

# A dual algorithm for a class of augmented convex models

Hui Zhang\*

Lizhi Cheng\*

Wotao Yin<sup>†</sup>

August 30, 2013

## Abstract

Convex optimization models find interesting applications, especially in signal/image processing and compressive sensing. We study some augmented convex models, which are perturbed by strongly convex functions, and propose a dual gradient algorithm. The proposed algorithm includes the linearized Bregman algorithm and the singular value thresholding algorithm as special cases. Based on fundamental properties of proximal operators, we present a concise approach to establish the convergence of both primal and dual sequences, improving the results in the existing literature.

**Keywords:** augmented convex model; Lagrange dual; primal-dual, proximal operator; signal recovery

## 1 Introduction

The past two decades have witnessed several successful convex models for signal processing. They include, but are not limited to, the total variation model [13] and the basis pursuit model [4], both of which have been widely applied in signal/image processing and compressed sensing. Recently, augmented convex models [19, 20, 5], obtained by adding strongly convex perturbations to the original objective functions, are introduced for fast computation [21] and for incorporating certain prior information regarding the underlying signal [22]. Well known in convex analysis [12], if the original problem is strongly convex, then the dual problem is differentiable and can thus take advantages of a rich set of gradient-based optimization techniques. In addition, the augmented term can reflect certain structures of the target signal, for example, the group structure through the term  $\|\cdot\|_2^2$  in the elastic net model [22].

This paper is devoted to analyzing a dual gradient algorithm for a class of augmented convex models. The proposed algorithm is motivated by and includes two well-known algorithms as its special cases: the linearized Bregman algorithm (LBreg) [17, 18] and the singular value thresholding algorithm (SVT) [3]. Based on some fundamental properties of proximal operators, we prove the convergence of both the primal and dual point sequences. This result is stronger than the previously shown vanishing distance between the dual sequence and the dual solution set (the dual sequence itself is not shown to converge).

## 2 Augmented convex models for signal recovery

Let  $x \in \mathbb{R}^n$ . Consider the convex problem

$$\min \mu \|x\|, \quad \text{subject to } \mathcal{A}x = b, \quad (1)$$

---

\*Department of Mathematics and Systems Science, College of Science, National University of Defense Technology, Changsha, Hunan, China, 410073. Email: [hhuuii.zhang@gmail.com](mailto:hhuuii.zhang@gmail.com)

<sup>†</sup>Department of Mathematics, University of California, Los Angeles, CA. Email: [wotaoyin@math.ucla.edu](mailto:wotaoyin@math.ucla.edu)

and its augmented model

$$\min P(x) \triangleq \mu \|x\| + \frac{\mu}{2\tau} \|x\|_2^2, \quad \text{subject to } \mathcal{A}x = b, \quad (2)$$

where  $\|\cdot\|$  is a norm whose dual norm is denoted by  $\|\cdot\|_\diamond$ ,  $\frac{\mu}{2\tau} \|x\|_2^2$  is the *augmented* term, linear operator  $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and observed data  $b \in \mathbb{R}^m$  are given, and  $\tau, \mu$  are positive parameters. Throughout the paper, we assume that  $\mathcal{A}x = b$  is consistent. Parameter  $\mu$  is redundant to both objectives and does not affect the solutions, but it is kept in order to unify the models and algorithms that appear in the previous literature. Parameter  $\tau$  weights the term  $\|\cdot\|_2^2$  and affects the solution to (2) when it falls in a certain range. In what follows, we give a few examples of (2) in signal recovery.

*Example 1* (Augmented  $\ell_1$  norm and nuclear-norm models) Paper [5] proposes the augmented  $\ell_1$  norm model for sparse signal recovery

$$\min \|x\|_1 + \frac{1}{2\tau} \|x\|_2^2, \quad \text{subject to } Ax = b \quad (3)$$

and the augmented nuclear-norm model for low-rank matrix recovery

$$\min \|X\|_* + \frac{1}{2\tau} \|X\|_F^2, \quad \text{subject to } \mathcal{A}(X) = b. \quad (4)$$

Suppose  $b = Ax^0$  where  $x^0$  is a sparse vector. Model (3) will recover  $x^0$  provided that  $\tau \geq 10\|x^0\|_\infty$  and the sensing matrix  $A$  satisfies certain conditions such as the null-space property and restricted isometry property. Similarly,  $\tau \geq 10\|X^0\|$  is used for recovering a low-rank matrix  $X^0$ , where  $\|X^0\|$  is its spectral norm.

*Example 2* (Strongly convex matrix completion model) Papers [19, 20] study the following strongly convex model for matrix completion

$$\min \|X\|_* + \frac{1}{2\tau} \|X\|_F^2, \quad \text{subject to } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M), \quad (5)$$

where  $M$  is a low-rank matrix,  $\Omega$  is the sample index set, and  $\mathcal{P}_\Omega$  is the corresponding element-selection operator. To recover the low-rank matrix  $M$ , the best known bound is  $\tau \geq \frac{4}{p} \|\mathcal{P}_\Omega(M)\|_F$  given in [20], where  $p$  is the sample ratio.

*Example 3* (Strongly convex RPCA model) Paper [20] studies the following strongly convex model for robust principle component analysis (RPCA):

$$\min_{L,S} \|L\|_* + \frac{1}{2\tau} \|L\|_F^2 + \lambda \|S\|_1 + \frac{1}{2\tau} \|S\|_F^2, \quad \text{subject to } D = L + S, \quad (6)$$

where  $D$  is the observed data matrix and  $\lambda$  is a given parameter. Bound  $\tau \geq \frac{8\sqrt{15}\|D\|_F}{3\lambda}$  guarantees the decomposition of the observed matrix  $D$  into its low-rank component and sparse component [16].

Below section 3 presents a dual gradient algorithm for problem (2) and section 4 studies its convergence. Section 5 extends these results to problems with gauge objective functions.

### 3 A dual gradient algorithm

In this section, we first introduce properties of proximal operators required for convergence analysis. Then, we derive a Lagrange dual problem and an iterative gradient algorithm for solving it.

### 3.1 Proximal operators

Let  $f : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$  be a closed proper convex function. The proximal operator [9]  $\mathbf{prox}_f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined by

$$\mathbf{prox}_f(v) = \arg \min_x \left( f(x) + \frac{1}{2} \|x - v\|_2^2 \right). \quad (7)$$

Since the objective function is strongly convex and proper,  $\mathbf{prox}_f(v)$  is properly defined for every  $v \in \mathbb{R}^n$ . The following properties [9, 11] will be used in our analysis.

**Lemma 1.** *Let  $f : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$  be a closed proper convex function. Then, for all  $x, y \in \mathbb{R}^n$  the proximal operator  $\mathbf{prox}_{f(\cdot)}$  satisfies the followings:*

(i). *Firmly nonexpansive:*  $\|\mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y)\|_2^2 \leq \langle x - y, \mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y) \rangle$

(ii). *Lipschitz continuous:*  $\|\mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y)\|_2 \leq \|x - y\|_2$

**Lemma 2.** *For any  $\tau > 0$  and norm  $\|\cdot\|$ , it holds that  $\tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\tau}v) = \mathbf{prox}_{\tau\|\cdot\|}(v)$*

*Proof.* Let  $u = \mathbf{prox}_{\tau\|\cdot\|}(v) = \arg \min_z \tau \cdot \|z\| + \frac{1}{2} \|z - v\|_2^2$ . Let  $w = \mathbf{prox}_{\|\cdot\|}(\frac{1}{\tau}v)$  and  $\bar{u} = \tau w$ . Since

$$\begin{aligned} w &= \arg \min \|x\| + \frac{1}{2} \|x - \frac{1}{\tau}v\|_2^2 \\ &= \arg \min \tau \cdot \|\tau x\| + \frac{1}{2} \|\tau x - v\|_2^2 \end{aligned}$$

after the change of variable  $\tau x \rightarrow z$ , we have  $u = \tau w = \bar{u}$ . This completes the proof.  $\square$

**Remark 1.** *Lemma 2 remains valid if  $\|\cdot\|$  is replaced by a closed proper convex function  $f$  that is one-homogeneous. We, however, restrict our attention to  $f(\cdot) = \|\cdot\|$  for brevity.*

**Lemma 3.** *Moreau decomposition: any  $v \in \mathbb{R}^n$  can be decomposed as  $v = \mathbf{prox}_f(v) + \mathbf{prox}_{f^*}(v)$ , where  $f^* = \sup_x (\langle y, x \rangle - f(x))$  is the convex conjugate of  $f$ .*

There is a close relationship between proximal and projection operators. Let  $\mathcal{B} = \{z : \|z\|_{\diamond} \leq 1\}$  and consider the projection onto  $\mathcal{B}$ :  $\Pi_{\mathcal{B}}(v) = \arg \min_{x \in \mathcal{B}} \|x - v\|_2$ . Applying the Moreau decomposition to  $\|\cdot\|$ , we have

$$v = \mathbf{prox}_{\|\cdot\|}(v) + \Pi_{\mathcal{B}}(v). \quad (9)$$

In order to derive the gradient of  $D(y)$ , we define the point-to-set function

$$h_{\mathcal{Z}}(x) = \min_{z \in \mathcal{Z}} \|x - z\|_2, \quad (10)$$

where  $\mathcal{Z}$  is a closed convex set. Following Example 2.79 in [14], it holds that

$$h_{\mathcal{Z}}(x) = \begin{cases} \|x - \Pi_{\mathcal{Z}}(x)\|_2 & x \notin \mathcal{Z} \\ 0 & x \in \mathcal{Z}, \end{cases} \quad (11)$$

and

$$\nabla h_{\mathcal{Z}}(x) = \begin{cases} \frac{x - \Pi_{\mathcal{Z}}(x)}{\|x - \Pi_{\mathcal{Z}}(x)\|_2} & x \notin \mathcal{Z} \\ 0 & x \in \mathcal{Z}. \end{cases} \quad (12)$$

### 3.2 Lagrange dual analysis

The Lagrangian of the augmented convex model (2) is

$$L(x, y) = \mu\|x\| + \frac{\mu}{2\tau}\|x\|_2^2 + \langle y, b - \mathcal{A}x \rangle. \quad (13)$$

Following  $\|x\| = \max_{\|z\|_\diamond \leq 1} \langle x, z \rangle$ , we derive the dual function as

$$\begin{aligned} L_D(y) &= \min_x L(x, y) = \langle y, b \rangle + \min_x \max_{\|z\|_\diamond \leq 1} \mu \langle x, z \rangle + \frac{\mu}{2\tau}\|x\|_2^2 - \langle \mathcal{A}^*y, x \rangle \\ &= \langle y, b \rangle + \max_{\|z\|_\diamond \leq 1} \min_x \mu \langle x, z \rangle + \frac{\mu}{2\tau}\|x\|_2^2 - \langle \mathcal{A}^*y, x \rangle \\ &= \langle y, b \rangle - \frac{\tau}{2\mu} \min_{\|z\|_\diamond \leq 1} \|\mathcal{A}^*y - \mu z\|_2^2 \\ &= \langle y, b \rangle - \frac{\tau\mu}{2} \min_{\|z\|_\diamond \leq 1} \left\| \frac{1}{\mu} \mathcal{A}^*y - z \right\|_2^2 \end{aligned}$$

where the  $x$ -minimization problem has solution  $x = \frac{\tau}{\mu}(\mathcal{A}^*y - \mu z)$ . Hence, the dual problem is

$$\max_y L_D(y) = -\min D(y), \quad \text{where } D(y) \triangleq -\langle y, b \rangle + \frac{\tau\mu}{2} \min_{\|z\|_\diamond \leq 1} \left\| \frac{1}{\mu} \mathcal{A}^*y - z \right\|_2^2. \quad (15)$$

The minimum over  $z$  is obtained at  $z = \Pi_{\mathcal{B}}(\frac{1}{\mu}\mathcal{A}^*y)$ . To distinguish  $L_D(y)$  and  $D(y)$ , we call the later dual objective. Following the definition of  $h_{\mathcal{Z}}(x)$  in (10),  $D(y)$  can be written equivalently as

$$D(y) = -\langle y, b \rangle + \frac{\tau\mu}{2} h_{\mathcal{B}}\left(\frac{1}{\mu}\mathcal{A}^*y\right)^2.$$

Following from (12), the gradient of  $D(y)$  is

$$\nabla D(y) = -b + \tau\mathcal{A} \left( \frac{1}{\mu}\mathcal{A}^*y - \Pi_{\mathcal{B}}\left(\frac{1}{\mu}\mathcal{A}^*y\right) \right), \quad (16)$$

and, due to (9),

$$\nabla D(y) = -b + \tau\mathcal{A} \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\mu}\mathcal{A}^*y \right). \quad (17)$$

We highlight the primal-dual relationship:  $x = \frac{\tau}{\mu}(\mathcal{A}^*y - \mu z)$  and  $z = \Pi_{\mathcal{B}}(\frac{1}{\mu}\mathcal{A}^*y)$ . If  $y$  is dual optimal, by standard convex analysis,  $x$  in the relationship is primal optimal.

**Lemma 4.** *Let  $\hat{x}$  be the unique solution to problem (2). Then the dual solution set to problem (15) is*

$$\mathcal{Y} = \left\{ y : \tau \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\mu}\mathcal{A}^*y \right) = \hat{x} \right\}, \quad (18)$$

*which is nonempty and convex.*

*Proof.* From convex analysis and (17) it follows that the dual solution set is  $\mathcal{Y}' = \{y : \nabla D(y) = 0\} = \{y : \tau\mathcal{A} \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y) = b\}$ . Comparing this to (18) and since  $\mathcal{A} \cdot \hat{x} = b$ , we have  $\mathcal{Y} \subset \mathcal{Y}'$ . Therefore, it suffices to show  $\mathcal{Y}' \subset \mathcal{Y}$ . Indeed, let  $\hat{y} \in \mathcal{Y}'$ ; following the primal-dual relationship,  $\hat{y}$  shall give optimal  $\hat{x}$ , i.e.,

$$\frac{\tau}{\mu} \left( \mathcal{A}^*\hat{y} - \mu \Pi_{\mathcal{B}}\left(\frac{1}{\mu}\mathcal{A}^*\hat{y}\right) \right) = \frac{\tau}{\mu}(\mathcal{A}^*\hat{y} - \mu\hat{z}) = \hat{x}.$$

Since the left-hand side equals,  $\tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*\hat{y})$ , by the definition  $\mathcal{Y}$ , we have  $\hat{y} \in \mathcal{Y}$ . The convexity of  $\mathcal{Y}$  follows from the convexity of primal problem; it can also be seen through:

$$\langle \nabla D(y_1) - \nabla D(y_2), y_1 - y_2 \rangle = \tau\mu \langle \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y_1) - \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y_2), \frac{1}{\mu}\mathcal{A}^*y_1 - \frac{1}{\mu}\mathcal{A}^*y_2 \rangle \quad (19a)$$

$$\geq \tau\mu \left\| \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y_1) - \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y_2) \right\|_2^2 \geq 0, \quad (19b)$$

where the inequality follows from Lemma 1. The consistency of  $\mathcal{A}x = b$  guarantees  $\mathcal{Y}$  to be nonempty.  $\square$

### 3.3 Algorithm and examples

Applying the gradient iteration to the dual objective  $D(y)$  gives:

$$y^{k+1} = y^k + h \left( b - \tau \mathcal{A} \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\mu} \mathcal{A}^* y^k \right) \right), \quad (20)$$

where  $h > 0$  is the step size whose range shall be studied later for convergence. By setting  $x^{k+1} = \tau \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\mu} \mathcal{A}^* y^k \right)$ , we obtain the equivalent iteration in the primal-dual form:

$$\begin{cases} x^{k+1} = \tau \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\mu} \mathcal{A}^* y^k \right) \\ y^{k+1} = y^k + h(b - \mathcal{A}x^{k+1}). \end{cases} \quad (21)$$

Recalling  $\tau \cdot \mathbf{prox}_{\|\cdot\|} \left( \frac{1}{\tau} v \right) = \mathbf{prox}_{\tau\|\cdot\|} (v)$  from Lemma 2 and setting  $\mu = \tau$ , we simplify it to

$$\begin{cases} x^{k+1} = \mathbf{prox}_{\tau\|\cdot\|} (\mathcal{A}^* y^k) \\ y^{k+1} = y^k + h(b - \mathcal{A}x^{k+1}). \end{cases} \quad (22)$$

*Example 1* (The LBreg algorithm) It is a well studied algorithm for solving the augmented  $\ell_1$ -norm model and has the following primal-dual form

$$\begin{cases} x^{k+1} = \tau \cdot \text{shrink}(A^T y^k) \\ y^{k+1} = y^{(k)} + h(b - Ax^{k+1}) \end{cases}$$

where  $\text{shrink}(\cdot)$  equals  $\mathbf{prox}_{\|\cdot\|_1}(\cdot)$ . It is a special case of (21). The iteration is proposed in [17] and its convergence analyzed in [1, 2, 18]. Then paper [5] establishes its global geometric convergence, whose rate is further improved in [21].

*Example 2* (The SVT algorithm) It is a well-known algorithm for matrix complete and has the following primal-dual form

$$\begin{cases} X^{k+1} = \mathcal{D}_\tau(Y^k) \\ Y^{k+1} = Y^k + h \cdot \mathcal{P}_\Omega(M - X^{k+1}) \end{cases}$$

where  $\mathcal{D}_\tau(\cdot)$  equals  $\mathbf{prox}_{\tau\|\cdot\|_*}(\cdot)$ . Generally, we take  $Y^0 = 0$  so that  $\mathcal{P}_\Omega(Y^k) = Y^k$  from  $Y^{k+1} = Y^k + h \cdot \mathcal{P}_\Omega(M - X^{k+1})$ . Moreover,  $\mathcal{P}_\Omega = \mathcal{P}_\Omega^*$ . Hence, it is a special case of (22) with  $\mathcal{A} = \mathcal{P}_\Omega$ . It is proposed in [3].

**Remark 2.** In paper [6], Nesterov's first-order methods [7] are applied to accelerate the LBreg and SVT algorithms. Further speedup is introduced in [21] by combining Nesterov's methods [7, 8] with an adaptive restart technique [10]. With little effort, they can be applied to the primal-dual algorithm (21).

## 4 Convergence analysis

In this part, we prove the convergence of primal sequence  $\{x^k\}$  and dual sequence  $\{y^k\}$  in iteration (21).

**Theorem 1.** Set step size  $h \in (0, \frac{2\mu}{\tau\|\mathcal{A}\|^2})$  and  $y^0 = 0$  in iteration (21). Let  $\hat{x}$  be the unique minimizer to problem (2) and  $\mathcal{Y}$  be the solution set to problem (15). Then,  $\lim_{k \rightarrow +\infty} x^k = \hat{x}$ , and there exists a point  $\bar{y} \in \mathcal{Y}$  such that  $\lim_{k \rightarrow +\infty} y^k = \bar{y}$ .

*Proof.* Let  $\hat{y} \in \mathcal{Y}$ . By Lemma 4, we have  $\hat{x} = \tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*\hat{y})$ . Together with  $x^{k+1} = \tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y^k)$  and Lemma 1, we derive that

$$\langle \frac{1}{\mu}\mathcal{A}^*y^k - \frac{1}{\mu}\mathcal{A}^*\hat{y}, x^{k+1} - \hat{x} \rangle \quad (23a)$$

$$= \tau \cdot \langle \frac{1}{\mu}\mathcal{A}^*y^k - \frac{1}{\mu}\mathcal{A}^*\hat{y}, \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y^k) - \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*\hat{y}) \rangle \quad (23b)$$

$$\geq \tau \cdot \left\| \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y^k) - \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*\hat{y}) \right\|_2^2 \quad (23c)$$

$$= \tau^{-1} \cdot \|x^{k+1} - \hat{x}\|_2^2 \quad (23d)$$

Using this inequality, we have

$$\|y^{k+1} - \hat{y}\|_2^2 = \|y^k - \hat{y} + h(b - \mathcal{A}x^{k+1})\|_2^2 \quad (24a)$$

$$= \|y^k - \hat{y} + h(\mathcal{A}\hat{x} - \mathcal{A}x^{k+1})\|_2^2 \quad (24b)$$

$$= \|y^k - \hat{y}\|_2^2 - 2h\mu \langle \frac{1}{\mu}\mathcal{A}^*y^k - \frac{1}{\mu}\mathcal{A}^*\hat{y}, x^{k+1} - \hat{x} \rangle + h^2 \|\mathcal{A}\hat{x} - \mathcal{A}x^{k+1}\|_2^2 \quad (24c)$$

$$\leq \|y^k - \hat{y}\|_2^2 - 2h\frac{\mu}{\tau} \|x^{k+1} - \hat{x}\|_2^2 + h^2 \|\mathcal{A}\|^2 \|x^{k+1} - \hat{x}\|_2^2 \quad (24d)$$

$$= \|y^k - \hat{y}\|_2^2 - h(\frac{2\mu}{\tau} - h\|\mathcal{A}\|^2) \|x^{k+1} - \hat{x}\|_2^2. \quad (24e)$$

Therefore, under the assumption  $0 < h < \frac{2\mu}{\tau\|\mathcal{A}\|^2}$  we can make the following claims:

**claim 1:**  $\|y^{k+1} - \hat{y}\|_2$  is monotonically nonincreasing in  $k$  and thus converges to a limit;

**claim 2:**  $\|x^{k+1} - \hat{x}\|_2$  converges to 0 as  $k$  tends to  $+\infty$ , i.e.,  $\lim_{k \rightarrow +\infty} x^{k+1} = \hat{x}$ .

From claim 1, it follows that  $\{y^k\}$  is bounded and thus has a converging subsequence  $y^{k_i}$ . Let  $\bar{y} = \lim_{i \rightarrow \infty} y^{k_i}$ . By the Lipschitz continuity of the proximal operator, proved in Lemma 1, we have

$$\hat{x} = \lim_{i \rightarrow \infty} x^{k_i+1} = \lim_{i \rightarrow \infty} \tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*y^{k_i}) = \tau \cdot \mathbf{prox}_{\|\cdot\|}(\frac{1}{\mu}\mathcal{A}^*\bar{y}),$$

so  $\bar{y} \in \mathcal{Y}$  by (18). Recall  $\hat{y} \in \mathcal{Y}$  is arbitrary. Hence, claim 1 holds for  $\hat{y} = \bar{y}$ . If  $\{y^k\}$  had another limit point, then  $\|y^{k+1} - \bar{y}\|_2$  would fail to be monotonic. So,  $y^k$  converges to  $\bar{y} \in \mathcal{Y}$  (in norm).  $\square$

**Remark 3.** *Being a dual gradient algorithm, it is well known that the dual objective sequence converges at a rate of  $O(1/k)$ . With Nesterov's acceleration [6], the rate improves to  $O(1/k^2)$ . For piece-wise linear norm  $\|\cdot\|$ , such as the 1-norm, the rate improves to  $O(e^{-k})$  and applies to both the sequence and primal/dual point sequences following the arguments in [5, 21].*

## 5 Extension to gauge

Some interesting models such as those based on total variation, analysis  $\ell_1$ , and fused Lasso use objective functions that are related to but more general than norms. To extend our results to these models, we study the gauge objective.

**Definition 1** (Gauge [12]). *Let  $C \subset \mathbb{R}^n$  be a closed convex set containing the origin. The gauge of  $C$  is the function  $\gamma_C : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by*

$$\gamma_C(x) = \inf\{\lambda > 0 : x \in \lambda C\}.$$

If  $C$  is bounded and symmetric and has a nonempty interior, then  $\gamma_C$  recovers a norm, whose unit ball is  $C$ . If such  $C$  is unbounded, then  $\gamma_C$  generalizes to a semi-norm. Recent paper [15] studies (strongly) piecewise regular gauges for signal recovery, which include analysis-type  $\ell_1$  semi-norms such as total variation and fused Lasso. We consider a general gauge function  $J$  in the following model

$$\min J(x), \quad \text{subject to } \mathcal{A}x = b, \quad (25)$$

and its augmented model

$$\min P(x) \triangleq J(x) + \frac{1}{2\tau} \|x\|_2^2, \quad \text{subject to } \mathcal{A}x = b. \quad (26)$$

## 5.1 Gauge and its polar

We collect the definitions of the polar set and polar gauge, as well as some useful properties from [12].

**Definition 2** (Polar set). *Let  $C \subset \mathbb{R}^n$  be a non-empty closed convex set. The polar of  $C$  is*

$$C^\circ = \{v : \langle v, x \rangle \leq 1, \forall x \in C\}.$$

**Definition 3** (Polar Gauge). *The polar of a gauge  $\gamma_C$  is the function  $\gamma_C^\circ : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by*

$$\gamma_C^\circ(u) = \inf\{\mu \geq 0 : \langle x, u \rangle \leq \mu\gamma_C(x), \forall x\}.$$

**Lemma 5.** *Let  $C \subset \mathbb{R}^n$  be a closed convex set containing the origin. Then,*

- (i).  $\gamma_C^\circ = \gamma_{C^\circ}$ , or equivalently  $C^\circ = \{x : \gamma_C^\circ(x) \leq 1\} = \{x : \gamma_{C^\circ}(x) \leq 1\}$ , which is a closed convex set.
- (ii).  $\gamma_C = \sigma_{C^\circ}$  and  $\gamma_{C^\circ} = \sigma_C$ , where  $\sigma_C(x)$  is the support function of  $C$ .
- (iii). Moreau decomposition:  $v = \mathbf{prox}_{r_C(\cdot)}(v) + \Pi_{C^\circ}(v)$  for any  $v \in \mathbb{R}^n$ .

*Proof.* Parts (i) and (ii) are given in corollaries 15.1.1 and 15.1.2 in [12], respectively. To show part (iii), let  $f(x) = \delta_{C^\circ}(x)$ , where  $\delta_C(x)$  is the indicator function

$$\delta_C(x) = \begin{cases} +\infty & x \notin C \\ 0 & x \in C. \end{cases}$$

Then, the convex conjugate of  $f(x)$  is

$$f^*(x) = \sup_y (\langle x, y \rangle - \delta_{C^\circ}(y)) = \max_{y \in C^\circ} \langle x, y \rangle \stackrel{\theta_1}{=} \sigma_{C^\circ}(x) \stackrel{\theta_2}{=} \gamma_C(x),$$

where  $\theta_1$  follows from the definition of support function and  $\theta_2$  follows from part (ii). From  $\mathbf{prox}_{f(\cdot)}(v) = \Pi_{C^\circ}(v)$  and  $v = \mathbf{prox}_{f(\cdot)}(v) + \mathbf{prox}_{f^*(\cdot)}(v)$ , the result follows.  $\square$

## 5.2 Dual analysis and algorithm

Let  $J(x) = \gamma_C(x)$  be a gauge function. Based on part (ii) of Lemma 5, we have  $\gamma_C(x) = \sigma_{C^\circ}(x) = \max_{z \in C^\circ} \langle x, z \rangle$ , from which we can follow subsection 3.2 and derive the dual problem of (26):

$$\min_y D_J(y) := -\langle y, b \rangle + \frac{\tau}{2} \min_{z \in C^\circ} \|\mathcal{A}^*y - z\|_2^2, \quad (27)$$

where the optimal  $z = \Pi_{C^\circ}(\mathcal{A}^*y)$  is a function of  $y$ . Following the definition of  $h_{\mathcal{Z}}(x)$  in (10), we have

$$D_J(y) = -\langle y, b \rangle + \frac{\tau}{2} h_{C^\circ}(\mathcal{A}^*y)^2.$$

By (12) and part (iii) of Lemma 5, we obtain

$$\nabla D_J(y) = -b + \tau \mathcal{A} \cdot \mathbf{prox}_{\gamma C(\cdot)}(\mathcal{A}^*y) = -b + \tau \mathcal{A} \cdot \mathbf{prox}_{J(\cdot)}(\mathcal{A}^*y). \quad (28)$$

We have the primal-dual relationship:  $x = \tau(\mathcal{A}^*y - z)$  and  $z = \Pi_{C^o}(\mathcal{A}^*y)$ . If  $y$  is dual optimal, the equations give  $x$  that is primal optimal. Similar to Lemma 4, we have the following result:

**Lemma 6.** *Let  $\hat{x}$  be the unique solution to problem (26). Then the dual solution set to problem (27) is*

$$\mathcal{W} = \left\{ y : \tau \cdot \mathbf{prox}_{J(\cdot)}(\mathcal{A}^*y) = \hat{x} \right\}, \quad (29)$$

which is nonempty and convex.

Based on (28), one can obtain the dual gradient ascent iteration for problem (26). From the above primal-dual relationship, we give the primal-dual form of this algorithm as follows:

$$\begin{cases} x^{k+1} = \tau \cdot \mathbf{prox}_{J(\cdot)}(\mathcal{A}^*y^k) \\ y^{k+1} = y^k + h(b - \mathcal{A}x^{k+1}). \end{cases} \quad (30)$$

Similar to Theorem 1, we can show:

**Theorem 2.** *Set step size  $h \in (0, \frac{2}{\tau \|\mathcal{A}\|^2})$  and  $y^0 = 0$  in iteration (30). Let  $\hat{x}$  be the unique minimizer to problem (26) and  $\mathcal{W}$  be the solution set to problem (27). Then,  $\lim_{k \rightarrow +\infty} x^k = \hat{x}$ , and there exists a point  $\bar{y} \in \mathcal{W}$  such that  $\lim_{k \rightarrow +\infty} y^k = \bar{y}$ .*

## Acknowledgements

We would like to thank Professor Jian-Feng Cai (U. Iowa) for suggestions and corrections. The work of H. Zhang is supported in part by the Graduate School of NUDT under Fund of Innovation B110202, Hunan Provincial Innovation Foundation for Postgraduate CX2011B008, and NSFC grant 61201328. The work of L. Cheng is supported in part by NSFC grants 61271014 and 61072118. The work of W. Yin is supported in part by NSF grants DMS-0748839 and ECCS-1028790.

## References

- [1] J.-F. Cai, S. Osher, and Z. Shen. Linearized Bregman iterations for compressed sensing. *Mathematics of Computation*, 78(267):1515-1536, 2009.
- [2] J.-F. Cai, S. Osher, and Z. Shen. Convergence of the linearized Bregman iteration for  $\ell_1$ -norm minimization. *Mathematics of Computation*, 78(268):2127-2136, 2009.
- [3] J.-F. Cai, E. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956-1982, 2010.
- [4] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33-61, 1999.
- [5] M. J. Lai and W. Yin. Augmented  $\ell_1$  and nuclear-norm models with a globally linearly convergent algorithm. *SIAM. J. Imaging Sci.*, 6(2):1059-1091, 2013.
- [6] B. Huang, S. Q. Ma, and D. Goldfarb. Accelerated Linearized Bregman Method. *Journal of Scientific Computation*, 54:428-453, 2013.
- [7] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*. Kluwer Academic Publishers, 2004.



- [8] Y. Nesterov. Gradient methods for minimizing composite objective function. CORE discussion paper, 2007.
- [9] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Ser. A Math. 255:2897-2899, 1962.
- [10] B. O’Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. To appear in Foundations of Computational Mathematics, 2012.
- [11] N. Parikh, and S. Boyd. Proximal algorithm. Foundations and Trends in optimization, 1-96, 2013.
- [12] R. T. Rockafellar. Convex analysis, Princeton, 1970.
- [13] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D, 60:259-268, 1992.
- [14] A. Ruszczyński. Nonlinear optimization. Princeton and Oxford, 2006.
- [15] S. Vaiter, M. Golbabaee, J. Fadili, and G. Peyré. Model selection with piecewise regular gauges, arXiv:1307.2342v1, 2013.
- [16] Y. You, Q. Wan, and Y. Liu. A short note on strongly convex programming for exact matrix completion and robust principal component analysis. Inverse Problem and Imaging, 7(1): 305-306, 2013.
- [17] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences, 1(1):143-168, 2008.
- [18] W. Yin. Analysis and generalizations of the linearized Bregman method. SIAM Journal on Imaging Sciences, 3(4):856-877, 2010.
- [19] H. Zhang, L. Cheng, and W. Zhu. A lower bound guaranteeing exact matrix completion via singular value thresholding algorithm. Applied and Computational Harmonic Analysis, 31(3):454-459, 2011.
- [20] H. Zhang, J.-F. Cai, L. Cheng, and J. Zhu. Strongly convex programming for exact matrix completion and robust principal component analysis. Inverse Problem and Imaging, 6(2): 357-372, 2012.
- [21] H. Zhang and W. Yin. Gradient methods for convex minimization: better rates under weaker conditions. UCLA CAM Report(13-17), 2013.
- [22] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B, 67:301-320, 2005.