ICIC Express Letters
Part B: Applications
Volume **8**, Number **3**, March **2017**

ICIC International ©2017 ISSN 2185-2766
pp. **563–569**

AN ACCELERATED AUGMENTED LAGRANGIAN MULTIPLIER METHOD FOR CONVEX OPTIMIZATION WITH APPLICATION TO IMAGE RECONSTRUCTION

Pibin Bing and Shichao Huang

Institute of Electric Power
North China University of Water Resources and Electric Power
No. 36, Beihuan Road, Zhengzhou 450045, P. R. China
bingpibin@126.com

Received August 2016; accepted November 2016

Abstract. *The augmented Lagrangian multiplier method (ALM) is a highly influential method for optimization problems, and it has been successfully applied in various fields, such as compressive sensing, matrix completion, and image recovery. However, its efficiency is dominated by how to efficiently solve the involved subproblem, which is often as hard as the original problem. In this paper, we propose an accelerated ALM (AALM) for a linearly constrained convex optimization, which incorporates two accelerating skills to make our AALM implementable for most optimization problems in practice. Under mild conditions, the global convergence of the new AALM is proved. Furthermore, we also extend the new AALM to solve the similar problem but with the linear inequality constraints. Some preliminary numerical experiments on image reconstruction are performed to show the improvements of the new method.*
Keywords: Augmented Lagrangian multiplier method, Accelerating skill, Global convergence, Image reconstruction

1. **Introduction.** In this paper, we consider a special but quite useful convex optimization, that is the linearly constrained convex optimization, which is to solve

$$\min\{f(x)|Ax = b, x \in \mathcal{X}\}, \tag{1}$$

where $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$, $\mathcal{X} \subseteq \mathcal{R}^n$ is a closed convex set and $f(x) : \mathcal{R}^n \to \mathcal{R}$ is a proper convex function. The solution set of (1), denoted by $\mathcal{X}^*$, is assumed to be nonempty and compact. Problem (1) is the mathematical model of many important problems arising from machine learning, statistical inference, image processing, and sparse signals recovery, see [1,2] and reference therein.

Let $\beta > 0$ be a given penalty parameter. The augmented Lagrangian function of problem (1) is defined by, for any $x \in \mathcal{X}$,

$$\mathcal{L}_\beta(x, \lambda) := f(x) - \langle\lambda, Ax - b\rangle + \frac{\beta}{2}\|Ax - b\|^2,$$

where $\lambda \in \mathcal{R}^m$ is the dual variable of (1). A benchmark solver for (1) is the augmented Lagrangian multiplier method (ALM) (see [2-7] for the development of ALM). Given λ^k, ALM takes the following scheme for $k = 0, 1, \ldots$,

$$\begin{cases} x^{k+1} \in \operatorname{argmin}_{x\in\mathcal{X}} \left\{\mathcal{L}_\beta(x, \lambda^k)\right\}, \\ \lambda^{k+1} = \lambda^k - \beta\left(Ax^{k+1} - b\right). \end{cases} \tag{2}$$

At each iteration, the computation of ALM for (2) is dominated by solving the following subproblem:

$$\min\left\{f(x) + \frac{\beta}{2}\|Ax - a\|^2 | x \in \mathcal{X}\right\}$$

with a certain known $a \in \mathcal{R}^m$. Obviously, the above problem is often difficult compared with the following problem:

$$\min\left\{f(x)+\frac{\beta}{2}\|x-a\|^2 | x\in\mathcal{X}\right\}.$$

In fact, as stated in [5], when $\mathcal{X}=\mathcal{R}^m$, the above subproblem is just the resolvent operator of $f(x)$, which is defined as $\left(I+\frac{1}{\beta}\partial f\right)^{-1}(a)$. Here, $\partial f(x)$ denotes the subdifferential of a convex but nonsmooth function f at the point x, which is defined by

$$\partial f(x)=\{\xi | f(y)-f(x)\geq\langle\xi, y-x\rangle, \forall y\in\mathcal{R}^n\}.$$

Fortunately, the resolvent operator of $f(x)$ in many practical applications often has closed form solution, such as

$$x^{k+1}=\text{shrink}_{1,2}(a,1/\beta)\doteq\text{sign}(a)\cdot\max\{0,|a|-1/\beta\},\ \text{for } f(x)=\|x\|_1,$$

or

$$x^{k+1}=\text{shrink}_{2,2}(a,1/\beta)\doteq\frac{a}{\|a\|}\cdot\max\{0,\|a\|-1/\beta\},\ \text{for } f(x)=\|x\|,$$

where $\text{sign}(\cdot)$ is the sign function.

In this paper, we further study the ALM and propose a new accelerated ALM (AALM) which adopts two acceleration skills. First, a variable regularization term is added to the subproblem of (2) to linearize the quadratic term $\frac{1}{2}\|Ax-b\|^2$, and this procedure often makes that the resulting subproblem has closed form solution in many practical applications. Then the computational load of the corresponding method is greatly reduced at each iteration; Second, we use an extrapolating step in the primal and dual variables (x,λ), which includes a problem-depending parameter $\tau\in(0,2)$, and $\tau>1$ often leads to better convergence behavior. Thus, we can choose some proper τ to accelerate the convergence speed of AALM. It is noteworthy that by setting different parameters, our new AALM includes the famous customized proximal point algorithm [9], the linearized augmented Lagrangian method [10], and the Bregman operator splitting algorithm [11] as special cases.

The outline of this paper is as follows. In Section 2, we review some preliminaries and deliver our proposed method. In Section 3, we analyze the global convergence of the new method. The extension of the new method is also presented in the section. In Section 4, we report some numerical results on image reconstruction to verify the efficiency of the new AALM. Some conclusions are drawn in Section 5.

2. **Preliminaries and the Algorithm.** Firstly, we characterize the optimal condition of (1) as a variational reformulation, which makes our analysis quite succinct. Obviously, the first-order optimality condition of (1) is: Finding a vector $w^*\in\mathcal{W}$ such that

$$f(x)-f(x^*)+(w-w^*)^\top F(w^*)\geq 0,\quad \forall w\in\mathcal{W}, \tag{3}$$

where $\mathcal{W}=\mathcal{X}\times\mathcal{R}^l$, and

$$w=\begin{pmatrix} x\\ \lambda\end{pmatrix} \text{ and } F(w)=\begin{pmatrix} -A^\top\lambda\\ Ax-b\end{pmatrix}. \tag{4}$$

Obviously, the mapping $F(w)$ defined in (4) is affine with a skew-symmetric matrix; it is easily verified that $F(w)$ is monotone. We denote by $\mathcal{W}^*$ the solution set of (3), which is nonempty under the aforementioned assumption of the solution set $\mathcal{X}^*$ of (1).

In [12], by introducing a free parameter $t\in(-\infty,+\infty)$, Ma and Ni proposed the following customized ALM:

$$\left\{\begin{array}{l} x^{k+1}\in\text{argmin}_{x\in\mathcal{X}}\left\{f(x)+\frac{r}{2}\left\|x-x^k-\frac{1}{r}A^\top\left(\lambda^k-\frac{1}{s}(1-t)(Ax^k-b)\right)\right\|^2\right\},\\ \lambda^{k+1}=\lambda^k-\frac{1}{s}\left((1+t)Ax^{k+1}-tAx^k-b\right),\end{array}\right. \tag{5}$$

where $r > 0$, $s > 0$, $rs > \|A^\top A\|$. Obviously, the subproblem in (5) only needs to solve the resolvent operator of $f(x)$.

Based on (5), we describe our accelerated ALM as follows.

Algorithm 2.1. (Accelerated ALM)

Step 0. Given any initial point $(x^0, \lambda^0) \in \mathcal{W}$, the parameters $r > 0$, $s > 0$ satisfying $rs > \|A^\top A\|$, and $t \in (-\infty, +\infty)$, $\tau \in (0, 2)$. Set $k = 0$.

Step 1. Compute the temporary point $\tilde{w}^k = \left(\tilde{x}^k, \tilde{\lambda}^k\right)$ by

$$\begin{cases} \tilde{x}^k \in \operatorname{argmin}_{\mathrm{x}\in\mathcal{X}} \left\{ f(x) + \frac{r}{2} \left\| x - x^k - \frac{1}{r} A^\top \left(\lambda^k - \frac{1}{s}(1-t)(Ax^k - b) \right) \right\|^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{s}\left((1+t)A\tilde{x}^k - tAx^k - b \right), \end{cases} \tag{6}$$

Step 2. If $w^k = \tilde{w}^k$, then stop; otherwise, go to Step 3.

Step 3. Compute the new iterate $w^{k+1} = \left(x^{k+1}, \lambda^{k+1}\right)$ by

$$w^{k+1} = w^k - \tau\left(w^k - \tilde{w}^k\right). \tag{7}$$

Set $k := k + 1$ and go to Step 1.

Remark 2.1. *If $\tau = 1$, Algorithm 2.1 reduces to the customized ALM in [12]. That is to say, our new method includes the customized ALM in [12] as a special case. Then, from [12], our new method also includes the methods in [9-11] as special cases.*

Remark 2.2. *If $w^k = \tilde{w}^k$, then $w^k \in \mathcal{W}^*$, i.e., it is a solution of (3), and x^k is a solution of (1). The reason is as follows. First, from the second equation of (6), $x^k = \tilde{x}^k$ and $\lambda^k = \tilde{\lambda}^k$, we can get*

$$Ax^k - b = 0. \tag{8}$$

Note that the first equation of (6) is equivalent to the following subproblem: find $\tilde{x}^k \in \mathcal{X}$, such that

$$f(x') - f\left(\tilde{x}^k\right) + \left(x' - \tilde{x}^k\right)^\top \left\{ -A^\top \tilde{\lambda}^k + \left(rI + \frac{1}{s}(t^2 - 1)A^\top A \right) \left(\tilde{x}^k - x^k\right) \right.$$
$$\left. + \, tA^\top \left(\tilde{\lambda}^k - \lambda^k \right) \right\} \geq 0, \quad \forall x' \in \mathcal{X}.$$

Therefore, when $w^k = \tilde{w}^k$, the above problem reduces to: find $\tilde{x}^k \in \mathcal{X}$, such that

$$f(x') - f\left(x^k\right) + \left(x' - x^k\right)^\top \left(-A^\top \lambda^k\right) \geq 0, \quad \forall x' \in \mathcal{X}. \tag{9}$$

Then, from (3), (8) and (9), we have $w^k \in \mathcal{W}^*$.

In the numerical experiment, we can use $\|w^k - \tilde{w}^k\| \leq \varepsilon$ as the stopping criterion, where $\varepsilon > 0$ is a prescribed sufficiently small parameter, such as $\varepsilon = 10^{-4}$, 10^{-6}.

3. **Global Convergence and Extension.** To prove the global convergence of Algorithm 2.1, we need to make the following matrix G to be positively definite:

$$G = \begin{pmatrix} rI + \frac{1}{s}(t^2 - 1)A^\top A & tA^\top \\ tA & sI \end{pmatrix}.$$

Obviously, we only need to impose the following parameter condition:

$$r > 0, \quad s > 0, \quad rs > \|A^\top A\|.$$

In fact, for any $(x, \lambda) \neq 0$, we have

$$w^\top G w = x^\top \left(rI + \frac{1}{s}(t^2 - 1)A^\top A \right) x + 2t(Ax)^\top \lambda + s\|\lambda\|^2.$$

- If $x = 0$, then $\lambda \neq 0$; therefore, $w^\top G w > 0$;

• If $x \neq 0$, then

$$\begin{aligned} w^\top G w &\geq r\|x\|^2 + \frac{1}{s}\big(t^2-1\big)\|Ax\|^2 - s\|\lambda\|^2 - \frac{t^2}{s}\|Ax\|^2 + s\|\lambda\|^2 \\ &= \left(r - \frac{\|A^\top A\|}{s}\right)\|x\|^2 \\ &> 0, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwartz inequality, and the second inequality follows from $r>0$, $s>0$, $rs > \left\|A^\top A\right\|$.

Then, by some simple manupulations, (6) can be rewritten as the following compact form:

$$f(x') - f\big(\tilde{x}^k\big) + \big\langle w' - \tilde{w}^k, F\big(\tilde{w}^k\big) - G\big(w^k - \tilde{w}^k\big)\big\rangle \geq 0, \quad \forall w' \in \mathcal{W}. \tag{10}$$

Let $w' = w^*$ in (10), we obtain

$$f(x^*) - f\big(\tilde{x}^k\big) + \big\langle w^* - \tilde{w}^k, F\big(\tilde{w}^k\big) - G\big(w^k - \tilde{w}^k\big)\big\rangle \geq 0.$$

That is

$$\begin{aligned} \big\langle \tilde{w}^k - w^*, G\left(w^k - \tilde{w}^k\right)\big\rangle &\geq f\big(\tilde{x}^k\big) - f\big(x^*\big) + \big\langle \tilde{w}^k - w^*, F\big(\tilde{w}^k\big)\big\rangle \\ &\geq f\big(\tilde{x}^k\big) - f\big(x^*\big) + \big\langle \tilde{w}^k - w^*, F(w^*)\big\rangle \\ &\geq 0, \end{aligned} \tag{11}$$

where the second inequality follows from the monotonicity of $F(\cdot)$, and the third inequality follows from $w^* \in \mathcal{W}^*$. Obviously, the result of (11) can be rewritten as

$$\big\langle w^k - w^*, G\big(w^k - \tilde{w}^k\big)\big\rangle \geq \big\|w^k - \tilde{w}^k\big\|_G, \tag{12}$$

which indicates that $-\big(w^k - \tilde{w}^k\big)$ is a descent direction of the merit function $\frac{1}{2}\big\|w - w^*\big\|_G^2$ at the point w^k, which also indicates that our correction step (7) is reasonable.

Lemma 3.1. *The sequence $\{w^k\}$ generated by Algorithm 2.1 satisfies*

$$\big\|w^{k+1} - w^*\big\|_G^2 \leq \big\|w^k - w^*\big\|_G^2 - \tau(2-\tau)\big\|w^k - \tilde{w}^k\big\|_G^2. \tag{13}$$

Proof: By (12), we have

$$\begin{aligned} \big\|w^{k+1} - w^*\big\|_G &= \big\|w^k - \tau\big(w^k - \tilde{w}^k\big) - w^*\big\|_G \\ &= \big\|w^k - w^*\big\|_G - 2\tau\big\langle w^k - w^*, G\big(w^k - \tilde{w}^k\big)\big\rangle + \tau^2\big\|w^k - \tilde{w}^k\big\|_G^2 \\ &\leq \big\|w^k - w^*\big\|_G - \tau(2-\tau)\big\|w^k - \tilde{w}^k\big\|_G^2, \end{aligned}$$

which completes the proof of the lemma.

Theorem 3.1. *The sequence $\{w^k\}$ generated by Algorithm 2.1 converges to a solution point in $\mathcal{W}^*$ globally.*

Proof: From (13) and $\tau \in (0,2)$, we can get

$$\big\|w^{k+1} - w^*\big\|_G^2 \leq \big\|w^k - w^*\big\|_G^2.$$

Therefore, $\left\{\big\|w^k - w^*\big\|_G^2\right\}$ is a descent sequence. Therefore, $\{w^k\}$ is a bounded sequence, and it has at least one cluster point, saying $\hat{w}$, and there is a subsequence $\{w^{k_i}\}$ such that

$$\lim_{i\to\infty} w^{k_i} \to \hat{w}.$$

Then, by (13) again, we have

$$\sum_{k=0}^{\infty} \big\|w^k - \tilde{w}^k\big\|_G^2 \leq \frac{1}{\tau(2-\tau)}\big\|w^0 - w^*\big\|_G,$$

which indicates that

$$\lim_{k\to\infty} \left\|w^k - \tilde{w}^k\right\| = 0.$$

Then, following the same analysis in Remark 2.2, we have

$$A\hat{x} = b, \text{ and } f(x') - f(\hat{x}) + (x' - \hat{x})^\top \left(-A^\top \hat{\lambda}\right) \geq 0, \quad \forall x' \in \mathcal{X}.$$

This shows that $\hat{w} \in \mathcal{W}^*$. That is to say, $\hat{w}$ is a solution of (3). Then the following proof follows immediately from the results in [13]. We omit the detail for conciseness.

Now, let us extend Algorithm 2.1 to solve the inequality constrained convex optimization problem:

$$\min\{f(x)|Ax \geq b, x \in \mathcal{X}\}.$$

Similarly, by introducing a dual variable $\lambda \in \mathcal{R}_+^m$, we can transform the above problem into an equivalent variational inequality with $\mathcal{W} = \mathcal{X} \times \mathcal{R}_+^m$. Then, based on Algorithm 2.1, we present the following algorithm for solving the above problem.

Algorithm 2.2.

Step 0. Given any initial point $(x^0, \lambda^0) \in \mathcal{W}$, the parameters $r > 0$, $s > 0$ satisfying $rs > \|A^\top A\|$, and $t \in (-\infty, +\infty)$, $\tau \in (0, 2)$. Set $k = 0$.

Step 1. Compute the temporary point $\tilde{w}^k = \left(\tilde{x}^k, \tilde{\lambda}^k\right)$ by

$$\begin{cases} \tilde{x}^k \in \operatorname{argmin}_{x\in\mathcal{X}} \left\{ f(x) + \frac{r}{2}\left\|x - x^k - \frac{1}{r}A^\top \left(\lambda^k - \frac{1}{s}(1-t)\left(Ax^k - b\right)\right)\right\|^2 \right\}, \\ \tilde{\lambda}^k = P_{\mathcal{R}^m}\left[\lambda^k - \frac{1}{s}\left((1+t)A\tilde{x}^k - tAx^k - b\right)\right], \end{cases}$$

Step 2. If $w^k = \tilde{w}^k$, then stop; otherwise, go to Step 3.

Step 3. Compute the new iterate $w^{k+1} = \left(x^{k+1}, \lambda^{k+1}\right)$ by

$$w^{k+1} = w^k - \tau\left(w^k - \tilde{w}^k\right).$$

Set $k := k + 1$ and go to Step 1.

Theorem 3.2. *The sequence $\{w^k\}$ generated by Algorithm 2.2 converges to a solution point in $\mathcal{W}^*$ globally.*

4. **Applications in Image Reconstruction.** In this section, we apply Algorithm 2.1 to testing a concrete application of the model (1), that is image reconstruction problem, and report the numerical results. All the codes were written by Matlab 7.10 and were conducted on ThinkPad notebook with Pentium(R) Dual-Core CPU T4400@2.2GHz, 2GB of memory.

Let $\mathbf{x} \in \mathcal{R}^l$ represent a digital image with $l = l_1 \times l_2$, and $W \in \mathcal{R}^{l\times n}$ be a wavelet dictionary. Set $\mathbf{x} = Wx$ with x being a spars vector. Then, the image reconstruction problem is to recover the clear image x based on some observation b, and its mathematical model is:

$$\min\{\|x\|_1|BWx = b\},$$

where $B \in \mathcal{R}^{m\times l}$ is a diagonal matrix whose elements are either 0 (missing pixels) or 1 (known pixels), and $B = SH$, where $S \in \mathcal{R}^{m\times l}$ is a downsampling matrix generated by the following subroutines

```
dtx = 64; dty = 64; S1 = rand(dtx,dty)>0.6;
```

and

```
S2 = ones(n1/dtx,n2/dty); S = kron(S1,S2).
```

Furthermore, we set $H =$ `fspecial('disk',7)` $\in \mathcal{R}^{l\times l}$ is a blurry matrix, and therefore H can be diagonalized by the discrete consine transform. For the special B, W, we have $\|A^\top A\| = 1$. More information about the image reconstruction problem can be found in [9,12].

In our numerical experiment, we set $r = 0.6$, $s = 1.5$, $t = 0$, $\tau = 1.9$ and the maximized iteration number is set 300. The initial point is set to be $(x^0, \lambda^0) = \big(W^\top(b), 0\big)$. We test 512×512 image of Lena.png, 256×256 images of Chart.tiff and House.png for the image reconstruction problem. To assess the restoration performance qualitatively, we use the signal to noise ratio (SNR) defined as

$$\mathrm{SNR} = 20\log_{10}\frac{\|\mathbf{x}\|}{\|\tilde{\mathbf{x}} - \mathbf{x}\|}.$$

Here $\mathbf{x}$ is the true image, and $\tilde{\mathbf{x}}$ is the restored image.

The original images, the degraded images and the images restored by Algorithm 2.1 are shown in Figure 1, and we also plot the evolutions of SNR with respect to the number of iterations for Algorithm 2.1 in Figure 2. From Figures 1 and 2, it is clear that Algorithm 2.1 can recover the degraded images successfully. Therefore, Algorithm 2.1 is an efficient method for the image reconstruction problem.

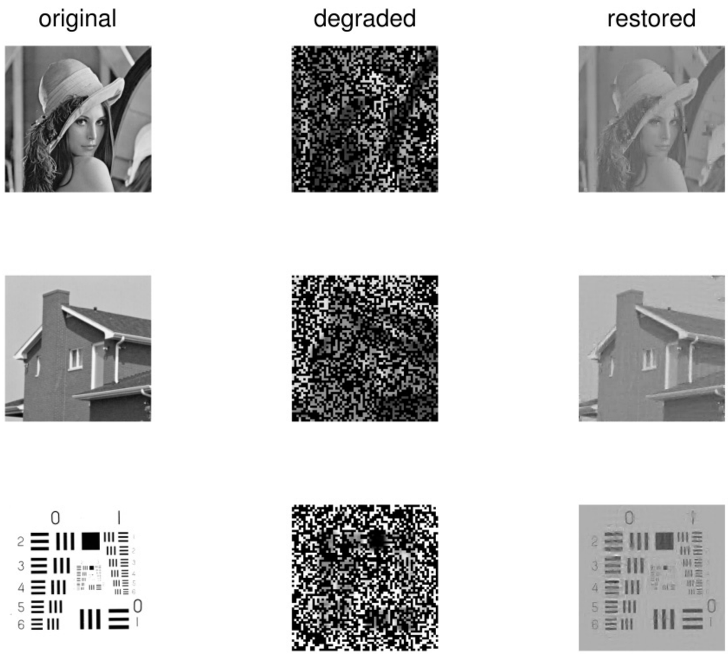

FIGURE 1. Original images, degraded images and restored images

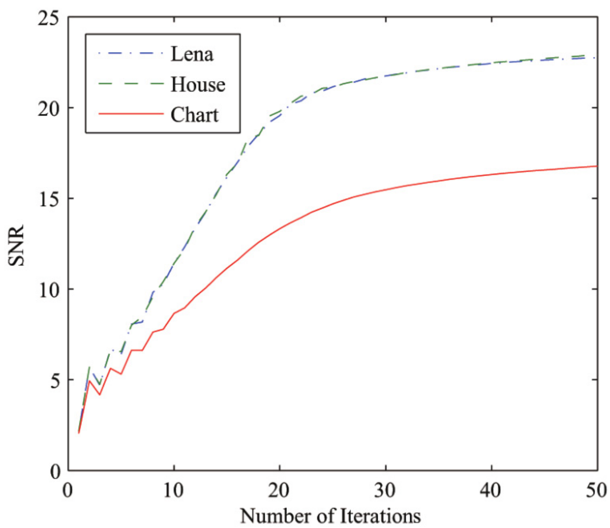

FIGURE 2. Evolutions of SNRs with respect to the number of iterations

5. **Conclusions.** In this paper, we have proposed an accelerated ALM method for the linear equality constrained convex optimization, and extend it to solve the convex optimization with the linear inequality constraints. The global convergence of the proposed methods is proved under quite mild conditions, and some numerical results on image reconstructions are presented, which indicate that our new method is quite efficient for the tested problems. In the future, the following two problems need to research: (I) the restriction $rs > \|A^\top A\|$ imposed on the two parameters r, s is too stringent, and relaxing or removing this restriction is quite meaningful; (II) the parameter τ is quite problem-depending, and some adaptive scheme should be designed to choose proper τ for different problems.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 61601183), the Major Science and Technology Project Education Department of Henan Province (Grant No. 17A510010) and the Basic Research Projects and Advanced Technology of Henan Province (Grant No. 112300410326).

REFERENCES

[1] W. T. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for l_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.*, vol.1, pp.143-168, 2008.
[2] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imag. Vis.*, vol.40, pp.120-145, 2011.
[3] M. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, vol.4, pp.303-320, 1969.
[4] B. S. He, L. S. Hou and X. M. Yuan, On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming, *SIAM J. Optim.*, vol.25, no.4, pp.2274-2312, 2015.
[5] N. Parikh and S. P. Boyd, Proximal algorithms, *Found. Trends Opt.*, vol.1, no.3, pp.123-231, 2014.
[6] Z. C. Lin, R. S. Liu and H. Li, Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning, *Machine Learning*, vol.95, no.2, pp.287-325, 2015.
[7] M. J. Powell, A method for non-linear constraints in minimization problems, in *Optimization*, R. Fletcher (ed.), Academic Press, Waltham Massachusetts, 1969.
[8] B. S. He, PPA-like contraction methods for convex optimization: A framework using variational inequality approach, *J. Oper. Res. Soc. China*, vol.3, pp.391-420, 2015.
[9] B. S. He, X. M. Yuan and W. X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, *Comput. Optim. Appl.*, vol.56, pp.559-572, 2013.
[10] J. F. Yang and X. M. Yuan, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, *Math. Comput.*, vol.82, no.281, pp.301-329, 2013.
[11] X. Q. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM J. Imaging Sci.*, vol.3, pp.253-276, 2010.
[12] F. Ma and M. F. Ni, A class of customized proximal point algorithms for linearly constrained convex optimization, *Comp. Appl. Math.*, 2016.
[13] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Berlin, 2011.