On Convergent Finite Difference Schemes for Variational - PDE Based Image Processing*

V. B. Surya Prasath†   Juan C. Moreno‡

Abstract

We study an adaptive anisotropic Huber functional based image restoration scheme. By 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 [25], 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 illustrate that the best convergence rates are obtained for our algorithm.

Keywords: Image restoration, Adaptive denoising, Finite differences, Convergence, Huber functional.

1 Introduction

Variational and partial differential differential equations (PDEs) based schemes are popular in image and video processing problems. In particular 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 [32] as the starting point in PDE based image processing and the connections to variational and robust statistics has also been considered later [3, 17, 51]. We refer to the recent monographs [1, 44] for an overview of these methods.

---

†Corresponding author. Department of Computer Science, University of Missouri-Columbia, MO 65211 USA. E-mail: prasaths@missouri.edu

‡IT, Department of Computer Science, University of Beira Interior, 6201–001, Covilhã, Portugal. E-mail: jmoreno@ubi.pt.

Based on the smoothness or regularity assumptions on the true image, various regularization functions can be used. The Tikhonov regularization function [49] 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 [43]. Recently, there are efforts to combine both the L^2 and L^1 based fundtionals into one common minimization problem such as the Huber function [4, 36], inf-sup convolution [11, 6]. Adaptive versions of the variational - PDE models are gaining popularity [15, 16, 38, 39, 34, 40] 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 have been studied [7, 18, 52, 13, 8, 47, 27, 53, 54]. Convergence of finite differences for various PDEs is a classic area within numerical analysis[1] and is still an active area of research in application areas such as image processing [10, 30, 5, 55].

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. By using piecewise continuous linear functions along with the discrete energy we study the convergence of discrete minimizer to the continuous solution. To solve 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 [45] etc. Here we use the split Bregman method studied by Goldstein and Osher [25, 24] for computing the discrete energy minimizer as it is the fastest in terms of computational complexity and then prove a convergence result for the class of weakly regular images. 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 visualization are provided to support split Bregman based solution.

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 Section 5 concludes the paper.

---

[1]Semen Aronovich Geršgorin's work [21] 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.

(a) Regularizers $\varphi(s)$ (b) Diffusivities $g(s) = \varphi'(s)/2s$
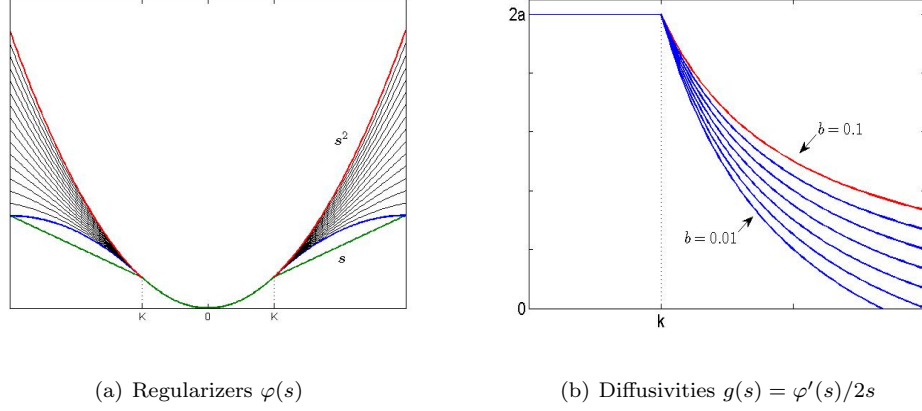
Figure 1: Regularization and diffusion functions. (a) The regularization function φ_S (6) lies between the quadratic curve s^2 and linear s when $|s| > k$ depending on the parameter $0 < b < 1$. (b) Corresponding diffusion functions g.

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],

$$\min_{u \in BV(\Omega)} E(u) = \int_\Omega \phi(\mathbf{x}, |\nabla u|) \, d\mathbf{x} + \frac{\lambda}{2} \int_\Omega |u - u_0|^2 \, d\mathbf{x} \tag{1}$$

The corresponding PDE can be written in term of the Euler-Lagrange equation,

$$\frac{\partial u}{\partial t} = div \left(\frac{\phi'(\mathbf{x}, |\nabla u|)\nabla u}{|\nabla u|} \right) - \lambda \, (u - u_0) \tag{2}$$

The adaptive discontinuity function $\phi(\cdot, |\nabla u(x)|) = W(\cdot) \times \varphi(|\nabla u(x)|)$ is chosen to be an even function. Note that the PDE in Eqn. (2) is a generalized Perona and Malik [32]

$$\frac{\partial u}{\partial t} = div \, (g(|\nabla u|)\nabla u) - \lambda \, (u - u_0), \tag{3}$$

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 choosing φ exists in the literature, see [19, 20, 31, 14] and [48] for a recent review. Note that under Gaussian noise assumption the data fidelity term (also called the likelihood term) in Eqn. (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

---

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

[3]Note we use the notation ∇ to denote the gradient and in the space of bounded variation functions BV it is infact a Radon measure and is understood in the sense of distributions. The equality $\int_\Omega |Du| = \int_\Omega |\nabla u| \, dx$ is true when $u \in W^{1,1}(\Omega)$.

51 non-convex [20, 31, 4, 42] with $\varphi(s) \sim s^2$ near 0 and asymptotically linear as $|s| \to +\infty$. This can cause

52 unstable behavior as the scheme can be plagued with local minima. In this paper, we concentrate on

53 convex regularization functions and study a stable and convergent scheme.

54 **Remark 1.** *There are other ways to incorporate adaptive weights inside the regularization function or*

55 *equivalently the diffusion coefficient. For example, as in adaptive total variation, i.e., with $\varphi(s) = s$,*

56 $\phi(x, |\nabla u(x)|) = |W(x) \cdot \nabla u(x)|$ *or in general* $\phi(\cdot, |\nabla u(x)|) = \varphi(W(\cdot) |\nabla u(x)|)$. *The main difference lies*

57 *in the way the regularization function φ is weighted anisotropically and the final results change according*

58 *to the formulation utilized. The main convergence result in Section 3 holds true for these type of adaptive*

59 *functions as well.*

Two of the most obvious choices for the regularization function φ are the Tikhonov or L^2-gradient
$\varphi(s) = s^2$ and the total variation (TV) or L^1-gradient $\varphi(s) = s$, see Figure 1(a). 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 Figure 3 a 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 the resultant
Fig. 2(c). On the other hand the TV regularization better preserves the edges but some additional
regions in the homogeneous parts can be enhances which is known as 'staircasing' artifact, see Fig. 2(d).
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 [26] and the Tukey's bisquare function [50] which are given by,

$$\varphi_H(s) = \begin{cases} s^2/2 & \text{if } |s| < k, \\ k(|s| - \dfrac{k}{2}) & \text{if } |s| > k, \end{cases} \tag{4}$$

$$\varphi_T(s) = \begin{cases} \frac{k^2}{6}\left(1 - [1 - s^2/k^2]^3\right) & \text{if } |s| < k, \\ \frac{k^2}{6} & \text{if } |s| > k, \end{cases} \tag{5}$$

60 respectively. Note that the parameter $k > 0$ determines the region of transition between low and high

61 gradients thereby providing a separation of homogeneous (flat) regions and edges (jumps). To study the

62 fine properties of the Huber and Tukey regularization functions on the final restoration result, we consider

63 a simple 1D signal which consist of a sharp peak like edge and ramp edges along with flat regions.

64 • The Huber function φ_H (4) is convex and has a linear response to noisy pixels (outliers) and is

65 strongly depends on the parameter k for that. Figure 4 shows how the dependence on k affects the

---

[4]Evolution of the *Step* edge synthetic image mesh under different schemes are available as movies in the supplementary
material.

(a) Original, $\sigma_n = 20$ (b) Noisy, $\sigma_n = 20$ (c) Tikhonov   (d) TV   (e) Our
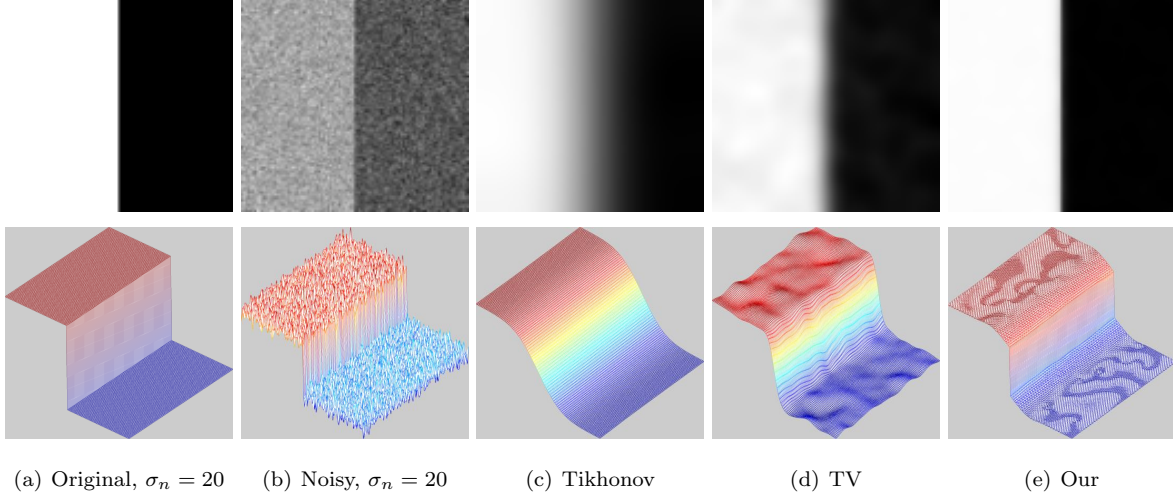
Figure 2: Synthetic *Step* image showing the effects of the choice of regularization function on the final restoration results. The L^2 - gradient scheme (Tikhonov) over-smoothes the edge whereas L^1 - gradient scheme (TV) though edge-preserving can introduce oscillations known as staircasing in homogeneous regions. An adaptive combination via (6) balances the smoothing along with edge preservation.
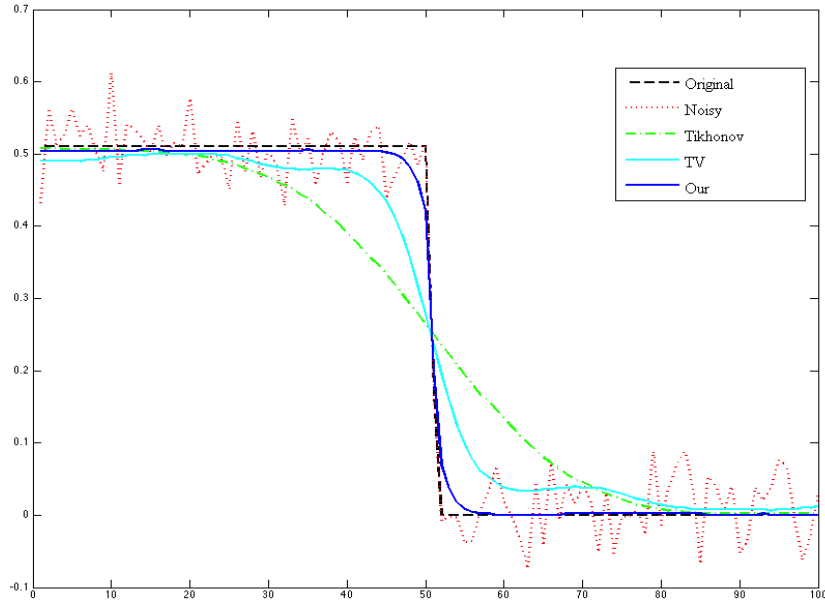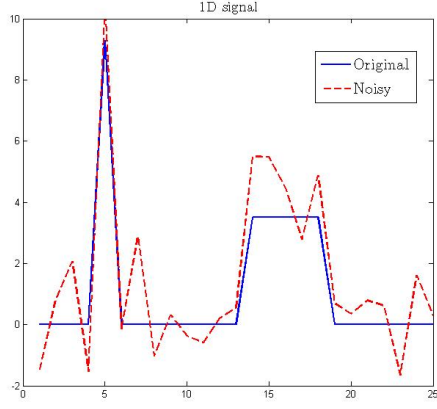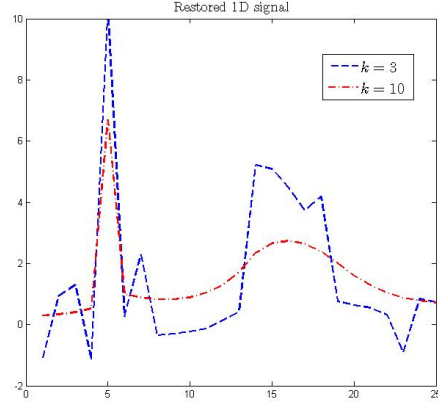

Figure 3: One dimensional signal (line) taken across the middle of synthetic *Step* image in Figure 2. The proposed adaptive scheme provides smoothing with edge preservation when compared with Tikhonov (over-smoothing) and TV (staircasing) regularization approaches.
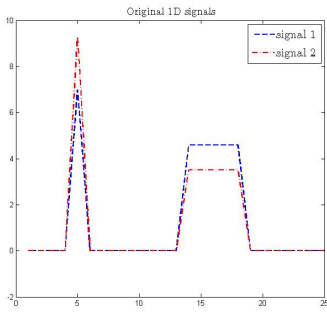
5

(a) Original & noisy signals

(b) Huber Restorations

Figure 4: Dependence on the parameter k for the Huber function φ_H give in (4). (a) Original signal with an impulse edge at 5 and a step edge in the range $[14-18]$ and additive Gaussian noise added (unit variance) (b) Restoration using the variational minimization (1) with Huber φ_H in (4) with $k = 3$ and $k = 10$.

(a) Original signals
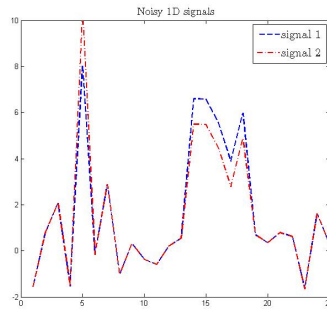
(b) Noisy signals
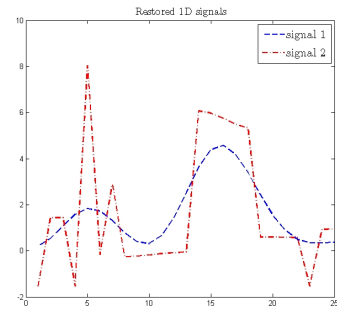
(c) Tukey Restorations

Figure 5: Instability using Tukey function given in φ_T (5). (a) Two perturbed signals with different magnitudes (b) Additive Gaussian noise added (unit variance) signals (c) Restoration using the minimization (1) with Tukey φ_T in (5).

final restoration strongly on a 1-D noisy signal ($\sigma_n = 1$) with two type of discontinuities given in Fig. 4(a). 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. 4(b)) 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.

- On the other hand the Tukey function φ_T (5) is non-convex and gives constant response to outliers (Fig. 1)(a), 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 perturb slightly to obtain another 1-D signal copy, see Fig. 5(a). 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. 5(b)) we use Tukey function (5) based minimization scheme (1) and obtain the results Fig. 5(c). 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 [36],

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

where the free parameters $1 \gg b > 0$ is chosen so as to make the function lie between quadratic case of Tikhonov and Huber's min-max function, see Fig. 1(a). This also makes the function to be in between both φ_H and φ_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 the set of all bounded variation functions [22] from $\Omega \to \mathbb{R}^m$ by $BV(\Omega; \mathbb{R}^m)$ where Ω 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 φ is also strictly convex then the well-posedness and maximum - minimum principle follows from [36]. □

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

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

$$W(x) = \frac{1}{1 + K \left| G_\rho \star \nabla u_0 \right|^2}, \tag{7}$$

where $K > 0$ and G_ρ is the Gaussian kernel with width $\rho > 0$, $G_\rho = (2\pi\sigma)^{-1} exp(-(|\mathbf{x}|^2/2\rho))$ and $\star$ is the convolution operation. Theorem 1 guarantees that the regularization function ϕ_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 λ in (1) can be made adaptive so that when we use an iterative scheme as in Section 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 λ in the numerical simulations is given in Section 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 λ is data adaptive, i.e., $\lambda = \lambda(u, \nabla u)$ (see Eqn. (1)) then the above theorem holds true if $\lambda \in C^\infty(\Omega)$ and continuous, in our case it is true, see Eqn (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 $\bar{\Omega}$ 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,

$$\nabla_+^x u_{ij} = \begin{cases} 0 & u_{1j} = 0, \ u_{Nj} = 0 \\ \frac{u_{i+1,j} - u_{ij}}{h} & i, j = 1, \ldots, N-1, \end{cases} \qquad \nabla_-^x u_{ij} = \begin{cases} 0 & u_{1j} = 0, \ u_{Nj} = 0 \\ \frac{u_{ij} - u_{i-1,j}}{h} & i, j = 1, \ldots, N-1. \end{cases} \tag{8}$$

and similarly for the y-direction gradients ∇_+^y, ∇_-^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,

$$E_h(u) = \sum_{1 \leq i,j \leq N} \phi_h(W_{ij}(\nabla u)_{i,j}) + \frac{h^2 \lambda}{2} \sum_{1 \leq i,j \leq N} (u_{i,j} - (D_h u_0)_{i,j})^2 \tag{9}$$

where D_h is the discrete operator applied to the input image u_0. The discrete regularizer in the above equations is,

$$\phi_h(W_{ij}(\nabla u)_{i,j}) = \frac{W_{ij} h^2}{2} \times \begin{cases} a\left(|\nabla_+ u_{i,j}|^2 + |\nabla_- u_{i,j}|^2\right) & \text{if } |s| < k, \\ b\left(|\nabla_+ u_{i,j}|^2 + |\nabla_- u_{i,j}|^2\right) + c\left(|\nabla_+ u_{i,j}| + |\nabla_- u_{i,j}|\right) & \text{if } |s| > k, \end{cases} \tag{10}$$

8

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

3.2 Split Bregman method

123 We recall the split Bregman method to solve the discrete energy functional in Eqn. (9). We sketch the
124 main parts of the algorithm here and we refer to [24] and [25] for the general treatment on Split Bregman
125 approach. This is a very fast scheme, faster than other numerical schemes reported in the literature, as
126 we will see for example in image denoising tasks, Section 4. An auxiliary variable $\vec{d} \leftarrow \nabla u$ is introduced
127 in the model with a quadratic L^2 penalty function. That is to solve the TV minimization,

$$\min_u TV(u) = \int_\Omega |\nabla u| \ dx, \tag{11}$$

128 we consider the following unconstrained minimization problem,

$$\min_{u,\vec{d}} \left\{ |\vec{d}| + \frac{\lambda}{2} \left\| \vec{d} - \nabla u \right\|_{L^2(\Omega)}^2 \right\}. \tag{12}$$

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

$$(u^{t+1}, \vec{d}^{t+1}) = arg \min_{0 \le u \le 1, \ \vec{d}} |\vec{d}| + \frac{\lambda}{2} \left\| \vec{d} - \nabla u - \vec{b}^t \right\|_{L^2(\Omega)}^2 \tag{13}$$

$$\vec{e}^{t+1} = \vec{e}^t + \nabla u^t - \vec{d}^t \tag{14}$$

129 First a minimization with respect to u is performed using a Gauss-Seidel method. Next a minimization
130 with respect to $\vec{d}$ is done using a shrinkage method. Finally, the vector $\vec{e}$ is updated using (14). The
131 following steps summarize the algorithm,

132 1. Initialize $d^0, e^0 \in (L^2(\Omega))^n$

133 2. For $t \ge 1$

134   (a) $(\mu I - \lambda \Delta) u^{t+1} = \mu u_0 - \nabla^T (d^t - e^t)$

135   (b) Compute

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

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

137 The shrinkage operation is given by,

$$shrink(x, \gamma) = \frac{x}{|x|} * \max \left(|x| - \gamma, 0 \right).$$

9

It can be shown that this algorithm converges very quickly even when an approximate solution is used in Eqn. (13). The split Bregman algorithm for solving our functional (9) can similarly be derived. Note that in our case the shrinkage becomes

$$d^{k+1} = shrink\left(\nabla u^t + e^t, \frac{W}{\lambda}\right), \tag{15}$$

where W is the adaptive edge indicator function given in Eqn. (7).

3.3 Convergence

The digital image $u \in \mathbb{R}^{N \times N}$ is interpolated using continuous piecewise linear functions on Ω,

$$\mathcal{P}_h U(x) = \sum_{1 \le i,j \le N} U_{i,j} \ell_{i,j}(x)$$

with $\ell_{i,j} : \Omega \to \mathbb{R}$ and $\ell_{i,j}(v_{i,j}) = 1$, $\ell_{i,j}(v) = 0$, $\omega \in \{v_{i,j}\}^c$. Similarly we define piecewise constant extension $\mathcal{C}_h U(x) = U_{i,j}$ for $x \in int(\Omega_{i,j})$, and the sampling operator

$$\mathcal{Q}_h U(x) = \frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} U(y)\, dy, \quad \text{for } x \in int(\Omega_{i,j}).$$

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 $L^p(\Omega)$ $(1 \le p \le \infty)$ and functions of bounded variation $BV(\Omega)$ spaces. We define the translation of a set and 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 $\phi \in L^p(\Omega)$, $\omega(\phi, t)_p = \sup_{|\tau| \le t} \|T^\tau \phi - \phi\|_{L^p(\Omega \cap T^{-\tau}\Omega)}$. 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 $0 < \mathcal{L} \le 1$. We say ϕ 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 $\mathbb{R}^{N \times N}$,*

$$E_h(u) = \sum_{1 \le i,j \le N} \phi_h(W_{ij}(\nabla u_{i,j})) + \frac{h^2 \lambda}{2} \sum_{1 \le i,j \le N} (u_{i,j} - (D_h u_0)_{i,j})^2$$

which is obtained using the split Bregman scheme in Section 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,*

$$\|\mathcal{P}_h(U) - u\|_{L^2(\Omega)} \to 0 \quad as \quad h \to 0.$$

(ii) $E_h(\mathcal{P}_h(U))$ *converges to* $E(u)$ *as* $h \to 0$.

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*

$$E(\mathcal{P}_h(\tilde{U})) - E_h(\tilde{U}) \leq \frac{\lambda}{2} C\omega(u_0, h)_2 [C\omega(u_0, h)_2 + 8 \|u_0\|_2] \tag{17}$$

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

(2) We first note that

$$\|\mathcal{P}_h\mathcal{Q}_h u_0 - u_0\|_2 \leq C\omega(u_0, h)_2 \quad \text{and} \quad \left\|\mathcal{P}_h(\tilde{U} - \mathcal{Q}_h u_0)\right\|_2^2 \leq 4 \|u_0\|_2.$$

Then the inequality (17) follows from,

$$\frac{2}{\lambda}(E(\mathcal{P}_h(\tilde{U})) - E_h(\tilde{U})) \leq \|\mathcal{P}_h\mathcal{Q}_h u_0 - u_0\|_2 \left\{ \|\mathcal{P}_h\mathcal{Q}_h u_0 - u_0\|_2 + 2 \left\|\mathcal{P}_h(\tilde{U} - \mathcal{Q}_h u_0)\right\|_2^2 \right\}.$$

$\square$

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* $u_\epsilon = G_\epsilon \star u$ *be the mollified extension of the image function* $u \in BV(\Omega)$ *to* $u \in BV(\mathbb{R}^2)$*. Then*

$$E_h(\tilde{U}) - E(u_\epsilon) \leq C \|u_0\|_\infty^2 + \mathcal{O}(h/\epsilon^2).$$

Proof. First note that

$$E_h(\tilde{U}) \leq E_h(u_\epsilon)$$
$$\leq \int_\Omega |\nabla \mathcal{P}_h u_\epsilon|^2 \, dx + \frac{\lambda}{2} \sum_{1 \leq i,j \leq N} h^2 |(u_\epsilon)_{i,j} - (\mathcal{Q}_h u_0)_{i,j}|^2$$

and

$$\|\mathcal{P}_h u_\epsilon - u_\epsilon\|_{W^{1,2}} \leq Ch \sum_{|\alpha|=2} \|D^\alpha u_\epsilon\|_2 \leq Ch/\epsilon^2$$

11

Then the inequality follows from,

$$\sum_{1 \leq i,j \leq N} h^2 \left| \mathcal{Q}_h (u\epsilon - u_0)_{i,j} \right|^2 \leq \|u_\epsilon - u_0\|_2^2 + C \|u_0\|_\infty^2 .$$

and

$$\sum_{1 \leq i,j \leq N} h^2 \left| (u_\epsilon)_{i,j} - (\mathcal{Q}_h u_0)_{i,j} \right|^2 \leq \sum_{1 \leq i,j \leq N} h^2 \left| (\mathcal{Q}_h u_\epsilon - \mathcal{Q}_h u_0)_{i,j} \right|^2 + C\mathcal{O}(h/\epsilon^2).$$

$\square$

Proof of Theorem 2:

Let $\epsilon > 0$ and $h \leq 1$. From Eqn. (16),

$$\|\mathcal{P}_h(U) - u\|_2^2 \leq \frac{2}{\lambda} \{E(\mathcal{P}_h U) - E(u)\}$$
$$\leq \frac{2}{\lambda} \{(E(\mathcal{P}_h U) - E_h(U)) + (E_h(U) - E(u))\}$$

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

$$\|\mathcal{P}_h(U) - u\|_2^2 \leq \omega(u_0, h)_2 \{\omega(u_0, h)_2 + C \|u_0\|_2\} + \frac{32h}{\lambda} \|u_0\|_\infty^2 + \frac{2Ch}{\lambda \epsilon^2} + \frac{2}{\lambda} \{E(u_\epsilon) - E(u)\}. \quad (18)$$

Let $\epsilon = h^{1/(2\mathcal{L}+1)}$ and since u_0 is weakly regular $\omega(u_0, h)_2 \leq \mathcal{O}(h^{\mathcal{L}})$, the above inequality becomes

$$\|\mathcal{P}_h(U) - u\|_2^2 \leq \frac{2}{\lambda} \{E(u_\epsilon) - E(u)\} + Ch^{\mathcal{L}/(\mathcal{L}+1)}$$

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

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 [41],

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

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 Eqns. (8). All the test images are normalized to the range $[0, 1]$.

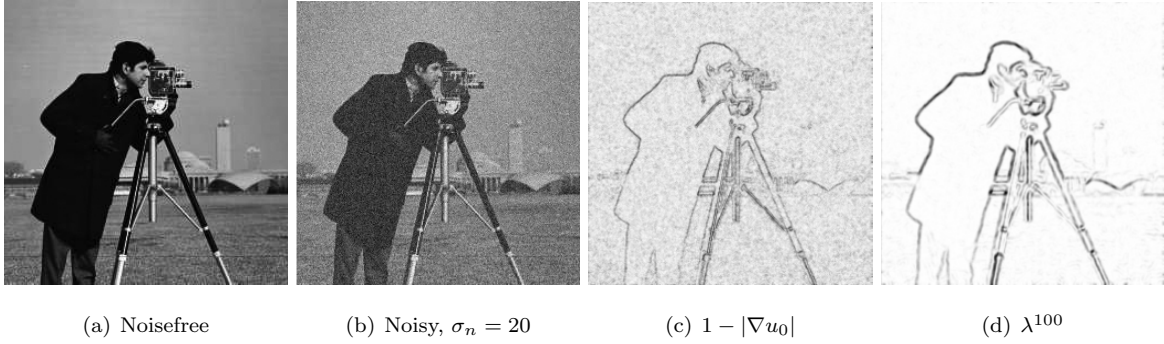|			
(a) Noisefree	(b) Noisy, $\sigma_n = 20$	(c) $1 - |\nabla u_0|$	(d) λ^{100}

Figure 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, $\sigma_n = 20$ (c) Gradient image, inverted $(1 - |\nabla u_0|)$ 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 homogeneous regions as the iterations are increased.

We further introduce an iteration and pixel adaptive $\lambda_{i,j}^{(t)}$ using the gradient information at iteration $(t-1)$ via

$$\lambda_{i,j}^{(t)} := \frac{1}{\epsilon^2 + \sqrt{(u_{i+1,j}^{(t-1)} - u_{i,j}^{(t-1)})^2 + (u_{i,j+1}^{(t-1)} - u_{i,j}^{(t-1)})^2}} \tag{19}$$

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 minmax 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). Fig. 7 we consider the same 1-D signal shown earlier in Fig. 4(a). 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 show the *Cameraman* gray-scale 256×256 size image used in our later comparison results. We add Gaussian white noise of standard deviation $\sigma_n = 20$ and mean zero[5]. Figure 6 (b) & (c) shows the gradient image (Computed using the formulae (8)) from the initial noisy image $|\nabla u_0|$ and adaptive λ parameter computed using Eqn. (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 picture (2 mega-pixels, image contains unknown amount of

---

[5]Using MATLAB command `imnoise($u_0$,'gaussian',0,$\sigma_n$)`.
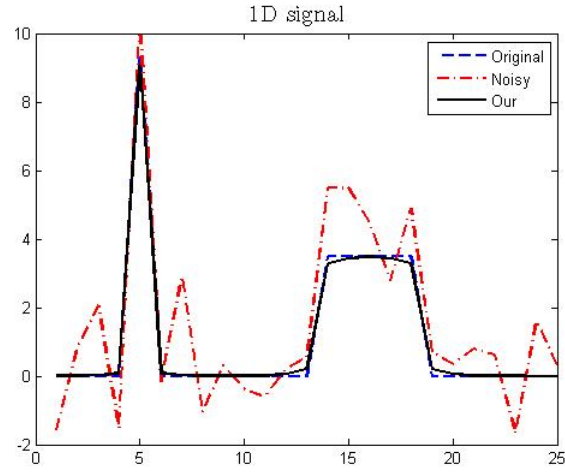
Figure 7: Restoration of a 1-D signal by our scheme: $(---)$ Original signal $(-\cdot-)$ noisy signal and (---) dark line is the restored curve using our adaptive minimization scheme (1) with function (6). Compare this with the corresponding results for Huber and Tukey functions in Figure 4(b) and Figure 5(c) respectively.

(a) Color *Movie* scene and *Kid*, *Goat* gray scale photo

(b) Smoothed images at iteration $t = 100$

Figure 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.

14

(a) Original   (b) Input   (c) Result   (d) Edges

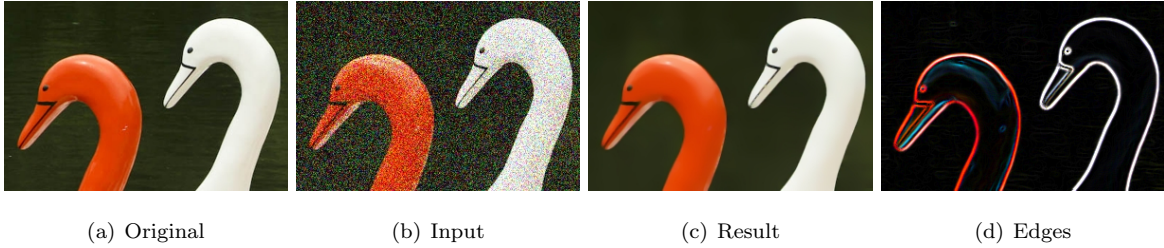Figure 9: *Ducks* color image $300 \times 200 \times 3$ restoration result. (a) Original RGB image (b) Noisy image, $\sigma_n = 20$, $PSNR = 12.56dB$ (c) Restored by our method, $PSNR = 23.15dB$ (d) Edges computed in all three channels (RGB) using the Canny edge detector.

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. 8 (b) exhibit marked improvements. Note that fine texture details are lost in Fig. 8 (b) (background wall, 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. 9 (a). This *Ducks* color image consists of flat background with strong curved edges and the result in Fig. 9 (c) indicates the local smoothing due to Gaussian filtering effect in regions where $|\nabla u| < k$ and edge preserving TV filtering in other areas. Figure 9 (d) shows the Canny edge map of computed in all the three color channels[6].

4.3   Comparison results

Figure 10 we show 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 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). On the other hand our function (6) is robust when compared to the other two classical functions as can be seen from the PSNR comparison Fig. 11 (b) as well. The topmost PSNR curve indicates that the scheme proposed in this paper surpasses the other two when the noise level increases $\sigma_n = 10 \rightarrow 25$. Note

---

[6]Using MATLAB command `edge(`u_0`,'canny')`. Note that the Canny edge detector employs non-maximal suppression to avoid small scale edges. The edges are computed for each of Red, Green, Blue channels and the final result is shown by combining them.

(a) True image u (b) Huber U^H (c) Tukey U^T (d) Our U^{our}

(e) Noisy image I (f)   (g)   (h)

(i) Noisy contour   (j)   (k)   (l)
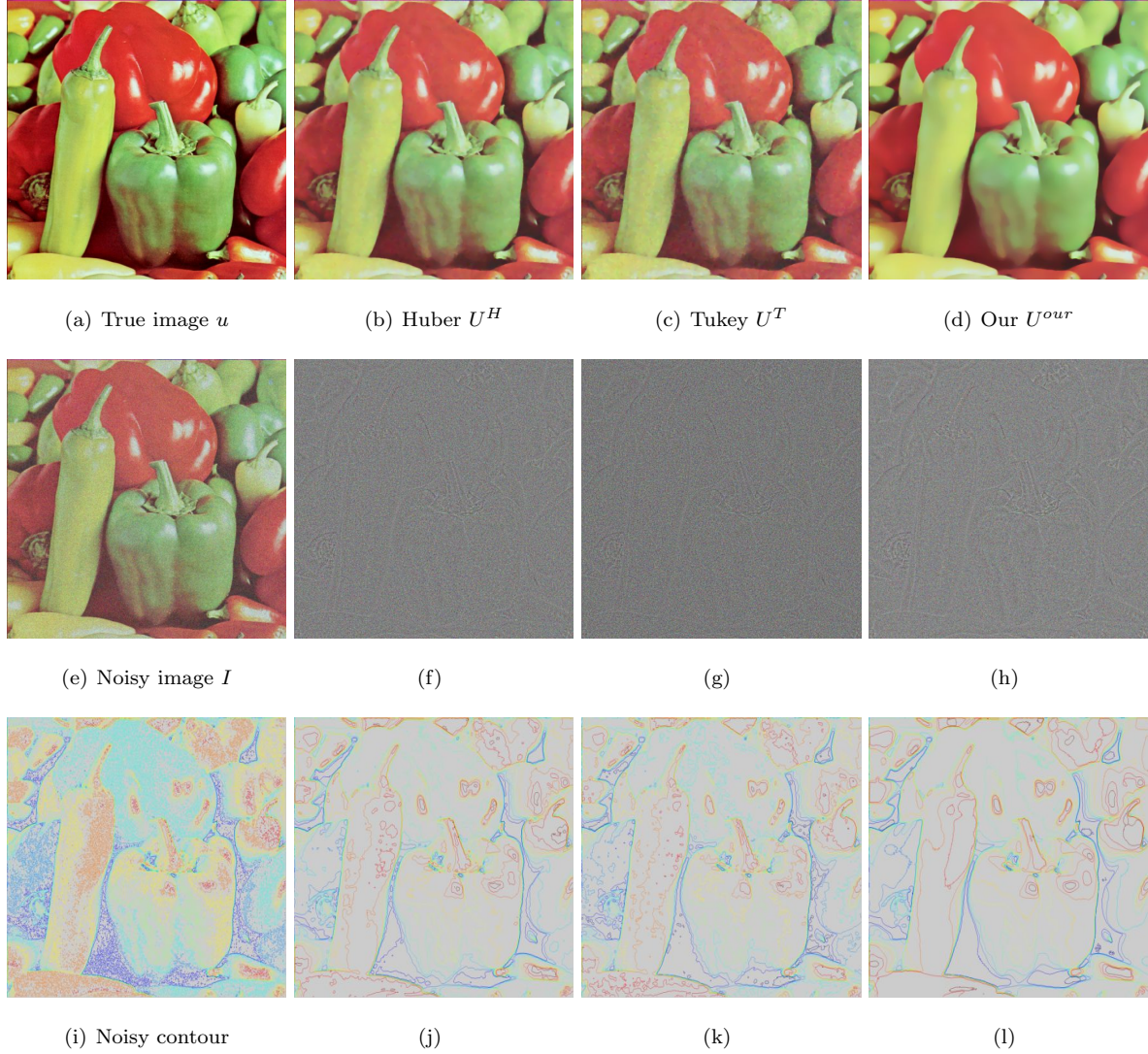
Figure 10: Restoration result for a color *Peppers* color image (size $512 \times 512 \times 3$) with our adaptive regularization scheme: (a) Original image (b-d) results of Huber, Tukey and our adaptive regularization function based scheme with tolerance $tol = 10^{-6}$ respectively (e) Gaussian noise ($\sigma_n = 20$) corrupted image (f-h) Residual noise/method noise image, $|u - U|^2$ (i-j) Contour map showing the restoration on level lines.

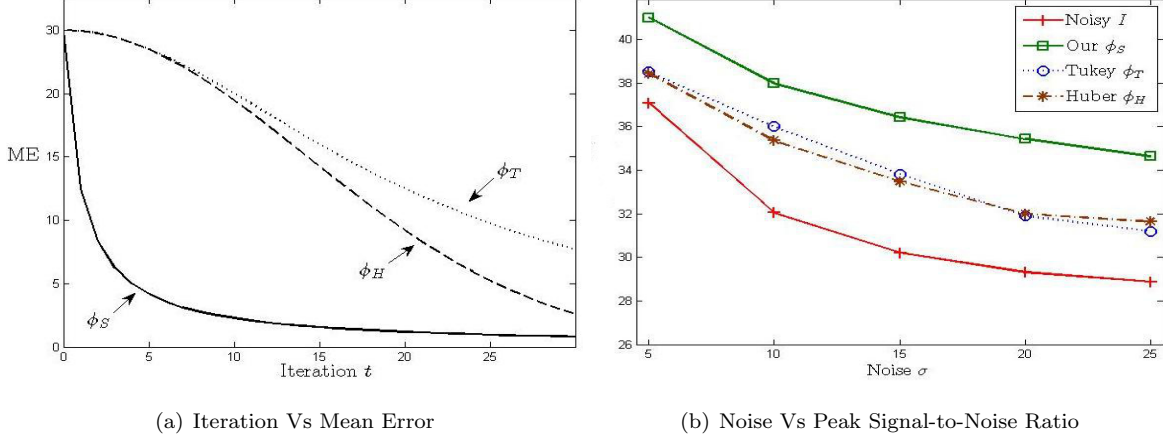(a) Iteration Vs Mean Error

(b) Noise Vs Peak Signal-to-Noise Ratio

Figure 11: Comparison of our proposed scheme with φ_S in (6) with Huber's φ_H (4) and Tukey's φ_T: (5) for the *Peppers* color image in Figure 10. (a) Number of iterations (t) Vs Mean error (ME) (b) Noise level (σ_n) Vs Peak signal-to-noise ratio ($PSNR$) for different noise levels.

that $\sigma_n^2 > 400$ is a high level noise and our scheme (1) does a good job in distinguishing between outliers correspond to noise and true edges due to the adaptive nature of $\lambda_j^{(t)}$ (19).

We next provide comparison with primal dual hybrid gradient (PDHG) [56], projected averaged gradient (Proj. Grad) [57], fast gradient projection (FGP) [2], alternating direction method of multipliers (ADMM) [23], 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 Eqn. (9).

- Relative duality gap:

$$\mathcal{R}(u,b) = \frac{E_{Primal}(u) - E_{Dual}(b)}{E_{Dual}(b)}, \tag{20}$$

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

- Peak Signal-to-Noise (PSNR) ratio,

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

The higher the PSNR the better the restoration result.

- The mean error (ME):

$$ME(u,I) := \frac{1}{MN} \sum_i |u_i - I_i|$$
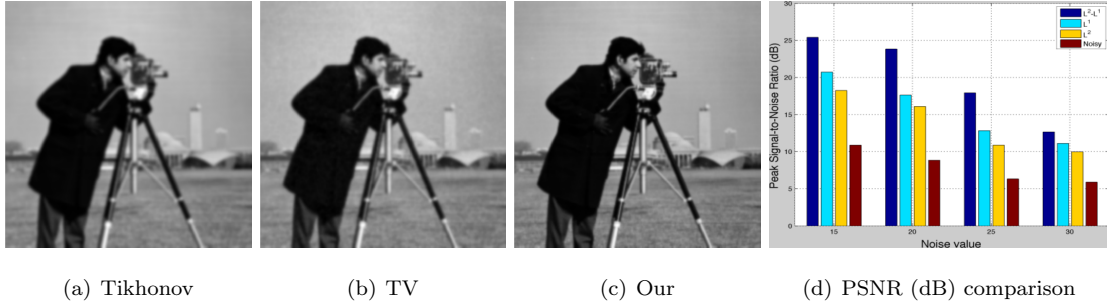
The mean error needs to be small for restored images.

17

(a) Tikhonov	(b) TV	(c) Our	(d) PSNR (dB) comparison

Figure 12: Image restoration of noisy *Cameraman* (Figure 6 (b)) and PSNR (dB) comparison of results for the (a) Tikhonov ($\varphi(s) = s^2$), (b) TV ($\varphi(s) = s$) and (c) our regularization function (6) based schemes (d) PSNR comparison shows that the proposed adaptive scheme performs better across different noise levels.

Algorithm	tol = 10^{-2}	tol = 10^{-4}	tol = 10^{-6}
PDHG	14	70	310
Proj. Grad.	46	721	14996
FGP	24	179	1264
ADMM	97	270	569
Split Breg.	**10**	**28**	**55**

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

First comparative example in Fig. 12 compares the restoration results for the noisy *Cameraman* gray scale image from Fig. 6 (b). As can be seen, adaptive Huber function performs better than the classical TV and Tikhonov schemes. Moreover, improvement in PSNR is $> 5dB$ (see Fig. 12 (d)) in different noise levels which indicates 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 a 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

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)

Table 2: Comparison of different algorithms in terms of noise level (σ_n) for the *Cameraman* gray scale image. The results are given in terms of best possible PSNR (computational time in seconds, maximum iterations). Each scheme is terminated if the maximum number of iterations exceeded 500 or when the duality gap is less than $\mathcal{R}(u, b) \leq 10^{-6}$.

for the image deblurring and deconvolution requires a delicate analysis of the boundary conditions [46] and is treated elsewhere. Other avenues of exploration are treating higher order models [54, 28], multi grid [47] and FEM [29] based schemes and their convergence analysis.

5   Conclusion

In this paper we considered adaptive Huber type regularization function based image restoration scheme. By 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 by using inter-channel correlations [37, 33, 35] and this defines our future work.

References

[1] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: Partial differential equation and calculus of variations.* Springer-Verlag, New York, USA, 2006.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(18):2419–2434, 2009.

[3] M. J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.

[4] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432, 1998.

[5] E. Carlini and R. Ferretti. A semi-Lagrangian approximation for the AMSS model of image processing. *Applied Numerical Mathematics*, 2012.

[6] V. Caselles, G. Sapiro, and D. H. Chung. Vector median filters, inf-sup operations, and coupled PDE's: Theoretical connections. *Journal of Mathematical Imaging and Vision*, 12(2):109–119, 2000.

[7] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM Journal on Applied Mathematics*, 55(3):827–863, 1995.

[8] A. Chambolle. Finite–differences discretizations of the Mumford–Shah functional. *M2AN Mathematical Modeling and Numerical Analysis*, 33(2):261–288, 1999.

[9] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004.

[10] A. Chambolle, S. Levine, and B. J. Lucier. An upwind finite-difference method for total variation based image smoothing. *SIAM Journal on Imaging Sciences*, 4(1):277–299, 2011.

[11] A. Chambolle and P. L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[12] 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.

[13] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.

[14] T. F. Chan and J. Shen. *Image processing and analysis: Variational, PDE, wavelet, and stochastic methods.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.

[15] K. Chen. Adaptive smoothing via contextual and local discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1552–1567, 2005.

[16] Y. Chen, S. Levine, and M. Rao. Variable exponent, linear growth functionals in image restoration. *SIAM Journal on Applied Mathematics*, 66(4):1383–1406, 2006.

[17] C. K. Chu, I. K. Glad, F. Godtliebsen, and J. S. Marron. Edge-preserving smoothers for image processing. *Journal of American Statistical Association*, 93(442):526–541, 1998.

[18] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34(5):1779–1791, 1997.

[19] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[20] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. In *In Proceedings of the 46-th Session of the ISI, Bulletin of the ISI*, volume 52, pages 22–26, 1987.

[21] S. A. Gersgorin. Fehlerabschätzung für das differenzverfahren zur lösung partieller differentialgleichungen. *J. Angew. Math. Mech.*, 10:373–382, 1930.

[22] F. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Birkhauser, Basel, Switzerland, 1984.

[23] R. Glowinski and A. Marrocco. 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, 1975.

[24] T. Goldstein, X. Bresson, and S. Osher. Geometric applications of the split Bregman method: segmentation and surface reconstruction. *Journal of Scientific Computation*, 45(1–3):272–293, 2010.

[25] T. Goldstein and S. Osher. The split Bregman algorithm for L1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[26] P. J. Huber. *Robust Statistics*. Wiley, New York, NY, USA, 1981.

[27] R.-Q. Jia, H. Q. Zhao, and W. Zhao. Convergence analysis of the bregman method for the variational model of image denoising. *Applied and Computational Harmonic Analysis*, 27(3):367379, 2009.

[28] Q. Jiang. Correspondence between frame shrinkage and high-order nonlinear diffusion. *Applied Numerical Mathematics*, 62(1):51–66, 2012.

[29] J. Kaccur and K. Mikula. Solution of nonlinear diffusion appearing in image smoothing and edge detection. *Applied Numerical Mathematics*, 17(1):47–59, 1995.

[30] M.-J. Lai and L. M. Messi. Piecewise linear approximation of the continuous Rudin–Osher–Fatemi model for image denoising. *SIAM Journal on Numerical Analysis*, 50(5):2446–2466, 2012.

[31] S. Z. Li. *Markov Field Random Modeling in Computer Vision*. Springer, Berlin, Germany, 1995.

[32] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[33] V. B. S. Prasath. Weighted Laplacian differences based multispectral anisotropic diffusion. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4042–4045, Vancouver BC, Canada, July 2011.

[34] V. B. S. Prasath. A well-posed multiscale regularization scheme for digital image denoising. *International Journal of Applied Mathematics and Computer Science*, 21(4):769–777, 2011.

[35] V. B. S. Prasath, J. C. Moreno, and K. Palaniappan. Color image denoising by chromatic edges based vector valued diffusion. *Preprint*, 2013. Available at http://arxiv.org/abs/1304.5587.

[36] V. B. S. Prasath and A. Singh. A hybrid convex variational model for image restoration. *Applied Mathematics and Computation*, 215(10):3655–3664, 2010.

[37] V. B. S. Prasath and A. Singh. Multispectral image denoising by well-posed anisotropic diffusion scheme with channel coupling. *International Journal of Remote Sensing*, 31(8):2091–2099, 2010.

[38] V. B. S. Prasath and A. Singh. Well-posed inhomogeneous nonlinear diffusion scheme for digital image denoising. *Journal of Applied Mathematics*, 2010:14pp, 2010. Article ID 763847.

[39] V. B. S. Prasath and A. Singh. An adaptive anisotropic diffusion scheme for image restoration and selective smoothing. *International Journal of Image and Graphics*, 12(1):18pp, 2012.

[40] V. B. S. Prasath and D. Vorotnikov. On a system of adaptive coupled pdes for image restoration. *Journal of Mathematical Imaging and Vision*, Online First, 2012. Available at arXiv:1112.2904.

[41] W. J. J. Rey. *Introduction to robust and quasirobust statistical methods*. Springer-Verlag, Berlin, Germany, 1983.

[42] M. Rivera and J. L. Marroquin. Efficient half-quadratic regularization with granularity control. *Image and Vision Computing*, 21(4):345–357, 2003.

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

[44] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer-Verlag, New York, USA, 2009.

[45] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 92(3):265–280, 2011.

[46] Y. Shi and Q. Chang. Acceleration methods for image restoration problem with different boundary conditions. *Applied Numerical Mathematics*, 58(5):602–614, 2008.

[47] R. M. Spitaleri, R. March, and D. Arena. A multigrid finite-difference method for the solution of Euler equations of the variational image segmentation. *Applied Numerical Mathematics*, 39(2):181–189, 2001.

[48] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tapen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[49] A.N. Tikhonov and V.Y. Aresenin. *Solutions of Ill-posed Problems*. John Wiley, New York, NY, USA, 1997.

[50] J. W. Tukey. *Exploratory data analysis*. Addison-Wesley Publishers, 1977.

[51] J. Weickert. *Anisotropic diffusion in image Processing*. B.G. Teubner-Verlag, Stuttgart, Germany, 1998.

[52] J. Weickert, B. M. H. Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, 1998.

[53] P. Weiss, L. Blanc-Feraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM Journal of Scientific Computing*, 31(3):2047–2080, 2009.

[54] T.-T. Wu, Y.-F. Yang, and Z.-F. Pang. A modified fixed-point iterative algorithm for image restoration using fourth-order PDE model. *Applied Numerical Mathematics*, 62(2):79–90, 2012.

[55] G. Xu. Consistent approximations of several geometric differential operators and their convergence. *Applied Numerical Mathematics*, 69:1–12, 2013.

[56] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical Report 08–34, UCLA CAM, 2008.

[57] M. Zhu, S. J. Wright, and T. F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications*, 47(3):377–400, 2010.