Research Article

A Line-Search-Based Partial Proximal Alternating Directions Method for Separable Convex Optimization

Yu-hua Zeng,[1,2] Yu-fei Yang,[1] and Zheng Peng[3]

[1] College of Mathematics and Econometrics, Hunan University, Changsha 410082, China
[2] Department of Mathematics, Hunan First Normal University, Changsha 410205, China
[3] College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China

Correspondence should be addressed to Yu-fei Yang; mayfyang@163.com

We propose an appealing line-search-based partial proximal alternating directions (LSPPAD) method for solving a class of separable convex optimization problems. These problems under consideration are common in practice. The proposed method solves two subproblems at each iteration: one is solved by a proximal point method, while the proximal term is absent from the other. Both subproblems admit inexact solutions. A line search technique is used to guarantee the convergence. The convergence of the LSPPAD method is established under some suitable conditions. The advantage of the proposed method is that it provides the tractability of the subproblem in which the proximal term is absent. Numerical tests show that the LSPPAD method has better performance compared with the existing alternating projection based prediction-correction (APBPC) method if both are employed to solve the described problem.

1. Introduction

In this paper, we consider a separable convex optimization problem of the form

$$\min \quad \theta_1(x) + \theta_2(y),$$
$$\text{s.t.} \quad Ax - y = 0, \tag{1}$$

where $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $\theta_1 : \mathcal{X} \rightarrow R$, and $\theta_2 : \mathcal{Y} \rightarrow R$ are convex functions; $\mathcal{X} \subset R^n$, $\mathcal{Y} \subset R^m$ are closed convex sets; $A \in R^{m \times n}$ is a given matrix; and $\text{rank}(A) = \min\{m, n\}$.

The augmented Lagrangian function associated with the problem (1) is

$$L_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax - y) + \frac{\beta}{2}\|Ax - y\|^2. \tag{2}$$

For simplicity of analysis, we assume the objective function is continuously differentiable. Let $f(x) = \nabla\theta_1(x)$ and $g(y) = \nabla\theta_2(y)$; by the convexity of the functions θ_1 and θ_2, f and g are monotone in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Thus, by

the optimality, the problem (1) is equivalent to the following monotone variational inequalities: find $(x, y, \lambda) \in \mathcal{W}$ such that

$$(x' - x)^T \left[f(x) - A^T(\lambda - \beta(Ax - y)) \right] \geq 0, \quad \forall x' \in \mathcal{X},$$

$$(y' - y)^T \left[g(y) + (\lambda - \beta(Ax - y)) \right] \geq 0, \quad \forall y' \in \mathcal{Y},$$

$$(\lambda' - \lambda)^T (Ax - y) \geq 0, \quad \forall \lambda' \in R^m, \tag{3}$$

where $\mathcal{W} = \mathcal{X} \times \mathcal{Y} \times R^m$.

Chen and Teboulle [1] investigated the problem (1) and proposed a proximal-based decomposition method. Tseng [2] interpreted Chen and Teboulle's approach as an alternating version of the proximal point method and the extragradient method. Furthermore, Tseng generalized their method to solve much broader classes of problems and to yield new decomposition methods for convex programming and variational inequalities.

Indeed, there are many methods to deal with the problem (1), or its equivalent version (3), in the literature. Among these

methods, proximal point method and alternating directions method are power tools, for example, [3–8].

Throughout this paper, we assume that the solution set of the problem (1) or, equivalently, the solution set of SVIs (3), denoted by $\mathscr{W}^*$, is nonempty. The notation $\| \cdot \|$ denotes the Euclidean norm, $\| \cdot \|_\infty$ denotes the infinite norm defined by $\| u \|_\infty = \max_i\{|u_i|\}$, and $\| \cdot \|_\infty$ denotes the l_1-norm defined by $\| u \|_1 = \sum_{i=1}^n |u_i|$.

From a given $w^k = (x^k, y^k, \lambda^k)$, the classical proximal alternating directions method produces the new iterate triple $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the following scheme.

(1) Find $\hat{x}^k$ via solving

$$x \in \mathscr{X},$$

$$\left(x' - x\right)^T \left\{ f\left(x^k\right) - A^T \left[\lambda^k - \beta_k \left(A x^k - y^k\right)\right] \right. \tag{4}$$

$$\left. + r_k \left(x - x^k\right)\right\} \geq 0, \quad \forall x' \in \mathscr{X}.$$

(2) Find $\hat{y}^k$ via solving

$$y \in \mathscr{Y}, \left(y' - y\right)^T$$

$$\times \left\{ g\left(y^k\right) + \left[\lambda^k - \beta_k \left(A \hat{x}^k - y^k\right)\right] \right. \tag{5}$$

$$\left. + s_k \left(y - y^k\right)\right\} \geq 0, \quad \forall y' \in \mathscr{Y}.$$

(3) Update $\hat{\lambda}^k$ via

$$\hat{\lambda}^k = \lambda^k - \beta_k \left(A \hat{x}^k - \hat{y}^k\right), \tag{6}$$

where β_k is a given penalty parameter of the linear constraint $Ax - y = 0$. The coefficients $r_k > 0$ in formulas (4) and $s_k > 0$ in (5) are referred to as proximal parameters. The method is convergent by setting $w^{k+1} = \hat{w}^k = (\hat{x}^k, \hat{y}^k, \hat{\lambda}^k)$ (for a proof see [3]).

However, finding a solution of the subproblem (4) or (5) is not an easy task, since each of them requires an implicit projection. He et al. [4] suggested solving inexactly the subproblems (4) and (5). Their method is referred to as alternating projection based prediction-correction (APBPC) method. In [4], the authors considered a generalized version of the problem (1) with the constraints $Ax + By = b$. The problem under consideration in this paper, that is, problem (1), has some features depending on $B = -I$. These features provide some advantages for constructing more efficient method.

This paper is organized as follows. Section 2 proposes a line-search-based partial proximal alternating directions method for solving the problem (1) and states some useful notation. The proposed method can be looked as to a proximal alternating directions method by setting $s_k = 0$ in (5) and admits inexact solution at all iterations. Thus, the proposed method is actually an inexact partial proximal alternating direction method using a certain line-search technique. In Section 3, we provide two descent directions of a given merit function and prove the descent property of the proposed method. In Section 4, the convergence property of the proposed method is established under some suitable conditions.

Section 5 gives some preliminary numerical experimental results on the compressed sensor problem. These results show that the proposed method has better performance compared with the existing APBPC method, when both are used to solve the problem with the described features. Finally, some concluding remarks are given in Section 6.

2. Method Description

In this section, we first describe the line-search-based partial proximal alternating directions method for the problem (3) in what follows.

Line-Search-Based Partial Proximal Alternating Directions (LSPPAD) Method. Let $\nu \in (0, 1)$. For a given $w^k = (x^k, y^k, \lambda^k) \in \mathscr{W}$, the LSPPAD method produces the new triple $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the following scheme.

(s1) Solving the following variational inequality, find $x \in \mathscr{X}$ such that

$$\left(x' - x\right)^T \left\{ f\left(x^k\right) - A^T \left[\lambda^k - \beta_k \left(A x^k - y^k\right)\right] \right.$$

$$\left. + r_k \left(x - x^k\right)\right\} \geq 0, \tag{7}$$

$$\forall x' \in \mathscr{X}.$$

And denote the solution by $\hat{x}^k$. Let $\xi_x^k = f(x^k) - f(\hat{x}^k) + \beta_k A^T A(x^k - \hat{x}^k)$, if

$$\left\| \xi_x^k \right\| \geq \nu r_k \left\| x^k - \hat{x}^k \right\|. \tag{8}$$

Let $r_k \leftarrow 2 * r_k$, and go back to (s1).

(s2) Solving the following variational inequality, find $y \in \mathscr{Y}$ such that

$$\left(y' - y\right)^T \left\{ g\left(y^k\right) + \left[\lambda^k - \beta_k \left(A \hat{x}^k - y\right)\right]\right\} \geq 0, \quad \forall y' \in \mathscr{Y}, \tag{9}$$

And denote the solution by $\hat{y}^k$. Let $\xi_y^k = g(y^k) - g(\hat{y}^k)$, if

$$\left\| \xi_y^k \right\| \geq \frac{\nu \beta_k}{2\sqrt{2}} \left\| y^k - \hat{y}^k \right\|. \tag{10}$$

Let $\beta_k \leftarrow 2 * \beta_k$, and go back to (s2).

(s3) Update $\hat{\lambda}^k$ via

$$\hat{\lambda}^k = \lambda^k - \beta_k \left(A \hat{x}^k - \hat{y}^k\right). \tag{11}$$

(s4) Compute the step length $\alpha_k > 0$. Let

$$M_k = \begin{pmatrix} r_k I & & \\ & \beta_k I & \\ & & \frac{1}{\beta_k} I \end{pmatrix}, \qquad \xi^k = \begin{pmatrix} \xi_x^k \\ \xi_y^k \\ 0 \end{pmatrix}, \tag{12}$$

$$w^{k+1} = w^k - \alpha_k \left(M_k \left(w^k - \hat{w}^k\right) - \xi^k\right). \tag{13}$$

Remark 1. (1) The main difference between the APBPC method [4] and the proposed LSPPAD method is that the proximal term $(s_k/2) \parallel y - y^k \parallel^2$ appears in the prediction step 2 of the APBPC method, while it vanishes in the LAPPAD method.

(2) In the LSPPAD method, (s1) is inherent one iteration of the projection gradient method for the x-subproblem based on the augmented Lagrangian function $L_\beta(x, y, \lambda)$ (see (2)) with the fixed $y = y^k$ and step length $1/r_k$. That is

$$\hat{x}^k = P_{\mathcal{X}} \left\{ x^k - \frac{1}{r_k} \left[f\left(x^k\right) - A^T \left(\lambda^k - \beta_k \left(Ax^k - y^k\right)\right) \right] \right\}. \tag{14}$$

And (s2) is one iteration of the same method for the y-subproblem with the fixed $x = \hat{x}^k$ and step length $1/\beta_k$; that is,

$$\hat{y}^k = P_{\mathcal{Y}} \left\{ y^k - \frac{1}{\beta_k} \left[g\left(y^k\right) + \left(\lambda^k - \beta_k \left(A\hat{x}^k - y^k\right)\right) \right] \right\}. \tag{15}$$

Combining (7) and (9) together with (11) and by a manipulation we have

$$\begin{pmatrix} x' - \hat{x}^k \\ y' - \hat{y}^k \\ \lambda' - \hat{\lambda}^k \end{pmatrix}^T$$

$$\times \begin{pmatrix} f\left(\hat{x}^k\right) - A^T\hat{\lambda}^k + \beta_k A^T\left(y^k - \hat{y}^k\right) & +r_k\left(\hat{x}^k - x^k\right) + \xi_x^k \\ g\left(\hat{y}^k\right) + \hat{\lambda}^k - \beta_k\left(y^k - \hat{y}^k\right) & +\beta_k\left(\hat{y}^k - y^k\right) + \xi_y^k \\ A\hat{x}^k - \hat{y}^k & +\frac{1}{\beta_k}\left(\hat{\lambda}^k - \lambda^k\right) \end{pmatrix} \geq 0. \tag{16}$$

The following notations are useful for convenience in the future discussion:

$$F(w) = \begin{pmatrix} f(x) - A^T\lambda \\ g(y) + \lambda \\ Ax - y \end{pmatrix}, \tag{17}$$

$$\eta(\beta, w, \hat{w}) = \begin{pmatrix} \beta A^T\left(y - \hat{y}\right) \\ -\beta\left(y - \hat{y}\right) \\ 0 \end{pmatrix}.$$

$$d_1\left(w^k, \hat{w}^k, \xi^k\right) = M_k\left(w^k - \hat{w}^k\right) - \xi^k. \tag{18}$$

$$d_2\left(w^k, \hat{w}^k\right) = F\left(\hat{w}^k\right) - \eta\left(\beta_k, w^k, \hat{w}^k\right). \tag{19}$$

$$\varphi\left(w^k, \hat{w}^k, \xi^k\right) = \left(\hat{\lambda}^k - \lambda^k\right)^T\left(y^k - \hat{y}^k\right) + \left(w^k - \hat{w}^k\right)^T d_1\left(w^k, \hat{w}^k, \xi^k\right). \tag{20}$$

By these notations, the variational inequality (16) can be rewritten into a compact form:

$$\hat{w}^k \in \mathcal{W}, \quad \left(w' - \hat{w}^k\right)^T \left[d_2\left(w^k, \hat{w}^k\right) - d_1\left(w^k, \hat{w}^k, \xi^k\right)\right] \geq 0,$$

$$\forall w' \in \mathcal{W}. \tag{21}$$

3. Descent Directions of the Merit Function $(1/2) \parallel w - w^* \parallel^2$

For the convenience of analysis, we ignore the index k of the matrices, vectors, and scalars in the section.

It is easy to show that, for all $w^* \in \mathcal{W}^*$, $(w - w^*)$ is the gradient of the unknown merit function $(1/2) \parallel w - w^* \parallel^2$ at the point w. The vector d is a descent direction of the merit function $(1/2) \parallel w - w^* \parallel^2$ if and only if $(w - w^*)^T d < 0$. In the section, under suitable conditions, we will show that both $-d_1(w, \hat{w}, \xi)$ and $-d_2(w, \hat{w})$ (see (18) and (19), resp.) are the descent directions of $(1/2) \parallel w - w^* \parallel^2$ while $w \in \mathcal{W} \setminus \mathcal{W}^*$.

Hereafter, we denote $u^T H u$ by the compact form $\parallel u \parallel_H^2$ whenever H is a semidefinite symmetric matrix. It is obvious that $\parallel u \parallel_H^2 \geq 0$ by the semidefinite property of matrix H.

Lemma 2. *For a given $w \in \mathcal{W}$, let $\hat{w} = (\hat{x}, \hat{y}, \hat{\lambda})$ be generated by (s1) to (s3) of the LSPPAD method, and let $\varphi(w, \hat{w}, \xi)$ be defined by (20). Then one has*

$$\varphi(w, \hat{w}, \xi) > \frac{\beta}{2} \parallel A\hat{x} - y \parallel^2 + \frac{\tau}{2} \parallel d_1(w, \hat{w}, \xi) \parallel^2, \tag{22}$$

where $\tau = \min\{1/r, 1/2\beta, \beta\} > 0$.

Proof. By a manipulation, we get

$$\varphi(w, \hat{w}, \xi)$$

$$= \left(\hat{\lambda} - \lambda\right)^T\left(y - \hat{y}\right) + \left(w - \hat{w}\right)^T\left[M\left(w - \hat{w}\right) - \xi\right]$$

$$= \left(\hat{\lambda} - \lambda\right)^T\left(y - \hat{y}\right) + \parallel w - \hat{w} \parallel_M^2 - \left(w - \hat{w}\right)^T\xi$$

$$= \left(\hat{\lambda} - \lambda\right)^T\left(y - \hat{y}\right) + r\parallel x - \hat{x} \parallel^2$$

$$\quad + \beta\parallel y - \hat{y} \parallel^2 + \frac{1}{\beta}\parallel \lambda - \hat{\lambda} \parallel^2$$

$$\quad - (x - \hat{x})^T\xi_x - (y - \hat{y})^T\xi_y$$

$$= \left[\left(\hat{\lambda} - \lambda\right)^T\left(y - \hat{y}\right) + \frac{1}{2}\beta\parallel y - \hat{y} \parallel^2 + \frac{1}{2\beta}\parallel \lambda - \hat{\lambda} \parallel^2\right]$$

$$\quad + \frac{1}{2\beta}\parallel \lambda - \hat{\lambda} \parallel^2$$

$$\quad + r\parallel x - \hat{x} \parallel^2 - (x - \hat{x})^T\xi_x + \frac{1}{2}\beta\parallel y - \hat{y} \parallel^2 - (y - \hat{y})^T\xi_y. \tag{23}$$

Substituting $\hat{\lambda} - \lambda = -\beta(A\hat{x} - \hat{y})$ into the first term of the right side of (23), we get

$$\left(\hat{\lambda} - \lambda\right)^T\left(y - \hat{y}\right) + \frac{1}{2}\beta\parallel y - \hat{y} \parallel^2 + \frac{1}{2\beta}\parallel \lambda - \hat{\lambda} \parallel^2$$

$$= \frac{\beta}{2}\left[-2(A\hat{x} - \hat{y})^T\left(y - \hat{y}\right) + \parallel y - \hat{y} \parallel^2 + \parallel A\hat{x} - \hat{y} \parallel^2\right]$$

$$= \frac{\beta}{2}\parallel A\hat{x} - y \parallel^2. \tag{24}$$

By (s1), we have $\| \xi_x^k \| \le \nu r_k \| x^k - \widehat{x}^k \|$. Note that $\nu \in (0, 1)$; we get

$$
\begin{aligned}
r\|x - \widehat{x}\|^2 &- (x - \widehat{x})^T \xi_x \\
&= \frac{1}{2} r\|x - \widehat{x}\|^2 - (x - \widehat{x})^T \xi_x + \frac{1}{2} r\|x - \widehat{x}\|^2 \\
&= \frac{1}{2r}\left[r^2\|x - \widehat{x}\|^2 - 2r(x - \widehat{x})^T \xi_x + r^2\|x - \widehat{x}\|^2 \right] \\
&> \frac{1}{2r}\left[r^2\|x - \widehat{x}\|^2 - 2r(x - \widehat{x})^T \xi_x + \|\xi_x\|^2 \right] \\
&= \frac{1}{2r}\|r(x - \widehat{x}) - \xi_x\|^2 .
\end{aligned} \tag{25}
$$

Similarly, we have

$$
\begin{aligned}
\frac{1}{2}\beta\|y - \widehat{y}\|^2 &- (y - \widehat{y})^T \xi_y \\
&= \frac{1}{4}\beta\|y - \widehat{y}\|^2 - (y - \widehat{y})^T \xi_y + \frac{1}{4}\beta\|y - \widehat{y}\|^2 \\
&= \frac{1}{\beta}\left[\left\|\frac{1}{2}\beta(y - \widehat{y})\right\|^2 - \frac{1}{2}\beta(y - \widehat{y})^T \xi_y + \left\|\frac{1}{2}\xi_y\right\|^2 \right. \\
&\quad \left. + \frac{1}{4}\beta^2\|y - \widehat{y}\|^2 - \left\|\frac{1}{2}\xi_y\right\|^2 - \frac{1}{2}\beta(y - \widehat{y})^T \xi_y \right] \\
&= \frac{1}{4\beta}\|\beta(y - \widehat{y}) - \xi_y\|^2 \\
&\quad + \frac{1}{\beta}\left[\frac{1}{4}\beta^2\|y - \widehat{y}\|^2 - \left\|\frac{1}{2}\xi_y\right\|^2 - \frac{1}{2}\beta(y - \widehat{y})^T \xi_y \right] .
\end{aligned} \tag{26}
$$

By (s2), we have $\| \xi_y^k \| \le (\nu \beta_k / 2\sqrt{2}) \| y^k - \widehat{y}^k \|$. Using Cauchy-Schwarz inequality, we get

$$
\frac{\beta}{2}(y - \widehat{y})^T \xi_y \le \frac{\beta}{2}\|y - \widehat{y}\| \cdot \|\xi_y\| \le \frac{\beta^2}{4\sqrt{2}}\|y - \widehat{y}\|^2 . \tag{27}
$$

Again using $\| \xi_y^k \| \le (\nu \beta_k / 2\sqrt{2}) \| y^k - \widehat{y}^k \|$, we get

$$
\begin{aligned}
\frac{1}{\beta}\left[\frac{1}{4}\beta^2\|y - \widehat{y}\|^2 - \left\|\frac{1}{2}\xi_y\right\|^2 - \frac{1}{2}\beta(y - \widehat{y})^T \xi_y \right] \\
\ge \frac{\beta}{4}\frac{7\sqrt{2} - 8}{8\sqrt{2}}\|y - \widehat{y}\|^2 \ge 0 .
\end{aligned} \tag{28}
$$

Thus

$$
\frac{1}{2}\beta\|y - \widehat{y}\|^2 - (y - \widehat{y})^T \xi_y \ge \frac{1}{4\beta}\|\beta(y - \widehat{y}) - \xi_y\|^2 . \tag{29}
$$

Substituting (24), (25), and (29) into (23), we have

$$
\begin{aligned}
\varphi(w, \widehat{w}, \xi) &> \frac{1}{2}\beta\|A\widehat{x} - y\|^2 + \frac{1}{2r}\|r(x - \widehat{x}) - \xi_x\|^2 \\
&\quad + \frac{1}{4\beta}\|\beta(y - \widehat{y}) - \xi_y\|^2 + \frac{1}{2\beta}\|\lambda - \widehat{\lambda}\|^2 .
\end{aligned} \tag{30}
$$

Recalling

$$
\tau = \min\left\{ \frac{1}{r}, \frac{1}{2\beta}, \beta \right\} , \tag{31}
$$

we obtain

$$
\varphi(w, \widehat{w}, \xi) > \frac{\beta}{2}\|A\widehat{x} - y\|^2 + \frac{\tau}{2}\|d_1(w, \widehat{w}, \xi)\|^2 . \tag{32}
$$

$\square$

Lemma 3. *By the same conditions of Lemma 2 one has*

$$
\begin{aligned}
\varphi(w, \widehat{w}, \xi) \ge\ & (1 - \nu) r\|x - \widehat{x}\|^2 \\
&+ \frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\nu\right)\beta\|y - \widehat{y}\|^2 + \frac{1}{2\beta}\|\lambda - \widehat{\lambda}\|^2 .
\end{aligned} \tag{33}
$$

Proof. By Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
(x - \widehat{x})^T \xi_x &\le \|x - \widehat{x}\| \|\xi_x\| , \\
(y - \widehat{y})^T \xi_y &\le \|y - \widehat{y}\| \|\xi_y\| .
\end{aligned} \tag{34}
$$

Noting that $\beta > 0$, we have

$$
\begin{aligned}
\frac{1}{2}\left(\beta\|y - \widehat{y}\|^2 + \frac{1}{\beta}\|\lambda - \widehat{\lambda}\|^2\right) \\
\ge \sqrt{\frac{1}{\beta}}\|\lambda - \widehat{\lambda}\| \sqrt{\beta}\|y - \widehat{y}\| \ge (\lambda - \widehat{\lambda})^T (y - \widehat{y}) .
\end{aligned} \tag{35}
$$

Then by (s1) and (s2), we get

$$
\begin{aligned}
\varphi(w&, \widehat{w}, \lambda) \\
&= (\widehat{\lambda} - \lambda)^T (y - \widehat{y}) + (w - \widehat{w})^T [M(w - \widehat{w}) - \xi] \\
&= (\widehat{\lambda} - \lambda)^T (y - \widehat{y}) + (x - \widehat{x})^T r(x - \widehat{x}) - (x - \widehat{x})^T \xi_x \\
&\quad + (y - \widehat{y})^T \beta(y - \widehat{y}) - (y - \widehat{y})^T \xi_y + (\lambda - \widehat{\lambda})^T \frac{1}{\beta}(\lambda - \widehat{\lambda}) \\
&\ge (1 - \nu) r\|x - \widehat{x}\|^2 + \left(1 - \frac{\nu}{2\sqrt{2}}\right)\beta\|y - \widehat{y}\|^2 \\
&\quad + \frac{1}{\beta}\|\lambda - \widehat{\lambda}\|^2 - (\lambda - \widehat{\lambda})^T (y - \widehat{y}) \\
&\ge (1 - \nu) r\|x - \widehat{x}\|^2 + \frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\nu\right)\beta\|y - \widehat{y}\|^2 \\
&\quad + \frac{1}{2\beta}\|\lambda - \widehat{\lambda}\|^2 .
\end{aligned} \tag{36}
$$

$\square$

Lemma 4. *For a given $w = (x, y, \lambda) \in \mathcal{W}$, let $\widehat{w} = (\widehat{x}, \widehat{y}, \widehat{\lambda})$ be generated by (s1) to (s3) of the LSPPAD method. Let $\varphi(w, \widehat{w}, \xi)$ be defined by (20), and let $d_1(w, \widehat{w}, \xi)$ and $d_2(w, \widehat{w})$ be defined by (18) and (19), respectively. One has*

$$
(\widehat{w} - w^*)^T d_2(w, \widehat{w}) \ge \varphi(w, \widehat{w}, \xi) - (w - \widehat{w})^T d_1(w, \widehat{w}, \xi) . \tag{37}
$$

Proof. Since $f(x)$ (resp., $g(x)$) is monotone with respect to $\mathcal{X}$ (resp., $\mathcal{Y}$), it follows that $F(w)$ (see (17)) is a monotone operator with respect to $\mathcal{W}$, we have

$$(\widehat{w} - w^*)^T [F(\widehat{w}) - F(w^*)] \geq 0, \qquad (38)$$

and, consequently,

$$(\widehat{w} - w^*)^T F(\widehat{w}) \geq (\widehat{w} - w^*)^T F(w^*). \qquad (39)$$

Note that w^* is a solution of (3), which implies $(\widehat{w} - w^*)^T F(w^*) \geq 0$ for $\widehat{w} \in \mathcal{W}$. Thus $(\widehat{w} - w^*)^T F(\widehat{w}) \geq 0$. We have

$$\begin{aligned}
(\widehat{w} &- w^*)^T d_2(w, \widehat{w}) \\
&= (\widehat{w} - w^*)^T F(\widehat{w}) - (\widehat{w} - w^*)^T \eta(\beta_k, w^k, \widehat{w}^k) \\
&\geq -(\widehat{w} - w^*)^T \eta(\beta_k, w^k, \widehat{w}^k) \\
&= -\beta(A\widehat{x} - Ax^*)^T (y - \widehat{y}) + (\widehat{y} - y^*)\beta(y - \widehat{y}) \\
&= (y - \widehat{y})^T \beta(\widehat{y} - A\widehat{x}) \\
&= (y - \widehat{y})^T (\widehat{\lambda} - \lambda).
\end{aligned} \qquad (40)$$

Recalling the definition of $\varphi(w, \widehat{w}, \xi)$, the inequality (37) follows from (40) directly. □

Theorem 5. *For a given $w = (x, y, \lambda) \in \mathcal{W}$, let $\widehat{w} = (\widehat{x}, \widehat{y}, \widehat{\lambda})$ be generated by (s1) to (s3) of the LSPPAD method. Then one has*

$$(w - w^*)^T d_1(w, \widehat{w}, \xi) \geq \varphi(w, \widehat{w}, \xi), \quad \forall w^* \in \mathcal{W}^*, \quad (41)$$

$$(w - w^*)^T d_2(w, \widehat{w}) \geq \varphi(w, \widehat{w}, \xi), \quad \forall w^* \in \mathcal{W}^*. \quad (42)$$

Proof. By (21), we get (by letting $w' = w^*$)

$$(\widehat{w} - w^*)^T d_1(w, \widehat{w}, \xi) \geq (\widehat{w} - w^*)^T d_2(w, \widehat{w}), \quad \forall \widehat{w} \in \mathcal{W}. \qquad (43)$$

Combining (43) and (37), we have

$$(\widehat{w} - w^*)^T d_1(w, \widehat{w}, \xi) \geq \varphi(w, \widehat{w}, \xi) - (w - \widehat{w}) d_1(w, \widehat{w}, \xi). \qquad (44)$$

Thus (41) follows from (44). Adding (37) and (21), we get

$$\begin{aligned}
(w' - w^*)^T d_2(w, \widehat{w}) &\geq \varphi(w, \widehat{w}, \xi) \\
&\quad - (w - w')^T d_1(w, \widehat{w}, \xi), \qquad (45) \\
&\forall w' \in \mathcal{W}, \quad \forall w^* \in \mathcal{W}^*.
\end{aligned}$$

By letting $w' = w$, we get (42) from (45) directly. □

4. Convergence Analysis

For establishing the convergence, we first prove the following contractility property.

Lemma 6. *For a given $w^k \in \mathcal{W}$, let w^{k+1} be generated by the LSPPAD method. Then one has*

$$\begin{aligned}
\|w^k - w^*\|^2 - \|w^{k+1} - w^*\|^2 &\geq 2\alpha_k \varphi(w^k, \widetilde{w}^k, \xi^k) \\
&\quad - \alpha_k^2 \|d_1(w^k, \widetilde{w}^k, \xi^k)\|^2.
\end{aligned} \qquad (46)$$

Proof. Recall the definition of $d_1(w^k, \widehat{w}^k, \xi^k)$ and the iteration formula (13); by a straightforward computation and Theorem 5, we get

$$\begin{aligned}
\|w^k &- w^*\|^2 - \|w^{k+1} - w^*\|^2 \\
&= \|w^k - w^*\|^2 - \|w^k - \alpha_k d_1(w^k, \widehat{w}^k, \xi^k) - w^*\|^2 \\
&= 2\alpha_k (w^k - w^*)^T d_1(w^k, \widehat{w}^k, \xi^k) \\
&\quad - \alpha_k^2 \|d_1(w^k, \widehat{w}^k, \xi^k)\|^2 \\
&\geq 2\alpha_k \varphi(w^k, \widetilde{w}^k, \xi^k) - \alpha_k^2 \|d_1(w^k, \widetilde{w}^k, \xi^k)\|^2.
\end{aligned} \qquad (47)$$

□

By Lemma 6 we have

$$\begin{aligned}
\|w^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 + \alpha_k^2 \|d_1(w^k, \widetilde{w}^k, \xi^k)\|^2 \\
&\quad - 2\alpha_k \varphi(w^k, \widetilde{w}^k, \xi^k).
\end{aligned} \qquad (48)$$

Thus, we can get the maximal drop-out value at each iteration by maximizing the following function:

$$\Delta_k(\alpha) = -\alpha^2 \|d_1(w^k, \widetilde{w}^k, \xi^k)\|^2 + 2\alpha\varphi(w^k, \widetilde{w}^k, \xi^k). \quad (49)$$

This quadratic function $\Delta_k(\alpha)$ reaches its maximum at the point

$$\alpha_k^* = \frac{\varphi(w^k, \widetilde{w}^k, \xi^k)}{\|d_1(w^k, \widetilde{w}^k, \xi^k)\|^2}, \qquad (50)$$

with its value

$$\Delta_k(\alpha_k^*) = \alpha_k^* \varphi(w^k, \widetilde{w}^k, \xi^k). \qquad (51)$$

Thus, by Lemma 2 we have the following.

Corollary 7. *If the constant sequences $\{r_k\}$ and $\{\beta_k\}$ are positive and bounded above for all iteration k, then one has*

$$\alpha_k^* > \frac{\tau}{2}. \qquad (52)$$

We will show later that the conditions of Corollary 7 can be satisfied in practice.

The numerical experiment [5] suggests that for fast convergence one can use a relaxation factor $\gamma \in [1, 2)$ and let $\alpha_k = \gamma \alpha_k^*$ at each iteration. To do so, we get

$$\Delta_k \left(\gamma \alpha_k^* \right) = \gamma \left(2 - \gamma \right) \Delta_k \left(\alpha_k^* \right). \tag{53}$$

By (46) or (49), we have

$$\left\| w^{k+1} - w^* \right\|^2 \le \left\| w^k - w^* \right\|^2 - \gamma \left(2 - \gamma \right) \alpha_k^* \varphi \left(w^k, \widetilde{w}^k, \xi^k \right). \tag{54}$$

Furthermore, by Corollary 7, we have

$$\left\| w^{k+1} - w^* \right\|^2 \le \left\| w^k - w^* \right\|^2 - \frac{\tau}{2} \gamma \left(2 - \gamma \right) \varphi \left(w^k, \widetilde{w}^k, \xi^k \right) \tag{55}$$

with

$$\tau = \min_k \left\{ \frac{1}{r_k}, \frac{1}{2\beta_k}, \beta_k \right\}. \tag{56}$$

The inequality (55) plays an important role in proving convergence of the proposed method.

Letting

$$N = \begin{pmatrix} 2 \left(1 - \nu \right) rI & 0 & 0 \\ 0 & \left(1 - \frac{\sqrt{2}}{2} \nu \right) \beta I & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}, \tag{57}$$

by Lemma 3 we have

$$\varphi \left(w, \widehat{w}, \xi \right) \ge \frac{1}{2} \| w - \widehat{w} \|_N^2. \tag{58}$$

Combing (55) and (58), we get the Fejér monotonicity of the generated sequence $\{ w^k \}$.

Theorem 8. *Suppose the sequence $\{ w^k \}$ be generated by the LSPPAD method and the conditions of Corollary 7 hold. Then the sequence $\{ w^k \}$ converges to w^∞, which is a solution of the problem* (3), *or, equivalently, the problem* (1).

Proof. By (55) and (58), we get

$$\left\| w^{k+1} - w^* \right\|^2 \le \left\| w^k - w^* \right\|^2 - \frac{\tau \gamma \left(2 - \gamma \right)}{4} \left\| w^k - \widehat{w}^k \right\|_N^2. \tag{59}$$

Consequently,

$$\left\| w^{k+1} - w^* \right\|^2 \le \left\| w^0 - w^* \right\|^2, \quad \forall k. \tag{60}$$

It follows that the sequence $\{ w^k \}$ is bounded. Moreover, we get

$$\frac{\tau \gamma \left(2 - \gamma \right)}{4} \sum_{k=0}^{\infty} \left\| w^k - \widehat{w}^k \right\|_N^2 \le \left\| w^0 - w^* \right\|^2, \tag{61}$$

from which it follows that

$$\lim_{k \to \infty} \left\| w^k - \widehat{w}^k \right\|_N^2 = 0. \tag{62}$$

Thus $\{ \widehat{w}^k \}$ is bounded and it has at least one cluster point. Let w^∞ be a cluster point of $\{ \widehat{w}^k \}$, and suppose the subsequence $\{ \widehat{w}^{k_j} \}$ converges to w^∞. It follows from (62) that

$$\lim_{k \to \infty} \left\| x^k - \widehat{x}^k \right\| = 0,$$
$$\lim_{k \to \infty} \left\| y^k - \widehat{y}^k \right\| = 0, \qquad \lim_{k \to \infty} \left\| \lambda^k - \widehat{\lambda}^k \right\| = 0, \tag{63}$$

and, consequently (by (s1) and (s2)),

$$\lim_{k \to \infty} \left\| \xi_x^k \right\| = 0, \qquad \lim_{k \to \infty} \left\| \xi_y^k \right\| = 0. \tag{64}$$

Then, by (7), (9), and (11), we get

$$\widehat{w}^k \in \mathscr{W}, \begin{cases} \lim_{k \to \infty} \left(x' - \widehat{x}^k \right)^T \left\{ f \left(\widehat{x}^k \right) - A^T \widehat{\lambda}^k \right\} \ge 0, \\ \lim_{k \to \infty} \left(y' - \widehat{y}^k \right)^T \left\{ g \left(\widehat{y}^k \right) + \widehat{\lambda}^k \right\} \ge 0, \\ \lim_{k \to \infty} \left(A \widehat{x}^k - \widehat{y}^k \right) = 0, \end{cases} \tag{65}$$
$$\forall w' \in \mathscr{W},$$

$$\widehat{w}^{k_j} \in \mathscr{W}, \begin{cases} \lim_{k_j \to \infty} \left(x' - \widehat{x}^{k_j} \right)^T \left\{ f \left(\widehat{x}^{k_j} \right) - A^T \widehat{\lambda}^{k_j} \right\} \ge 0, \\ \lim_{k_j \to \infty} \left(y' - \widehat{y}^{k_j} \right)^T \left\{ g \left(\widehat{y}^{k_j} \right) + \widehat{\lambda}^{k_j} \right\} \ge 0, \\ \lim_{k_j \to \infty} \left(A \widehat{x}^{k_j} - \widehat{y}^{k_j} \right) = 0, \end{cases}$$
$$\forall w' \in \mathscr{W},$$
$$\tag{66}$$

which implies that

$$\widehat{w}^\infty \in \mathscr{W}, \begin{cases} \left(x' - \widehat{x}^\infty \right)^T \left\{ f \left(\widehat{x}^\infty \right) - A^T \widehat{\lambda}^\infty \right\} \ge 0, \\ \left(y' - \widehat{y}^\infty \right)^T \left\{ g \left(\widehat{y}^\infty \right) + \widehat{\lambda}^\infty \right\} \ge 0, \\ \left(A \widehat{x}^\infty - \widehat{y}^\infty \right) = 0, \end{cases} \tag{67}$$
$$\forall w' \in \mathscr{W}.$$

Thus w^∞ solves the variational inequalities (3).

Since $\{ \widehat{w}^{k_j} \} \to w^\infty$, by (62), for any $\varepsilon > 0$, there exists an integer $l > 0$ such that

$$\left\| w^{k_l} - \widehat{w}^{k_l} \right\|_N < \frac{\varepsilon}{2}, \qquad \left\| \widehat{w}^{k_l} - w^\infty \right\|_N < \frac{\varepsilon}{2}. \tag{68}$$

Therefore, for any $k \ge k_l$, we have

$$\left\| w^k - w^\infty \right\|_N \le \left\| w^{k_l} - w^\infty \right\|_N$$
$$\le \left\| w^{k_l} - \widehat{w}^{k_l} \right\|_N + \left\| \widehat{w}^{k_l} - w^\infty \right\|_N < \varepsilon, \tag{69}$$

which implies that the sequence $\{ w^k \}$ converges to w^∞, which is a solution of the problem (3) or the problem (1). □

5. Practical Implementation and Numerical Results

For the implementation of the proposed method, we have to give some rules to determine the proximal parameter r_k and penalty parameter β_k.

It follows from (56) that if the sequences $\{r_k\}, \{\beta_k\}$ are positive and bounded, then τ is well defined and $\tau > 0$, which guarantees that the conditions of Corollary 7 hold.

On the penalty parameter sequence $\{\beta_k\}$, He et al. [9] proposed a self-adaptive rule at each iteration based on the iterate information, which is referred to as self-adaptive penalty parameters method, and showed that the sequence $\{\beta_k\}$ generated by the method is bounded and away from zero. In the proposed method of this paper, we use directly the self-adaptive penalty parameters method (Method 3, Strategy S3 in [9]). The self-adaptive rule is

$$
\begin{cases}
2 * \beta_k, & \text{if} \quad \left\|y_{k+1} - y_k\right\|_\infty < 0.25 \left\|\lambda_{k+1} - \lambda_k\right\|_\infty, \\
0.5 * \beta_k, & \text{if} \quad 0.25 \left\|y_{k+1} - y_k\right\|_\infty > \left\|\lambda_{k+1} - \lambda_k\right\|_\infty, \\
\beta_k, & \text{otherwise.}
\end{cases} \tag{70}
$$

Combing (70) and the line search in (s2) of the LSPPAD method, we have that the sequence $\{\beta_k\}$ is positive, away from zero and bounded above.

We next focus on the proximal parameter sequence $\{r_k\}$.

Lemma 9. *If f is monotone and Lipschitz continuous with constant L_f, that is,*

$$
\left\|f(x) - f(\widehat{x})\right\| \le L_f \left\|x - \widehat{x}\right\|, \quad \forall x, \widehat{x} \in \mathcal{X}, \tag{71}
$$

then, whenever

$$
r_k \ge \frac{L_f + \beta_k \left\|A^T A\right\|}{\nu} \tag{72}
$$

one has $\| \xi_x \| \le \nu r_k \| x - x^k \|$.

Proof. Recall the definition of ξ_x^k in (s1) of the LSPPAD method; by (71) we get

$$
\begin{aligned}
\left\|\xi_x^k\right\| &= \left\|f\left(x^k\right) - f(x) + \beta_k A^T A \left(x^k - x\right)\right\| \\
&\le \left\|f\left(x^k\right) - f(x)\right\| + \left\|\beta_k A^T A \left(x^k - x\right)\right\| \\
&\le L_f \left\|x^k - x\right\| + \beta_k \left\|A^T A\right\| \left\|x^k - x\right\| \\
&\le \nu r_k \left\|x - x^k\right\|.
\end{aligned} \tag{73}
$$

$\square$

Lemma 9 indicates that the line search in (s1) will terminate in finite number of iterations.

In practice we use a self-adaptive rule to guarantee r_k not tend to infinity under the line search in (s1) of the LSPPAD method. The self-adaptive rule is letting $r_k = r_k * \kappa * \rho$ whenever $\kappa = \| \xi_x^k \| / r_k \| x^k - \widehat{x}^k \| < 0.5$, where $\rho = 1.85$.

By this rule and the step (s1) of the LAPPAD method, $\{r_k\}$ is obviously bounded away from zero; thus the condition on $\{r_k\}$ of Corollary 7 holds.

Finally, we give some numerical experiments on compressed sensor problem; see [10]. These experiments are tested on a laptop with 166 GHz CPU, 2.5 GB RAM, and Matlab 6.5.

The separable convex optimization formulation of compressed sensor problem is

$$
\begin{aligned}
\min \quad & \frac{1}{2}\|Dx - b\|_2^2 + \mu\|y\|_1, \\
\text{s.t.} \quad & x - y = 0,
\end{aligned} \tag{74}
$$

where $x, y \in R^n, D \in R^{m \times n}, b \in R^m$, and $\mu > 0$. In the test problem, $\theta_2(y) = \mu \| y \|_1$ is nonsmooth. However, the special structure of the problem (74) admits the equivalence relation between the problems (1) and (3) if the subderivative of the absolute value function $|u|$ is specialized as follows:

$$
\partial(|u|) \ni g(u) = \begin{cases} 1, & \text{if } u > 0, \\ 0, & \text{if } u = 0, \\ -1, & \text{if } u < 0. \end{cases} \tag{75}
$$

So the proposed method is available for this problem.

The test data of the problem (74) is generated by the following style: $D \in R^{m \times n}$ is a random matrix with $D_{ij} \in (1, 2)$ according to uniform distribution, and $b \in R^m$ is a given signal vector with a random noise, $\mu = 0.2$. The practical parameters of the LSPPAD method are given in the following: $\nu = 0.95, r_0 = 0.51 * \max(\sigma(AA^T))$ (where $\sigma(AA^T)$ denotes the eigenvalue of AA^T), $\beta_0 = 1.5, x^0 = 0_{n\times1}, y^0 = 1_{n\times1}, \lambda^0 = 1_{n\times1}$, and $\gamma = 1.5$. The experimental results are stated in Table 1. The notation in Table 1 is as follows: m, n, and cputime are clear; sp denotes the required sparse degree; function evaluations is the number of computation of product on matrix and vector (which is the main cost of the proposed method). The process stops whenever $\max\{\| x^k - \widehat{x}^k \|_\infty, \| y^k - \widehat{y}^k \|_\infty, \| \lambda^k - \widehat{\lambda}^k \|_\infty\} < 1.0 \times 10^{-3}$.

The numerical results show that the proposed LSPPAD method has better performance compared with the APBPC method for the test problem.

6. Conclusions

In this paper, we present a line-search-based partial proximal alternating directions (LSPPAD) method for solving a class of structured convex optimization problems. This kind of problems is common in practice. The LSPPAD method makes full use of the special structure of the described problem. In the LSPPAD method, two subproblems are solved in an alternative fashion. The subproblem of y variable is solved without the proximal term, while the x-subproblem is solved by a proximal point method; both used line-search technique to guarantee the convergence. To do so provides some advantages for the tractability. The convergence property of the LSPPAD method is established. Numerical tests show that, due to the described problem, the LSPPAD method

TABLE 1: Numerical results on compressed sensor problems.

m	n	sp	Iteration number		Function evaluations		Cputime (in second)	
			LSPPAD	APBPC	LSPPAD	APBPC	LSPPAD	APBPC
1024	4096	160/4096	95	152	291	496	3.92	7.24
1600	8192	320/8192	104	212	318	632	13.0	24.4
2048	12000	400/12000	104	238	318	756	24.8	56.8
3052	16438	512/16438	106	252	324	812	49.1	120.6

has better performance compared with the APBPC method which solves both subproblems by the inexact proximal point method.

Conflict of Interests

All the authors of the paper declare that they do not have any conflict of interests and there are no financial or personal relationships with other people or organizations that can inappropriately influence their work in this paper.

Acknowledgments

References

[1] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming*, vol. 64, no. 1, Ser. A, pp. 81–101, 1994.

[2] P. Tseng, "Alternating projection-proximal methods for convex programming and variational inequalities," *SIAM Journal on Optimization*, vol. 7, no. 4, pp. 951–965, 1997.

[3] B. He, L.-Z. Liao, D. Han, and H. Yang, "A new inexact alternating directions method for monotone variational inequalities," *Mathematical Programming*, vol. 92, no. 1, pp. 103–118, 2002.

[4] B.-S. He, L.-Z. Liao, and M.-J. Qian, "Alternating projection based prediction-correction methods for structured variational inequalities," *Journal of Computational Mathematics*, vol. 24, no. 6, pp. 693–710, 2006.

[5] B.-S. He, S.-l. Wang, and H. Yang, "A modified variable-penalty alternating directions method for monotone variational inequalities," *Journal of Computational Mathematics*, vol. 21, no. 4, pp. 495–504, 2003.

[6] B. He and J. Zhou, "A modified alternating direction method for convex minimization problems," *Applied Mathematics Letters*, vol. 13, no. 2, pp. 123–130, 2000.

[7] L. A. Parente, P. A. Lotito, and M. V. Solodov, "A class of inexact variable metric proximal point algorithms," *SIAM Journal on Optimization*, vol. 19, no. 1, pp. 240–260, 2008.

[8] P. Smoczynski and M. A. Tawhid, "Two numerical schemes for general variational inequalities," *Journal of Industrial and Management Optimization*, vol. 4, no. 2, pp. 393–406, 2008.

[9] B. S. He, H. Yang, and S. L. Wang, "Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities," *Journal of Optimization Theory and Applications*, vol. 106, no. 2, pp. 337–356, 2000.

[10] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for l_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.