**Noname manuscript No.**
(will be inserted by the editor)

# Image denoising by generalized total variation regularization and least squares fidelity

**Jie Yan · Wu-Sheng Lu**

Received: date / Accepted: date

**Abstract** Inspired by the ability of $\ell_p$-regularized algorithms and the close connection of total variation (TV) to the $\ell_1$ norm, a $p$th-power type TV denoted as $\mathrm{TV}_p$ is proposed for $0 \leq p \leq 1$. The $\mathrm{TV}_p$-regularized problem for image denoising is nonconvex thus difficult to tackle directly. Instead, we deal with the problem by proposing a weighted TV (WTV) minimization where the weights are updated iteratively to locally approximate the $\mathrm{TV}_p$-regularized problem. The difficulty of WTV minimization is dealt with in a modified split Bregman framework. Numerical results are presented to demonstrate improved denoising performance of the new algorithm with $p < 1$ relative to that obtained by the standard TV minimization and several recent denoising methods from the literature on a variety of images.

**Keywords** TV · WTV · image denoising

## 1 Introduction

Denoising is an important procedure in image processing in which the noise that corrupts a digital image is removed or reduced. Removing white Gaussian noise from an image is not trivial because the white noise occupies the entire spectrum (the power spectrum of white noise is a nonzero constant versus frequency) hence the spectrum of a typical image cannot be separated from that of the white noise. This makes band-selective filtering based techniques ineffective as long as the corrupting noise is Gaussian white. A breakthrough for denoising (especially piecewise smooth) images is made in the work of Rudin, Osher and Fatemi (ROF) [Rudin et al., 1992] in which the standard $\ell_2$-norm fidelity optimization is regularized by the total variatoin (TV) of the image.

Variants of the ROF algorithm with improved performance and complexity are available [Chambolle, 2004, Chan et al., 2005, Beck and Teboulle, 2009]. In [Chan

J. Yan · W.-S. Lu
Department of Electrical and Computer Engineering, University of Victoria
3800 Finnerty Rd., Victoria, BC V8P 5C2, Canada
E-mail: jyan@ece.uvic.ca

and Esedoglu, 2005], a TV-$\ell_1$ model for image denoising is proposed and the model is shown to be contrast invariant which in turn suggests a data driven scale selection technique, see also [Allard, 2008, Allard, 2009] for indepth mathematical analysis of TV-regularized denoising algorithms. Reference [Wohlberg and Rodriguez, 2007] deals with the denoising problem based on a model that involves a generalized TV$_q$ regularizer with $q \in [1, 2)$ and an $\ell_p$-norm fidelity term with $p \geq 1$, and an iteratively-reweighted-norm algorithm is proposed to carry out the TV-$\ell_p$ minimization. Adaptive TV denoising has also been developed [Chen et al., 2010] in that the regularization term acts like a TV norm at object edges while approximating $\ell_2$-norm in flat and ramp regions so as to avoid staircase effect. In [Chopra and Lian, 2010], recovery of blocky images with improved performance over existing PDE-based approaches is addressed using a spatially adaptive TV model where a carefully designed penalization function is incorporated. Recently, the methodology for TV-based denoising is extended to a local TV-filtering scheme by employing the ROF model in a given neighbourhood of each pixel [Louchet and Moisan, 2011], and the concept of higher degree total variation (HDTV) is introduced in [Hu and Jacob, 2012] in order to deal with the staircase and ringing artifacts that are common in TV and wavelet based schemes. It is also interesting to note that the anisotropic HDTV regularizer is found to provide consistently better reconstruction performance over the isotropic counterpart [Hu and Jacob, 2012]. Very recently, a computational framework was proposed in [Lee et al., 2013] that incorporates a TV minimization formulation into a moving least squares method for image denoising in order to overcome drawbacks of both approaches.

Apart from the TV-based denoising techniques, the basis pursuit (BP) [Chen et al., 2001] based $\ell_1$ regularization is another major approach which is found effective in signal denoising. Recognizing the close connection of the BP algorithm to the compressive sensing methods for the reconstruction of sparse signals whose sparsity is measured by the $\ell_0$ "norm", several authors [Chartrand, 2007, Chartrand and Yin, 2008, Yan and Lu, 2011b, Yan and Lu, 2012] have proposed $\ell_p$ (with $p < 1$) regularization methods which are found to provide improved reconstruction and denoising results relative to those obtained by their $\ell_1$-regularization counterparts. In [Candes et al., 2008], an iteratively reweighted $\ell_1$ minimization (IRL1) is developed for enhanced sparsity. With the weights appropriately chosen in each round of iteration, the technique allows to carry out essentially $\ell_0$ regularization without having to perform nonconvex optimization. The IRL1 algorithm is a technique that connects an $\ell_0$-regularized (nonconvex) problem directly to an $\ell_1$-regularized (convex) problem without additional mechanism to avoid a degraded suboptimal solution. Finally, we mention reference [Fu et al., 2006] where models with mixed $\ell_2$-$\ell_1$ and $\ell_1$-$\ell_1$ norms for image restoration are proposed.

The main contributions of this paper are twofold. First, the standard TV is generalized to a $p$th-power TV with $0 \leq p \leq 1$ for further promoting gradient sparsity and a generalized TV (GTV) regularized least squares problem for image denoising is proposed. The GTV-regularized least squares problem is nonconvex, thus the second contribution of the paper is the development of a two-step solution method to solve the problem at hand - by introducing weighted TV (WTV) in which each element of discretized TV is weighted along the horizontal and vertical directions, then approximating the solution of the GTV-regularized problem by solving iteratively reweighted TV (IRTV) convex subproblems. In addition, new techniques are also developed to deal with technical difficulties encountered. These include a power-iterative warm-start technique for the proposed IRTV algorithm to reach a target power value $p_t < 1$ and a modified Split Bregman method in order to overcome the difficulties arising from the presence of the nontrivial weights in WTV. Numerical results are presented to demonstrate improved denoising performance in comparison with BPDN, IRL1, TV-$\ell_1$ as well as several more recent denoising methods.

## 2 Generalized total variation regularization

### 2.1 Total variation and denoising by TV minimization

Image denoising is probably the most successful application of TV minimization [Rudin et al., 1992]. Let the image model be given by

$$\mathbf{b} = \mathbf{u}^* + \mathbf{w} \tag{1}$$

where $\mathbf{b}$ denotes noisy measurement of original image $\mathbf{u}^* \in \mathbb{R}^{m\times n}$ and $\mathbf{w}$ is the noise term with independently and identically distributed (i.i.d) Gaussian entries of zero mean and variance $\sigma^2$. The denoising of $\mathbf{b}$ is carried out by solving the convex TV-regularized problem

$$\min_{\mathbf{u}} \ \mathrm{TV}(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{u}-\mathbf{b}\|_{\mathrm{F}}^2 \tag{2}$$

where $\mu > 0$ is a regularization parameter and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. The discretized anisotropic and isotropic TV of image $\mathbf{u}$ are defined as [Beck and Teboulle, 2009]

$$\mathrm{TV}^{(\mathrm{A})}(\mathbf{u}) = \sum_{i=1}^{m-1}\sum_{j=1}^{n}|u_{i,j}-u_{i+1,j}| + \sum_{i=1}^{m}\sum_{j=1}^{n-1}|u_{i,j}-u_{i,j+1}| \tag{3}$$

and

$$\begin{aligned}\mathrm{TV}^{(\mathrm{I})}(\mathbf{u}) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\sqrt{|u_{i,j}-u_{i+1,j}|^2+|u_{i,j}-u_{i,j+1}|^2}\\ &+\sum_{i=1}^{m-1}|u_{i,n}-u_{i+1,n}| + \sum_{j=1}^{n-1}|u_{m,j}-u_{m,j+1}|\end{aligned} \tag{4}$$

respectively.

### 2.2 Generalized $p$th power total variation

TV can be related to the $\ell_1$ norm in two ways. Conceptually, both constrained $\ell_1$-norm and TV minimization are found effective in signal processing, including denoising [Chen et al., 2001], deblurring [Mallat, 2009, Beck and Teboulle, 2009], and signal reconstruction [Candès et al., 2006]. Analytically, the anisotropic total variation $\mathrm{TV}^{(\mathrm{A})}(\mathbf{u})$ may be interpreted as the $\ell_1$ norm of the discretized gradient of $\mathbf{u}$ [Candès et al., 2006], namely, $\mathrm{TV}^{(\mathrm{A})}(\mathbf{u}) = \|\nabla_x\mathbf{u}\|_1 + \|\nabla_y\mathbf{u}\|_1$ where $(\nabla_x\mathbf{u})_{i,j} =$

$u_{i,j} - u_{i+1,j}$, $(\nabla_y \mathbf{u})_{i,j} = u_{i,j} - u_{i,j+1}$ and $\|\cdot\|_1$ denotes the sum of magnitudes of the matrix's entries. Moreover, if we define $(D\mathbf{u})_{i,j} = (u_{i,j} - u_{i+1,j}) + \sqrt{-1}(u_{i,j} - u_{i,j+1})$, then the isotropic TV becomes $\|D\mathbf{u}\|_1$. In addition, when the size of $\mathbf{u}$ is reduced to $m \times 1$ or $1 \times n$, both $\mathrm{TV}^{(\mathrm{A})}(\mathbf{u})$ and $\mathrm{TV}^{(\mathrm{I})}(\mathbf{u})$ become $\|D\mathbf{u}\|_1$.

We now propose a generalized $p$th power total variation, denoted as $\mathrm{TV}_p$, as follows. For the anisotropic case, $\mathrm{TV}_p^{(\mathrm{A})}$ is defined as

$$\mathrm{TV}_p^{(\mathrm{A})}(\mathbf{u}) = \sum_{i=1}^{m-1}\sum_{j=1}^{n} |u_{i,j} - u_{i+1,j}|^p + \sum_{i=1}^{m}\sum_{j=1}^{n-1} |u_{i,j} - u_{i,j+1}|^p \tag{5}$$

Note that $\mathrm{TV}_p^{(\mathrm{A})}(\mathbf{u})$ is related to $\ell_p$ "norm" by $\mathrm{TV}_p^{(\mathrm{A})}(\mathbf{u}) = \|\nabla_x \mathbf{u}\|_p^p + \|\nabla_y \mathbf{u}\|_p^p$. For the isotropic case, the generalized TV is defined by

$$\begin{aligned}\mathrm{TV}_p^{(\mathrm{I})}(\mathbf{u}) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1} \sqrt{|u_{i,j} - u_{i+1,j}|^{2p} + |u_{i,j} - u_{i,j+1}|^{2p}} \\ &+ \sum_{i=1}^{m-1} |u_{i,n} - u_{i+1,n}|^p + \sum_{j=1}^{n-1} |u_{m,j} - u_{m,j+1}|^p\end{aligned} \tag{6}$$

The value of power $p$ in (5) and (6) is in the range $0 < p \le 1$, and a target value of $p$ is always set to $p_t < 1$. Obviously, with $p = 1$, $\mathrm{TV}_p^{(\mathrm{A})}$ and $\mathrm{TV}_p^{(\mathrm{I})}$ recover the standard $\mathrm{TV}^{(\mathrm{A})}$ and $\mathrm{TV}^{(\mathrm{I})}$ respectively. We remark that reference [Wohlberg and Rodriguez, 2007] presents a $\mathrm{TV}_q$-$\ell_p$ model for image restoration, hence is relevant to this present paper. The main difference between the two is that the model in [Wohlberg and Rodriguez, 2007] is with $q \in [1, 2)$ and $p \ge 1$, hence it is a convex model, while the present paper investigates a nonconvex $\mathrm{TV}_p$ regularizer with $0 < p < 1$, which turns out to offer improved performance. We shall revisit this point in Sec. 4 where experimental results are reported.

With the generalized $\mathrm{TV}_p$ defined, we propose to study image denoising problem with a model that incorporates $\mathrm{TV}_p$ as the regularizer, namely,

$$\min_{\mathbf{u}} \ \mathrm{TV}_p(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{b}\|_{\mathrm{F}}^2 \tag{7}$$

where $\mathrm{TV}_p(\mathbf{u})$ can be taken as $\mathrm{TV}_p^{(\mathrm{A})}(\mathbf{u})$ or $\mathrm{TV}_p^{(\mathrm{I})}(\mathbf{u})$ with $0 < p < 1$. With $p < 1$ solving the nonconvex problem in (7) is far from trivial and only local solutions can be retrieved. In what follows, an iterative reweighting technique is proposed to address the problem at hand.

### 2.3 Weighted TV and Iterative Reweighting

The idea of reweighting in the context of sparsity enhancement is originated in [Candes et al., 2008] where an iteratively reweighted $\ell_1$-minimization (IRL1) technique is developed that connects an $\ell_0$-regularized nonconvex problem directly to an $\ell_1$-regularized convex problem. With $p < 1$ the generalized $p$th power TV proposed in Sec. 2.2 is also nonconvex. Here we propose an iterative reweighting which is tailored to tackle the $\mathrm{TV}_p$-regularized denoising problem in (7) in a convex setting. As such, the proposed reweighting technique may be regarded as a natural extension of the method in [Candes et al., 2008] to models with total variation regularizers.

We proceed by introducing weighted TV (WTV), denoted by $\mathrm{TV_w}(\mathbf{u})$, which for anisotropic TV is defined as

$$\mathrm{TV}_{\mathrm{w}}^{(\mathrm{A})}(\mathbf{u}) = \sum_{i=1}^{m-1}\sum_{j=1}^{n} \alpha_{i,j}|u_{i,j} - u_{i+1,j}| + \sum_{i=1}^{m}\sum_{j=1}^{n-1} \beta_{i,j}|u_{i,j} - u_{i,j+1}| \tag{8}$$

and for isotropic TV is defined as

$$\begin{aligned}\mathrm{TV}_{\mathrm{w}}^{(\mathrm{I})}(\mathbf{u}) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1} \sqrt{(\alpha_{i,j}|u_{i,j} - u_{i+1,j}|)^2 + (\beta_{i,j}|u_{i,j} - u_{i,j+1}|)^2} \\ &+ \sum_{i=1}^{m-1} \alpha_{i,n}|u_{i,n} - u_{i+1,n}| + \sum_{j=1}^{n-1} \beta_{m,j}|u_{m,j} - u_{m,j+1}|\end{aligned} \tag{9}$$

In the above definitions, $\alpha_{i,j} > 0$ and $\beta_{i,j} > 0$ are weights to weigh the first order differences along the vertical and horizontal directions, respectively. Apparently $\mathrm{TV_w}(\mathbf{u})$ becomes $\mathrm{TV}(\mathbf{u})$ when all the weights are set to unity.

To see the critical role the $\mathrm{TV_w}$ plays in the new algorithm, notice that function $\mathrm{TV_w}(\mathbf{u})$ remains convex as long as the weights are fixed. Now if in the $l$th round of $\mathrm{TV_w}(\mathbf{u})$ minimization one assigns the weights to

$$\alpha_{i,j} = (|u_{i,j}^{(l)} - u_{i+1,j}^{(l)}| + \varepsilon)^{p-1}, \beta_{i,j} = (|u_{i,j}^{(l)} - u_{i,j+1}^{(l)}| + \varepsilon)^{p-1} \tag{10}$$

where $\varepsilon > 0$ is a small constant to prevent the weights from being zero, then for $\mathbf{u}$ in a small neighborhood of iterate $\mathbf{u}^{(l)}$ (8) and (10) in conjunction with the continuity of $\mathrm{TV_w}(\mathbf{u})$ imply that

$$\begin{aligned}\mathrm{TV}_{\mathrm{w}}^{(\mathrm{A})}(\mathbf{u}) &\approx \mathrm{TV}_{\mathrm{w}}^{(\mathrm{A})}(\mathbf{u}^{(l)}) \\ &= \sum_{i=1}^{m-1}\sum_{j=1}^{n} \frac{|u_{i,j}^{(l)} - u_{i+1,j}^{(l)}|}{(|u_{i,j}^{(l)} - u_{i+1,j}^{(l)}| + \varepsilon)^{1-p}} + \sum_{i=1}^{m}\sum_{j=1}^{n-1} \frac{|u_{i,j}^{(l)} - u_{i,j+1}^{(l)}|}{(|u_{i,j}^{(l)} - u_{i,j+1}^{(l)}| + \varepsilon)^{1-p}} \\ &\approx \sum_{i=1}^{m-1}\sum_{j=1}^{n} |u_{i,j}^{(l)} - u_{i+1,j}^{(l)}|^p + \sum_{i=1}^{m}\sum_{j=1}^{n-1} |u_{i,j}^{(l)} - u_{i,j+1}^{(l)}|^p \\ &= \mathrm{TV}_p^{(\mathrm{A})}(\mathbf{u}^{(l)})\end{aligned} \tag{11}$$

With a similar argument, we can also see that $\mathrm{TV}_{\mathrm{w}}^{(\mathrm{I})}(\mathbf{u}) \approx \mathrm{TV}_p^{(\mathrm{I})}(\mathbf{u}^{(l)})$ for $\mathbf{u}$ in a small vicinity of $\mathbf{u}^{(l)}$. Consequently, nonconvex minimization of $\mathrm{TV}_p(\mathbf{u})$ can practically be carried out by convex minimization of $\mathrm{TV_w}(\mathbf{u})$ with reweighting strategy (10) for each iteration. In other words, the introduction of weighted TV with the weights in (10) allows an effective convexification of the nonconvex $\mathrm{TV}_p$. We note that in [Louchet and Moisan, 2011], the ROF model is examined in a given neighborhood of each image pixel with a local filter and weights are introduced to a local window to construct a locally weighted TV denoiser in order to reduce staircase effect, while the reweighting in the proposed algorithm is performed globally on the TV norm to enhance the image's gradient sparsity.

**Algorithm 1** Algorithm for IRTV Regularized Minimization

1: Select parameters $\mu$, $p_t$ (a target value for power $p$) and $\varepsilon$. Set iteration count $l = 0$ and $\alpha_{i,j} = \beta_{i,j} = 1$.
2: Solve the WTV-regularized problem

$$\min_{\mathbf{u}} \ \mathrm{TV}_{\mathrm{w}}(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{b}\|_{\mathrm{F}}^2 \tag{12}$$

for $\mathbf{u}^{(l+1)}$.
3: With $\mathbf{u}^{(l+1)}$ update the weights $\{\alpha_{i,j}\}$ and $\{\beta_{i,j}\}$ in (10).
4: Terminate if $l+1$ reaches a maximum number of iterations $L$. Otherwise, set $l = l+1$ and repeat from step 2.

Based on above analysis, we propose an iteratively reweighted TV (IRTV) algorithm as Algorithm 1.

The core of the proposed algorithm is the WTV-regularized problem (12) for given sets of weights $\{\alpha_{i,j}\}$ and $\{\beta_{i,j}\}$. We reiterate that problem (12) is convex. The algorithmic details of solving (12) for both weighted anisotropic and isotropic TV by split Bregman type iterations are given in Sec. 3.

### 2.4 Implementation issues

Suppose we are given a target value $p_t < 1$ for power $p$. Although the subproblem (12) involved in each round of iteration is convex, the $\mathrm{TV}_p$-regularized problem with $p = p_t$ is nonconvex that admits many local minimizers.

The nonconvex nature of the problem at hand is reflected in Algorithm 1 by the way the initial weights $\{\alpha_{i,j}\}$ and $\{\beta_{i,j}\}$ are selected. The IRTV algorithm proposed in Sec. 2.3 is implemented using a power-iterative warm-start strategy, which may be explained by looking at the case of $p_t = 0.4$. The power-iterative strategy suggests that we start the algorithm with $p = 1$. So the problem at hand is convex and the solution is unique and global regardless of the initial point used. Now we use the solution just obtained as the initial point for a slightly reduced power value, say $p = 0.8$. Although $\mathrm{TV}_p$-regularization with $p = 0.8$ is nonconvex, its global solution cannot be too far away from the solution with $p = 1$. So with an $\ell_1$ solution as a "warm" initial point, the IRTV algorithm will converge to a reasonably good local solution if not the global solution. Once the second solution is obtained, it is used as the initial point for the problem in (12) with a further reduced power value, say $p = 0.6$. Doing the same thing one more time will reach the target power value $p = 0.4$. Numerical experiments have indicated that decreasing the power value each time by 0.2 is adequate to secure an excellent local solution for any target power between 0 and 1. Additional implementation details will be given in Sec. 4.

## 3 Split Bregman type iterations for the WTV-regularized problem

It turns out that efficient methods such as those based on split Bregman (SB) iterations [Goldstein and Osher, 2009] are not applicable to (12) as long as nontrivial weights $\{\alpha_{i,j}\}$ and $\{\beta_{i,j}\}$ are present. This section describes algorithmic details of a technique for solving (12) for both the weighted anisotropic and isotropic TVs based on *modified* SB iterations where the standard SB iterations are revised in a major way to make the proposed algorithms work.

### 3.1 Solving (12) for anisotropic TV

By the definition of anisotropic WTV in (8), the problem in (12) can be explicitly expressed as

$$\min_{\mathbf{u}} \ \{\sum_{i=1}^{m-1}\sum_{j=1}^{n}\alpha_{i,j}|u_{i,j}-u_{i+1,j}| + \sum_{i=1}^{m}\sum_{j=1}^{n-1}\beta_{i,j}|u_{i,j}-u_{i,j+1}| + \frac{\mu}{2}\|\mathbf{u}-\mathbf{b}\|_{\mathrm{F}}^2\} \tag{13}$$

If we introduce $\mathbf{d}_x$ and $\mathbf{d}_y$ with each element in the vector defined as

$$(\mathbf{d}_x)_{i,j} = (\nabla_x^{\mathrm{w}}\mathbf{u})_{i,j} = \alpha_{i,j}(u_{i,j}-u_{i+1,j}) \tag{14a}$$

$$(\mathbf{d}_y)_{i,j} = (\nabla_y^{\mathrm{w}}\mathbf{u})_{i,j} = \beta_{i,j}(u_{i,j}-u_{i,j+1}) \tag{14b}$$

where $\nabla_x^{\mathrm{w}}\mathbf{u}$ and $\nabla_y^{\mathrm{w}}\mathbf{u}$ denote the weighted first-order differences along the $x$ and $y$ directions, respectively, then we can write

$$\|\mathbf{d}_x\|_1 = \sum_{i=1}^{m-1}\sum_{j=1}^{n}\alpha_{i,j}|u_{i,j}-u_{i+1,j}|, \qquad \|\mathbf{d}_y\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n-1}\beta_{i,j}|u_{i,j}-u_{i,j+1}| \tag{15}$$

Consequently, a splitting strategy can be applied using (14) and (15) to formulate the problem at hand as

$$\min_{\mathbf{u}} \ \{\|\mathbf{d}_x\|_1 + \|\mathbf{d}_y\|_1 + \frac{\mu}{2}\|\mathbf{u}-\mathbf{b}\|_{\mathrm{F}}^2\} \tag{16a}$$

$$\text{s.t. } \ \mathbf{d}_x = \nabla_x^{\mathrm{w}}\mathbf{u}, \ \mathbf{d}_y = \nabla_y^{\mathrm{w}}\mathbf{u} \tag{16b}$$

Enforcing the constraints in (16b), we obtain

$$\min_{\mathbf{u},\mathbf{d}_x,\mathbf{d}_y} \ \big\{\|\mathbf{d}_x\|_1 + \|\mathbf{d}_y\|_1 + \frac{\mu}{2}\|\mathbf{u}-\mathbf{b}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\|\mathbf{d}_x - \nabla_x^{\mathrm{w}}\mathbf{u} - \mathbf{e}_x^{(k)}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\|\mathbf{d}_y - \nabla_y^{\mathrm{w}}\mathbf{u} - \mathbf{e}_y^{(k)}\|_{\mathrm{F}}^2\big\}$$

where $\mathbf{e}_x^{(k)}$ and $\mathbf{e}_y^{(k)}$ are updated through Bregman iterations as follows. In the $k$th iteration, we solve three subproblems to obtain $\mathbf{u}^{(k+1)}$, $\mathbf{d}_x^{(k+1)}$, and $\mathbf{d}_y^{(k+1)}$ as

$$\mathbf{u}^{(k+1)} = \underset{\mathbf{u}}{\operatorname{argmin}}\big\{\frac{\mu}{2}\|\mathbf{u}-\mathbf{b}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\|\nabla_x^{\mathrm{w}}\mathbf{u} + \mathbf{e}_x^{(k)} - \mathbf{d}_x^{(k)}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\|\nabla_y^{\mathrm{w}}\mathbf{u} + \mathbf{e}_y^{(k)} - \mathbf{d}_y^{(k)}\|_{\mathrm{F}}^2\big\} \tag{17}$$

$$\mathbf{d}_x^{(k+1)} = \underset{\mathbf{d}_x}{\operatorname{argmin}} \|\mathbf{d}_x\|_1 + \frac{\lambda}{2}\|\mathbf{d}_x - \nabla_x^{\mathrm{w}}\mathbf{u}^{(k+1)} - \mathbf{e}_x^{(k)}\|_{\mathrm{F}}^2 \tag{18a}$$

$$\mathbf{d}_y^{(k+1)} = \underset{\mathbf{d}_y}{\operatorname{argmin}} \|\mathbf{d}_y\|_1 + \frac{\lambda}{2}\|\mathbf{d}_y - \nabla_y^{\mathrm{w}}\mathbf{u}^{(k+1)} - \mathbf{e}_y^{(k)}\|_{\mathrm{F}}^2 \tag{18b}$$

and $\mathbf{e}_x^{(k)}$ and $\mathbf{e}_y^{(k)}$ are updated by

$$\mathbf{e}_x^{(k+1)} = \mathbf{e}_x^{(k)} + \nabla_x^{\mathrm{w}} \mathbf{u}^{(k+1)} - \mathbf{d}_x^{(k+1)} \tag{19a}$$

$$\mathbf{e}_y^{(k+1)} = \mathbf{e}_y^{(k)} + \nabla_y^{\mathrm{w}} \mathbf{u}^{(k+1)} - \mathbf{d}_y^{(k+1)} \tag{19b}$$

Initially, $\mathbf{d}_x^{(0)} = \mathbf{d}_y^{(0)} = \mathbf{e}_x^{(0)} = \mathbf{e}_y^{(0)} = \mathbf{0}$.

The problems in (18) can be solved effectively by soft shrinkage [Zibulevsky and Elad, 2010] as

$$\mathbf{d}_x^{(k+1)} = \mathcal{T}_{1/\lambda}(\nabla_x^{\mathrm{w}} \mathbf{u}^{(k+1)} + \mathbf{e}_x^{(k)}) \tag{20a}$$

$$\mathbf{d}_y^{(k+1)} = \mathcal{T}_{1/\lambda}(\nabla_y^{\mathrm{w}} \mathbf{u}^{(k+1)} + \mathbf{e}_y^{(k)}) \tag{20b}$$

where $\mathcal{T}_{1/\lambda}$ applies pointwisely as

$$\mathcal{T}_{1/\lambda}(z) = \mathrm{sgn}(z) \cdot \max\{|z| - 1/\lambda, 0\} \tag{21}$$

Solving problem (17) is however far from trivial as the conventional method in [Goldstein and Osher, 2009] is not applicable to the WTV due to the presence of nontrivial weights $\{\alpha_{i,j}\}$ and $\{\beta_{i,j}\}$. The technique we propose to solve (17) starts by writing the first-order optimality condition for problem (17) as

$$\mu \mathbf{u} + \lambda \left[ (\nabla_x^{\mathrm{w}})^T \nabla_x^{\mathrm{w}} + (\nabla_y^{\mathrm{w}})^T \nabla_y^{\mathrm{w}} \right] \mathbf{u} = \mathbf{c}^{(k)} \tag{22}$$

where

$$\mathbf{c}^{(k)} = \mu \mathbf{b} + \lambda [(\nabla_x^{\mathrm{w}})^T (\mathbf{d}_x^{(k)} - \mathbf{e}_x^{(k)}) + (\nabla_y^{\mathrm{w}})^T (\mathbf{d}_y^{(k)} - \mathbf{e}_y^{(k)})], \tag{23}$$

$(\nabla_x^{\mathrm{w}})^T$ and $(\nabla_y^{\mathrm{w}})^T$ denote the adjoint operators of $\nabla_x^{\mathrm{w}}$ and $\nabla_y^{\mathrm{w}}$, respectively, and

$$\begin{aligned} &\left[ (\nabla_x^{\mathrm{w}})^T \nabla_x^{\mathrm{w}} \mathbf{u} \right]_{i,j} \\ &= \begin{cases} \alpha_{1,j}^2 (u_{1,j} - u_{2,j}) & \text{for } i = 1, 1 \le j \le n \\ -\alpha_{m-1,j}^2 (u_{m-1,j} - u_{m,j}) & \text{for } i = m, 1 \le j \le n \\ \alpha_{i,j}^2 (u_{i,j} - u_{i+1,j}) - \alpha_{i-1,j}^2 (u_{i-1,j} - u_{i,j}) & \text{elsewhere} \end{cases} \end{aligned}$$

and

$$\begin{aligned} &\left[ (\nabla_y^{\mathrm{w}})^T \nabla_y^{\mathrm{w}} \mathbf{u} \right]_{i,j} \\ &= \begin{cases} \beta_{i,1}^2 (u_{i,1} - u_{i,2}) & \text{for } 1 \le i \le m, j = 1 \\ -\beta_{i,n-1}^2 (u_{i,n-1} - u_{i,n}) & \text{for } 1 \le i \le m, j = n \\ \beta_{i,j}^2 (u_{i,j} - u_{i,j+1}) - \beta_{i,j-1}^2 (u_{i,j-1} - u_{i,j}) & \text{elsewhere} \end{cases} \end{aligned}$$

As the matrix involved in the linear system in (22) is symmetric and diagonally dominant, using a standard argument (see Section 10.2 of [Larson, 2013]) the iterations generated by the Gauss-Seidel method as applied to (22) can be shown to converge to the solution of (22). From the above analysis, it follows that the linear system in (22) can be expressed componentwisely as

$$\begin{aligned} &[\mu + \lambda(\alpha_{i,j}^2 + \alpha_{i-1,j}^2 + \beta_{i,j}^2 + \beta_{i,j-1}^2)] u_{i,j} = c_{i,j}^{(k)} + \\ &\lambda(\alpha_{i,j}^2 u_{i+1,j} + \alpha_{i-1,j}^2 u_{i-1,j} + \beta_{i,j}^2 u_{i,j+1} + \beta_{i,j-1}^2 u_{i,j-1}) \end{aligned}$$

which naturally suggests a Gauss-Seidel iteration scheme for $u_{i,j}$ as

$$u_{i,j}^{(r+1)} = \frac{\rho_{i,j}^{(r,r+1)}}{\tau_{i,j}} \quad \text{for } r = 0, 1, \cdots, R \tag{24}$$

where

$$\begin{aligned}\rho_{i,j}^{(r,r+1)} &= c_{i,j}^{(k)} + \\ \lambda &\left[\alpha_{i,j}^2 u_{i+1,j}^{(r)} + \alpha_{i-1,j}^2 u_{i-1,j}^{(r+1)} + \beta_{i,j}^2 u_{i,j+1}^{(r)} + \beta_{i,j-1}^2 u_{i,j-1}^{(r+1)}\right] \\ \tau_{i,j} &= \mu + \lambda(\alpha_{i,j}^2 + \alpha_{i-1,j}^2 + \beta_{i,j}^2 + \beta_{i,j-1}^2)\end{aligned}$$

In words, $u_{i,j}^{(r+1)}$ is computed using iterates $u_{i+1,j}^{(r)}$, $u_{i,j+1}^{(r)}$ from the preceding iteration while iterates $u_{i-1,j}^{(r+1)}$, $u_{i,j-1}^{(r+1)}$ are computed from the current iteration. The convergence of the proposed Gauss-Seidel iterations in (24) is guaranteed by the diagonal dominance of the linear system involved.

### 3.2 Solving (12) for isotropic TV

We now consider solving the convex problem in (12) with weighted isotropic TV (see (9) for its definition) using the split Bregman technique. The weighted TV in this case can be written as $\text{TV}_{\text{w}}^{(\text{I})}(\mathbf{u}) = \|(\mathbf{d}_x, \mathbf{d}_y)\|_F$ where $\|(\mathbf{d}_x, \mathbf{d}_y)\|_F = \sum_{i,j} \sqrt{(\mathbf{d}_x)_{i,j}^2 + (\mathbf{d}_y)_{i,j}^2}$. By using the splitting strategy, the problem at hand can be formulated as

$$\min_{\mathbf{u}} \|(\mathbf{d}_x, \mathbf{d}_y)\|_F + \frac{\mu}{2}\|\mathbf{u} - \mathbf{b}\|_{\text{F}}^2 \tag{25a}$$

$$\text{s.t. } \mathbf{d}_x = \nabla_x^{\text{w}}\mathbf{u}, \ \mathbf{d}_y = \nabla_y^{\text{w}}\mathbf{u} \tag{25b}$$

By enforcing the equality constraint in (25b), the problem becomes

$$\begin{aligned}\min_{\mathbf{u},\mathbf{d}_x,\mathbf{d}_y} &\left\{\|(\mathbf{d}_x, \mathbf{d}_y)\|_F + \frac{\mu}{2}\|\mathbf{u} - \mathbf{b}\|_{\text{F}}^2\right. \\ &\left.+\frac{\lambda}{2}\|\mathbf{d}_x - \nabla_x^{\text{w}}\mathbf{u} - \mathbf{e}_x^{(k)}\|_{\text{F}}^2 + \frac{\lambda}{2}\|\mathbf{d}_y - \nabla_y^{\text{w}}\mathbf{u} - \mathbf{e}_y^{(k)}\|_{\text{F}}^2\right\}\end{aligned} \tag{26}$$

Note that unlike the weighted anisotropic TV, here the variables $\mathbf{d}_x$ and $\mathbf{d}_y$ are coupled. We propose to tackle the problem as follows. In the $k$th iteration, we find a set of new iterate $\{\mathbf{u}^{(k+1)}, \mathbf{d}_x^{(k+1)}, \mathbf{d}_y^{(k+1)}\}$ for the problem in (26) in two steps: First, with fixed $\mathbf{d}_x^{(k)}$ and $\mathbf{d}_y^{(k)}$ we solve the subproblem

$$\begin{aligned}\mathbf{u}^{(k+1)} =& \arg\min_{\mathbf{u}} \left\{\frac{\mu}{2}\|\mathbf{u} - \mathbf{b}\|_{\text{F}}^2\right. \\ &\left.+\frac{\lambda}{2}\|\nabla_x^{\text{w}}\mathbf{u} + \mathbf{e}_x^{(k)} - \mathbf{d}_x^{(k)}\|_{\text{F}}^2 + \frac{\lambda}{2}\|\nabla_y^{\text{w}}\mathbf{u} + \mathbf{e}_y^{(k)} - \mathbf{d}_y^{(k)}\|_{\text{F}}^2\right\}\end{aligned} \tag{27}$$

Then we fix $\mathbf{u} = \mathbf{u}^{(k+1)}$ in (26) and solve it for $\mathbf{d}_x^{(k+1)}$ and $\mathbf{d}_y^{(k+1)}$, namely,

$$\begin{aligned}(\mathbf{d}_x^{(k+1)}, \mathbf{d}_y^{(k+1)}) = \arg\min_{\mathbf{d}_x,\mathbf{d}_y} &\left\{\|(\mathbf{d}_x, \mathbf{d}_y)\|_F + \frac{\lambda}{2}\|\mathbf{d}_x - \nabla_x^{\text{w}}\mathbf{u}^{(k+1)} - \mathbf{e}_x^{(k)}\|_{\text{F}}^2\right. \\ &\left.+ \frac{\lambda}{2}\|\mathbf{d}_y - \nabla_y^{\text{w}}\mathbf{u}^{(k+1)} - \mathbf{e}_y^{(k)}\|_{\text{F}}^2\right\}\end{aligned} \tag{28}$$

Despite the fact that the variables $\mathbf{d}_x$ and $\mathbf{d}_y$ are coupled, the subproblem (28) above can still be explicitly solved using a generalized shrinkage formula [Wang et al., 2008, Goldstein and Osher, 2009]

$$\mathbf{d}_x^{(k+1)} = \max\ (\mathbf{s}^{(k)} - 1/\lambda, 0)\frac{\nabla_x^{\mathrm{w}}\mathbf{u}^{(k+1)} + \mathbf{e}_x^{(k)}}{\mathbf{s}^{(k)}} \tag{29a}$$

$$\mathbf{d}_y^{(k+1)} = \max\ (\mathbf{s}^{(k)} - 1/\lambda, 0)\frac{\nabla_y^{\mathrm{w}}\mathbf{u}^{(k+1)} + \mathbf{e}_y^{(k)}}{\mathbf{s}^{(k)}} \tag{29b}$$

where the divisions are pointwisely operated and

$$\mathbf{s}^{(k)} = \sqrt{|\nabla_x^{\mathrm{w}}\mathbf{u}^{(k+1)} + \mathbf{e}_x^{(k)}|^2 + |\nabla_y^{\mathrm{w}}\mathbf{u}^{(k+1)} + \mathbf{e}_y^{(k)}|^2 + \varepsilon^2}$$

where the term $\varepsilon^2$ is to prevent the elements of $\mathbf{s}^{(k)}$ from being zeros. On comparing the modified SB iteration for weighted isotropic TV with its anisotropic counterpart, the main difference is the way the next iterate $\{\mathbf{d}_x^{(k+1)}, \mathbf{d}_y^{(k+1)}\}$ is obtained: the anisotropic case uses standard soft shrinkage while the isotropic case requires general shrinkage formulas.

## 4 Experimental studies

The IRTV algorithm proposed in Secs. 2 and 3 was applied to a variety of synthetic and natural images and its performance was evaluated in comparison with several algorithms in the literature that are known to be effective for image denoising. As argued in Sec. 2.2, both $\mathrm{TV}^{(\mathrm{A})}$ and $\mathrm{TV}^{(\mathrm{I})}$ are closely related to the $\ell_1$ norm of the image gradient. As such, denoising performance based on them are expected to be similar to each other. As a matter of fact, in [Beck and Teboulle, 2009], both TVs were found to yield very similar results, while the simulations reported in [Hu and Jacob, 2012] were slightly in favor of $\mathrm{TV}^{(\mathrm{A})}$. Under these circumstances, the simulations reported below were carried out using anisotropic type of TV and $\mathrm{TV}_p$.

The denoising performance was evaluated by *peak signal-to-noise* ratio (PSNR) which is defined by

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right)\ (\mathrm{dB}) \tag{30}$$

with

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[u(i,j) - u^*(i,j)]^2 \tag{31}$$

where $\{u(i,j)\}$ and $\{u^*(i,j)\}$ denote the processed and desired (clean) images, respectively. Note that MSE (31) represents the variance of the noise remained in the processed image. Consequently, if the variance of the residual noise can be estimated (e.g. using a certain patch of the denoised image), PSNR remains to be a feasible measure in case the clean image is not available.

In what follows, the term “standard TV” (as well as term “TV” in Tables 1-3) is referred to as an algorithm that solves problem (2). In our simulations, this problem was solved by the Split Bregman method [Goldstein and Osher, 2009].

Throughout the section, the term "TV-$\ell_1$" is referred to as an algorithm that solves the problem

$$\min_{\mathbf{u}} \ \mathrm{TV}(\mathbf{u}) + \mu\|\mathbf{u} - \mathbf{b}\|_1 \tag{32}$$

Since (32) is convex, in our simulations (32) was solved using CVX [Grant and Boyd, 2008, Grant and Boyd, 2012].

## 4.1 Experiment 1: Denoising with different rounds of reweighting

In this part of experimental studies, effectiveness of the reweighting strategy (10) on denoising performance was examined. Shepp-Logan phantom of size $256 \times 256$ was chosen as the test image in the study, where the image was normalized to within $[0, 1]$ and corrupted with zero-mean additive white Gaussian noise with standard deviation 0.1. The PSNR of the noisy phantom was found to be 20.00 dB. To test the reweighting performance, power $p$ was fixed to 0.9, $\epsilon$ was set to $10^{-3}$, and $\mu$ in (12) varied in the range $[4, 25]$. For each $\mu$ in that range, $\lambda$ was set to $2\mu$ and the maximum number of Bregman iterations was set to 30.

For the first round of iterations of the IRTV algorithm, i.e., $l = 0$, the weights were all set to one, and the image computed from (12) simply corresponds to the standard TV denoising result. The weights in the following rounds of iterations were updated using (10), and the WTV-regularized problem was solved accordingly. Note that in solving (12) at the $(l + 1)$th iteration, we passed the previous iterate $\mathbf{u}^{(l)}$ as the initial point $\mathbf{u}_0$. The IRTV algorithm was applied for each $\mu$ with the number of reweighting rounds set to 1, 2, and 3, respectively. The three PSNR-versus-$\mu$ curves together with the PSNR curve of the standard TV denoising (i.e., $l = 0$) are depicted in Fig. 1. We observe that the algorithm practically converges in 3 rounds of reweighting, yielding a peak PSNR of 36.44 dB relative to 34.32 dB offered by the conventional TV denoising. Also note that the first round of reweighting (i.e., $l = 1$) yields largest performance gain.

## 4.2 Experiment 2: Denoising with different power $p$

Next, the IRTV algorithm was applied to the same phantom image with the value of power $p$ reduced from 1 by 0.2 each time down to 0. The experiment's set-up was the same as in Experiment 1 except that the range of $\mu$ was extended to $[4, 50]$ to cope with the wide range of power values, and only one round of reweighting was used for each given power $p$ since it provides most performance gain (see Experiment 1). In the experiment, the IRTV algorithm for a given power value $p$ starts with the solution obtained with a slightly larger $p$ as an initial point so as to facilitate the algorithm to converge to a satisfactory solution with a small number of iterations.

The PSNR curves associated with $p = 1, 0.8, 0.6, 0.4, 0.2$ and 0 are plotted in Fig. 2. We observe that the algorithm yields better performance for smaller power $p$, with the best performance of 39.39 dB achieved when $p = 0$ and $\mu$ set to 34

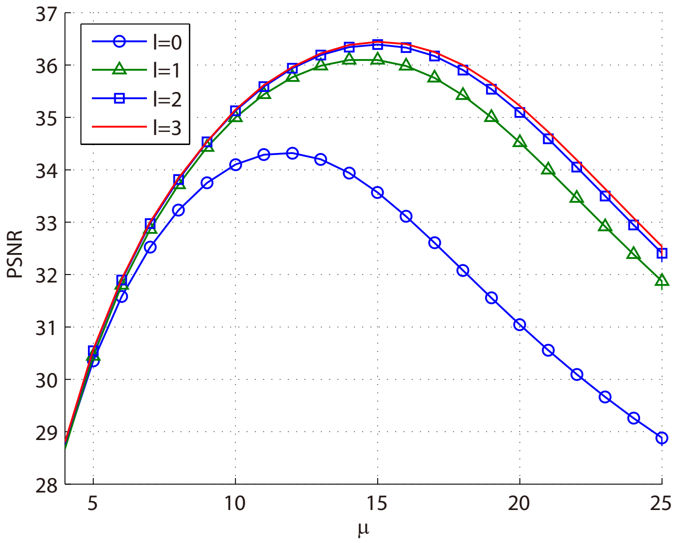


Fig. 1: Denoising Shepp-Logan phantom with 3 rounds of reweighting for $p = 0.9$, compared with the standard TV denoising which corresponds to the curve with $l = 0$.

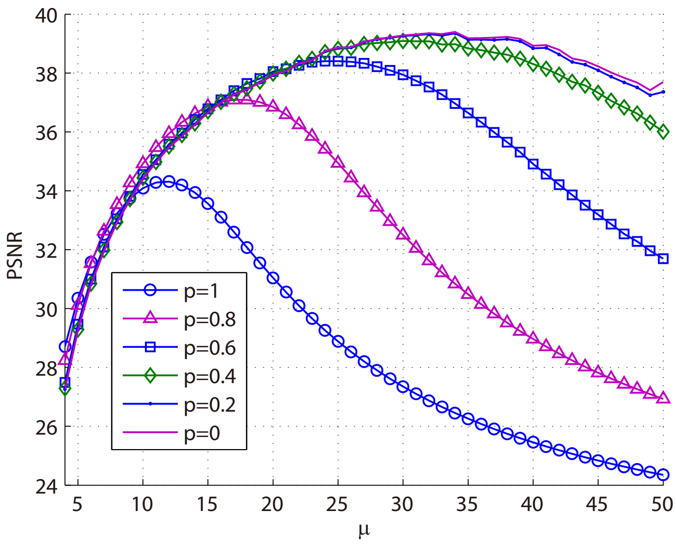


Fig. 2: Denoising Shepp-Logan phantom with one round of reweighting and $p = 1, 0.8, 0.6, 0.4, 0.2$ and 0. The PSNR with $p = 1$ coincides the standard TV denoising.

that is considerably higher than the peak PSNR of 34.32 dB by the standard TV denoising. The denoised images by the standard TV and IRTV algorithms are depicted in Fig. 3(a) and (c), respectively. To visualize the performance difference between the two algorithms, the difference between the original and denoised phantoms from the standard TV and IRTV algorithms are shown in Fig. 3(b) and (d), respectively. It is noticed that the difference after IRTV denoising is considerably less visible relative to that by the standard TV-based denoising.

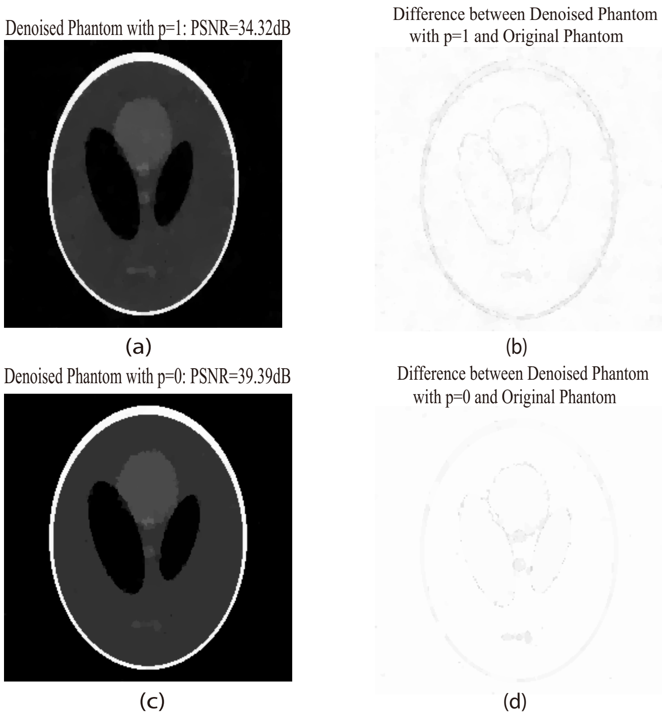


Fig. 3: Denoising Shepp-Logan phantom (a) by the standard TV minimization (i.e. $p = 1$), and (c) by the IRTV algorithm with $p = 0$. Difference between the denoised and original images are shown in (b) for the standard TV and (d) for IRTV.

### 4.3 Experiment 3: Image denoising of 18 natural and synthetic images

The proposed algorithm was tested on a variety of images of size 256 by 256 for different power $p$. Each image was normalized to be within $[0, 1]$, and was corrupted by zero-mean Gaussian noise with standard deviation 0.1, which corresponds to a PSNR of 20.00 dB.

For comparison purposes, several known denoising methods were also applied to the same set of test images. These include the well-known basis-pursuit denoising (BPDN) [Chen et al., 2001], iteratively reweighted $\ell_1$ algorithm (IRL1) [Candes et al., 2008], $\ell_2$-$\ell_p$ minimization based denoising [Yan and Lu, 2012, Chartrand, 2007, Chartrand and Yin, 2008, Foucart and Lai, 2009, Yan and Lu, 2011a], and the standard TV-regularized algorithm. For fair comparisons, when an algorithm was applied, the parameter(s) involved were tuned to optimize the performance of that particular algorithm. As was in Experiments 1 and 2, here the proposed IRTV algorithm was implemented using the power-iterative and warm-start strategy described in Sec. 2.4. The performance of the five algorithms in terms of PSNR before and after denoising for a total of sixteen natural and synthetic test images are shown in Table 1. For each test image, the best PSNR achieved is indicated with a bold-faced number. The pair of numbers in the columns $\ell_p$-$\ell_2/p$ and IRTV/$p$ denote the PSNR (in dB) achieved and the value of power $p$ utilized to obtain that PSNR, respectively. For visual inspection, Figs. 4-8 depict the IRTV-denoised natural images of "`axe`", "`church`", "`jet`", as well as a synthetic image "`phantomdisk`" and texture image "`fence`", respectively, as compared with their original versions and noisy counterparts.

| Image | PSNR in dB | | | | | |
|---|---|---|---|---|---|---|
| | IRL1 | BPDN | $\ell_p$-$\ell_2/p$ | TV-$\ell_1$ | TV | IRTV/$p$ |
| `axe` | 25.22 | 28.51 | 28.77/0.6 | 31.09 | 33.03 | **33.21**/0.6 |
| `building` | 21.57 | 25.42 | 25.54/0.0 | 26.47 | 27.85 | **28.08**/0.8 |
| `bird` | 24.92 | 28.07 | 28.20/0.4 | 30.33 | 31.87 | **31.94**/0.6 |
| `circles` | 24.05 | 28.35 | 28.71/0.2 | 33.87 | 37.35 | **43.65**/0.2 |
| `church` | 21.43 | 25.34 | 25.49/0.6 | 27.10 | 28.42 | **28.60**/0.6 |
| `camera` | 21.83 | 25.46 | 25.54/0.4 | 26.02 | 27.74 | **27.98**/0.8 |
| `crosses` | 21.45 | 26.33 | 26.39/0.8 | 23.54 | 29.23 | **32.20**/0.4 |
| `dome` | 20.00 | 22.47 | 19.68/0.6 | 18.18 | 24.00 | **24.22**/0.8 |
| `fence` | 21.55 | 25.40 | 25.52/0.2 | 26.75 | 28.24 | **28.45**/0.8 |
| `jet` | 21.40 | 24.96 | 25.03/0.4 | 26.05 | 27.45 | **27.62**/0.8 |
| `liftingbody` | 25.63 | 28.56 | 28.71/0.4 | 30.49 | 32.06 | **32.27**/0.8 |
| `phantomdisk` | 23.38 | 27.23 | 27.51/0.2 | 31.12 | 32.87 | **33.20**/0.8 |
| `pool` | 22.84 | 25.83 | 25.93/0.6 | 27.45 | 28.55 | **28.62**/0.6 |
| `reschart` | 20.56 | 25.56 | 25.86/0.4 | 25.45 | 29.11 | **30.51**/0.8 |
| `satellite` | 21.69 | 25.50 | 25.63/0.4 | 25.95 | 27.95 | **28.36**/0.8 |
| `squares` | 40.62 | 40.58 | 46.36/0.0 | 37.20 | 42.44 | **47.41**/0.0 |
| `tower` | 22.93 | 25.78 | 25.86/0.4 | 26.74 | 27.88 | **27.91**/0.6 |
| `text` | 20.36 | 25.88 | 26.22/0.6 | 27.31 | 30.38 | **33.84**/0.0 |

Table 1: PSNRs of test images denoised by the proposed algorithm and several existing denoising algorithms. In each row, the boldfaced numerical value indicates the best PSNR for the image.

Based on the experimental results, we remark that
(1) The IRTV algorithm consistently outperforms the other algorithms. Especially for the synthetic images such as "`circles`", "`squares`" and "`text`", its performance gain over the second best performer, the standard TV denoising, is found to be significant. We note that for these images the best power $p$ is zero or close to zero, while the standard TV algorithm corresponds to the case of $p = 1$. For natural images, although IRTV is found to offer decent gains over the standard TV for some images such as "`satellite`" (0.41 dB) and "`camera`" (0.24 dB), the

improvement over natural images are obviously less impressive relative to that achieved for synthetic images. The reason for this is that for most natural images the optimal power $p$, is found to be fairly close to 1, which implies that for natural images the standard TV denoising performs nearly optimally. In this regard the proposed IRTV algorithm tries to push the envelop to offer even better performance.

(2) The IRL1 algorithm does not seem to perform well relative to the other algorithms evaluated. There are several reasons for the outcome. Recall that the IRL1 essentially performs $\ell_0$-regularization without leaving convex programming environment. As argued earlier, however, this is not suited for denoising natural images as the best power for these image appears to be close to 1. This explains why the BPDN, which is an $\ell_1$-regularized algorithm, outperforms IRL1 for most instances. For the synthetic images, the IRL1 was expected to perform well. The reason it fails to do so has to do with its way to perform $\ell_0$-regularization - it jumps from $\ell_1$-regularization directly to $\ell_0$-regularization as opposed to the proposed IRTV algorithm where $\mathrm{TV}_p$ regularization is implemented with a power-iterative warm-start strategy. As a result, it leads to a suboptimal solution with degraded performance.

(3) By visual inspection of Figs. 4-8, the IRTV algorithm is found to well preserve edges as well as textures (see Fig. 8) relative to the TV-$\ell_1$ algorithm. Note that in this regard IRTV also performs better than the standard TV algorithm as can be seen from Fig. 3.

### 4.4 Experiment 4: Comparisons of IRTV algorithm with TV-$\ell_1$ and BM3D

In the literature, there are other known approaches to the image denoising. In [Chan and Esedoglu, 2005], a version of the ROF model that uses $\ell_l$-norm as a measure of fidelity was proposed, and it was argued that the TV-$\ell_1$ model offers several advantages: (1) the $\ell_1$ regularization is more geometric; (2) the TV-$\ell_1$ model is contrast invariant; and (3) the $\ell_1$-regularized model suggests a data-driven scale selection mechanism. Moreover, Wohlberg and Rodriguez have investigated a model that includes a generalized TV functional and an $\ell_p$ norm, where parameter $p$ is chosen to be greater than or equal to 1 [Wohlberg and Rodriguez, 2007]. In particular, they reported experiments on TV-$\ell_1$ denoising and deblurring. Another novel image denoising strategy named BM3D based on enhanced sparse representation in transform domain was first proposed in [Dabov et al., 2007]. The enhancement of the sparsity is achieved by grouping similar 2D image fragments into 3D data arrays, and a procedure of collaborative filtering is developed to deal with these 3D groups. As such, BM3D was found to offer state-of-the-art denoising performance. In what follows we report a set of experiments where the proposed IRTV algorithm was carried out in comparison with the well-established TV-$\ell_1$ and BM3D algorithms.

As in Experiment 3, the algorithms were applied to a variety of images of size 256 by 256 where each normalized image was corrupted by zero-mean Gaussian noise with standard deviation 0.1. The noisy images were of PSNR 20.00 dB. By passing the standard deviation as a known parameter to the BM3D algorithm, we obtain the denoising results using MATLAB toolbox BM3D. The PSNR of the denoised images obtained by these and IRTV algorithms are given in Table 2.

| Image | PSNR in dB | | | |
|---|---|---|---|---|
| | TV-$\ell_1$ | TV | BM3D | IRTV/$p$ |
| `axe` | 31.09 | 33.03 | 34.75 | **34.80**/0.4 |
| `building` | 26.47 | 27.85 | 29.31 | **29.37**/0.2 |
| `bird` | 30.33 | 31.87 | **33.66** | 33.47/0.4 |
| `circles` | 33.87 | 37.35 | 36.38 | **45.87**/0.4 |
| `church` | 27.10 | 28.42 | **30.35** | 30.09/0.4 |
| `camera` | 26.02 | 27.74 | **29.36** | 29.13/0.6 |
| `crosses` | 23.54 | 29.23 | **39.71** | 39.67/0.4 |
| `jet` | 26.05 | 27.45 | 28.40 | **28.51**/0.6 |
| `liftingbody` | 30.49 | 32.06 | **33.75** | 33.72/0.4 |
| `phantomdisk` | 31.12 | 32.87 | 34.10 | **34.77**/0.2 |
| `pool` | 27.45 | 28.55 | **29.99** | 29.72/0.4 |
| `reschart` | 25.45 | 29.11 | 32.34 | **32.71**/0.4 |
| `satellite` | 25.95 | 27.95 | 29.09 | **29.48**/0.6 |
| `squares` | 37.20 | 42.44 | 41.26 | **52.26**/0.2 |
| `tower` | 26.74 | 27.88 | **28.61** | 27.73/0.4 |
| `text` | 27.31 | 30.38 | 32.18 | **40.23**/0.0 |
| `phantom` | 31.95 | 34.32 | 35.49 | **40.66**/0.2 |

Table 2: Comparisons of TV-$\ell_1$, BM3D and IRTV applied to test images corrupted by Gaussian noise. In each row, the boldfaced numerical value indicates the best PSNR for the image.

From Table 2, it is observed that all three algorithms produce good results in PSNR. The BM3D algorithm was found to outperform TV-$\ell_1$ considerably. Out of eighteen images tested, eight of them were in favor of BM3D, with IRTV as a close second, while the remaining ten images were in favor of IRTV, especially several synthetic images for which IRTV performs significantly better than BM3D.

### 4.5 Experiment 5: Comparisons of IRTV algorithm with TV-$\ell_1$ on denoising images corrupted by speckle noise

The purpose of the experiment is to demonstrate that the IRTV algorithm is also able to denoise images corrupted by speckle noise. In the last decade, the modified TV functional with an $\ell_1$ data fidelity term has attracted much attention [Nikolova, 2002, Chan and Esedoglu, 2005] due to a number of advantages, including superior denoising performance with salt and pepper (speckle) noise, as shown in [Wohlberg and Rodriguez, 2007]. On the other hand, the BM3D algorithm was developed to remove Gaussian noise but not for speckle noise. In the following, we report experimental results where IRTV was applied to images corrupted by speckle noise with performance comparison to that of the TV-$\ell_1$ algorithm.

There are different types of speckle noise namely salt and pepper type of noise and random valued impulse noise [JayaManmadhaRao et al., 2010]. In salt and pepper type of noise, the noisy pixels takes either salt value (gray level 0) or pepper value (gray level 255) which appear as black and white spots on the image respectively. Assume $p$ as the total noise density, we consider salt noise and pepper noise with equal noise density of $p/2$ respectively. In our experiment setting, each image was corrupted by 10% salt and pepper noise, i.e., 5% of the image pixels have gray level 0 and another 5% have gray level 255.

The comparisons are made among the conventional TV denoising, the TV-$\ell_1$ denoising and the IRTV algorithm. As the TV-$\ell_1$ minimization is a convex model, we resort to the CVX model to find its global minimizer. The IRTV algorithm was applied to remove the speckle noise by starting from an initial point based on TV-$\ell_1$ solution. Denoising was carried out on a variety of synthetic and natural images of size 128 by 128 and the performance of the algorithms in PSNR are shown in Table 3. It can be observed that the IRTV method has comparable performance with the TV-$\ell_1$ model in removing speckle noise. In particular for images "squares" and "phantom", almost 1dB improvement in SNR was achieved by the IRTV algorithm.

| Image | PSNR in dB | | | |
|---|---|---|---|---|
| | Noisy | TV | TV-$\ell_1$ | IRTV/$p$ |
| `axe` | 21.28 | 30.39 | 32.60 | **32.83**/0.0 |
| `building` | 20.52 | 26.43 | **31.54** | 31.51/0.0 |
| `bird` | 21.90 | 31.40 | 34.81 | **34.99**/0.0 |
| `circles` | 20.42 | 29.47 | 30.32 | **30.36**/0.0 |
| `church` | 20.31 | 26.44 | 30.09 | **30.13**/0.0 |
| `camera` | 21.42 | 27.09 | **29.87** | 29.85/0.0 |
| `crosses` | 25.29 | 26.06 | 27.66 | **27.72**/0.0 |
| `jet` | 21.09 | 27.33 | **31.12** | 31.11/0.0 |
| `liftingbody` | 21.86 | 31.92 | 34.11 | **34.14**/0.0 |
| `phantomdisk` | 21.01 | 28.83 | 31.07 | **31.18**/0.0 |
| `pool` | 21.46 | 28.79 | **32.00** | 31.97/0.0 |
| `reschart` | 19.35 | 23.79 | **26.82** | 26.80/0.0 |
| `satellite` | 19.58 | 26.67 | **29.92** | 29.81/0.0 |
| `squares` | 21.25 | 34.99 | 39.26 | **40.52**/0.2 |
| `tower` | 21.43 | 29.39 | 33.08 | **33.12**/0.0 |
| `text` | 19.56 | 23.35 | **26.34** | 26.31/0.0 |
| `phantom` | 20.01 | 26.88 | 28.13 | **28.80**/0.4 |

Table 3: Comparisons of TV-$\ell_1$ and IRTV applied to test images corrupted by speckle noise. In each row, the boldfaced numerical value indicates the best PSNR for the image.

### 4.6 Comparisons of IRTV algorithm with MLS-TV

Very recently, a computational framework that combines a regularity based variational method with a moving least squares (MLS) method was proposed as MLS-TV in [Lee et al., 2013]. It is demonstrated that by incorporating a TV minimization formulation into the MLS method, drawbacks of both the standard TV and MLS based denoising schemes may be overcome.

Both the IRTV and MLS-TV algorithms were applied to an image `cameraman` which was normalized to be within the range $[0, 1]$ and was corrupted by zero-mean Gaussian noise with standard deviation 0.0773 (this was the same simulation set-up as in [Lee et al., 2013]). The PSNR of the noisy image was found to be 22.25 dB. With $p = 0.8$, the image denoised by the IRTV algorithm possesses a PSNR of 29.25 dB, while the image produced by the MLS-TV algorithm has a PSNR of 28.86 dB. The denoising results are illustrated in Fig. 9.

## 5 Conclusion

The concept of TV has been generalized to a $p$th power TV and an iteratively reweighted TV algorithm has been proposed to solve nonconvex $\mathrm{TV}_p$-regularized problem in a convex-programming setting for image denoising. Numerical results have indicated with an appropriate power $p < 1$, the proposed IRTV algorithm enhances denoising performance relative to several recent denoising algorithms from the literature. We remark that the concept of generalized TV may also be applied to several other problems including image deconvolution and compressive sampling. We remark that the present paper is focused on image denoising, nevertheless the idea and techniques developed here can be extended to a broad range of applications including signal deconvolution, image deblurring, and other related problems.

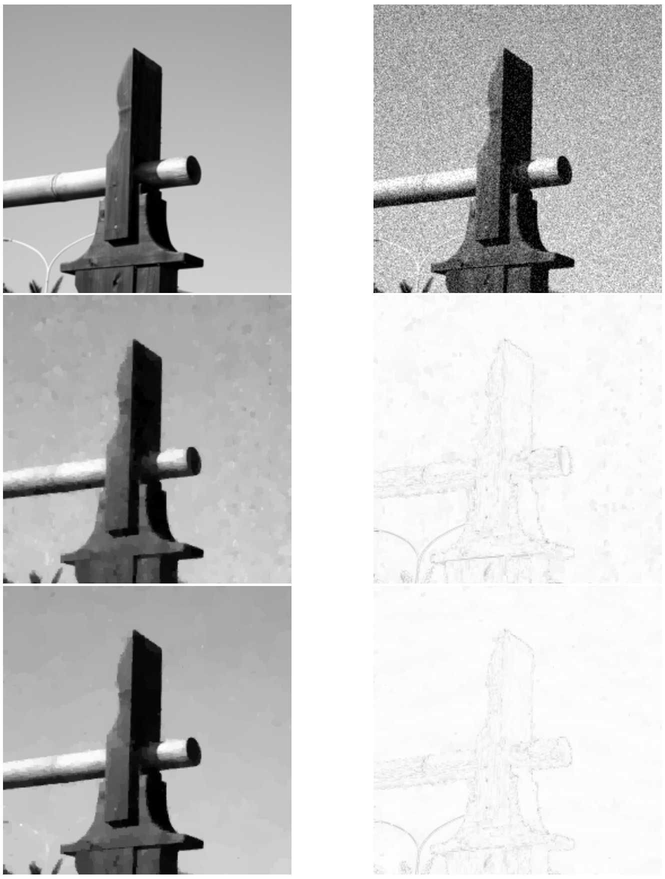

Fig. 4: Denoising image “axe” with IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the TV-$\ell_1$ (d) difference image between the TV-$\ell_1$ denoised image and original image (e) denoised image by the IRTV algorithm (f) difference image between the IRTV denoised image and original image

## References

[Allard, 2008] Allard, W. K. (2008). Total variation regularization for image denoising, ii. examples. *SIAM J. Imaging Sciences*, 1(4):400–417.

[Allard, 2009] Allard, W. K. (2009). Total variation regularization for image denoising, iii. examples. *SIAM J. Imaging Sciences*, 2(2):532–568.

[Beck and Teboulle, 2009] Beck, A. and Teboulle, M. (2009). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Image Processing, IEEE Transactions on*, 18(11):2419–2434.

[Candès et al., 2006] Candès, E., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2):489–509.

[Candes et al., 2008] Candes, E., Wakin, M., and Boyd, S. (2008). Enhancing sparsity by reweighted $\ell_1$ minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905.

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

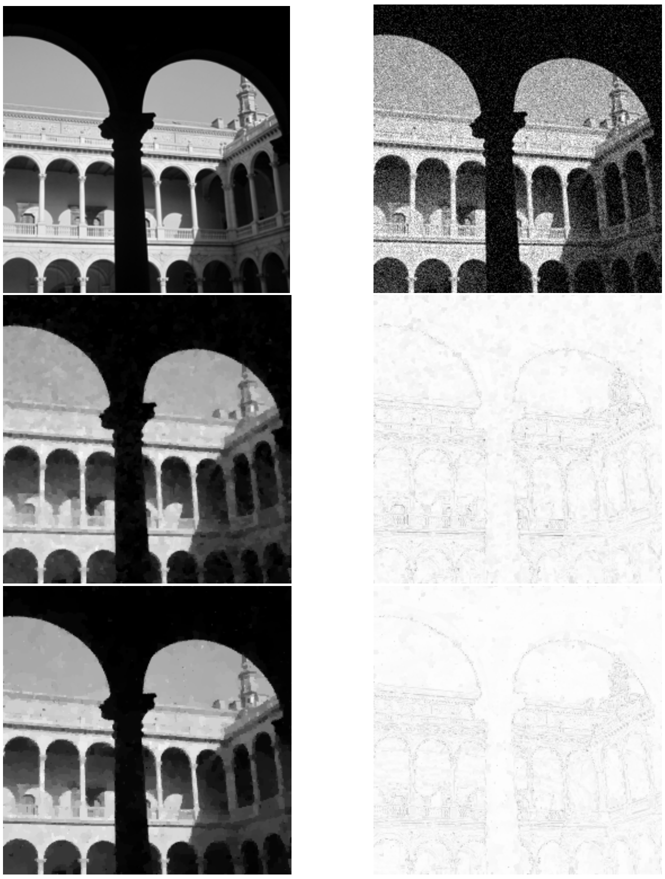

Fig. 5: Denoising image "church" with IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the TV-$\ell_1$ (d) difference image between the TV-$\ell_1$ denoised image and original image (e) denoised image by the IRTV algorithm (f) difference image between the IRTV denoised image and original image

[Chan et al., 2005] Chan, T., Esedoglu, S., Park, F., and Yip, A. (2005). Recent developments in total variation image restoration. *Mathematical Models of Computer Vision*, 17.

[Chan and Esedoglu, 2005] Chan, T. F. and Esedoglu, S. (2005). Aspects of total variation regularized l1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837.

[Chartrand, 2007] Chartrand, R. (2007). Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters, IEEE*, 14(10):707–710.

[Chartrand and Yin, 2008] Chartrand, R. and Yin, W. (2008). Iteratively reweighted algorithms for compressive sensing. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3869–3872. IEEE.

[Chen et al., 2010] Chen, Q., Montesinos, P., Sun, Q. S., Heng, P. A., and Xia, D. S. (2010). Adaptive total variation denoising based on difference curvature. *Image and Vision Computing*, 28(3):298–306.

[Chen et al., 2001] Chen, S., Donoho, D., and Saunders, M. (2001). Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159.

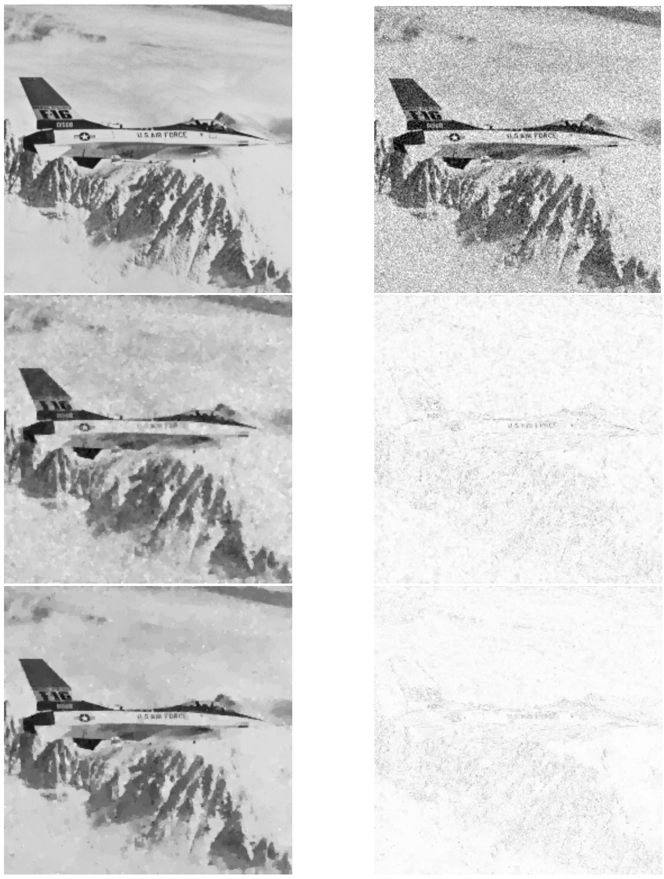

Fig. 6: Denoising image "jet" with IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the TV-$\ell_1$ (d) difference image between the TV-$\ell_1$ denoised image and original image (e) denoised image by the IRTV algorithm (f) difference image between the IRTV denoised image and original image

[Chopra and Lian, 2010] Chopra, A. and Lian, H. (2010). Total variation, adaptive total variation and nonconvex smoothly clipped absolute deviation penalty for denoising blocky images. *Pattern Recognition*, 43(8):2609–2619.

[Dabov et al., 2007] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on*, 16(8):2080–2095.

[Foucart and Lai, 2009] Foucart, S. and Lai, M. (2009). Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q < 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407.

[Fu et al., 2006] Fu, H., Ng, M. K., Nikolova, M., and Barlow, J. L. (2006). Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration. *SIAM Journal on Scientific computing*, 27(6):1881–1902.

[Goldstein and Osher, 2009] Goldstein, T. and Osher, S. (2009). The split bregman method for l1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.

[Grant and Boyd, 2008] Grant, M. and Boyd, S. (2008). Graph implementations for nonsmooth convex programs. In Blondel, V., Boyd, S., and Kimura, H., editors, *Recent Advances*

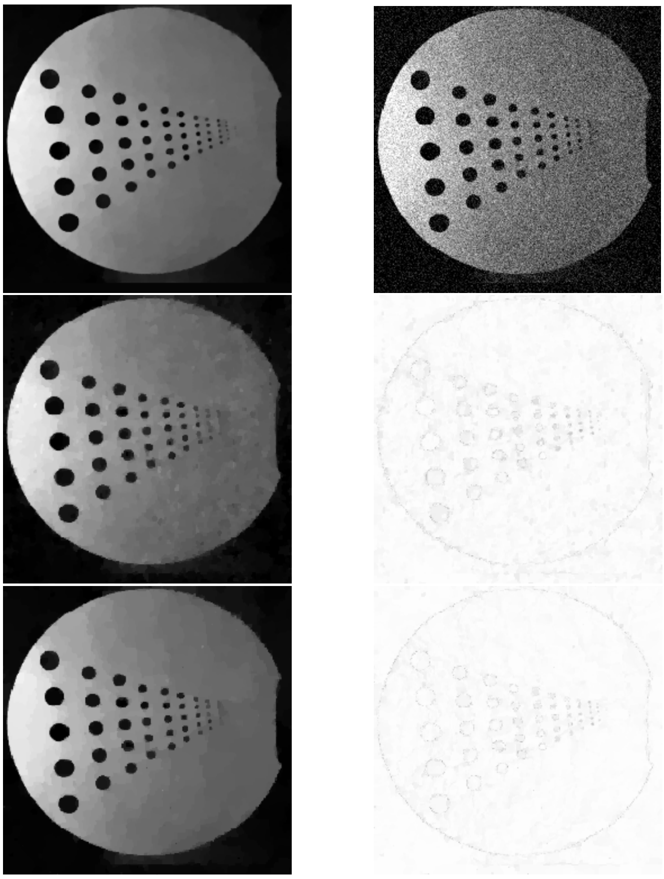

Fig. 7: Denoising image "phantomdisk" with IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the TV-$\ell_1$ (d) difference image between the TV-$\ell_1$ denoised image and original image (e) denoised image by the IRTV algorithm (f) difference image between the IRTV denoised image and original image

*in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited. `http://stanford.edu/~boyd/graph_dcp.html`.

[Grant and Boyd, 2012] Grant, M. and Boyd, S. (2012). CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`.

[Hu and Jacob, 2012] Hu, Y. and Jacob, M. (2012). Higher degree total variation (hdtv) regularization for image recovery. *Image Processing, IEEE Transactions on*, 21(5):2559–2571.

[JayaManmadhaRao et al., 2010] JayaManmadhaRao, M., Anuradha, S., and Reddy, K. (2010). Impulse noise removal in digital images. *International Journal of Computer Applications*, 10(8).

[Larson, 2013] Larson, R. (2013). *Elementary Linear Algebra, 7th Edition*. Brooks/Cole, Cengage Learning.

[Lee et al., 2013] Lee, Y. J., Lee, S., and Yoon, J. (2013). A framework for moving least squares method with total variation minimizing regularization. *Journal of Mathematical Imaging and Vision*. DOI 10.1007/s10851-013-0428-5.

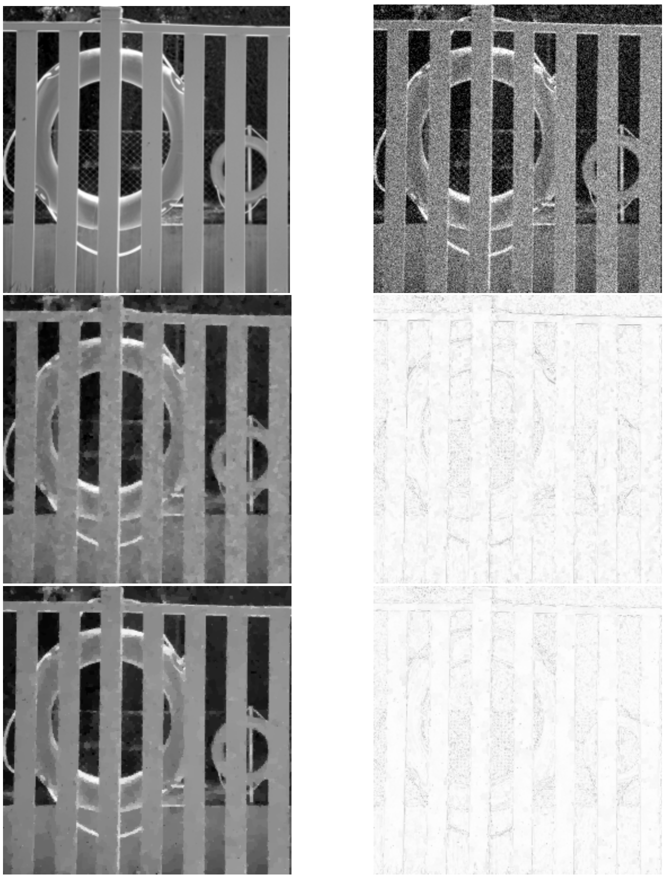

Fig. 8: Denoising image “fence” with IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the TV-$\ell_1$ (d) difference image between the TV-$\ell_1$ denoised image and original image (e) denoised image by the IRTV algorithm (f) difference image between the IRTV denoised image and original image

[Louchet and Moisan, 2011] Louchet, C. and Moisan, L. (2011). Total variation as a local filter. *SIAM Journal on Imaging Sciences*, 4(2):651–694.

[Mallat, 2009] Mallat, S. (2009). *A Wavelet Tour of Signal Processing: The Sparse Way, Third Edition.* Elsevier Inc.

[Nikolova, 2002] Nikolova, M. (2002). Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994.

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

[Wang et al., 2008] Wang, Y., Yang, J., Yin, W., and Zhang, Y. (2008). A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272.

[Wohlberg and Rodriguez, 2007] Wohlberg, B. and Rodriguez, P. (2007). An iteratively reweighted norm algorithm for minimization of total variation functionals. *Signal Processing Letters, IEEE*, 14(12):948–951.

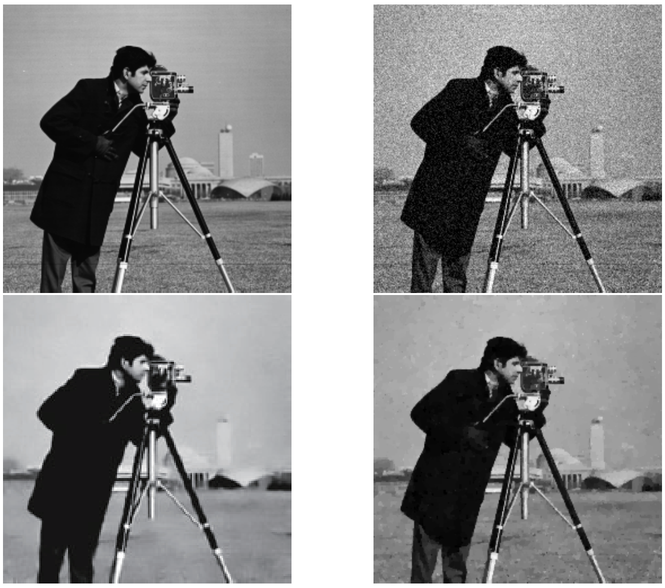

Fig. 9: Denoising image "cameraman" with MLS-TV and IRTV (left to right, up to bottom) (a) original image (b) noisy image (c) denoised image by the MLS-TV algorithm (d) denoised image by the IRTV algorithm

[Yan and Lu, 2011a] Yan, J. and Lu, W.-S. (2011a). New algorithms for sparse representation of discrete signals based on $\ell_p$-$\ell_2$ optimization. In *Proc. IEEE PacRim Conf. 2011*, pages 73–78.

[Yan and Lu, 2011b] Yan, J. and Lu, W.-S. (2011b). Power-iterative strategy for $\ell_p$-$\ell_2$ optimization for compressive sensing: towards global solution. In *Proc. IEEE Asilomar Conf. 2011*, pages 1153–1157.

[Yan and Lu, 2012] Yan, J. and Lu, W.-S. (2012). Smoothed $\ell_p$-$\ell_2$ Solvers for Signal Denoising. In *Proc. IEEE ICASSP Conf. 2012*, pages 3801–3804.

[Zibulevsky and Elad, 2010] Zibulevsky, M. and Elad, M. (2010). L1-L2 optimization in signal and image processing. *Signal Processing Magazine, IEEE*, 27(3):76–88.