On convergent finite difference schemes for variational–PDE-based image processing

V. B. Surya Prasath1 · Juan C. Moreno2

Received: 19 October 2016 / Revised: 9 December 2016 / Accepted: 15 December 2016 / Published online: 2 January 2017 © SBMAC - Sociedade Brasileira de Matemática Aplicada e Computacional 2016

Abstract We study an adaptive anisotropic Huber functional-based image restoration scheme. Using a combination of L2–L1 regularization functions, an adaptive Huber functional-based energy minimization model provides denoising with edge preservation in noisy digital images. We study a convergent finite difference scheme based on continuous piecewise linear functions and use a variable splitting scheme, namely the Split Bregman (In: Goldstein and Osher, SIAM J Imaging Sci 2(2):323–343, 2009) algorithm, to obtain the discrete minimizer. Experimental results are given in image denoising and comparison with additive operator splitting, dual fixed point, and projected gradient schemes illustrates that the best convergence rates are obtained for our algorithm.

Keywords Image restoration · Adaptive denoising · Finite differences · Convergence · Huber functional

Mathematics Subject Classification 65M06 · 65M12 · 68U10

1 Introduction

Variational and partial differential equations (PDEs)-based schemes are popular in image and video processing problems. Particularly in image restoration, adaptive edge preserving smoothing can be achieved by choosing regularizing functions or equivalently diffusion...
coefficients carefully. This has been the object of study for the last three decades and we mention the seminal work of Perona and Malik [34] as the starting point in PDE-based image processing and the connections to variational and robust statistics have also been considered later [3,17–19,53]. We refer to the recent monographs [1,46] for an overview of these methods.

Based on the smoothness or regularity assumptions on the true image, various regularization functions can be used. The Tikhonov regularization function [51], which is based on the quadratic growth, $L^2$-gradient minimization, suppresses gradients and thus is effective in removing noise. Unfortunately, gradients can also represent edges which are important for further pattern recognition tasks. To avoid the over smoothing total variation or the $L^1$-gradient minimization, which is widely known as the total variation (TV) regularization model, has been advocated [45]. Recently, there are efforts to combine both the $L^2$ and $L^1$ based functionals into one common minimization problem such as the Huber function [4,38], inf-sup convolution [6,11]. Adaptive versions of the variational–PDE models are gaining popularity [15,16,36,40–42] and can give better restoration results than non-adaptive schemes in terms of edge preservation. The discrete approximation to the continuous variational–PDE schemes from image processing using finite difference and finite element-based schemes has been studied [7,8,13,20,30,49,54–56]. Convergence of finite differences for various PDEs is a classic area within numerical analysis and is still an active area of research in application areas such as image processing [5,10,57].

In this paper, we consider convergent finite difference schemes for an adaptive Huber type functional-based energy minimization model. We provide comparison with other convex variational regularization functions and use an edge indicator function guided regularization model. Using piecewise continuous linear functions along with a dedicated discrete energy functional, we study the convergence of discrete minimizers to the continuous solution. To solve the corresponding discrete convex optimization problem, various solvers exist, such as the dual minimization [9], primal–dual [12] alternating direction method of multipliers and operator splitting [47]. Here, we use the split Bregman algorithm studied by Goldstein and Osher [26,27] for computing the discrete energy minimizer of $L^1$ regularization problems. This algorithm is proven to be the fastest in terms of computational complexity for solving various image processing problems. We prove a convergence result for the class of weakly regular images using split Bregman method. We utilize an image adaptive inverse gradient-based regularization parameter for better denoising without destroying salient edges. Experimental results on real and synthetic noisy images are given to highlight the noise removal property of the proposed model. Comparison results with different discrete optimization models in undertaken and further visualizations are provided to support split Bregman-based solutions.

The rest of the paper is organized as follows. Section 2 provides the background on an adaptive Huber variational–PDE model along with some basic results on bounded variation space. Section 3 details a convergent numerical scheme for the variational scheme. Section 4 provides comparative numerical results on noisy images and Sect. 5 concludes the paper.

---

1 Semen Aronovich Geršgorin’s work [23] in 1930 was the first paper to treat the important topic of the convergence of finite-difference approximations to the solution of Laplace-type equations.
2 Continuous \( L^2 - L^1 \) variational–PDE model

Let \( u_0 : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R} \) be the input (noisy\(^2\) image. We consider the following continuous variational–PDE scheme for image restoration:\(^3\)

\[
E(u) = \int_{\Omega} \phi(x, |\nabla u|) \, dx + \frac{\lambda}{2} \int_{\Omega} |u - u_0|^2 \, dx
\]  

The corresponding PDE can be written in terms of the Euler–Lagrange equation,

\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\phi'(x, |\nabla u|) \nabla u}{|\nabla u|} \right) - \lambda (u - u_0)
\]  

The adaptive discontinuity function \( \phi(x, |\nabla u(x)|) = W(x) \times \varphi(|\nabla u(x)|) \) is chosen to be an even function. Note that the PDE in Eq. (2) is a generalized Perona and Malik [34]

\[
\frac{\partial u}{\partial t} = \text{div}(g(|\nabla u|) \nabla u) - \lambda (u - u_0),
\]  

where the diffusion function \( g \) is related with \( \varphi'(s) = 2sg(s) \). The diffusion coefficient function \( g(\cdot) \) decides how much smoothness occurs and helps in noisy pixels (outlier) rejection. Various choices for \( \varphi \) exist in the literature, see [14,21,22,33] and [50] for a recent review. Note that under Gaussian noise assumption the data fidelity term (also called the likelihood term) in Eq. (1) is quadratic and hence convex in \( u \). Thus, if the regularization term is also convex in \( u \) then we are guaranteed of the well-posedness of the energy minimization scheme given in (1). There are functions which are non-convex [4,22,33,44] with \( \varphi(s) \sim s^2 \) near 0 and asymptotically linear as \( |s| \rightarrow +\infty \). This can cause unstable behavior as the scheme can be plagued with local minima. In this paper, we concentrate on convex regularization functions and study a stable and convergent scheme.

\(^2\) We assume Gaussian noise, i.e., \( n \sim \mathcal{N}(0, \sigma_n) \).

\(^3\) Note we use the notation \( \nabla \) to denote the gradient and in the space of bounded variation functions \( BV \) it is in fact a Radon measure and is understood in the sense of distributions. The equality \( \int_{\Omega} |Dw| = \int_{\Omega} |\nabla u| \, dx \) is true when \( u \in W^{1,1}(\Omega) \).
Fig. 2 Synthetic Step image showing the effects of the choice of regularization function on the final restoration results. The $L^2$-gradient scheme (Tikhonov) over-smoothes the edge whereas $L^1$-gradient scheme (TV) though edge-preserving can introduce oscillations known as staircasing in homogeneous regions. An adaptive combination via (6) balances the smoothing along with edge preservation.

**Remark 1** There are other ways to incorporate adaptive weights inside the regularization function or equivalently the diffusion coefficient. For example, as in adaptive total variation, i.e., with $\varphi(s) = s$, $\phi(x, |\nabla u(x)|) = |W(x) \cdot \nabla u(x)|$ or in general $\phi(\cdot, |\nabla u(x)|) = \varphi(W(\cdot) |\nabla u(x)|)$. The main difference lies in the way the regularization function $\varphi$ is weighted anisotropically and the final results change according to the formulation utilized. The main convergence result in Sect. 3 holds true for this type of adaptive functions as well.

Two of the most obvious choices for the regularization function $\varphi$ are the Tikhonov or $L^2$-gradient $\varphi(s) = s^2$ and the total variation (TV) or $L^1$-gradient $\varphi(s) = s$, see Fig. 1a. Both these functions have their advantages and drawbacks as illustrated by a synthetic noisy step image restoration example given in Fig. 2. To further highlight the smoothing properties, we show in Fig. 3a line taken across the Step image and corresponding results.\(^4\) The Tikhonov regularization though effective in removing noise penalizes higher gradients and, hence, can smooth the step edge excessively as can be seen in Fig. 2c. In contrast, the TV regularization better preserves edges but some additional regions in the homogeneous parts can be enhanced, which is known as ‘staircasing’ artifact, see Fig. 2d. Hence, a robust regularizer is required for effective smoothing for denoising while edges are preserved. For example, motivated from the robust statistics, we consider the classical M-estimators Huber’s min–max function [29] and the Tukey’s bisquare function [52] which are given by

\[
\varphi_H(s) = \begin{cases} 
    s^2/2 & \text{if } |s| < k, \\
    k \left(|s| - \frac{k}{2}\right) & \text{if } |s| > k,
\end{cases}
\]

and

\[
\varphi_T(s) = \begin{cases} 
    \frac{k^2}{6} \left(1 - \left[\frac{1 - s^2}{k^2}\right]^3\right) & \text{if } |s| < k, \\
    \frac{k^2}{6} & \text{if } |s| > k,
\end{cases}
\]

respectively. Note that the parameter $k > 0$ determines the region of transition between low and high gradients, thereby providing a separation of homogeneous (flat) regions and edges.

\(^4\) Evolution of the Step edge synthetic image mesh under different schemes is available as movies in the supplementary material.
Fig. 3 One-dimensional signal (line) taken across the middle of synthetic Step image in Fig. 2. The proposed adaptive scheme provides smoothing with edge preservation when compared with Tikhonov (over-smoothing) and TV (staircasing) regularization approaches.

(jumps). To study the fine properties of the Huber and Tukey regularization functions on the final restoration result, we consider a simple 1D signal which consists of a sharp peak-like edge and ramp edges along with flat regions.

- The Huber function $\varphi_H (4)$ is convex and has a linear response to noisy pixels (outliers) and strongly depends on the parameter $k$ for that. Figure 4 shows how the dependence on $k$ affects the final restoration strongly on a 1-D noisy signal ($\sigma_n = 1$) with two types of discontinuities given in Fig. 4a. If $k$ is smaller ($k = 3$) much of the noise remains and there is no smoothing, whereas if $k$ is bigger ($k = 10$) then smoothing occurs indiscriminately (Fig. 4b) and edges are blurred like the quadratic regularization (equivalent to Gaussian filtering) case. From this, we can conclude that setting a small value for the threshold $k$ captures edges as well as outliers corresponding to noise. Since we do not a priori know when and where $|\nabla u|$ jumps (edges) occur and the input image $u_0$ is corrupted with additive noise, there is a need to include an image adaptive measurement for choosing $k$.

- The Tukey function $\varphi_T (5)$ is non-convex and gives constant response to outliers (Fig. 1a). This can be a drawback in a scenario where the edges and outliers have same high-frequency content. To illustrate, we consider the same 1-D signal but slightly perturb to obtain another 1-D signal copy, see Fig. 5a. The two original signals are of same type but of different amplitude. After adding additive Gaussian noise of strength $\sigma_n = 1$ to both signals (Fig. 5b), we use Tukey function (5)-based minimization scheme (1) and obtain the results Fig. 5c. This shows that a even slight perturbation of the input signal can produce a very different output due to instability associated with the non-convexity nature of the regularization function.
Motivated by the above arguments and to avoid both the over-under smoothing, and local minima issues, in this paper we use the following regularization function [38]:

$$\varphi_S(s) = \begin{cases} 
  as^2 & \text{if } |s| < k, \\
  bs^2 + c|s| & \text{if } |s| > k,
\end{cases} \quad (6)$$

where the free parameters $1 \gg b > 0$ are chosen so as to make the function lie between quadratic case of Tikhonov and Huber’s min–max function, see Fig. 1a. This also makes the function to be in between both $\varphi_H$ and $\varphi_T$ and strictly convex. Thus, the energy minimization of $E$ in (1) is well posed. For completeness, we outline the theorem here. We denote the set of all bounded variation functions [24] from $\Omega \rightarrow \mathbb{R}^m$ by $BV(\Omega; \mathbb{R}^m)$ where $\Omega$ is the image domain, usually a rectangle in $\mathbb{R}^2$.

**Theorem 1** (Well-posedness) Let $u_0 \in BV(\Omega; \mathbb{R}^m)$ be the initial image. If the regularization function $\varphi(\cdot)$ is strictly convex, then the energy minimization problem $E(u)$ in (1) is well posed in $BV(\Omega; \mathbb{R}^m)$. Moreover, the maximum and minimum principle holds true.
Proof From (1), the first term \((u - I)^2\) is strictly convex in \(u\). Thus, if \(\varphi\) is also strictly convex then the well-posedness and maximum–minimum principle follow from [38]. \qed

Remark 2 Note that if \(b \to -1\) in (6) we approach the Tukey’s bisquare \(\phi_T\) function continuously but we lose the convexity, see Fig. 1a. Hence, we stick to \(0 < b < 1\) and use an adaptive selection of the threshold parameter \(k\), see Sect. 4.1.

Further, to reduce the dependence on the threshold \(k\) we use the following adaptive edge indicator function:

\[
W(x) = \frac{1}{1 + K \left| G_\rho \ast \nabla u_0 \right|^2},
\]

where \(K > 0\) and \(G_\rho\) is the Gaussian kernel with width \(\rho > 0\), \(G_\rho = (2\pi \sigma)^{-1} \exp\left(-|x|^2/2\rho\right)\) and \(\ast\) is the convolution operation. Theorem 1 guarantees that the regularization function \(\phi_S\) in (6) with the continuous variational minimization problem (1) is well posed in the sense of Hadamard. Note that the data fidelity or the lagrangian parameter \(\lambda\) in (1) can be made adaptive so that when we use an iterative scheme as in Sect. 3 it is made smaller as the iteration increases. This helps in reducing the regularization as the noise level decreases. An adaptive way to select \(\lambda\) in the numerical simulations is given in Sect. 4. As we will see in denoising examples, this makes our scheme to adjust according to the image information at the current iteration and gives better restoration results overall. If the parameter \(\lambda\) is data adaptive, i.e., \(\lambda = \lambda(u, \nabla u)\) (see Eq. 1), then the above theorem holds true if \(\lambda \in C^\infty(\Omega)\) and continuous; in our case, it is true, see Eq. (19) below.

3 A convergent finite difference scheme

3.1 Discretized functional

The digital image has a natural rectangular grid and without loss of generality we assume that the image \(u: \Omega \subset \mathbb{R}^2 \to \mathbb{R}\) has size \(N \times N\). Then, the domain \(\Omega\) is divided into \(N^2\) subdomains of side length \(h\). We let the vertices \(\{v_{i,j}: 1 \leq i, j \leq N\}\) so that the \((i, j)\)th square subdomains are \(\Omega_{i,j} = v_{i,j} + [-h/2, h/2]^2\). Then, we use the following finite difference approximations for the gradients:

\[
\nabla^+_x u_{ij} = \begin{cases} 0, & u_{1j} = 0, \ u_{Nj} = 0 \\ \frac{u_{i+1,j} - u_{ij}}{h}, & i, j = 1, \ldots, N - 1, \end{cases}
\]

\[
\nabla^-_x u_{ij} = \begin{cases} 0, & u_{ij} = 0, \ u_{Nj} = 0 \\ \frac{u_{ij} - u_{i-1,j}}{h}, & i, j = 1, \ldots, N - 1, \end{cases}
\]

and similarly for the \(y\)-direction gradients \(\nabla^+_y\), \(\nabla^-_y\), to obtain the forward and backward discrete gradients \(\nabla_+ = (\nabla^+_x, \nabla^+_y)\), and \(\nabla_- = (\nabla^-_x, \nabla^-_y)\) respectively. Then, the discretized functional over \(\mathbb{R}^{N \times N}\) is written as:

\[
E_h(u) = \sum_{1 \leq i,j \leq N} \phi_h(W_{ij}(\nabla u)_{i,j}) + \frac{h^2\lambda}{2} \sum_{1 \leq i,j \leq N} (u_{i,j} - (D_hu_0)_{i,j})^2
\]

\(\odot\) Springer
where $D_h$ is the discrete operator applied to the input image $u_0$. The discrete regularizer in the above equations is

$$\phi_h(W_{ij}(\nabla u)_{i,j}) = \frac{W_{ij}h^2}{2} \times \begin{cases} a(|\nabla_+ u_{i,j}|^2 + |\nabla_- u_{i,j}|^2) & \text{if } |s| < k, \\ b(|\nabla_+ u_{i,j}|^2 + |\nabla_- u_{i,j}|^2) + c(|\nabla_+ u_{i,j}| + |\nabla_- u_{i,j}|) & \text{if } |s| > k, \end{cases}$$

(10)

with $W_{ij}$ the discrete version of the edge indicator function (7) using the discrete gradient and the discrete window-based Gaussian function.

### 3.2 Split Bregman method

We recall the split Bregman method to solve the discrete energy functional in Eq. (9). We sketch the main parts of the algorithm here and we refer to [26] and [27] for the general treatment on Split Bregman approach. This is a very fast scheme, faster than other numerical schemes reported in the literature, as we will see for example in image denoising tasks, Sect. 4. An auxiliary variable $d \leftarrow \nabla u$ is introduced in the model with a quadratic $L^2$ penalty function. That is to solve the TV minimization,

$$\min_u TV(u) = \int_{\Omega} |\nabla u| \, dx,$$

(11)

we consider the following unconstrained minimization problem:

$$\min_{u,d} \left\{ |d| + \frac{\lambda}{2} \|d - \nabla u\|_{L^2(\Omega)}^2 \right\}.$$

(12)

The above problem is solved using an alternating minimization scheme, which includes the addition of a vector $e$, inside the quadratic functional. That is, the algorithm reduces to the following sequence of unconstrained problems:

$$(u^{t+1}, d^{t+1}) = \arg \min_{0 \leq u \leq 1, d} \left\{ |d| + \frac{\lambda}{2} \|d - \nabla u - b^t\|_{L^2(\Omega)}^2 \right\}$$

(13)

$$e^{t+1} = e^t + \nabla u^t - d^t$$

(14)

First a minimization with respect to $u$ is performed using a Gauss–Seidel method. Next, a minimization with respect to $d$ is done using a shrinkage method. Finally, the vector $e$ is updated using (14). The following steps summarize the algorithm:

1. Initialize $d^0, e^0 \in (L^2(\Omega))^n$
2. For $t \geq 1$
   
   (a) $(\mu I - \lambda \Delta)u^{t+1} = \mu u_0 - \nabla^T (d^t - e^t)$
   
   (b) Compute

$$d^{k+1} = \text{shrink} \left( \nabla u^t + e^t, \frac{1}{\lambda} \right)$$

3. $e^{t+1} = e^t + \nabla u^{t+1} - d^{t+1}$

The shrinkage operation is given by

$$\text{shrink}(x, \gamma) = \frac{x}{|x|} \max(|x| - \gamma, 0).$$
It can be shown that this algorithm converges very quickly even when an approximate solution is used in Eq. (13). The split Bregman algorithm for solving our functional (9) can similarly be derived. Note that in our case the shrinkage becomes
\[ a_{k+1} = \text{shrink} \left( \nabla u' + e', \frac{W}{\lambda} \right), \] (15)
where \( W \) is the adaptive edge indicator function given in Eq. (7).

### 3.3 Convergence

The digital image \( u \in \mathbb{R}^{N \times N} \) is interpolated using continuous piecewise linear functions on \( \Omega \):
\[ \mathcal{P}_h U(x) = \sum_{1 \leq i, j \leq N} U_{i,j} \ell_{i,j}(x) \]
with \( \ell_{i,j} : \Omega \rightarrow \mathbb{R} \) and \( \ell_{i,j}(v_{i,j}) = 1, \ell_{i,j}(v) = 0, \omega \in \{v_{i,j}\} \). Similarly, we define piecewise constant extension \( C_h U(x) = U_{i,j} \) for \( x \in \text{int}(\Omega_{i,j}) \), and the sampling operator
\[ Q_h U(x) = \frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} U(y) \, dy, \quad \text{for } x \in \text{int}(\Omega_{i,j}). \]

To prove the convergence of the interpolated function to the continuous solution, we first introduce some basic notations. In what follows, we use the standard notations on Lebesgue \( L^p(\Omega) \) (\( 1 \leq p \leq \infty \)) and functions of bounded variation \( BV(\Omega) \) spaces. We define the translation of a set and a function with vector \( \tau \in \mathbb{R}^2 \) as \( T^\tau \Omega = \{ x + \tau : x \in \Omega \} \), \( T^\tau \phi(x) = \phi(x + \tau) \) for \( x \in T^{-\tau} \Omega \), respectively. Let us recall the definition of \( p \)-modulus of continuity of order \( t > 0 \) for a function \( \phi \in L^p(\Omega) \), \( \omega(\phi, t)_p = \sup_{|\tau| \leq t} \| T^\tau \phi - \phi \|_{L^p(\Omega \cap T^{-\tau} \Omega)} \). Note that the modulus of continuity gives a quantitative account of the continuity property of \( L^p(\Omega) \) functions.

**Definition 1** (Weakly regular functions) Let \( \phi \in L^p(\Omega) \) and \( 0 < \mathcal{L} \leq 1 \). We say \( \phi \) is weakly regular (\( \mathcal{L} \)-Lipschitz) function if it satisfies the condition \( \sup_{0 < t < 1} \frac{\omega(\phi, t)}{t^{\mathcal{L}}} < \infty \).

The main convergence theorem is stated as follows.

**Theorem 2** (Convergence) Let \( u_0 \in L^\infty(\Omega) \), weakly regular (\( \mathcal{L} \)-Lipschitz, \( \mathcal{L} \in (0, 1] \)) and \( D_h u_0 \) be the discretization with respect to a uniform quadrangulation \( Q_h \). Let \( U \) be the minimizer of the discretized functional over \( \mathbb{R}^{N \times N} \),
\[ E_h(u) = \sum_{1 \leq i,j \leq N} \phi_h(W_{ij}(\nabla u_{i,j})) + \frac{h^2 \lambda}{2} \sum_{1 \leq i,j \leq N} (u_{i,j} - (D_h u_0)_{i,j})^2 \]
which is obtained using the split Bregman scheme in Sect. 3.2, and \( u \) be the minimizer of the continuous functional (1). Then
(i) The interpolated solution of the discrete model converges to the continuous solution,
\[ \| \mathcal{P}_h(U) - u \|_{L^2(\Omega)} \rightarrow 0 \text{ as } h \rightarrow 0. \]
(ii) \( E_h(\mathcal{P}_h(U)) \) converges to \( E(u) \) as \( h \rightarrow 0. \)

We derive some preliminary results required for proving the main theorem. We use a generic constant \( C \) which can change in line to line.
Lemma 1 (Bounds on solutions)

1. Continuous: Let \( \tilde{u} \in BV(\Omega; \mathbb{R}^m) \) be a solution of the energy minimization (1) with the adaptive regularization function (6). If \( u^* \in BV(\Omega; \mathbb{R}^m) \), then
   \[
   \| \tilde{u} - u^* \|_2^2 \leq \frac{2}{\lambda} |E(\tilde{u}) - E(u^*)| \tag{16}
   \]

2. Discrete: Let \( \tilde{U} \in \mathbb{R}^{N \times N} \) be the minimizer of the discretized functional \( E_h \) in (9). Then
   \[
   E(P_h(\tilde{U})) - E_h(\tilde{U}) \leq \frac{\lambda}{2} C \omega(u_0, h) \| u_0 \|_2 \tag{17}
   \]

Proof (1) The inequality follows from the fact that for the adaptive regularization (6)-based energy minimization functional \( E \) in Eq. (1) is \( L^2 \)-subdifferentiable.

(2) We first note that
   \[
   \| P_h Q_h u_0 - u_0 \|_2 \leq C \omega(u_0, h) \text{ and } \| P_h (\tilde{U} - Q_h u_0) \|_2 \leq 4 \| u_0 \|_2.
   \]

Then, the inequality (17) follows from
   \[
   \frac{2}{\lambda} (E(P_h(\tilde{U})) - E_h(\tilde{U})) \leq \| P_h Q_h u_0 - u_0 \|_2 \| P_h Q_h u_0 - u_0 \|_2 + 2 \| P_h (\tilde{U} - Q_h u_0) \|_2^2.
   \]

Lemma 2 (Convolution bound) Let \( \tilde{U} \in \mathbb{R}^{N \times N} \) be the minimizer of the discretized functional \( E_h \) in (9). Let \( u_\epsilon = G_\epsilon * u \) be the mollified extension of the image function \( u \in BV(\Omega) \) to \( u \in BV(\mathbb{R}^2) \). Then
   \[
   E_h(\tilde{U}) - E(u_\epsilon) \leq C \| u_0 \|_\infty^2 + O(h/\epsilon^2).
   \]

Proof First note that
   \[
   E_h(\tilde{U}) \leq E_h(u_\epsilon)
   \]
   \[
   \leq \int_{\Omega} |\nabla P_h u_\epsilon|^2 \, dx + \frac{\lambda}{2} \sum_{1 \leq i, j \leq N} h^2 \| (u_\epsilon)_{i,j} - (Q_h u_0)_{i,j} \|^2
   \]
   and
   \[
   \| P_h u_\epsilon - u_\epsilon \|_{W^{1,2}} \leq C h \sum_{|\alpha|=2} \| D^\alpha u_\epsilon \|_2 \leq C h/\epsilon^2
   \]

Then, the inequality follows from
   \[
   \sum_{1 \leq i, j \leq N} h^2 \| Q_h(u_\epsilon - u_0)_{i,j} \|^2 \leq \| u_\epsilon - u_0 \|_2^2 + C \| u_0 \|_\infty^2.
   \]
   \[
   \sum_{1 \leq i, j \leq N} h^2 \| (u_\epsilon)_{i,j} - (Q_h u_0)_{i,j} \|^2 \leq \sum_{1 \leq i, j \leq N} h^2 \| (Q_h u_\epsilon - Q_h u_0)_{i,j} \|^2 + C O(h/\epsilon^2).
   \]
Fig. 6 Original Cameraman gray scale test image of size 256 × 256 used in our experiments and its edge map computed with gradients. a Noise free image. b Gaussian noise added image, σn = 20. c Gradient image, inverted (1 − |∇u0|) for better visualization. d Adaptive λ from Eqn. (19) at iteration t = 100. Notice that the edges are preserved, whereas the noise is removed in the homogeneous regions as the iterations are increased.

Proof of Theorem 2: Let ε > 0 and h ≤ 1. From Eq. (16),

\[ \|P_h(U) - u\|_2^2 \leq \frac{2}{\lambda} \{E(P_h U) - E(u)\} \]

\[ \leq \frac{2}{\lambda} [ (E(P_h U) - E_h(U)) + (E_h(U) - E(u))] \]

Using Lemmas 1, 2, respectively, for the two difference terms, we obtain

\[ \|P_h(U) - u\|_2^2 \leq \omega(u_0, h)_2 [\omega(u_0, h) + C \|u_0\|_2] + \frac{32h}{\lambda} \|u_0\|_\infty^2 + \frac{2Ch}{\lambda} + \frac{2}{\lambda} \{E(u_\epsilon) - E(u)\} \]  

(18)

Let ε = h^{1/(2ζ+1)} and since u_0 is weakly regular \( \omega(u_0, h) \leq O(h^\zeta) \), the above inequality becomes

\[ \|P_h(U) - u\|_2^2 \leq \frac{2}{\lambda} \{E(u_\epsilon) - E(u)\} + Ch^{\zeta/(2\zeta+1)} \]

Since \( E(u_\epsilon) - E(u) \to 0 \) as \( \epsilon \to 0 \) we have the result.  

4 Experimental results and discussion

4.1 Parameters

We set the step size \( h = \delta t = 0.20 \), \( a = 1 \), and parameters in our regularization function in (6) to \( b = 0.05 \), \( \rho = 2 \), and the thresholding parameter \( k \) is determined using the mean absolute deviation (MAD) from robust statistics \[43\],

\[ k = 1.4826 \times \text{MAD}(\nabla u) \]

\[ = 1.4826 \times \text{median}_u[|\nabla u - \text{median}(|\nabla u|)|] \]

where the constant is derived from the fact that the MAD of a zero-mean normal distribution with unit variance is \( 0.6745 = 1/1.4826 \). For the discrete functional (9), the parameter \( k \) is computed using the gradient magnitude \( |\nabla u| \) for which we used the same finite difference approximations introduced before, see Eq. (8). All the test images are normalized to the range [0, 1].
We further introduce an iteration and pixel adaptive $\lambda_{i,j}^{(t)}$ using the gradient information at iteration $(t-1)$ via

$$
\lambda_{i,j}^{(t)} := \frac{1}{\epsilon^2 + \sqrt{(u_{i+1,j}^{(t-1)} - u_{i,j}^{(t-1)})^2 + (u_{i,j+1}^{(t-1)} - u_{i,j}^{(t-1)})^2}}
$$

(19)

where $\epsilon^2 = 10^{-6}$ is added to avoid numerical instabilities. Note that $\lambda_{i,j}^{(t)} \in [0, 1]$ reduces the influence of the regularization term at edges and makes the scheme an image adaptive method. This also reduces the dependence on the threshold $k$ to decide upon the outliers part [compare this with Huber’s min–max function (4) and Fig. 4]. Since Theorem 1 implies stability we are guaranteed of a good reconstruction even if the input is perturbed significantly [compare this with Tukey bisquare function (5) and Fig. 5]. In Fig. 7, we consider the same 1-D signal as shown earlier in Fig. 4a. The restoration result exhibits strong smoothing property of our adaptive regularization function with edge preservation. Exact locations of the true discontinuities are preserved and noise is completely removed in homogenous regions.

Figure 6 shows the Cameraman gray-scale $256 \times 256$ size image used in our later comparison results. We add Gaussian white noise of standard deviation $\sigma_n = 20$ and mean zero. Figure 6b, c shows the gradient image [computed using the formulae (8)] from the initial noisy image $|\nabla u_0|$ and adaptive $\lambda$ parameter computed using Eq. (19) at iteration 100 showing the improvement in the edge map.

4.2 Restoration results

In Fig. 8, we restore three real images, original color Movie still (film grain noise, medium granularity), a Kid image taken by a mobile camera (2 mega-pixels, image contains unknown amount of shot noise), and Goat an old gray-scale photograph (noise type unknown), respectively. Note that for (RGB) color images we use the scheme (1) for each of the channels red, green, and blue and combine the final restoration result. The restored results in Fig. 8b exhibit marked improvements. Note that fine texture details are lost in Fig. 8b (background wall,

---

5 Using MATLAB command `imnoise(u0,'gaussian',0,\sigma_n)`
Fig. 8  Restoration by our adaptive regularization scheme on some real images with unknown noise strength. 
*Top row* original images. *Bottom row* our adaptive regularization scheme results

Fig. 9  *Ducks* color image $300 \times 200 \times 3$ restoration result. 
* a* Original RGB image. 
* b* Noisy image, $\sigma_n = 20$, PSNR $= 12.56$ dB. 
* c* Restored by our method, PSNR $= 23.15$ dB. 
* d* Edges computed in all three channels (RGB) using the Canny edge detector.

goat, hair and shirt), we may need to include further statistical information about textures in our scheme. Apart from this, our scheme overall performs well and has strong edge preserving smoothing properties. The strong smoothing nature of our adaptive regularization (6) can be seen in another piecewise smooth image shown in Fig. 9a. This *Ducks* color image consists of flat background with strong curved edges and the result in Fig. 9c indicates the local smoothing due to Gaussian filtering effect in regions where $|\nabla u| < k$ and edge preserving TV filtering in other areas. Figure 9d shows the Canny edge map computed from the three color channels.6

4.3 Comparison results

Figure 10 shows a comparison of restoration results for the *Peppers* color image. As can be seen from the method noise and a contour maps, our adaptive regularization scheme

6 Using MATLAB command `edge(u0,'canny')`. The edges are computed from each of the red, green, and blue channels (images) and the final result is shown by combining them again into a color image. Note that the Canny edge detector employs non-maximal suppression to avoid small scale edges.
Fig. 10 Restoration result for a color Peppers color image (size $512 \times 512 \times 3$) with our adaptive regularization scheme: a original image, b–d Results of Huber, Tukey and our adaptive regularization function-based scheme with tolerance $\text{tol} = 10^{-6}$, respectively. e Gaussian noise ($\sigma_n = 20$) corrupted image. f–h Residual noise/method noise image, $|u - U|^2$. i Contour map of the noisy image. j–l Contour map showing the restoration on level lines.

Fig. 11 Comparison of our proposed scheme with $\phi_S$ in (6) with Huber’s $\phi_H$ (4) and Tukey’s $\phi_T$: (5) for the Peppers color image in Fig. 10. a Number of iterations ($t$) vs mean error (ME). b Noise level ($\sigma_n$) vs peak signal-to-noise ratio (PSNR) for different noise levels.
outperforms other schemes in terms of noise removal and edge preservation. The level lines are smoothed without reducing their edginess and flat regions are preserved without staircasing artifacts. Figure 11 shows the ME and PSNR comparisons illustrating the versatility of our adaptive scheme (1) with the proposed regularization function (6) against other functions. Also note that the ME error curve for our method outperforms Huber and Tukey functions-based regularization and quickly converges to a desired solution (usually \( t = 50 \) is sufficient). Our function (6) is robust when compared to the other two classical functions as can be seen from the PSNR comparison Fig. 11b as well. The topmost PSNR curve indicates that the scheme proposed in this paper surpasses the other two when the noise level increases \( \sigma_n \rightarrow 25 \). Note that \( \sigma_n^2 > 400 \) is a high level noise and our scheme performs well in distinguishing between outliers corresponding to noise and true edges due to the adaptive nature of \( \lambda(t) \) (19).

We next provide comparison with primal–dual hybrid gradient (PDHG) [58], projected averaged gradient (Proj. Grad) [59], fast gradient projection (FGP) [2], alternating direction method of multipliers (ADMM) [25], and split Bregman (Split Breg.)-based schemes. The following error metrics are used to compare the convergence and performance of different algorithms for the discrete minimization Eq. (9).

- Relative duality gap:

  \[
  R(u, b) = \frac{E_{\text{Primal}}(u) - E_{\text{Dual}}(b)}{E_{\text{Dual}}(b)}, \tag{20}
  \]

  where \( E_{\text{Primal}}, E_{\text{Dual}} \) represent the primal and dual objective functions, respectively. This is used as stopping criteria for the iterative schemes.

- Peak signal-to-noise (PSNR) ratio,

  \[
  \text{PSNR} = 20 \log_{10} \left( \frac{255}{\sqrt{\sum_{1 \leq i, j \leq N} (u - u_0)^2}} \right) \text{dB} \tag{21}
  \]

  The higher the PSNR, the better is the restoration result.

- The mean error (ME):

  \[
  \text{ME}(u, I) := \frac{1}{MN} \sum_i \| u_i - I_i \|
  \]

  The mean error needs to be small for restored images.
Table 1 Comparison with primal–dual hybrid gradient (PDHG), projected averaged gradient (Proj. Grad), fast gradient projection (FGP), alternating direction method of multipliers (ADMM), and split Bregman-based scheme

| Algorithm  | tol = $10^{-2}$ | tol = $10^{-4}$ | tol = $10^{-6}$ |
|------------|-----------------|-----------------|-----------------|
| PDHG       | 14              | 70              | 310             |
| Proj. Grad. | 46              | 721             | 14,996          |
| FGP        | 24              | 179             | 1264            |
| ADMM       | 97              | 270             | 569             |
| Split Breg. | 10              | 28              | 55              |

Iterations required for denoising of the *Cameraman* image (256 × 256, noise level $\sigma_n = 20$) with different numerical schemes for the relative duality gap $R(u, b) \leq \text{tol}$

Bold indicates the best result

Table 2 Comparison of different algorithms in terms of noise level ($\sigma_n$) for the *Cameraman* gray scale image. The results are given in terms of best possible PSNR (computational time in seconds, maximum iterations)

| Noise | PDHG | Proj. Grad. | FGP | ADMM | Split Breg. |
|-------|------|-------------|-----|------|-------------|
| 15    | 21.61 (28s, 100) | 21.58 (30s, 85) | 20.21 (20s, 70) | 21.85 (24s, 73) | **25.40** (10s, 55) |
| 20    | 20.46 (28s, 86) | 20.29 (30s, 80) | 20.12 (20s, 80) | 20.05 (24s, 70) | **23.82** (10s, 67) |
| 25    | 17.01 (28s, 75) | 16.88 (30s, 80) | 16.26 (20s, 75) | 17.73 (24s, 70) | **17.92** (10s, 65) |
| 30    | 10.77 (28s, 90) | 11.71 (30s, 90) | 11.93 (20s, 73) | 11.05 (24s, 70) | **12.67** (10s, 62) |

Each scheme is terminated if the maximum number of iterations exceeded 500 or when the duality gap is less than $R(u, b) \leq 10^{-6}$

Bold indicates the best result

The first comparative example in Fig. 12 compares the restoration results for the noisy *Cameraman* gray scale image from Fig. 6b. As can be seen, adaptive Huber function performs better than the classical TV and Tikhonov schemes. Moreover, improvement in PSNR is $> 5 \text{ dB}$ (see Fig. 12d) in different noise levels indicating the success of our scheme in terms of noise removal. Table 1 shows the number of iterations taken by different optimization schemes for solving the discrete regularization scheme (9) with respect to the relative duality gap error (20) as stopping criteria. The split Bregman-based implementation outperforms all the other schemes by reducing the relative duality gap within very few iterations. Next, Table 2 provides a comparison of PSNR (time in seconds, maximum iterations) for different noise levels and for different optimization schemes for the noisy *Cameraman* image. The experiments were performed on a Mac Pro Laptop with 2.3GHz Intel Core i7 processor, 8Gb memory, and MATLAB R2012a was used for visualizations. The split Bregman minimization outperforms all the related schemes in terms of PSNR (dB) as well as in timing as can be seen from the table. Similar analysis for the image deblurring and deconvolution requires a delicate analysis of the boundary conditions [48] and is treated elsewhere. Other avenues of exploration are treating higher order models [31,56], multi-grid [49], graphical processing unit implementation [28], and FEM [32]-based schemes and their convergence analysis.

5 Conclusion

In this paper, we considered adaptive Huber type regularization function-based image restoration scheme. Using discrete split Bregman scheme, we proved the convergence to continuous
formulation. Experimental results on real images are given to illustrate the results presented. Compared with other schemes, the splitting-based scheme provides faster convergence as well as good restoration results. The scheme can be extended to handle multispectral images using inter-channel correlations [35,37,39] and this defines our future work.

References

1. Aubert G, Kornprobst P (2006) Mathematical problems in image processing: partial differential equation and calculus of variations. Springer-Verlag, New York
2. Beck A, Teboulle M (2009) Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans Image Process 18(18):2419–2434
3. Black MJ, Rangarajan A (1996) On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. Int J Comput Vis 19(1):57–91
4. Black MJ, Sapiro G, Marimont DH, Heeger D (1998) Robust anisotropic diffusion. IEEE Trans Image Process 7(3):421–432
5. Carlini E, Ferretti R (2012) A semi-Lagrangian approximation for the AMSS model of image processing. Appl Numer Math 73:16–32
6. Caselles V, Sapiro G, Chung DH (2000) Vector median filters, inf-sup operations, and coupled PDE’s: theoretical connections. J Math Imaging Vis 12(2):109–119
7. Chambolle A (1995) Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. SIAM J Appl Math 55(3):827–863
8. Chambolle A (1999) Finite-differences discretizations of the Mumford–Shah functional. M2AN Math Model Numer Anal 33(2):261–288
9. Chambolle A (2004) An algorithm for total variation minimization and applications. J Math Imaging Vis 20(1–2):89–97
10. Chambolle A, Levine S, Lucier BJ (2011) An upwind finite-difference method for total variation based image smoothing. SIAM J Imaging Sci 4(1):277–299
11. Chambolle A, Lions PL (1997) Image recovery via total variation minimization and related problems. Numer Math 76(2):167–188
12. Chambolle A, Pock T (2011) A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vis 40(1):120–145
13. Chan TF, Mulet P (1999) On the convergence of the lagged diffusivity fixed point method in total variation image restoration. SIAM J Numer Anal 36(2):354–367
14. Chan TF, Shen J (2005) Image processing and analysis: variational, PDE, wavelet, and stochastic methods. Society for Industrial and Applied Mathematics, Philadelphia
15. Chen K (2005) Adaptive smoothing via contextual and local discontinuities. IEEE Trans Pattern Anal Mach Intell 27(10):1552–1567
16. Chen Y, Levine S, Rao M (2006) Variable exponent, linear growth functionals in image restoration. SIAM J Appl Math 66(4):1383–1406
17. Chu CK, Glad IK, Godtliebsen F, Marron JS (1998) Edge-preserving smoothers for image processing. J Am Stat Assoc 93(442):526–541
18. de Araujo AF, Constantinou CE, Tavares JMRS (2014) New artificial life model for image enhancement. Expert Syst Appl 41(13):5892–5906
19. de Araujo AF, Constantinou CE, Tavares JMRS (2016) Smoothing of ultrasound images using a new selective average filter. Expert Syst Appl 60:96–106
20. Dobson DC, Vogel CR (1997) Convergence of an iterative method for total variation denoising. SIAM J Numer Anal 34(5):1779–1791
21. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6(6):721–741
22. Geman S, McClure D (1987) Statistical methods for tomographic image reconstruction. In: Proceedings of the 46th session of the ISI, bulletin of the ISI, vol 52, pp 22–26
23. Gersgorin SA (1930) Fehlerabschätzung für das differenzverfahren zur lösung partieller differentialgleichungen. J Angew Math Mech 10:373–382
24. Giusti F (1984) Minimal surfaces and functions of bounded variation. Birkhauser, Basel
25. Glowinski R, Marrocco A (1975) Sur lapproximation par elements finis dordre un, et la resolution par penalisation-dualite dune classe de problemes de dirichlet nonlineaires. Rev Francaise dAut Inf Rech Oper R(2):41–76
26. Goldstein T, Bresson X, Osher S (2010) Geometric applications of the split Bregman method: segmentation and surface reconstruction. J Sci Comput 45(1–3):272–293
27. Goldstein T, Osher S (2009) The split Bregman algorithm for L1 regularized problems. SIAM J Imaging Sci 2(2):323–343
28. Gulio CASJ, de Arruda HF, de Araujo AF, Sementille AC, Tavares JMRS (2016) Efficient parallelization on gpu of an image smoothing method based on a variational model. J Real-Time Image Proc 1–13. doi:10.1007/s11554-016-0623-x
29. Huber PJ (1981) Robust statistics. Wiley, New York
30. Jia R-Q, Zhao HQ, Zhao W (2009) Convergence analysis of the Bregman method for the variational model of image denoising. Appl Comput Harmon Anal 27(3):367–379
31. Jiang Q (2012) Correspondence between frame shrinkage and high-order nonlinear diffusion. Appl Numer Math 62(1):51–66
32. Kaccour J, Mikula K (1995) Solution of nonlinear diffusion appearing in image smoothing and edge detection. Appl Numer Math 17(1):47–59
33. Li SZ (1995) Markov field random modeling in computer vision. Springer, Berlin
34. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12(7):629–639
35. Prasath VBS (2011) Weighted Laplacian differences based multispectral anisotropic diffusion. In: IEEE international geoscience and remote sensing symposium (IGARSS), pp 4042–4045, Vancouver BC, Canada
36. Prasath VBS (2011) A well-posed multiscale regularization scheme for digital image denoising. Int J Appl Math Comput Sci 21(4):769–777
37. Prasath VBS, Moreno JC, Palaniappan K (2013) Color image denoising by chromatic edges based vector valued diffusion. Preprint, 2013. Available at http://arxiv.org/abs/1304.5587
38. Prasath VBS, Singh A (2010) A hybrid convex variational model for image restoration. Appl Math Comput 215(10):3655–3664
39. Prasath VBS, Singh A (2010) Multispectral image denoising by well-posed anisotropic diffusion scheme with channel coupling. Int J Remote Sens 31(8):2091–2099
40. Prasath VBS, Singh A (2010) Well-posed inhomogeneous nonlinear diffusion scheme for digital image denoising. J Appl Math 14 :763847. doi: 10.1155/2010/763847
41. Prasath VBS, Singh A (2012) An adaptive anisotropic diffusion scheme for image restoration and selective smoothing. Int J Image Graph 12(1):18
42. Prasath VBS, Vorotnikov D (2012) On a system of adaptive coupled PDES for image restoration. J Math Imaging Vis, Online First, 2012. Available at arXiv:1112.2904
43. Rey WJJ (1983) Introduction to robust and quasirobust statistical methods. Springer-Verlag, Berlin
44. Rivera M, Marroquin JL (2003) Efficient half-quadratic regularization with granularity control. Image Vis Comput 21(4):345–357
45. Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Phys D 60(1–4):259–268
46. Scherzer O, Grasmair M, Grossauer H, Haltmeier M, Lenzen F (2009) Variational methods in imaging. Springer-Verlag, New York
47. Setzer S (2011) Operator splittings, Bregman methods and frame shrinkage in image processing. Int J Comput Vis 92(3):265–280
48. Shi Y, Chang Q (2008) Acceleration methods for image restoration problem with different boundary conditions. Appl Numer Math 58(5):602–614
49. Spitaleri RM, March R, Arena D (2001) A multigrid finite-difference method for the solution of Euler equations of the variational image segmentation. Appl Numer Math 39(2):181–189
50. Széll E R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tapen M, Rother C (2008) A comparative study of energy minimization methods for Markov random fields with smoothness based priors. IEEE Trans Pattern Anal Mach Intell 30(6):1068–1080
51. Tikhonov AN, Aresenin VY (1997) Solutions of ill-posed problems. Wiley, New York
52. Tukey JW (1977) Exploratory data analysis. Addison-Wesley Publishers, Reading, MA, USA
53. Weickert J (1998) Anisotropic diffusion in image processing. B.G. Teubner-Verlag, Stuttgart
54. Weickert J, Romeny BMH, Viergever MA (1998) Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans Image Process 7(3):398–410
55. Weiss P, Blanc-Feraud L, Aubert G (2009) Efficient schemes for total variation minimization under constraints in image processing. SIAM J Sci Comput 31(3):2047–2080
56. Wu T-T, Yang Y-F, Pang Z-F (2012) A modified fixed-point iterative algorithm for image restoration using fourth-order PDE model. Appl Numer Math 62(2):79–90
57. Xu G (2013) Consistent approximations of several geometric differential operators and their convergence. Appl Numer Math 69:1–12
58. Zhu M, Chan TF (2008) An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical report 08–34, UCLA CAM
59. Zhu M, Wright SJ, Chan TF (2010) Duality-based algorithms for total-variation-regularized image restoration. Comput Optim Appl 47(3):377–400