. It is because that numerical schemes also depend on the image structure, the numerical accuracy and details of convergence for these two methods. Here we still consider to restore images shown in Fig 1 (a)-(c) and also add to the white Gaussian noise with $\sigma = 0.01$ for degrading these images. Related results are shown in Table I. Here it can be seen that our proposed method is slightly superior to the LRLS in the restoration qualities based on values of the SNR and $SSIM$.

Images	Fig1 (a)		Fig1 (b)		Fig1 (c)	
Model	LRLS	Our	LRLS	Our	LRLS	Our
p	0.34	0.21	0.95	0.92	0.91	0.94
λ	54	13.5	15.8	16.2	29.2	26.1
SNR	26.1875	26.3294	21.9607	21.9659	18.2902	18.2816
SSIM	0.8842	0.8906	0.7906	0.7948	0.7209	0.7199
Time	167.9688	27.8125	20.6406	16.2813	14.2813	14.1719

TABLE I
RELATED DATA AND PARAMETERS FOR RESTORING DEGRADED IMAGES BY USING THE ITERATIVELY REWEIGHTED LEAST SQUARE (IRLS) SCHEME AND ALGORITHM 2(OUR).

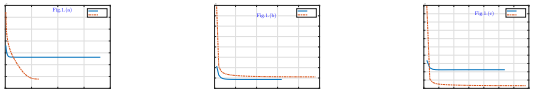


Fig. 4. Changing curves with the iteration for the objective function in the model (4) by using the LRLS and our proposed method.

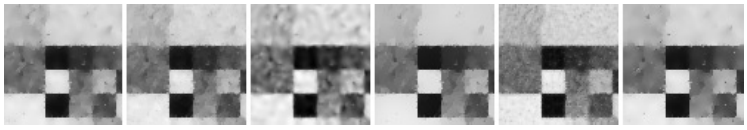
For more comparisons, it is also instructive to examine the energy changing for the objective function in the model (4), where related parameters λ and p for these two numerical methods are shown in Table I. It can be noticed that our proposed method is more robust when dealing with the approximation constant image as Fig 1(a). Furthermore, we shall notice that the nonincreasing value of objective function actually implies the convergence of these two numerical methods in the case of iteration. In addition, a more important fact is that our proposed method required significantly less CPU time elapsed as compared to the LRLS. These mentioned facts imply that our proposed method is more suitable for dealing with the larger scale image.

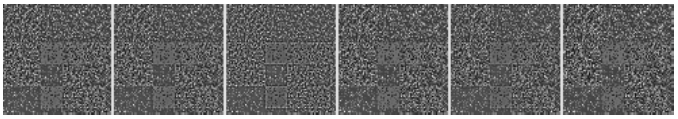
C. Image Denoising

Image denoising is an importance image restoration problem where the matrix $\mathcal{A}$ is assumed to be the identity operator $\mathcal{I}$. Here we consider the white Gaussian noise at three different noise levels (low, medium, and high) as $\sigma = 0.01, 0.05, 0.1$, respectively. In Table II, we report the denoising results in terms of the SNR and $SSIM$ for the corresponding testing images and noise level. It can be observed that that our proposed method almost gives higher $SSIM$ or SNR than other models for different level noises expect for the NLTV to deal with the Mandrill image. In fact the NLTV is more

Noise	0.01		0.05		0.1	
Images	Fig 1 (a):Texmos					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	23.9227	0.7934	17.5243	0.6585	14.6018	0.5794
ATV	23.4834	0.7710	17.5136	0.6445	14.6365	0.5680
HOTV	22.4652	0.6796	16.6996	0.5206	14.1927	0.4635
TGV	24.4355	0.8422	17.8302	0.7210	14.8175	0.6479
NLTV	24.2494	0.7840	17.6014	0.5630	14.5683	0.4501
Our	25.7133	0.8799	18.1812	0.7712	14.9147	0.6913
Image	Fig 1 (b):Lena					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	21.8957	0.7829	17.5775	0.6441	15.4964	0.5743
ATV	21.7439	0.7677	17.5756	0.6338	15.5468	0.5602
HOTV	21.8068	0.7524	17.5640	0.6096	15.6144	0.5472
TGV	22.1107	0.7833	17.8561	0.6551	15.8759	0.5955
NLTV	22.4826	0.7890	17.7090	0.5978	15.3845	0.4938
Our	21.9687	0.7928	17.5930	0.6461	15.5442	0.5797
Image	Fig 1 (c):Mandrill					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	20.3515	0.8085	18.3016	0.7232	17.1538	0.6591
ATV	20.2842	0.8047	18.2710	0.7205	17.1514	0.6575
HOTV	20.2214	0.8056	18.1252	0.7196	16.9607	0.6558
TGV	20.3549	0.8084	18.3100	0.7234	17.1653	0.6591
NLTV	20.7340	0.8222	18.6658	0.7390	17.5062	0.6763
Our	20.3764	0.8097	18.3184	0.7240	17.1729	0.6593

TABLE II
SNR AND SSIM IN DENOISING EXPERIMENTS WITH DIFFERENT LEVELS OF GAUSSIAN NOISE σ BY USING DIFFERENT MODELS AND TESTING IMAGES.





suitable to restore the texture image since it heavily depends on the structure similarity.

Next, in order to show the quantitative comparisons for different models, we also consider the visual assessment of the restoration course and results by only employing the medium noise image with $\sigma = 0.05$. Related comparisons including the part of restored images with size 81×81 and the difference images between restored images and original images are shown in Figure 5. For the comparison of the computational efficiency in Figure 6, CPU-time ratios are presented with our method as the base reference. For each test, it easily observes that the TV-model, the HOTV model and our proposed model need a less CPU time, other models consume more due to needing typical hundreds of CG iterations as the ATV model, solving hybrid-based model as the TGV model and finding a suitable weighted neighbourhood as the NLTV model. Though our proposed model costs more CPU time than the TV and the HOTV, we can observe that our proposed model can efficiently captures thin structures, piecewise constant/smoothing regions and weak signals quite well. Above observations imply that our proposed method also owns competitive computation cost.

D. Image Deblurring

This experiment mainly focuses on the image deblurring problem and uses 'Cameraman', 'Boat' and 'Barbara' shown in Figure 7 to be the testing images. Here we consider three kinds of blurring kernels: Gaussian, Motion, and Average. For the convenience of description, we denote the Gaussian blur with a blurring size 'S' and a standard deviation "σ_G" as (G, S, σ_G). The motion blur with a motion length "Len" and an angle "θ" is denoted as (M, Len, θ). Similarly, the average blur with a blurring size 'S' is denoted by (A, S). We generate these kernels using Matlab function "$fspecial$" with various

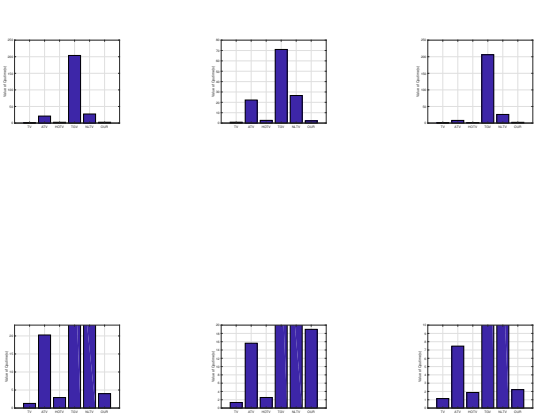


Fig. 6. Comparisons of CPU time by using different models. Top: Original bar image; Bottom: Zoom above bar image for efficient comparisons.

testing parameters and also add to the white Gaussian noise with $\sigma = 0.01$ to these blurring images.

Blurring	$(G, 15, 0.5)$		$(M, 10, 5)$		$(A, 5)$	
Images	Fig 7 (a):Cameraman					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	20.5131	0.7255	16.3945	0.6418	17.5102	0.6654
ATV	20.2521	0.6926	16.2640	0.6180	17.4166	0.6431
HOTV	20.0529	0.6804	16.0146	0.5940	16.9602	0.6142
TGV	20.3524	0.7471	16.3865	0.6423	17.4890	0.6675
NLTV	21.0138	0.6992	16.2449	0.5406	17.2337	0.5455
Our	20.5713	0.7313	16.4237	0.6493	17.5453	0.6725
Image	Fig 7 (b):Boat					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	20.5199	0.7211	17.3214	0.5967	17.8032	0.5987
ATV	20.4515	0.7079	17.3774	0.5928	17.8729	0.5952
HOTV	20.4837	0.7117	17.3978	0.5938	17.7509	0.5856
TGV	20.5297	0.7227	17.3120	0.5977	17.8055	0.6017
NLTV	20.9172	0.7179	17.3754	0.5689	17.8499	0.5534
Our	20.5291	0.7231	17.3475	0.6011	17.8664	0.6041
Image	Fig 7 (c):Mandrill					
Model	SNR	SSIM	SNR	SSIM	SNR	SSIM
TV	17.0517	0.6491	15.1423	0.5320	15.7158	0.5735
ATV	17.0348	0.6391	15.1559	0.5275	15.7731	0.5744
HOTV	16.9747	0.6358	15.1047	0.5169	15.6773	0.5619
TGV	17.0553	0.6492	15.1455	0.5340	15.7287	0.5767
NLTV	17.5552	0.6562	15.2047	0.5084	15.8275	0.5544
Our	17.0632	0.6520	15.1599	0.5355	15.7562	0.5784

TABLE III
SNR AND SSIM IN DEBLURRING EXPERIMENTS WITH DIFFERENT BLUR OPERATORS BY USING DIFFERENT MODELS AND TESTING IMAGES.

The quantitative indicators of the SNR and $SSIM$ are arranged in Table III, and the computational times of the different algorithms are listed in Table IV. These results show that our proposed model and numerical method significantly outperform other models according to $SSIM$, SNR or CPU time. In addition, for efficient vision comparisons, we also show a part of the restored images with the size 101×101 in Fig 7. It is obvious that all of models can efficiently suppress the blurring, especially for the Gaussian blur. Furthermore, our proposed model from the enlarge area produces very satisfying

Models	TV	ATV	HOTV	TGV	NLTV	Our
Image	$(G, 15, 0.5)$					
Fig 7(a)	2.6094	143.5000	1.3438	122.7656	608.0781	7.1563
Fig 7(b)	1.8438	108.7031	1.8125	205.7969	522.8906	17.9375
Fig 7(c)	1.8281	98.6563	0.9688	140.2188	529.2344	17.6094
Image	$(M, 10, 5)$					
Fig 7(a)	2.3125	187.7344	0.9688	181.6719	509.8438	19.5156
Fig 7(b)	2.3438	175.7500	0.7813	200.4063	509.9531	18.2500
Fig 7(c)	2.2344	185.1406	0.7969	70.9531	514.9531	6.6719
Image	$(A, 5)$					
Fig 7(a)	2.4063	211.1719	1.0781	202.6875	511.0000	19.5625
Fig 7(b)	2.2344	188.4844	1.5469	203.4688	511.2656	18.2188
Fig 7(c)	2.0938	235.7500	0.7969	120.7813	512.2813	19.8594

TABLE IV
CPU TIME IN DEBLURRING EXPERIMENTS WITH DIFFERENT BLUR OPERATORS BY USING DIFFERENT MODELS AND TESTING IMAGES.

results with clear contours, sharp edges and fine image detail.

IV. Conclusion

This paper proposed an adaptive $p(x)$-(quasi)norm regularization model based on the total variation functional to the image restoration problem. In order to efficiently describe the local structures such as the edges and smoothing regions, we first computed related eigenvalues of structure tensor based on the contaminated image and then coupled them to establish the adaptive weighted exponent $p(x) \in (0, 2)$ by choosing some robust truncated parameters. Since the proposed model coupled the cases of convexity and non-convexity when choosing $p(x) \in (0, 1)$ or $p(x) \in [1, 2)$, we first used the ADMM to obtain some solvable subproblems and then employed the half-quadratic scheme to obtain an approximated solution of one of subproblems. In addition, we also proved the convergence due to the fact that the proposed numerical methods is asymptotic. In further demonstrating the merits of our proposed model and justifying its remarkable performance, we carried out extensive numerical comparisons with other classic variation-based models. The experimental by dealing with the standard testing images shown that the proposed method outperforms other state-of-the-art variational-based restoration approaches, in terms of CPU time or the restored measured standard such as the SNR and SSIM. our proposed model and numerical method can obtain the efficient restored images. Our future work will focus on extending the proposed method to the medical image reconstruction problem and the low rank matrix completion problem.

Acknowledgment

This work is partially supported by the National Basic Research Program of China (973Program) (No. 2015CB856003) and National Natural Science Foundation of China (Nos.11401170). The first authors also gratefully acknowledge financial support from China Scholarship Council (CSC) as a research scholar to visit the University of Liverpool.

References

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

[2] H. Attouch, J. Bolte, and B. Svaiter. *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forwardCbackward splitting, and regularized gaussCseidel methods.* Mathematical Programming, 137(1-2):91-129, 2013.

[3] G. Aubert and P. Kornprobst. *Mathematical Problem in Image Processing: partial differential equations and the calculus of variations.* Springer, New York, 2008.

[4] J. Bai and X. Feng. *Fractional-order anisotropic diffusion for image denoising.* IEEE Transactions on Image Processing, 16(10):2492-2502, 2007.

[5] P. Blomgren, T. Chan, P. Mulet, and C. Wong. *Total variation image restoration: numerical methods and extensions.* In IEEE International Conference on Image Processing, pages III, 384-387, 1997.

[6] 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, 3(1):1-122, 2011

[7] K. Bredies, K. Kunisch, and T. Pock. *Total generalized variation.* SIAM Journal on Imaging Sciences, 3(3):492-526, 2010.

[8] A. Buades, B. Coll, and J. Morel. *A review of image denoising methods, with a new one*. Multiscale Modeling and Simulation, 4(2):490-530, 2006.

[9] A. Chambolle and T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging.* Journal of Mathematical Imaging and Vision, 40(1):120-145, 2011.

[10] R. Chan, T. Chan, and H. Zhou. *Continuation method for total variation denoising problems.* Proceedings of SPIE, 2563:314-325, 1995.

[11] R. Chan and H. Liang. *Half-Quadratic algorithm for ℓ_q-ℓ_p problems with application to TV-ℓ_1 image restoration and compressive sensing.* Global Optimization Methods, LNCS, 78-103, 2014.

[12] T. Chan, A. Marquina, and P. Mulet. *High-order total variation-based image restoration.* SIAM Journal on Scientific Computing, 22(2):503-516, 2000.

[13] T. Chan and J. Shen. *Image Processing and Analysis-Variational, PDE, Wavelet, and Stochastic Methods.* SIAM, Philadephia, 2005.

[14] R Chartrand. *Nonconvex regularization for shape preservation*. IEEE International Conference on Image Processing, 293-296, 2007.

[15] Y. Chen, S. Levine, and M. Rao. *Variable exponent, linear growth functionals in image restoration.* SIAM Journal on Applied Mathematics, 64(4):1383-1406, 2006.

[16] Q. Chena, P. Montesinosb, Q. Suna, P. Hengc, and D. Xia. *Adaptive total variation denoising based on difference curvature.* Image and Vision Computing, 28(3):298-306, 2010.

[17] P. Combettes, L. Condat, J. Pesquet. and B. Vũ. *A forward-backward view of some primal-dual optimization methods in image recovery.* IEEE International Conference on Image Processing (ICIP), 27-30, 2014.

[18] R. Courant, K. Friedrichs, and H. Lewy. *On the partial difference equations of mathematical physics.* IBM Journal of Research and Development, 11(2):215-234, 1967.

[19] I. Daubechies, M. Defrise, and C. Mol. *Sparsity-enforcing regularisation and ISTA revisited.* Inverse Problems, 32(10):104001, 2016.

[20] Y. Dong, M. Hintermller, and M. Rincon-Camacho. *Automated regularization parameter selection in multi-scale total variation models for image restoration.* Journal of Mathematical Imaging and Vision, 40(1):82-104, 2011.

[21] V. Estellers, S. Soatto, and X. Bresson. *Adaptive regularization with the structure tensor.* IEEE Transactions on Image Processing, 24(6):1777-1790, 2015.

[22] D. Gabay and B. Mercier. *A dual algorithm for the solution of nonlinear variational problems via finite element approximation.* Computers and Mathematics with Applications, 2(1):17-40, 1976.

[23] N. Galatsanos and A. Katsaggelos. *Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation*. IEEE Transactions on Image Processing, 1(3):322-336, 1992.

[24] D. Geman and C. Yang. *Nonlinear image recovery with half-quadratic regularization.* IEEE Transactions on Image Processing, 47(2):932-946, 1995.

[25] G. Gilboa and S. Osher. *Nonlocal operators with applications to image processing.* Multiscale Modeling and Simulation, 7(3):1005-1028, 2008.

[26] R. Glowinski and A. Marroco. *Sur lapproximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problémes de Dirichlet non linéares.* Revue Francaise d'Automatique, Informatique, Recherche Opérationelle, 9(R-2):41-76, 1975.

[27] T. Goldstein and S. Osher. *The Split Bregman method for L1-regularized problems.* SIAM Journal on Imaging Sciences, 2(2):323-343, 2009.

[28] R. Gonzalez and R. Woods. *Digital Image Processing.* Hardcover, 2007.

[29] K. Guo, D. Han, and T. Wu. *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints.* International Journal of Computer Mathematics, 1-17, 2016.

[30] P. Hansen and D. Leary. *The use of the L-curve in the regularization of discrete ill-posed problems.* SIAM Journal on Scientific Computing, 14(6):1487-1503, 1993.

[31] B. He, Y. You, and X. Yuan. *On the convergence of primal-dual hybrid gradient algorithm.* SIAM Journal on Imaging Sciences, 7(4):2526-2537, 2014.

[32] A. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, 1989.

[33] A. Lanza, S. Morigi, and F. Sgallari. *Constrained $TV_p - \ell_2$ model for image restoration.* Journal of Scientific Computing, 68(1):64-91, 2016.

[34] A. Lanza, S. Morigi, L. Reichel, and F. Sgallari. *A generalized Krylov subspace method for $\ell_p - \ell_q$ minimization.* SIAM Journal on Scientific Computing, 37(5):S30-S50, 2015.

[35] T. Lindeberg. *Scale-Space Theory in Computer Vision.* Berlin, Springer-Verlag, 1994.

[36] J. Lv, L. Shen, X. Xu, and Y. Xu. *Multiplicative noise removal in imaging: An exp-model and its fixed-point proximity algorithm* . Applied and Computational Harmonic Analysis, 4(2):518C539, 2016.

[37] H. Ma and Y. Nie. *An edge fusion scheme for image denoising based on anisotropic diffusion models.* Journal of Visual Communication and Image Representation, 40(B):406-417, 2016.

[38] C. Miao and H. Yu. *A General-thresholding solution for $\ell_p(0 < p < 1)$ regularized CT reconstruction.* IEEE Transactions on Image Processing, 24(2):5455-5468, 2015.

[39] J. Moreau. *Propriét propriés des applications "prox"*. Comptes Rendus de l'Académie des Sciences Paris, 256:1069-1071,1963.

[40] V. Morozov. *Methods for solving incorrectly posed problems*. Springer-Verlag, New York, 1984.

[41] M. Nikolova and T. Chan. *The equivalence of the half-quadratic minimization and the gradient linearization iteration.* IEEE Transactions on Image Processing 16(6):1623-1627, 2007.

[42] L. Rudin, S. Osher, and E. Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D, 60(1-4):259-268, 1992.

[43] Y. Hu, G. Ongie, S. Ramani, and M. Jacob. *Generalized higher degree total variation (HDTV) regularization.* IEEE Transactions on Image Processing, 23(6):2423-35, 2014.

[44] J. Oliveira, J. Bioucas-Dias, and M. Figueiredo. *Adaptive total variation image deblurring: A majorizationCminimization approach*. Signal Processing, 89(9):1683-1693, 2009.

[45] V. Prasath, D. Vorotnikov, R. Pelapur, G. Seetharaman, and K. Palaniappan. *Multiscale Tikhonov-total variation image restoration using spatially varying edge coherence exponent.* IEEE Transactions on Image Processing, 24(12):5220-5235, 2015.

[46] R. Rockafellar and R. Wets. *Variational Analysis.* Springer, 1998.

[47] P. Rodriguez and B. Wohlberg. *Efficient minimization method for a generalized total variation functional*. IEEE Transactions on Image Processing, 18(2):322-332, 2009.

[48] O. Scherzer. *Handbook of Mathematical Methods in Imaging*. Springer, New York, 2015.

[49] I. Selesnick. *Generalized total variation: tying the knots.* IEEE Signal Processing Letters, 22(11):2009-2013, 2015.

[50] S. Setzer. *Operator splittings, Bregman methods and frame shrinkage in image processing.* International Journal of Computer Vision. 92(3):265-280, 2009.

[51] H. Shen, L. Peng, L. Yue, Q. Yuan, and L. Zhang. *Adaptive norm selection for regularized image restoration and super-resolution.* IEEE Transactions on Cybernetics, 46(6):1388-1399, 2016.

[52] E. Sidky, R. Chartrand, J. Boone, and X. Pan. *Constrained TV^p minimization for enhanced exploitation of gradient sparsity: application to CT image reconstruction.* IEEE Journal of Translational Engineering in Health and Medicine, 30(2):1800418, 2004.

[53] X. Wang, S. Huang, and H. Xu. *An efficient local Chan-Vese model for image segmentation.* Pattern Recognition, 43(3):603-618, 2010.

[54] J. Weikert, B. Romeny, and M. Viergever. *Efficient and reliable schemes for nonlinear diffusion.* IEEE Transactions on Image Processing, 7(3):398-410, 1998.

[55] J. Woodworth and R. Chartrand. *Compressed sensing recovery via nonconvex shrinkage penalties.* Inverse Problems, 32(7):075004, 2016.

[56] C. Wu and X. Tai. *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models.* SIAM Journal on Imaging Sciences, 3(3):300-339, 2010.

[57] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang. *$L_{1/2}$ regularization:convergence of iterative half thresholding algorithm.* IEEE Transactions on Signal Processing, 69(2):2317-2329, 2014.

[58] J. Zhang and Z. Wei. *A class of fractional-order multi-scale variational models and alternating projection algorithm for image denoising*. Applied Mathematical Modelling, 35(5):2516-2528, 2011.