A WEIGHTED DIFFERENCE OF ANISOTROPIC AND ISOTROPIC TOTAL VARIATION MODEL FOR IMAGE PROCESSING

YIFEI LOU*, TIEYONG ZENG†, STANLEY OSHER‡, AND JACK XIN§

Abstract. We propose a weighted difference of anisotropic and isotropic total variation (TV) as a regularization for image processing tasks, based on the well-known TV model and natural image statistics. Due to the difference form of our model, it is natural to compute via a difference of convex algorithm (DCA). We draw its connection to the Bregman iteration for convex problems, and prove that the iteration generated from our algorithm converges to a stationary point with the objective function values decreasing monotonically. A stopping strategy based on the stable oscillatory pattern of the iteration error from the ground truth is introduced. In numerical experiments on image denoising, image deblurring, and magnetic resonance imaging (MRI) reconstruction, our method improves on the classical TV model consistently, and is on par with representative start-of-the-art methods.

Key words. Anisotropic TV, Isotropic TV, Weighted Difference, Difference of Convex Algorithm, Convergence to Stationary Points, Stable Oscillatory Errors, Bregman and Split Bregman Iterations.

AMS subject classifications. 90C90, 65K10, 49N45, 49M20

1. Introduction.

Many image processing tasks can be formulated as an inverse problem, in which the data f is assumed to be obtained approximately by applying a linear operator A on an image u with additive noise. For example, A is the identity matrix for image denoising, a convolution matrix for deblurring, and subsampling of Fourier transform for a magnetic resonance image (MRI) reconstruction problem. In most scenarios, solving u from $Au = f$ is ill-posed in the sense that directly inverting A would result in bad and possibly multiple solutions. It is necessary and even desirable to constrain the solutions through regularization, with the help of prior knowledge of images that one wants to reconstruct. A general model for such inverse problem is

$$\hat{u} := \text{argmin}_u J(u) + \frac{\mu}{2}\|Au - f\|_2^2 \ , \tag{1.1}$$

where $J(u)$ is the regularization term, μ is a positive parameter to balance $J(u)$ and the data fidelity term $\|Au - f\|_2^2$, and $\hat{u}$ is an optimal solution of the model or a reconstructed result. A classical regularization is the total variation (TV) proposed by Rudin-Osher-Fatemi [33]. It is widely used in image processing applications, such as deconvolution [7,16,25], inpainting [6] and super-resolution [26], just to name a few. The TV model originated in [33] is isotropic, and later an anisotropic formulation has been addressed in the literature [10, 30] among others. We give mathematical definition for both the isotropic and anisotropic TV in the discrete setting. Denoting u as the column vector by a lexicographical ordering of a 2D image, we have

$$J_{iso}(u) := \|\nabla u\|_2 = \|\sqrt{|D_x u|^2 + |D_y u|^2}\|_1, \tag{1.2}$$

$$J_{ani}(u) := \|D_x u\|_1 + \|D_y u\|_1 \ , \tag{1.3}$$

where D_x, D_y denote the horizontal and vertical partial derivative operators. Throughout this paper, we shall use notations $\|\nabla u\|_2$ and $\|\sqrt{|D_x u|^2 + |D_y u|^2}\|_1$ interchangeably.

Another interpretation of TV can be given from the perspective of compressive sensing (CS) [3, 12], which is to reconstruct a signal from an under-determined system provided that the signal is

---

*Department of Mathematical Sciences, University of Texas at Dallas, Dallas, TX 75080. Email: yifei.lou@utdallas.edu. YL is partially supported by NSF grants DMS-0928427 and DMS-1222507.

† Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. Email: zeng@hkbu.edu.hk. TZ is partially supported by NSFC 11271049, RGC 211911, 12302714 and RFGs of HKBU.

‡Department of Mathematics, UCLA, Los Angeles, CA 90095. Email: sjo@math.ucla.edu. SO is supported by the Keck Foundation, ONR N000141410683, N000141110749, and NSF DMS-1118971.

§Department of Mathematics, UC Irvine, Irvine, CA 92697. Email: jxin@math.uci.edu. JX is partially supported by NSF grants DMS-0928427 and DMS-1222507.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE
A Weighted Difference of Anisotropic and Isotropic Total Variation Model for Image Processing

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles,Department of Mathematics,Los Angeles,CA,90095

8. PERFORMING ORGANIZATION REPORT NUMBER
CAM14-69

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We propose a weighted di erence of anisotropic and isotropic total variation (TV) as a regularization for image processing tasks, based on the well-known TV model and natural image statistics. Due to the di erence form of our model, it is natural to compute via a di erence of convex algorithm (DCA). We draw its connection to the Bregman iteration for convex problems, and prove that the iteration generated from our algorithm converges to a stationary point with the objective function values decreasing monotonically. A stopping strategy based on the stable oscillatory pattern of the iteration error from the ground truth is introduced. In numerical experiments on image denoising, image deblurring, and magnetic resonance imaging (MRI) reconstruction, our method improves on the classical TV model consistently, and is on par with representative start-of-the-art methods.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**17**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

sufficiently sparse or sparse in a transform domain. For example, a natural image is mostly sparse after taking gradient. Mathematically, it amounts to minimizing the L_0 norm of the image gradient, *i.e.*, $J(u) = \|\nabla u\|_0$. To bypass the NP-hard L_0 norm, the convex relaxation approach in CS is to replace L_0 by L_1, and L_1 on the gradient is the total variation. The restricted isometry property (RIP) condition [3] theoretically guarantees the exact recovery of sparse solutions by L_1. The RIP regime is where the sensing matrix is *incoherent*, such as a random Gaussian matrix. Several non-convex penalties have been proposed and studied as alternatives to L_1, [19]. A few notable examples are L^p for $p \in (0,1)$ [8,21,39], L_1/L_2 (scale invariant L_1) and $L_1 - L_2$ [13,22,23,40,41]. In particular, $L_1 - L_2$ penalty is found to be the best among existing methods for recovering sparse solutions when the sensing matrix is highly coherent or significantly violating the RIP condition [23,41].

The TV-regularization has been a very active research topic in the past two decades. Though a gradient descent approach in the original paper can be slow to converge, a projection algorithm is later proposed by Chambolle [5] to speed up convergence based on duality. More recently, the Bregman and split Bregman methodology [9,15,29] offers another line of fast algorithms equivalent to the role of alternating direction method of multipliers (ADMM) and Douglas-Rachford splitting algorithm in the optimization literature dating back to the 1970's. There are also a few approaches to solve the L_0 minimization directly. In [38], a special alternating minimization strategy with half-quadratic splitting is adopted for image smoothing. Image restoration via L_0 is considered in [31], which uses hard shrinkage for L_0 as opposed to soft shrinkage for L_1. In addition, the L_0 on the gradient can be interpreted as the length of the partition boundaries, which leads to the classical Potts model [32] or piece-wise constant Mumford-Shah model [28] for image segmentation or partition. Recently, Storath *et. al.* [34] propose a hybrid ADMM and dynamic programming method to solve the Potts model.

Motivated from $L_1 - L_2$ minimization of coherent CS [23,41], we propose the following weighted difference of convex regularization,

$$J(u) := J_{ani} - \alpha J_{iso} = \|D_x u\|_1 + \|D_y u\|_1 - \alpha \|\sqrt{|D_x u|^2 + |D_y u|^2}\|_1 \ , \tag{1.4}$$

where $\alpha \in [0,1]$ is a parameter for a more general model. When $\alpha = 1$, $J(u)$ is to apply $L_1 - L_2$ on the gradient vector. Two advantages of $L_1 - L_2$ over other nonconvex measures are its Lipschitz regularity, and guaranteed convergence via the difference of convex algorithm (DCA) [35,36]. We find that the DCA requires solving the L_1 type of minimization as a subproblem, which can be handled efficiently by utilizing the split Bregman technique. We prove that the DCA approach converges to stationary points, a typical situation for nonconvex problems. In practice, the DCA iterations, when properly stopped, are often close to global minima and produce excellent results. The stopping issue is discussed later based on the oscillatory pattern of the iteration errors.

The rest of the paper is organized as follows. Section 2 describes our model in detail including numerical algorithms and convergence analysis. Section 3 is devoted to numerical experiments, where three image processing applications (denoising, deblurring and MRI reconstruction) are examined. Finally, discussions and conclusions are given in Section 4 and Section 5 respectively.

2. Our model.

Let (u_{jx}, u_{jy}) be gradient vector at pixel j. Then equation (1.4) can be rewritten as

$$J(u) = \sum_j \left(|u_{jx}| + |u_{jy}| - \alpha \sqrt{u_{jx}^2 + u_{jy}^2} \right) \ . \tag{2.1}$$

This point-wise formulation suggests that sparsity is enforced on every gradient vector. More specifically, we encourage the gradient to be 1-sparse at every pixel, which implies that horizontal or vertical edges are more preferable in this model. In order to understand the image gradient and 1-sparsity, we plot the histogram of gradient angles over the range of $[0, 90]$ degree in Figure 2.1 for a large number of natural images. The angle distribution in other quadrants is similar. As shown in Figure 2.1, the

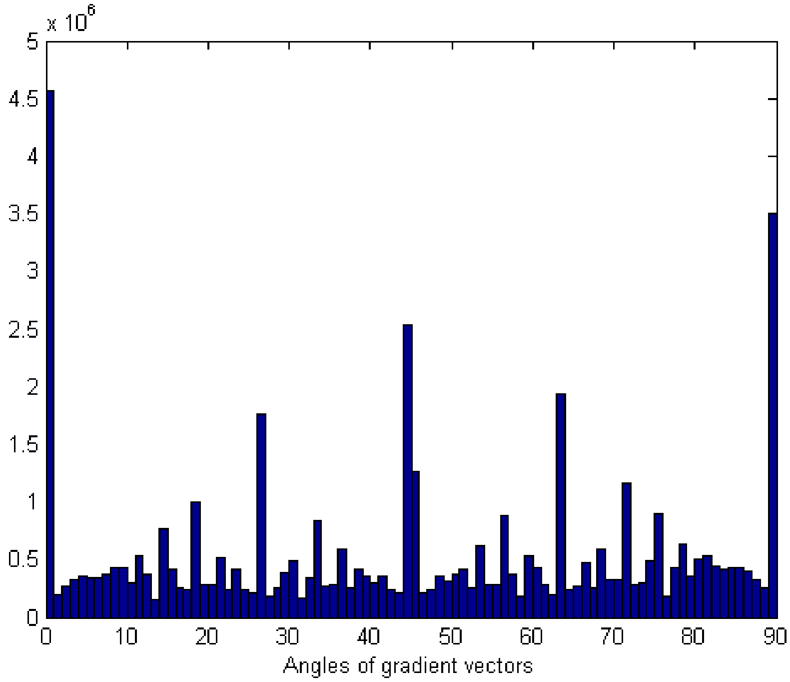

FIG. 2.1. *The histogram of gradient angles over 300 images from Berkeley segmentation dataset [27]. Two largest peaks are at 0 and 90 degrees, indicating that gradient vectors are mostly 1-sparse.*

two largest peaks are at 0 and 90 degrees, which implies that gradient vectors are 1-sparse at a fairly good chance, with non-sparse occurrences also at positive probability. Hence we insert a constant α in (1.4) to reflect such behavior in the histogram. In Figure 2.2, we plot the level lines of L_0 norm on the gradient, whose value is 0 at origin, 1 at axes, and 2 elsewhere. The level lines corresponding to $\alpha < 1$ in (1.4) is closer to L_0 than that of $\alpha = 1$ in the sense that the latter yields 0 at both axes.

Let us derive the value of α based on the gradient distribution. Suppose that the gradient value $D_x u$ follows the distribution [21], $\frac{p}{2\Gamma(\frac{1}{p})} e^{-|x|^p}$ where $\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x}$. It is Gaussian distribution for $p = 2$, Laplacian distribution for $p = 1$, and hyper-Laplacian for $0 < p < 1$. We have

$$E_1 = E|D_x u| = \frac{p}{2\Gamma\left(\frac{1}{p}\right)} \int_{-\infty}^{+\infty} e^{-|x|^p} |x| \mathrm{d}x = \frac{1}{\Gamma(\frac{1}{p})} \int_0^{+\infty} e^{-t} t^{\frac{2}{p}-1} \mathrm{d}t = \frac{\Gamma(\frac{2}{p})}{\Gamma(\frac{1}{p})} , \tag{2.2}$$

$$E_2 = E|D_x u|^2 = \frac{p}{2\Gamma\left(\frac{1}{p}\right)} \int_{-\infty}^{+\infty} e^{-|x|^p} |x|^2 \mathrm{d}x = \frac{1}{\Gamma(\frac{1}{p})} \int_0^{+\infty} e^{-t} t^{\frac{3}{p}-1} \mathrm{d}t = \frac{\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})} . \tag{2.3}$$

The value of α corresponds to the ratio of L_1 and L_2, *i.e.*,

$$\alpha = \frac{E_1}{\sqrt{E_2}} = \frac{\Gamma(2/p)}{\sqrt{\Gamma(3/p)\Gamma(1/p)}} . \tag{2.4}$$

Table 2.1 lists the values of α based on gradient distributions for $p = 0.5, 1, 2$. We analyze the gradient distribution in Figure 2.3 which shows that the distribution of image gradient data matches the $p = 1/2$ distribution better than classical Gaussian ($p = 2$) or Laplacian ($p = 1$) distribution. This observation is consistent with the choice of hyper-Laplacian [4, 21] for image processing ($p \in [0.5, 0.8]$). In the rest of the paper, we shall fix the weighting coefficient $\alpha = 1/2$ to approximate the desired value in Table 2.1.

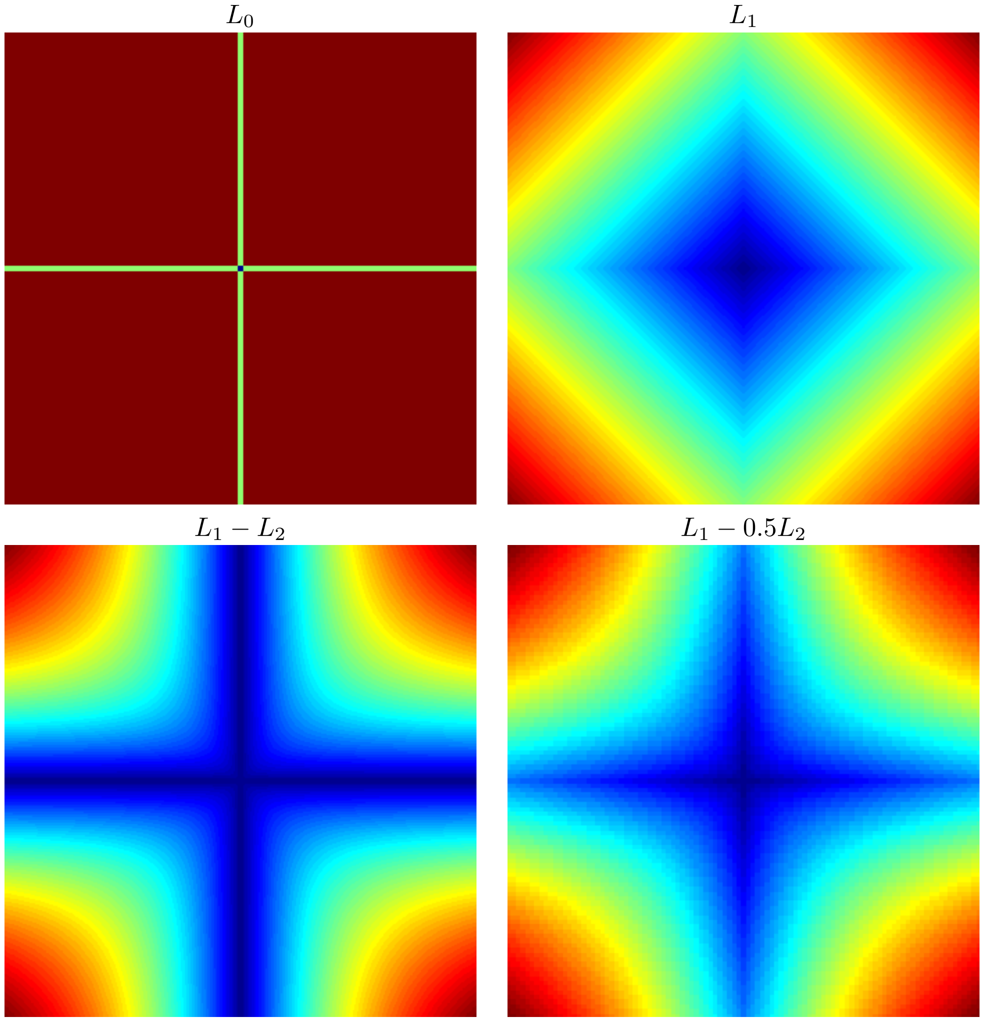

FIG. 2.2. *Level curves of different metrics. The level lines corresponding to $\alpha < 1$ in (1.4) is closer to L_0 than that of $\alpha = 1$ in the sense that the latter yields 0 at both axes.*

TABLE 2.1
The value of α based on the gradient distribution.

p	α
0.5	0.5477
1	0.7071
2	0.7979

2.1. Numerical algorithms. To solve (1.1) with $J(u)$ defined in (1.4), we apply the technique of difference of convex algorithm (DCA) by linearizing the isotropic term

$$u^{n+1} = \arg\min_u \|D_x u\|_1 + \|D_y u\|_1 - \alpha\langle \nabla u, q^n\rangle + \frac{\mu}{2}\|Au - f\|_2^2 \,, \tag{2.5}$$

for $q^n = (q_x^n, q_y^n) = (D_x u^n, D_y u^n)/\sqrt{|D_x u^n|^2 + |D_y u^n|^2}$ at step u^n. Note that q^n is a point-wise calculation; and if the denominator is zero at some point, the corresponding q^n value is set to be zero. Each DCA subproblem, eq. (2.5), amounts to solving a TV type of minimization. We employ the split

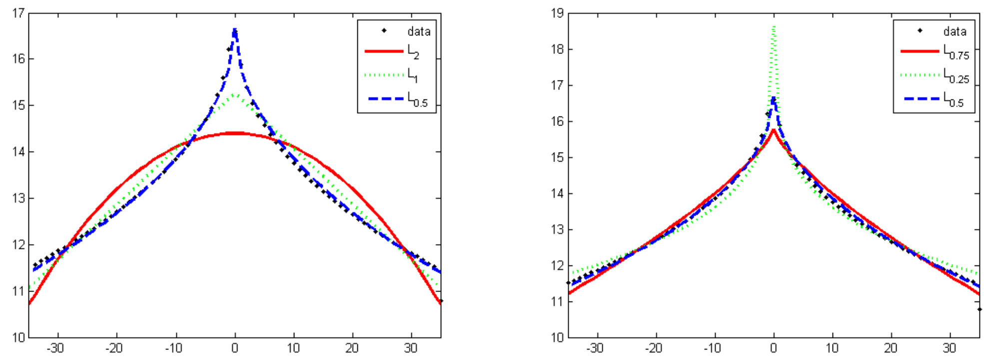

FIG. 2.3. *The plot of log probability v.s. gradient in comparison with different distributions, indicating that the gradient distribution of a large natural image dataset matches $L_{1/2}$ or the $p = 1/2$ hyper-Laplacian distribution better than classical Gaussian or Laplacian distribution.*

Bregman technique [15] to do the job. Specifically, we introduce two auxiliary variables and split the anisotropic term in the following way,

$$\begin{aligned} u^{k+1} = \arg \min_{u,d_x,d_y} \; & \|d_x\|_1 + \|d_y\|_1 - \alpha(d_x^T \cdot q_x^n + d_y^T \cdot q_y^n) \\ & + \frac{\mu}{2}\|Au - f\|_2^2 + \frac{\lambda}{2}\|d_x - D_x u\|_2^2 + \frac{\lambda}{2}\|d_y - D_y u\|_2^2 \,, \end{aligned} \tag{2.6}$$

in which d_x, d_y can be updated via soft shrinkage, defined as

$$\text{shrink}(s, \gamma) = \text{sgn}(s) \max\{|s| - \gamma, 0\} \,. \tag{2.7}$$

The pseudo-code is summarized in Algorithm 1. The algorithm is efficient for many applications where the matrix to be inverted is diagonal or can be diagonalized by Fourier transform, which is true for image denoising, deconvolution, and MRI reconstruction.

Algorithm 1 for solving unconstrained problem (2.5)

Define $u = q_x = q_y = 0, z = f$ and $MaxDCA, MaxBregman$
for 1 **to** $MAXDCA$ **do**
  $b_x = b_y = 0$
  for 1 **to** $MAXBregman$ **do**
    $u = (\mu A^T A - \lambda \triangle)^{-1}(\mu \mathcal{A} z + \lambda D_x^T(d_x - b_x) + \lambda D_y^T(d_y - b_y))$
    $d_x = \text{shrink}(D_x u + b_x + \alpha q_x/\lambda, 1/\lambda)$
    $d_y = \text{shrink}(D_y u + b_y + \alpha q_y/\lambda, 1/\lambda)$
    $b_x = b_x + D_x u + d_x$
    $b_y = b_y + D_y u + d_y$
  end for
  $(q_x, q_y) = (D_x u, D_y u)/\sqrt{|D_x u|^2 + |D_y u|^2}$
end for

For the corresponding constrained problem,

$$\min \|D_x u\|_1 + \|D_y u\|_1 - \alpha\|\nabla u\|_2 \qquad \text{s.t.} \qquad Au = f \,, \tag{2.8}$$

the DCA is expressed as

$$u^{n+1} = \arg\min_u \left\{ \|D_x u\|_1 + \|D_y u\|_1 - \alpha\langle \nabla u, q^n \rangle \quad \text{s.t.} \quad Au = f \right\} \,. \tag{2.9}$$

Each DCA subproblem could be reduced to a sequence of unconstrained problems of the form

$$u_{k+1} = \arg\min_u \|D_x u\|_1 + \|D_y u\|_1 - \alpha\langle \nabla u, q^n\rangle + \frac{\mu}{2}\|Au - z_k\|_2^2 \ , \tag{2.10}$$

$$z_{k+1} = z_k + f - Au_{k+1} \ . \tag{2.11}$$

Again the first equation can be solved by the split Bregman method. Algorithm 2 for solving the constrained problem (2.9) is almost the same as Algorithm 1, except for an additional update on z.

Algorithm 2 for solving constrained problem (2.9)

Define $u = q_x = q_y = 0, z = f$ and $MaxDCA, MaxBregmanInner, MaxBregmanOuter$
for 1 **to** $MAXDCA$ **do**
  $b_x = b_y = 0$
  for 1 **to** $MaxBregmanOuter$ **do**
    for 1 **to** $MAXBregmanInner$ **do**
      $u = (\mu A^T A - \lambda\triangle)^{-1}(\mu\mathcal{A}z + \lambda D_x^T(d_x - b_x) + \lambda D_y^T(d_y - b_y))$
      $d_x = \text{shrink}(D_x u + b_x + \alpha q_x/\lambda, 1/\lambda)$
      $d_y = \text{shrink}(D_y u + b_y + \alpha q_y/\lambda, 1/\lambda)$
      $b_x = b_x + D_x u + d_x$
      $b_y = b_y + D_y u + d_y$
    end for
    $z = z + f - Au$
  end for
  $(q_x, q_y) = (D_x u, D_y u)/\sqrt{|D_x u|^2 + |D_y u|^2}$
end for

2.2. Convergence analysis. We want to show that the sequence of $\{u^n\}$ obtained from the DCA iterations, *i.e.*, eq. (2.5), converges to a stationary point. The standard DCA requires strong convexity to prove convergence [35], here we can get rid of this requirement using the fact that L_1 is convex (not strictly though), and its subgradient[1] is a close set. We first prove two lemmas saying that the objective function is coercive and monotonically non-increasing for the minimizing sequence; and then complete the convergence proof.

LEMMA 2.1. *Suppose $\mu > 0, 0 < \alpha < 1$, and $ker(A)\bigcap ker(D) = \{\mathbf{0}\}$, where $D = [D_x; D_y]$. Then the objective function*

$$F(u) := \|D_x u\|_1 + \|D_y u\|_1 - \alpha\|\sqrt{|D_x u|^2 + |D_y u|^2}\|_1 + \frac{\mu}{2}\|Au - f\|_2^2 \ ,$$

is coercive.

Proof. It suffices to show that for any fixed $u \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, $F(\gamma u) \to \infty$ as $\gamma \to \infty$. We discuss two cases separately.

- If $u \notin ker(D)$, then we have

$$F(\gamma u) > (1-\alpha)\gamma(\|D_x u\|_1 + \|D_y u\|_1) \to \infty \quad \text{as} \quad \gamma \to \infty \ ,$$

  since $\|\sqrt{|D_x u|^2 + |D_y u|^2}\|_1 < \|D_x u\|_1 + \|D_y u\|_1$.
- If $u \in ker(D)$, then $Au \neq 0$, since $ker(A)\bigcap ker(D) = \{\mathbf{0}\}$ and $u \neq 0$. Therefore, we have

$$F(\gamma u) \geqslant \frac{\mu}{2}\|\gamma Au - f\|_2^2 \geqslant \frac{\mu}{2}(\gamma\|Au\|_2^2 - \|f\|_2^2) \to \infty \quad \text{as } \gamma \to \infty \ .$$

[1]We say a vector g is a *subgradient* of f at $x \in \mathbf{dom}(f)$ if $f(z) \geqslant f(x) + g^T(z - x)$ for all $z \in \mathbf{dom}(f)$.

□

Lemma 2.2. *If the sequence $\{u^n\}$ is generated by the DCA algorithm (2.5),* i.e.,

$$u^{n+1} = \arg\min \|D_x u\|_1 + \|D_y u\|_1 - \alpha\langle \frac{\nabla u^n}{\|\nabla u^n\|}, \nabla u\rangle + \frac{\mu}{2}\|Au - f\|_2^2 \ ,$$

then

$$F(u^n) - F(u^{n+1}) \geqslant 0. \tag{2.12}$$

Proof. It follows from the first-order optimality condition at u^{n+1} that there exist $p^{n+1} \in \partial\|Du^{n+1}\|_1$ such that

$$p^{n+1} - \alpha q^n + \mu A^T(Au^{n+1} - f) = 0 \ . \tag{2.13}$$

A simple calculation shows that

$$\begin{aligned} & F(u^n) - F(u^{n+1}) \\ = & \frac{\mu}{2}\|A(u^n - u^{n+1})\|_2^2 + \mu\langle A(u^n - u^{n+1}), Au^{n+1} - f\rangle + \|Du^n\|_1 - \|Du^{n+1}\|_1 - \alpha(\|\nabla u^n\|_2 - \|\nabla u^{n+1}\|_2) \\ = & \frac{\mu}{2}\|A(u^n - u^{n+1})\|_2^2 - \langle p^{n+1} - \alpha q^n, u^n - u^{n+1}\rangle + \|Du^n\|_1 - \|Du^{n+1}\|_1 - \alpha(\|\nabla u^n\|_2 - \|\nabla u^{n+1}\|_2) \\ = & \frac{\mu}{2}\|A(u^n - u^{n+1})\|_2^2 + (\|Du^n\|_1 - \langle p^{n+1}, u^n\rangle) + \alpha(\|\nabla u^{n+1}\|_2 - \langle q^n, \nabla u^{n+1}\rangle) \ . \end{aligned}$$

The second equality above is obtained from left multiplying (2.13) by $(u^n - u^{n+1})^T$, and third one uses the fact that $\langle p^{n+1}, u^{n+1}\rangle = \|Du^{n+1}\|_1$.

The chain rule of subgradient [18] suggests that $\partial\|Du\|_1 = D^T\partial\|Du\|_1$, where

$$\partial|r|_1 = \begin{cases} [-1,1] & r = 0 \ , \\ \mathrm{sign}(r) & \text{otherwise} \ . \end{cases}$$

It implies that $\|Du^n\|_1 - \langle p^{n+1}, u^n\rangle = \|Du^n\|_1 - \langle p^{n+1}, Du^n\rangle \geqslant 0$, since $p^{n+1} \leqslant 1$ for each component. Using the definition of subgradient and the fact that $q^n \in \partial\|\nabla u\|_2$, we have $\|\nabla u^{n+1}\|_2 - \langle q^n, \nabla u^{n+1}\rangle \geqslant 0$, which concludes the proof. □

Theorem 2.3. *Under the assumptions in Lemma 2.1, any non-zero limit point u^* of $\{u^n\}$ satisfies the first-order optimality condition, which means u^* is a stationary point.*

Proof. It follows from Lemma 2.1 and Lemma 2.2 that the objective function F is coercive, monotonically decreasing, and convergent. As a result, the sequence $\{u^n\}$ is bounded, and

$$\|A(u^n - u^{n+1})\|_2 \to 0 \ , \tag{2.14}$$

$$\|\nabla u^{n+1}\|_2 - \langle q^n, \nabla u^{n+1}\rangle \to 0 \ . \tag{2.15}$$

We start our algorithm from $u^0 = \mathbf{0}$, and u^1 is obtained from

$$u^1 = \arg\min \|Du\|_1 + \frac{\mu}{2}\|Au - f\|_2^2 \ .$$

If u^1 is a constant vector, we stop our algorithm; otherwise $F(u^1) \leqslant F(v)$ for any constant vector v. Since F monotonically decreases, any subsequent solution u^n for $n \geqslant 1$ is not constant. Then we can properly define

$$c^n := \frac{\langle \nabla u^n, \nabla u^{n+1}\rangle}{\|\nabla u^n\|_2\|\nabla u^{n+1}\|_2} \ .$$

Eq. (2.15) suggests that $(1-c^n)\|\nabla u^{n+1}\| \to 0$. Therefore, $c^n \to 1$. It follows from the minimizing sequence that $\|\nabla u^n\|_1 - \|\nabla u^{n+1}\|_1 \to 0$, and so $\nabla(u^n - u^{n+1}) \to 0$. Combining (2.14) and $ker(A) \bigcap ker(D) = \{\mathbf{0}\}$, we get $u^n - u^{n+1} \to 0$.

This implies that there exists a subsequence of $\{u^n\}$ converging to u^*, denoted as $\{u^{n_k}\}$. The optimality condition at the n_k-th step of DCA reads

$$\mathbf{0} \in \partial\|Du^{n_k}\|_1 + \alpha\nabla \cdot \frac{\nabla u^{n_k-1}}{\|\nabla u^{n_k-1}\|_2} + \mu A^T(Au^{n_k} - f)\ , \tag{2.16}$$

or

$$-\alpha\nabla \cdot \frac{\nabla u^{n_k-1}}{\|\nabla u^{n_k-1}\|_2} - \mu A^T(Au^{n_k} - f) \in \partial\|Du^{n_k}\|_1\ . \tag{2.17}$$

We can show Du^{n_k} converges to Du^*, as

$$\|Du^{n_k} - Du^*\| \leqslant \|D\| \cdot \|u^{n_k} - u^*\| \to 0 \quad \text{as} \quad n_k \to \infty\ .$$

When n_k is sufficiently large, $\text{supp}(Du^*) \subseteq \text{supp}(Du^{n_k})$ and $\text{sign}(Du^{n_k}) = \text{sign}(Du^*)$. Using the chain rule of subgradient, we have $\partial|Du^{n_k}|_1 \subseteq \partial|Du^*|_1$, and $D^T\partial|Du^{n_k}|_1 \subseteq D^T\partial|Du^*|_1$. Consequently, it follows from (2.17) that

$$-\nabla \cdot \frac{\nabla u^{n_k-1}}{\|\nabla u^{n_k-1}\|} - \mu A^T(Au^{n_k} - f) \in D|Du^*|_1\ .$$

We assume $\frac{\nabla u}{\|\nabla u\|} \doteq 0$ if $\|\nabla u\| = 0$ at some points. Letting $n_k \to \infty$, we obtain

$$-\nabla \cdot \frac{\nabla u^*}{\|\nabla u^*\|} - \mu A^T(Au^* - f) \in D|Du^*|_1\ ,$$

which means that u^* satisfies the first-order optimality condition. □

3. Experiments. We apply the proposed method to three applications: image denoising, deconvolution, and the MRI construction. The matrix A in these examples can be diagonalized by Fourier transform, and hence Algorithm 1 or Algorthm 2 can be efficiently implemented. We compare L_1 and $L_1 - \alpha L_2$ for $\alpha = 0.5$ or 1 with some existing methods, such as L_0 for image smoothing in [38], L_0 in [31], L_p for $p = 2/3$ in [21], and $L_1 + L_2^2$ in [2] for image deblurring. We use structural similarity (SSIM) index [37] as a quantitative measure for image quality. Let us first define *local* similarity index computed on windows x and y,

$$\text{ssim}(x, y) := \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}\ , \tag{3.1}$$

where μ_x, μ_y are the average of x, y, σ_x^2, σ_y^2 are the variance, σ_{xy} is covariance of x, y, and c_1, c_2 are two variables to stabilize the division with weak denominator. The overall SSIM is the mean of local similarity indexes, *i.e.*,

$$\text{SSIM}(X, Y) := \frac{1}{N}\sum_{i=1}^{N} \text{ssim}(x_i, y_i)\ , \tag{3.2}$$

where X is a reference image, Y is a distorted one, x_i, y_i are corresponding windows indexed by i, and N is the number of windows. Here we consider windows of size 8×8.

Image denoising. We examine the problem of image denoising using an artificial piece-wise constant image in Figure 3.1 and a Lena image in Figure 3.2. We assume zero-mean additive Gaussian noise with standard deviations being 0.2 and 0.05 respectively. Not only does our method work

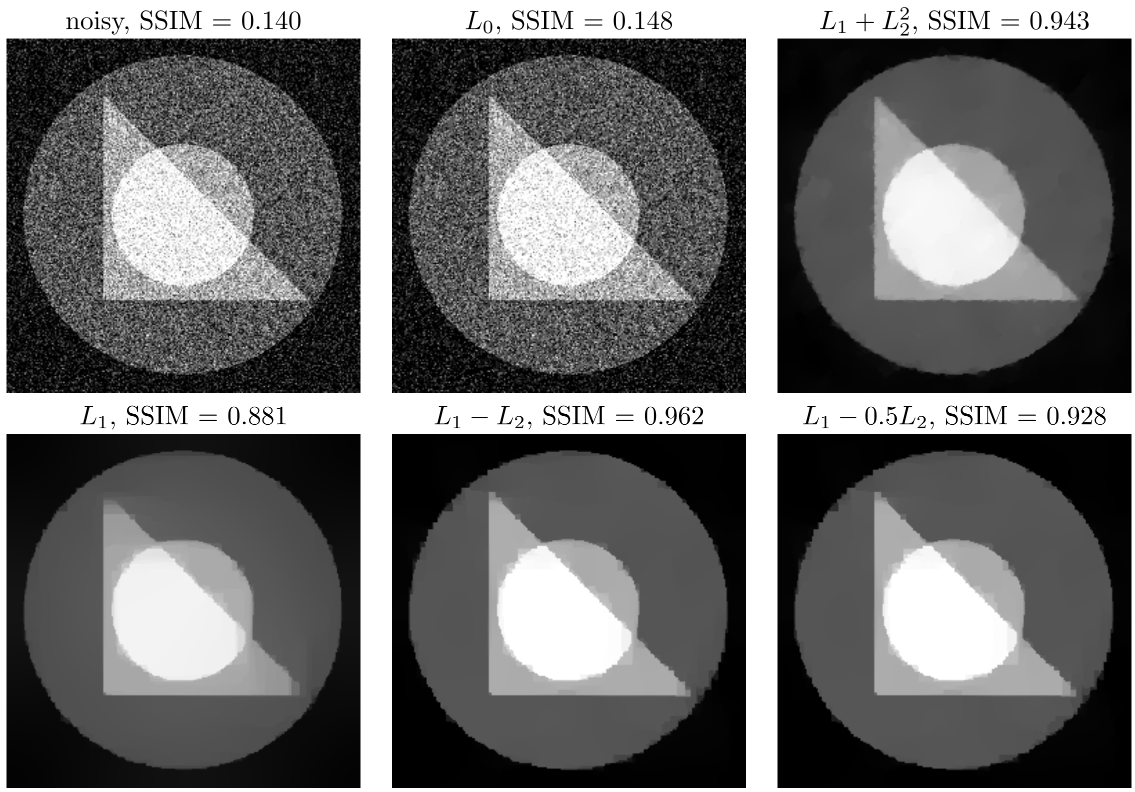

FIG. 3.1. *Denoising results with comparison to L_0 in [38] and $L_1 + L_2^2$ in [2].*

particularly well on horizonal or vertical edges by design, it can deal with natural images as well. To verify convergence analysis, the difference of u^n and u^{n-1} versus iterations is plotted in logarithm scale for both denoising examples shown in Figure 3.3, which shows that $L_1 - 0.5L_2$ converges faster than $L_1 - L_2$. As the ground-truth is available, we plot the relative errors versus cpu runtime for $L_1, L_1 - L_2, L_1 - 0.5L_2$ in Figure 3.4. This figure implies that our solutions oscillate around the ground truth due to the nonconvex nature of our model. Additionally we observe that the larger α is (say approaching 1), the less well-behaved DCA becomes due to more weight on the nonconvex term. On the other hand, $L_1 - L_2$ yields better results than $L_1 - 0.5L_2$ for the first few DCA iterates. The denoising results presented in Figure 3.1 and Figure 3.2 are from stopping DCA after 2 iterations.

Image deblurring. In Figure 3.5, a binary image is vertically blurred by motion blur of 15 pixels plus Gaussian additive noise with zero mean and standard deviation 0.1. Our method outperforms L_0 in [31], L_p for $p = 2/3$ in [21], $L_1 + L_2$ in [2], and the state-of-the-art deblurring method BM3D [11]. In Figure 3.6, we present deblurring results for a natural image: Cameraman. The original image is blurred by 15×15 Gaussian blur whose standard deviation is 1.5 plus Gaussian additive noise with zero mean and standard deviation 0.05. Although $L_0, L_{2/3}$ and BM3D are better than ours in terms of SSIM, their results have some ringing artifacts. In both deblurring examples, our method is better than the classical L_1 approach. The relative errors versus computational time is plotted in Figure 3.7 for both examples. It shows similar behavior as in the denoising problem that $L_1 - L_2$ tends to worsen beyond certain iterations while $L_1 - 0.5L_2$ is more stable. The deblurring results presented in Figure 3.5 and Figure 3.6 are from stopping DCA after 2 and 10 iterations for $L_1 - 0.5L_2$ and $L_1 - L_2$ respectively. A discussion on stopping criterion is given later.

MRI reconstruction. In Figure 3.8, we investigate the MRI reconstruction problem using a Shepp-Logan phantom from 7 and 8 radial projections. There is no noise when we synthesize the data. Consequently we adopt the constrained formulation, *i.e.*, Algorithm 2 for solving eq.(2.9). Due to the

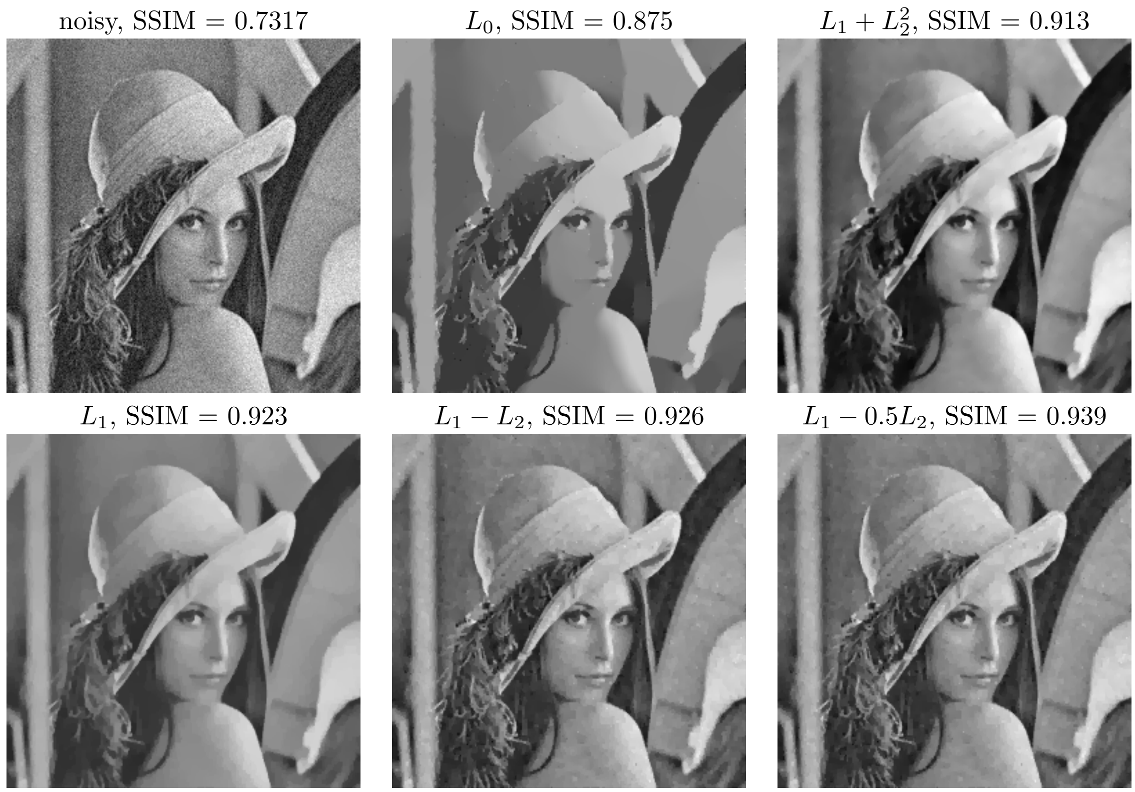

FIG. 3.2. *Denoising results with comparison to L_0 in [38] and $L_1 + L_2^2$ in [2].*

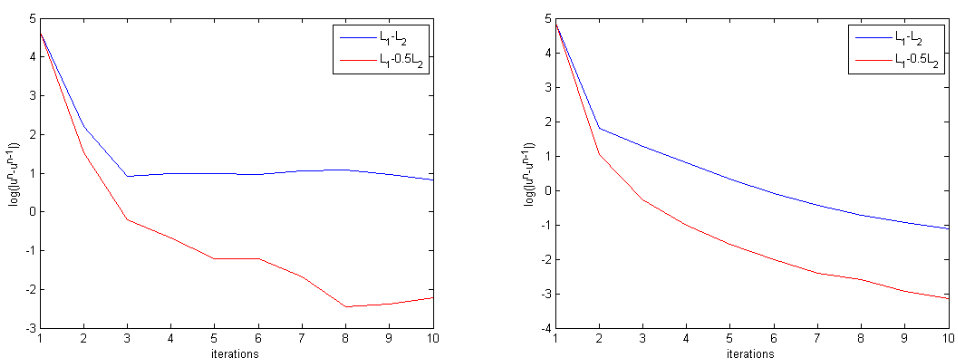

FIG. 3.3. *The difference of u^n and u^{n-1} versus iterations is plotted in logarithm scale for denoising examples in Figure 3.1 (left) and Figure 3.2 (right). $L_1 - 0.5L_2$ converges faster than $L_1 - L_2$.*

presence of complex values in MRI reconstruction problem, SSIM is no longer applicable; instead we use root-mean-square (RMS) error to measure the performance quantitatively. RMS between reference and distorted images X, Y is defined as $\text{RMS}(X, Y) = \frac{1}{\sqrt{M}}\|X - Y\|_2$ where M is the number of pixels in images X, Y. Figure 3.8 shows that our method can get a perfect reconstruction using only 8 projections, while a similar work [8] reports that 10 projections are required. When the number of projections is down to 7, $L_1 - 0.5L_2$ is much better than L_1 and $L_1 - L_2$ visually as well as in terms of RMS. The relative errors versus cpu time is plotted in Figure 3.9. The relative errors of $L_1 - L_2$ iterations in the constrained formulation appear as stable oscillations in contrast to the unstable

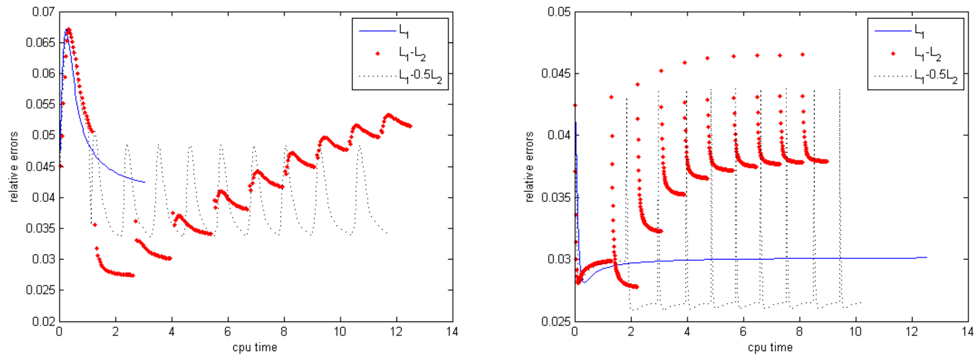

FIG. 3.4. *The relative errors versus runtime for methods $L_1, L_1-L_2, L_1-0.5L_2$ for denoising examples in Figure 3.1 (left) and Figure 3.2 (right). Our model solutions are seen to oscillate around the ground truth due to nonconvexity.*

oscillations in the unconstrained problems.

4. Discussions. Let us draw some connections of this work to two existing methods, Lysaker-Osher-Tai (LOT) model [24] and Bregman iterations [29]. Additionally, we comment on the stopping criterion.

4.1. Relation to existing methods. At first, the iterative scheme (2.6) for $\alpha = 1$ resembles the work of denoising the normals, proposed by Lysaker-Osher-Tai [24],

$$u^{n+1} = \text{argmin}_u \ \|\nabla u\|_2 - q^n \cdot \nabla u + \frac{\mu}{2}\|Au - f\|_2^2 \ , \tag{4.1}$$

where $q^n = \dfrac{\nabla u^n}{|\nabla u^n|}$ is the surface normal. Notice that the TV norm in (4.1) is isotropic, while the first term in our model is the anisotropic TV; and hence $L_1 - L_2$ applied to the gradient with linearized L_2 term is different from the LOT.

On the other hand, the LOT model leads to the discovery of Bregman iterations [29], which relates to the DCA as well. Specifically, the Bregman distance [1] based on a convex functional $J(\cdot)$ between two points u and v is defined as

$$D_J^p(u, v) := J(u) - J(v) - \langle p, u - v\rangle \ , \tag{4.2}$$

where $p \in \partial J(v)$ is the subgradient of J at the point v. Osher *et. al.* [29] suggest an iterative refinement procedure to update u as follows,

$$u^{n+1} = \text{argmin} \ D_J^{p^n}(u, u^n) + \frac{\mu}{2}\|Au - f\|_2^2 \ , \tag{4.3}$$

$$= \text{argmin} \ J(u) - \langle p^n, u\rangle + \frac{\mu}{2}\|Au - f\|_2^2 \ , \tag{4.4}$$

which is referred to as the Bregman iterations. Let $J(u) = \|\nabla u\|_2$ be the isotropic TV as in the LOT model, and its subgradient has the form $-\nabla \cdot \dfrac{\nabla u}{|\nabla u|}$. Consequently, we rewrite the second term in Eq. (4.4) as

$$\langle p^n, u\rangle = \langle -\nabla \cdot \frac{\nabla u^n}{|\nabla u^n|}, u\rangle = \langle \frac{\nabla u^n}{|\nabla u^n|}, \nabla u\rangle \ , \tag{4.5}$$

which coincides with the second term in the LOT model (4.1).

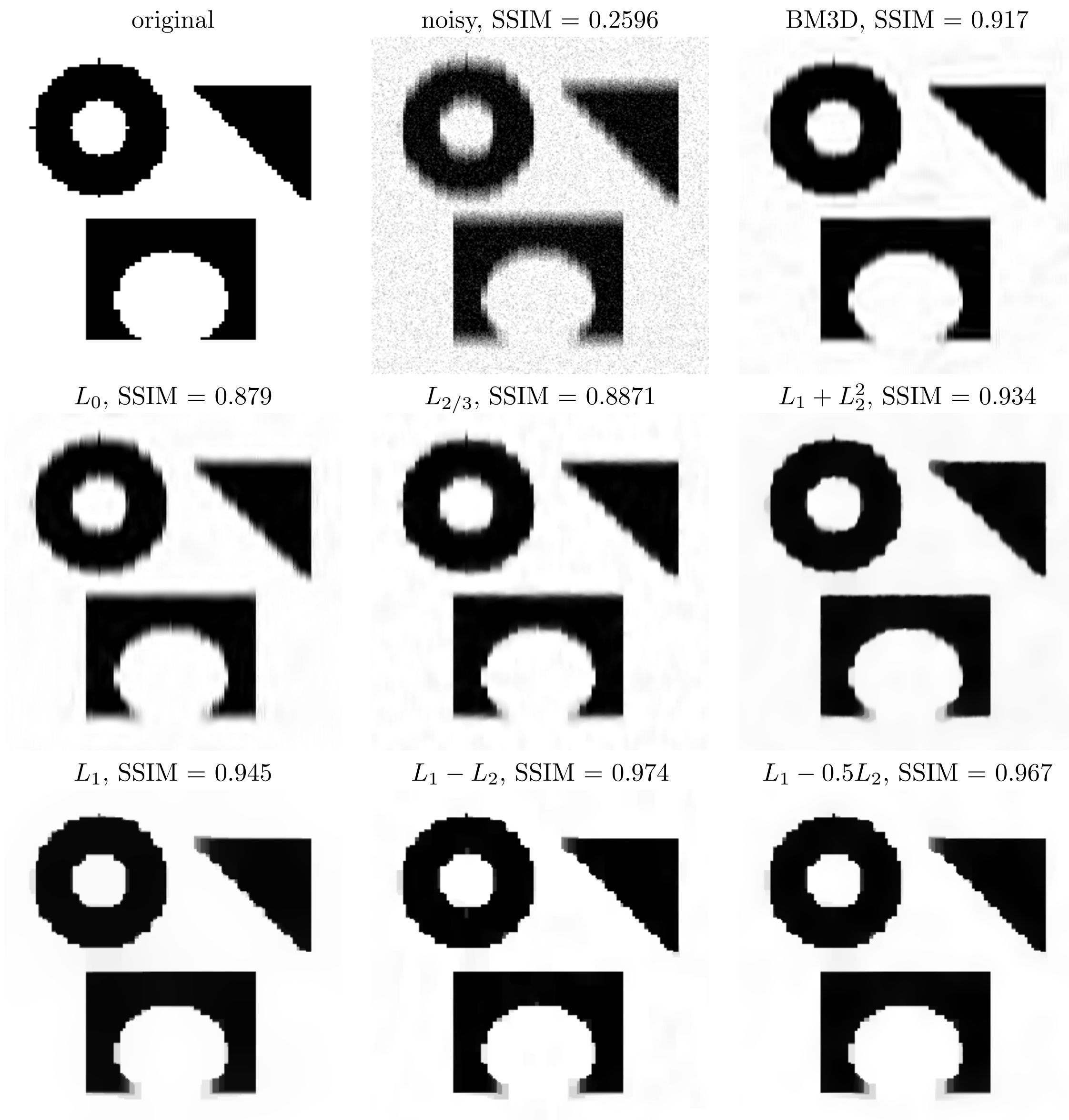

FIG. 3.5. *Deblurring results with comparison to L_0 in [31], L_p for $p = 2/3$ in [21], $L_1 + L_2^2$ in [2] and the state-of-the-art deblurring method BM3D [11].*

Bregman iterations can be viewed as an optimization technique. Computing the optimality condition for each subproblem (4.4), we obtain

$$p^{n+1} - p^n + \mu A^T(Au^{n+1} - f) = 0 \ . \tag{4.6}$$

Summing up to $n+1$, we have $p^{n+1} - \mu A^T(u^{n+1} - z^n)$ for $p^0 = 0$ and $z^{n+1} = z^n + (f - Au^n)$. It is the optimality condition for solving u^{n+1} from argmin $J(u) + \frac{\mu}{2}\|Au - z^n\|_2^2$. In short, the Bregman iterations can be rewritten as

$$u^{n+1} = \text{argmin } J(u) + \frac{\mu}{2}\|Au - z^n\|_2^2 \ , \tag{4.7}$$

$$z^{n+1} = z^n + (f - Au^n) \ . \tag{4.8}$$

FIG. 3.6. *Deblurring results with comparison to L_0 in [31], L_p for $p = 2/3$ in [21], $L_1 + L_2^2$ in [2] and the state-of-the-art deblurring method BM3D [11].*

The DCA for solving $L_1 - L_2$ minimization can be derived from a similar way of the Bregman iterations. Let p and q be the subgradient of anisotropic J_{ani} and isotropic J_{iso} respectively. Lagging the isotropic term gives us

$$p^{n+1} - p^n - \alpha(q^n - q^{n-1}) + \mu A^T(Au^{n+1} - f) = 0 \ . \tag{4.9}$$

We apply the same summation technique as in (4.6) and obtain

$$p^{n+1} - \alpha q^n + \mu A^T(Au^{n+1} - z^{n+1}) = 0 \ , \tag{4.10}$$

$$z^{n+1} = z^n + (f - Au^n) \ . \tag{4.11}$$

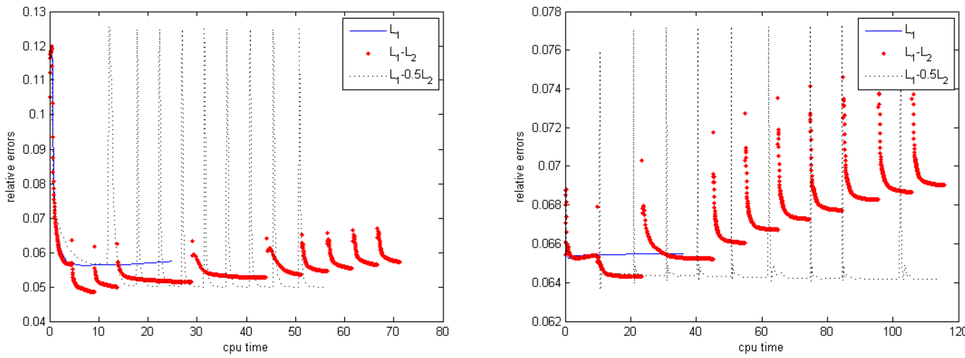

FIG. 3.7. *The relative errors versus runtime for methods $L_1, L_1 - L_2, L_1 - 0.5L_2$ for deblurring examples in Figure 3.5 (left) and Figure 3.6 (right).*

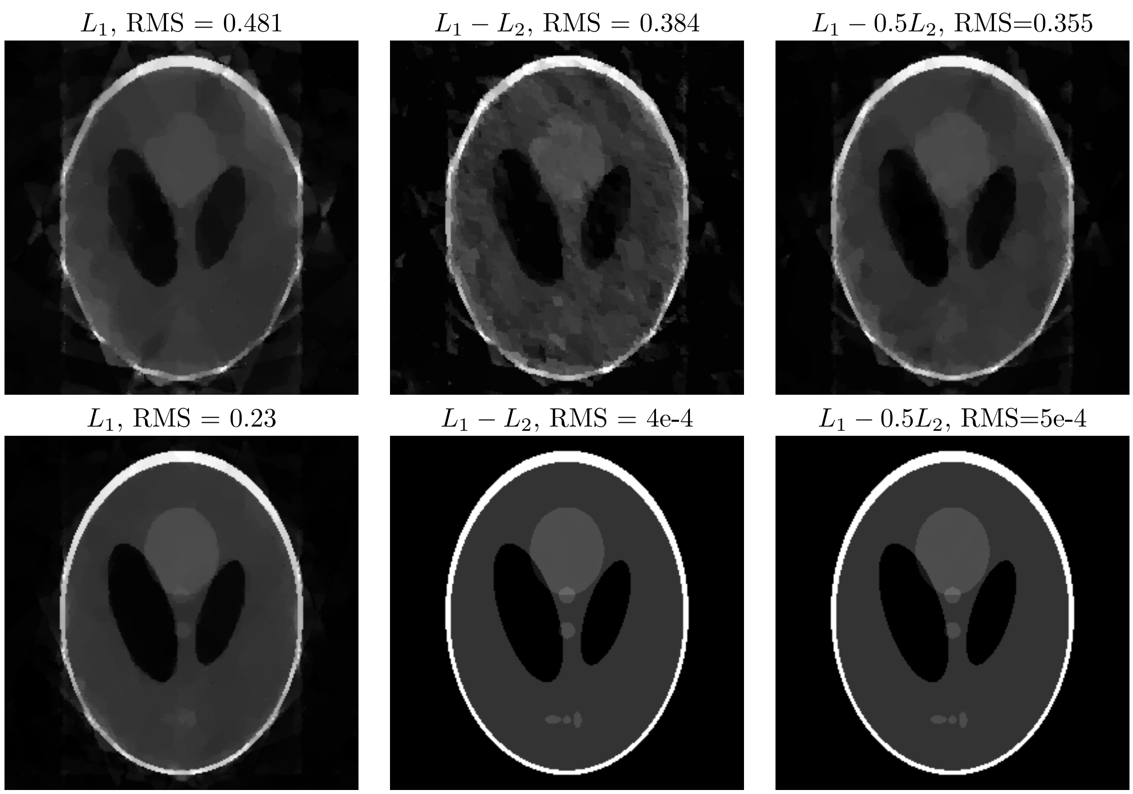

FIG. 3.8. *MRI reconstruction using 7 (top) and 8 projections (bottom). The root-means-error (RMS) is provided for comparison.*

for $p^0 = q^0 = z^0 = 0$. The subproblem (4.10) is equivalent to

$$u^{n+1} = \arg\min J_{ani}(u) - \alpha\langle q^n, u\rangle + \frac{\mu}{2}\|Au - z^n\|_2^2 \ , \tag{4.12}$$

which looks very similar to applying the DCA for a constrained problem, eq. (2.5). The algorithm derived from the Bregman iterations is summarized in Algorithm 3. Its difference to Algorithm 2 lies in the update of z and q. For Algorithm 2, z is updated *MaxBregmanOuter* iterations and then q is updated, while Algorithm 3 is to update z and q simultaneously. The comparison between the

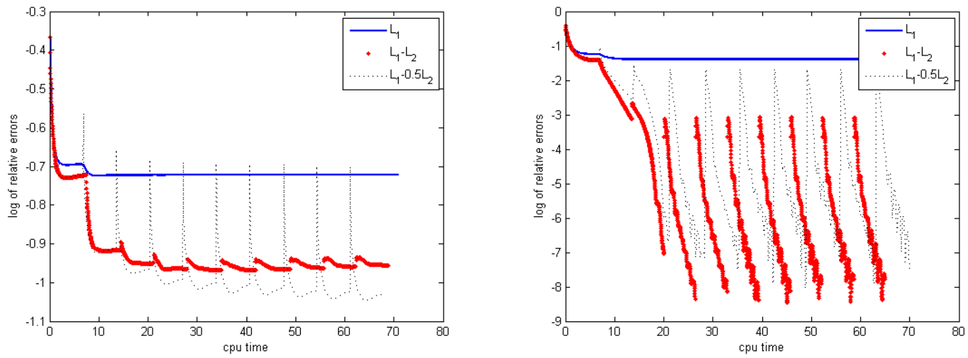

FIG. 3.9. *The logarithm of relative errors versus runtime for methods $L_1, L_1 - L_2, L_1 - 0.5L_2$ in MRI reconstruction problem using 7 (left) and 8 (right) projections. All are solved under constrained formulation.*

Bregman and DCA iterations for solving such constrained nonconvex problems is a subject of further study.

Algorithm 3 for solving constrained problem (2.9) using Bregman method

Define $u = q_x = q_y = 0, z = f$ and $MaxDCA, MaxBregman$
for 1 **to** $MAXDCA$ **do**
  $b_x = b_y = 0$
  for 1 **to** $MaxBregman$ **do**
    $u = (\mu A^T A - \lambda\triangle)^{-1}(\mu\mathcal{A}z + \lambda D_x^T(d_x - b_x) + \lambda D_y^T(d_y - b_y))$
    $d_x = \text{shrink}(D_x u + b_x + \alpha q_x/\lambda, 1/\lambda)$
    $d_y = \text{shrink}(D_y u + b_y + \alpha q_y/\lambda, 1/\lambda)$
    $b_x = b_x + D_x u + d_x$
    $b_y = b_y + D_y u + d_y$
  end for
  $z = z + f - Au$
  $(q_x, q_y) = (D_x u, D_y u)/\sqrt{|D_x u|^2 + |D_y u|^2}$
end for

4.2. Stopping criterion. We discuss the stopping conditions of Algorithm 1 and Algorithm 2 for unconstrained and constrained problems respectively. Both algorithms have an outer DCA loop, which iteratively updates q, and inner iterations for updating u. We use u^n and u_k to specify the outer and inner outputs of u.

The inner loop is easier to impose a proper stopping criterion for, because the inner loop solves a convex subproblem. Some standard stopping criteria are either the relative error being small or objective function being stagnant or both *i.e.*,

$$\frac{\|u_{k+1} - u_k\|}{\|u_k\|} < \epsilon_u \quad \text{and/or} \quad \frac{|F(u_{k+1}) - F(u_k)|}{|F(u_k)|} < \epsilon_F \tag{4.13}$$

with pre-defined tolerance values ϵ_u, ϵ_F. In this paper, we choose to stop the inner iteration when the relative error is smaller than $1e^{-6}$.

As for the outer iterations, Figures 3.4, 3.7, and 3.9 show that the relative error develops an oscillatory pattern. One can estimate the onset time t_b of the oscillation stage of the error based on training images. In the denoising (deblurring) example, $t_b = 2$ $(= 10)$. Hence, a good stopping time for the outer iteration is at the end of an inner loop when the cpu time exceeds t_b.

More generally, if the error does not follow a clear oscillatory pattern, one could inject random perturbations with slowly reduced magnitudes to steer away from unstable stationary points or directions to help convergence towards the ground truth [17]. This approach is closely related to simulated annealing [14, 20].

5. Conclusion. We proposed a weighted difference of anisotropic and isotropic total variation as a regularization term for image processing applications. We presented a difference of convex algorithm (DCA) for both the constrained and unconstrained formulations. We proved the convergence of the algorithm to ensure that each limiting point is a stationary point and the values of the objective function monotonically decrease. The behavior of the iterations was observed numerically to be oscillatory around the ground truth. The deviation occurs at the beginning of outer loops of DCA. A stopping criterion was introduced based on such oscillatory pattern of the errors.

In the numerical experiments, we examined three particular applications: image denoising, deblurring and MRI reconstruction. By design, our method works particularly well for piecewise constant images. For natural images, it improved the classical TV model, and is comparable to the state-of-the-art methods. In future work, we plan to carry out a detailed comparison between the DCA and Bregman methods, and further study the error pattern and the resulting stopping criterion for other imaging science problems.

Acknowledgments. YL would like to thank Dr. Ernie Esser at University of British Columbia for helpful discussions. JX would like to thank Profs. Krishna Nayak and Angel Pineda for their hospitality during a visit to USC in March 2014, and their suggestion to consider a weighted variant of $L_1 - L_2$ for compressed sensing and the SSIM measure for image quality evaluation.

REFERENCES

[1] L. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problmes in convex programming*, USSR COMP MATH MATH+, (1967), pp. 200–217.

[2] X. Cai, R. Chan, and T. Zeng, *A two-stage image segmentation method using a convex variant of the mumford-shah model and thresholding*, SIAM J. Imaging Sci., 6 (2013), pp. 368–390.

[3] E. Candes, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[4] W. Cao, J. Sun, and Z. Xu, *Fast image deconvolution using closed-form thresholding formulas of lq(q=1/2,2/3) regularization*, J. Vis. Comun. Image Represent., 24 (2013), pp. 31–41.

[5] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math Imaging Vision, 20 (2004), pp. 89–97.

[6] T. F. Chan and J. Shen, *Mathematical models for local non-texture inpainting*, SIAM J. Appl. Math, 62 (2002), pp. 1019–1043.

[7] T. F. Chan and C. K. Wong, *Total variation blind deconvolution*, IEEE Trans. on Image Process., 7 (1998), pp. 370–375.

[8] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett, 10 (2007), pp. 707–710.

[9] Y. Chen, W. W. Hager, M. Yashtini, X. Ye, and H. Zhang, *Bregman operator splitting with variable stepsize for total variation image reconstruction*, comput. Optim. Appl., 54 (2013), pp. 317–342.

[10] R. Choksi, Y. van Gennip, and A. Oberman, *Anisotropic total variation regularized l^1-approximation and denoising/deblurring of 2d bar codes*, Inverse Probl. and Imaging, 3 (2011), pp. 591–617.

[11] K. Dabov, A. Foi, and K. Egiazarian, *Image restoration by sparse 3D transform-domain collaborative filtering*, in Proc. SPIE Electronic Imaging, vol. 6812-07, San Jose, California, USA, January 2008.

[12] D. L. Donoho, *Compressed sensing*, IEEE Trans. on Inform. Theory, 52 (2006).

[13] E. Esser, Y. Lou, and J. Xin, *A method for finding structured sparse solutions to non-negative least squares problems with applications*, SIAM J. Imaging Sci., 6 (2013), pp. 2010–2046.

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

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

[16] L. He, A. Marquina, and S. Osher, *Blind deconvolution using TV regularization and Bregman iteration*, International Journal of Imaging Systems and Technology, 15 (2005), pp. 74–83.

[17] Q. He and J. Xin, *A randomly perturbed INFOMAX algorithm for blind source separation*, in Internatinal conference on Acoustics, speech, and signal processing (ICASSP), 2013, pp. 3218–3222.

[18] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer Verlag, Heidelberg, 1996. Two volumes - 2nd printing.
[19] N. Hurley and S. Rickard, *Comparing measures of sparsity*, IEEE Trans. on Inform. Theory, 55 (2009), pp. 4723–4741.
[20] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Optimization by simmulated annealing*, Science, 220 (1983), pp. 671–680.
[21] D. Krishnan and R. Fergus, *Fast image deconvolution using hyper-laplacian priors*, in Advances in Neural Information Processing Systems (NIPS), 2009, pp. 1033–1041.
[22] D. Krishnan, T. Tay, and R. Fergus, *Blind deconvolution using a normalized sparsity measure*, in Computer Vision and Pattern Recognition (CVPR), 2011.
[23] Y. Lou, P. Yin, Q. He, and J. Xin, *Computing sparse representation in a highly coherent dictionary based on difference of l1 and l2*, tech. report, UCLA, 2014. CAM Report [14-02].
[24] M. Lysaker, S. Osher, and X. C. Tai, *Noise removal using smoothed normals and surface fitting*, IEEE Trans. Image Process., 13 (2004), pp. 1345–1357.
[25] A. Marquina, *Nonlinear inverse scale space methods for total variation blind deconvolution*, SIAM J. Imaging Sci., 2 (2009), pp. 64–83.
[26] A. Marquina and S. Osher, *Image super-resolution by TV-regularization and Bregman iteration*, J. Sci. Comput., 37 (2008), pp. 367–382.
[27] D. Martin, C. Fowlkes, D. Tal, and J. Malik, *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*, in Int'l Conf. Computer Vision, vol. 2, July 2001, pp. 416–423.
[28] D. Mumford and J. Shah, *boundary detection by minimizing functionals*, in Computer Vision and Pattern Recognition (CVPR), 1985, pp. 137–154.
[29] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterated regularization method for total variation-based image restoration*, Multiscale Model. Simul, (2005), pp. 460–489.
[30] S. J. Osher and S. Esedoglu, *Decomposition of images by the anisotropic rudin-osher-fatemi model*, Comm. Pure Appl. Math, 57 (2003), pp. 1609–1626.
[31] J. Portilla, *Image restoration through l0 analysis-based sparse optimization in tight frames*, in Int. conf. image. proc., 2009, pp. 3909–3912.
[32] R. B. Potts, *Some generalized order-disorder transformations*, Math. Proc. Cambridge Philos. Soc, (1952), pp. 106–109.
[33] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[34] M. Storath, A. Weinmann, and L. Demaret, *Jump-sparse and sparse recovery using potts functionals*, IEEE Trans. on Signal Process., 62 (2014), pp. 3654 – 3666.
[35] P. Tao and L.T.H. An, *A d.c. optimization algorithm for solving the trust-region subproblem*, SIAM J. Optim., 8 (1998), pp. 476–505.
[36] P. D. Tao and L. T. H. An, *Convex analysis approach to d.c. programming: Theory, algorithms and applications*, Acta Mathematica Vietnamica, 22 (1997), pp. 289–355.
[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Transa. on Image Process., 13 (2004), pp. 600–612.
[38] L. Xu, C. Lu, Y. Xu, and J. Jia, *Image smoothing via l0 gradient minimization*, ACM Trans. Graphics (SIGGRAPH Asia), (2011).
[39] Z. Xu, X. Chang, F. Xu, and H. Zhang, $L_{1/2}$ *Regularization: A Thresholding Representation Theory and a Fast Solver*, IEEE Trans. on Neural Networks, 23 (2012), pp. 1013–1027.
[40] P. Yin, E. Esser, and J. Xin, *Ratio and difference of l1 and l2 norms and sparse representation with coherent dictionaries*, tech. report, UCLA, CAM Report [13-21], 2013. Comm. Info. Systems, 2014, to appear.
[41] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of* $l_1 - l_2$ *for compressed sensing*, tech. report, UCLA, 2014. CAM Report [14-01].