RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2018.44.034

ScienceAsia 44 (2018): **34–39**

A prediction-correction primal-dual hybrid gradient method for convex programming with linear constraints

Pibin Bing[a], **Jialei Sui**[a], **Min Sun**[b,c,*]

[a] Institute of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou, 450045, China
[b] School of Mathematics and Statistics, Zaozhuang University, Shandong, 277160, China
[c] School of Management, Qufu Normal University, Shandong, 276826, China

*Corresponding author, e-mail: ziyouxiaodou@163.com

Received 10 Jul 2017
Accepted 25 Nov 2017

ABSTRACT: In recent years, the primal-dual hybrid gradient (PDHG) method has been widely used. However, the original PDHG method may diverge without additional conditions. Here we propose a convergent prediction-correction PDHG (PD-PDHG) method for canonical convex programming with linear constraints. The most important characteristic of the PD-PDHG method is that it adopts a new descent direction in the correction step, which does not converge to zero in general. Convergence of the new method is proved under mild assumptions. Finally, its efficiency is verified by compressive sensing.

KEYWORDS: iteration method, global convergence, compressive sensing

MSC2010: 90C25 90C30

INTRODUCTION

Let $\theta(x):\mathbb{R}^n\to\mathbb{R}$ be a convex (not necessarily smooth) function, $A\in\mathbb{R}^{m\times n}$, and $b\in\mathbb{R}^m$. We consider the following canonical convex programming problem with linear constraints:

$$\min\{\theta(x)\,|\,Ax=b \text{ (or } \geqslant b),x\in\mathscr{X}\}, \tag{1}$$

where $\mathscr{X}\subseteq\mathbb{R}^n$ is a closed convex set. Throughout, we assume that the solution set of (1) is nonempty, and the set $\mathscr{X}$ is simple in the sense that the orthogonal projection onto it has a closed-form expression or is simple enough to compute numerically (e.g., the nonnegative orthant, spheroidal or box areas). Problem (1) is the mathematical expression of many problems arising from signal processing, machine learning, wireless networking, and so on[1,2]. For example, it includes the following basis pursuit model of compressive sensing as a special case:

$$\min_x \theta(x)=\|x\|_1 \text{ s.t. } Ax=b, \tag{2}$$

where $A\in\mathbb{R}^{m\times n}(m\ll n)$ is the sensing matrix, $b\in\mathbb{R}^m$ is the observed signal, and the ℓ_1-norm of the vector x is defined as $\|x\|_1=\sum_{i=1}^n|x_i|$.

If $Ax=b$ in (1), we set $\Lambda=\mathbb{R}^m$; otherwise, we set $\Lambda=\mathbb{R}^m_+$. By introducing the Lagrangian multiplier $\lambda\in\Lambda$ to the linear constraints of (1) we obtain its Lagrangian function,

$$\mathscr{L}(x,\lambda)=\theta(x)-\langle\lambda,Ax-b\rangle.$$

To solve (1), two benchmark solvers are the augmented Lagrangian multiplier (ALM) method[3] and the primal-dual hybrid gradient (PDHG) method[4]. The x-related subproblem in the ALM method is often difficult to solve exactly, and some inner iteration is usually needed to obtain an inexact solution[5]. However, the x-related subproblem in the PDHG method has a closed-form solution if the convex set $\mathscr{X}$ in (1) equals $\mathbb{R}^n$, and the proximal operator of the objective function $\theta(x)$ of (1), defined by

$$\left(I_n+\frac{1}{r}\partial\theta\right)^{-1}(a):=\arg\min\left\{\theta(x)+\frac{r}{2}\|x-a\|^2\right\},$$

has a closed-form solution (e.g., $\theta(x)=\|x\|_1$ or $\|x\|$ [5–8]). This is exactly the case in many contemporary applications. More specifically, set $w=(x,\lambda)$, and for given $w^k=(x^k,\lambda^k)$, the iterative scheme of the PDHG method reads as

$$\begin{aligned} x^{k+1}&=\arg\min_{x\in\mathscr{X}}\left\{\mathscr{L}(x,\lambda^k)+\frac{r}{2}\|x-x^k\|^2\right\},\\ \lambda^{k+1}&=\arg\max_{\lambda\in\Lambda}\left\{\mathscr{L}(x^{k+1},\lambda)-\frac{s}{2}\|\lambda-\lambda^k\|^2\right\},\end{aligned} \tag{3}$$

where the two parameters $r, s > 0$. The numerical results in Ref. 4 indicate that the PDHG method is an effective approach for total variation image restoration. However, He[9] demonstrated via a counterexample that the PDHG method could diverge without additional conditions. He[10] proposed a prediction-correction version of the PDHG method, whose prediction is just the iterative scheme (3) except that it sets the output $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ of (3) as $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, and in the correction the new iterative w^{k+1} is generated by

$$w^{k+1} = w^k - \gamma \alpha_k M(w^k - \tilde{w}^k), \tag{4}$$

where $\gamma \in (0, 2)$, α_k is a judiciously chosen step length, and M is a predefined matrix. Although this method is convergent, the descent direction $-M(w^k - \tilde{w}^k)$ in its correction step converges to zero when the sequence $\{w^k\}$ tends to the solution w^*, which often slows down its convergence speed (see the numerical results of this work).

In this work, to deal with the above issue, we propose a new prediction-correction PDHG (PC-PDGH) method for (1). Our motivation can be summarized as follows. The monotone variational inequality problem denoted by VIP($F, \mathscr{U}$) is to find $u^* \in \mathscr{U}$ such that

$$(u - u^*)^{\mathrm{T}} F(u^*) \geqslant 0, \quad \forall u \in \mathscr{U}, \tag{5}$$

where $\mathscr{U} \subseteq \mathbb{R}^n$ is a nonempty closed convex set, and $F(u) : \mathbb{R}^n \to \mathbb{R}^n$ is a monotone mapping. He[9–11] has proposed a geminate ascent direction

$$d_1(u^k, \tilde{u}^k) = \begin{cases} (u^k - \tilde{u}^k), & F(u) \text{ is the gradient} \\ & \text{of a certain function,} \\ (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)], & \\ & \text{otherwise,} \end{cases}$$

$$d_2(\tilde{u}^k) = \beta_k F(\tilde{u}^k),$$

where $\beta_k > 0$ and $\tilde{u}^k = P_{\mathscr{U}}[u^k - \beta_k F(u^k)]$. Many numerical results in Ref. 11 indicate that $d_2(\tilde{u}^k)$ often performs better than $d_1(u^k, \tilde{u}^k)$. The reasons maybe: (i) $\{d_1(u^k, \tilde{u}^k)\}$ eventually converges to zero, while $\{d_2(\tilde{u}^k)\}$ converges to $d_2(u^*)$, which is not equal to zero in general (and numerical results given in Ref. 12 indicate that such directions usually perform worse than the directions that do not converge to zero); (2) the tight lower bound of the expression $\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2$ generated by $d_2(\tilde{u}^k)$ is usually larger than that of $d_1(u^k, \tilde{u}^k)$ in general[10].

In Ref. 10, He developed an open problem: how does one extend the above geminate ascent directions to convex optimization? Clearly, from (4), the method of Ref. 10 adopts a descent direction similar to $-d_1(u^k, \tilde{u}^k)$. Hence to answer this open problem, we only need to see if we can design a numerical method for (1) such that $d_2(\tilde{u}^k)$ is an ascent direction at the current iterate u^k. In this paper, we shall answer the above open problem. In fact, we design a PC-PDHG method for problem (1) which adopts $-d_2(\tilde{u}^k)$ with $d_2(\tilde{u}^k) \in \partial F(\tilde{u}^k)$ as the descent direction in the correction step.

PRELIMINARIES

In this section, we review some basic concepts, and give the relevant properties for further analysis. Problem (1) is then characterized as a structured variational inequality problem.

If $G \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, we denote by $\|x\|_G = \sqrt{x^{\mathrm{T}} G x}$ the G-norm of the vector x. Let $F : \mathbb{R}^n \Longrightarrow \mathbb{R}^n$ be a multifunction. Then F is said to be monotone on $\mathbb{R}^n$ if

$$(x - y)^{\mathrm{T}}(\xi - \zeta) \geqslant 0, \quad \forall \xi \in F(x), \quad \forall \zeta \in F(y).$$

Let $\partial f(x)$ denote the subdifferential of a nonsmooth function $f(\cdot)$ at x which is multifunction and is defined by

$$\partial f(x) = \{\xi \mid f(y) - f(x) \geqslant \langle \xi, y - x \rangle, \quad \forall y \in \mathbb{R}^n\}.$$

It is well known that $\partial f(x)$ is a monotone multifunction for any convex function $f(x)$[13], that is,

$$(x - y)^{\mathrm{T}}(\xi - \zeta) \geqslant 0, \quad \forall \xi \in \partial f(x), \quad \forall \zeta \in \partial f(y). \tag{6}$$

Let $\mathscr{C} \subset \mathbb{R}^n$. Then we use $N_{\mathscr{C}}(x)$ to denote the normal cone of $\mathscr{C}$ at $z \in \mathscr{C}$, which is defined by

$$N_{\mathscr{C}}(x) = \{v \in \mathbb{R}^n : v^{\mathrm{T}}(y - x) \leqslant 0, \quad \forall y \in \mathscr{C}\}.$$

Set $\mathscr{W} = \mathscr{X} \times \Lambda$. Let $P_{\mathscr{W}}(\cdot)$ denote the orthogonal projection mapping from $\mathbb{R}^{n+m}$ onto $\mathscr{W}$, i.e.,

$$P_{\mathscr{W}}(v) = \arg\min\{|v - w| \mid w \in \mathscr{W}\}.$$

The orthogonal projection mapping $P_{\mathscr{W}}(\cdot)$ has the following nice property:

$$\|P_{\mathscr{W}}(v) - w\|^2 \leqslant \|v - w\|^2 - \|v - P_{\mathscr{W}}(v)\|^2, \quad \forall w \in \mathscr{W}. \tag{7}$$

Let $\mathscr{H} = \{x : Ax = b\}$. Throughout this paper, we make the following assumption.

(A_1) The pair $\{\mathscr{H}, \mathscr{X}\}$ has the strong conical hull intersection property. That is, for any $z \in \mathscr{H} \cap \mathscr{X}$, $N_{\mathscr{H} \cap \mathscr{X}} = N_{\mathscr{H}}(z) + N_{\mathscr{X}}(z)$[14,15].

Now let us characterize problem (1) as a structured variational inequality problem. Set

$$F(w)=\begin{pmatrix}\partial\theta(x)-A^{\mathrm{T}}\lambda\\ Ax-b\end{pmatrix}. \tag{8}$$

Then under Assumption A_1, the vector x^* is a solution of problem (1) iff there exists $\lambda^*\in\mathbb{R}^m$ such that $w^*=(x^*,\lambda^*)$ is a solution of the structured variational inequality problem VIP$(F,\mathscr{W})$. That is, there exists $\xi^*\in\partial\theta(x^*)$ such that[15]

$$\begin{pmatrix}x-x^*\\ \lambda-\lambda^*\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\xi^*-A^{\mathrm{T}}\lambda^*\\ Ax^*-b\end{pmatrix}\geqslant 0,\quad \forall\begin{pmatrix}x\\ \lambda\end{pmatrix}\in\mathscr{W}. \tag{9}$$

The solution set of VIP$(F,\mathscr{W})$ is denoted by $\mathscr{W}^*$, which is nonempty since the solution set of problem (1) is nonempty. From the convexity of $\theta(x)$, the multifunction $F(w)$ is clearly monotone in the sense of (6).

THE PC-PDHG METHOD AND ITS GLOBAL CONVERGENCE

In this section, we shall present a convergent PDHG method which we will call the PC-PDHG method for solving problem (1) and prove its global convergence under Assumption A_1. At each iteration, the PC-PDHG method is composed of two steps: the prediction step and the correction step. In the prediction step, the PC-PDHG method firstly generates a trial iterate $\tilde{w}^k$ via the iterative scheme (10). In the contraction step, the new iterate w^{k+1} is generated along the descent direction

$$d_2(\tilde{w}^k)=\begin{pmatrix}\tilde{\xi}^k-A^{\mathrm{T}}\tilde{\lambda}^k\\ A\tilde{x}^k-b\end{pmatrix} \tag{10}$$

with $\tilde{\xi}^k\in\partial\theta(\tilde{x}^k)$. Then $d_2(\tilde{w}^k)\in\partial F(\tilde{w}^k)$. The iterative scheme of the PC-PDHG method for solving problem (1) is as follows.

Algorithm 1 (PC-PDHG method)

Step 1: Initialization. Select four constants: $r>0$, $s>0$ with $rs>\|A^{\mathrm{T}}A\|/4$, and $\gamma\in(0,2)$, $\varepsilon>0$. Choose an initial point $w^0=(x^0,\lambda^0)\in\mathscr{W}$ arbitrarily, and set $k=0$.

Step 2: Prediction step. Compute the trial iterate $\tilde{w}^k=(\tilde{x}^k,\tilde{\lambda}^k)$ via

$$\begin{aligned}\tilde{x}^k&=\arg\min_{x\in\mathscr{X}}\left\{\mathscr{L}(x,\lambda^k)+\frac{r}{2}\|x-x^k\|^2\right\},\\ \tilde{\lambda}^k&=\arg\max_{\lambda\in\Lambda}\left\{\mathscr{L}(\tilde{x}^k,\lambda)-\frac{s}{2}\|\lambda-\lambda^k\|^2\right\}.\end{aligned} \tag{11}$$

Step 3: Stopping condition. If $\|w^k-\tilde{w}^k\|\leqslant\varepsilon$ then stop.

Step 4: Correction step. Set

$$w^{k+1}=P_{\mathscr{W}}[w^k-\gamma\alpha_k d_2(\tilde{w}^k)], \tag{12}$$

where

$$\alpha_k=\frac{\|w^k-\tilde{w}^k\|^2_{(Q+Q^{\mathrm{T}})/2}}{\|Q(w^k-\tilde{w}^k)\|^2},\quad Q=\begin{pmatrix}rI_n & A^{\mathrm{T}}\\ 0 & sI_m\end{pmatrix}.$$

Set $k=k+1$. Go to Step 1.

Remark 1 When $rs>\|A^{\mathrm{T}}A\|/4$, the matrix $Q+Q^{\mathrm{T}}$ is positive definite. Note that from the definition of the step length α_k, we have

$$\alpha_k\geqslant\frac{\lambda_{\min}(\frac{1}{2}[Q+Q^{\mathrm{T}}])}{\|Q^{\mathrm{T}}Q\|}\doteq\alpha>0, \tag{13}$$

which shows that the step length sequence $\{\alpha_k\}$ has a positive lower bound uniformly.

The following lemma indicates that the stopping criterion of Algorithm 1 is reasonable, and its proof is motivated by the corresponding results in Ref. 9.

Lemma 1 *The sequences $\{w^k\}$ and $\{\tilde{w}^k\}$ generated by Algorithm 1 satisfy*

$$(w-\tilde{w}^k)^{\mathrm{T}}d_2(\tilde{w}^k)\geqslant(w-\tilde{w}^k)^{\mathrm{T}}Q(w^k-\tilde{w}^k),\quad\forall w\in\mathscr{W}. \tag{14}$$

Proof: By the first-order optimality condition for the two subproblems of (11), there exists a vector $\tilde{\xi}^k\in\theta(\tilde{x}^k)$ such that

$$(x-\tilde{x}^k)^{\mathrm{T}}\{\tilde{\xi}^k-A^{\mathrm{T}}\lambda^k+r(\tilde{x}^k-x^k)\}\geqslant 0,\quad\forall x\in\mathscr{X}, \tag{15}$$

and

$$(\lambda-\tilde{\lambda}^k)^{\mathrm{T}}\{(A\tilde{x}^k-b)+s(\tilde{\lambda}^k-\lambda^k)\}\geqslant 0,\quad\forall\lambda\in\Lambda. \tag{16}$$

Then, adding (15) and (16), and by a simple manipulation for all $w\in\mathscr{W}$, we obtain

$$\begin{pmatrix}x-\tilde{x}^k\\ \lambda-\tilde{\lambda}^k\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\tilde{\xi}^k-A^{\mathrm{T}}\tilde{\lambda}^k\\ A\tilde{x}^k-b\end{pmatrix}\geqslant(w-\tilde{w}^k)^{\mathrm{T}}Q(w^k-\tilde{w}^k),$$

which is just the assertion (14) by the definition of $d_2(\tilde{w}^k)$ in (10). □

Remark 2 If $\|w^k-\tilde{w}^k\|=0$ then from (14) we have

$$\begin{pmatrix}x-\tilde{x}^k\\ \lambda-\tilde{\lambda}^k\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\tilde{\xi}^k-A^{\mathrm{T}}\tilde{\lambda}^k\\ A\tilde{x}^k-b\end{pmatrix}\geqslant 0,\quad\forall w\in\mathscr{W},$$

which together with (9) implies that $\tilde{w}^k$ is a solution of VIP$(F,\mathscr{W})$, i.e., $\tilde{x}^k$ is a solution of (1). Hence the stopping criterion of Algorithm 1 is reasonable.

The following lemma gives that $d_2(\tilde{w}^k)$ is a descent direction of the merit function $\frac{1}{2}\|w-w^*\|^2$, where $w^* \in \mathscr{W}^*$ is an unknown solution.

Lemma 2 *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the two sequences generated by Algorithm 1. Then we have*

$$(w-w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \geqslant (w-\tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k), \quad \forall w \in \mathscr{W}. \tag{17}$$

Proof: Set

$$d_2(w^*) = \begin{pmatrix} \xi^* - A^{\mathrm{T}}\lambda^* \\ Ax^* - b \end{pmatrix}$$

with $\xi^* \in \theta(x^*)$. Then $d_2(w^*) \in \partial F(w^*)$. From the monotonicity of the multifunction $F(w)$, we have $(\tilde{w}^k - w^*)^{\mathrm{T}}(d_2(\tilde{w}^k) - d_2(w^*)) \geqslant 0$, i.e.,

$$(\tilde{w}^k - w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \geqslant (\tilde{w}^k - w^*)^{\mathrm{T}} d_2(w^*) \geqslant 0,$$

where the second inequality follows from $\tilde{w}^k \in \mathscr{W}$, $w^* \in \mathscr{W}^*$. Then adding the above inequality to (14), we obtain the assertion (17). □

Remark 3 Setting $w = w^k$ in (17), we have

$$(w^k-w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \geqslant (w^k-\tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k), \quad \forall w \in \mathscr{W},$$

i.e., $\forall w \in \mathscr{W}$,

$$(w^k-w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \geqslant (w^k-\tilde{w}^k)^{\mathrm{T}} \frac{Q+Q^{\mathrm{T}}}{2} (w^k-\tilde{w}^k),$$

Hence the above inequality indicate that $-d_2(\tilde{w}^k)$ is a descent direction at the iterate w^k.

The following theorem indicates that Algorithm 1 is a contraction method.

Theorem 1 *For the sequence $\{w^k\}$ generated by Algorithm 1, we have*

$$\|w^{k+1}-w^*\|^2 \leqslant \|w^k-w^*\|^2 - c_0\|w^k-\tilde{w}^k\|^2_{(Q+Q^{\mathrm{T}})/2}, \tag{18}$$

where $c_0 = \alpha\gamma(2-\gamma) > 0$.

Proof: Since $w^* \in \mathscr{W}^*$, it follows from (7) and (12) that

$$\begin{aligned}
&\|w^{k+1}-w^*\|^2 \\
&\quad \leqslant \|w^k - \gamma\alpha_k d_2(\tilde{w}^k) - w^*\|^2 \\
&\qquad - \|w^k - \gamma\alpha_k d_2(\tilde{w}^k) - w^{k+1}\|^2 \\
&\quad = \|w^k-w^*\|^2 - 2\gamma\alpha_k(w^{k+1}-w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \\
&\qquad - \|w^k - w^{k+1}\|^2.
\end{aligned} \tag{19}$$

Setting $w = w^{k+1} \in \mathscr{W}$ in (17), we obtain

$$(w^{k+1}-w^*)^{\mathrm{T}} d_2(\tilde{w}^k) \geqslant (w^{k+1}-\tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k).$$

Substituting the above inequality into the right-hand side of (19), we have

$$\begin{aligned}
&\|w^{k+1}-w^*\|^2 \\
&\quad \leqslant \|w^k-w^*\|^2 - 2\gamma\alpha_k(w^{k+1} \\
&\qquad - \tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k) - \|w^k-w^{k+1}\|^2 \\
&\quad = \|w^k-w^*\|^2 - 2\gamma\alpha_k(w^k-\tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k) \\
&\qquad + \gamma^2\alpha_k^2\|Q(w^k-\tilde{w}^k)\|^2 \\
&\qquad - \|w^k - w^{k+1} + \gamma\alpha_k Q(w^k-\tilde{w}^k)\|^2 \\
&\quad \leqslant \|w^k-w^*\|^2 - 2\gamma\alpha_k(w^k-\tilde{w}^k)^{\mathrm{T}} Q(w^k-\tilde{w}^k) \\
&\qquad + \gamma^2\alpha_k^2\|Q(w^k-\tilde{w}^k)\|^2 \\
&\quad = \|w^k-w^*\|^2 - \gamma(2-\gamma)\alpha_k\|w^k-\tilde{w}^k\|^2_{(Q+Q^{\mathrm{T}})/2} \\
&\quad \leqslant \|w^k-w^*\|^2 - \gamma(2-\gamma)\alpha\|w^k-\tilde{w}^k\|^2_{(Q+Q^{\mathrm{T}})/2},
\end{aligned}$$

where the last inequality comes from (13). □

Theorem 2 *The sequence $\{w^k\}$ generated by Algorithm 1 converges to a solution in $\mathscr{W}^*$ globally.*

Proof: From (18) and $c_0 > 0$, the sequence $\{w^k\}$ satisfies the strictly contractive property. That is

$$\|w^{k+1}-w^*\|^2 \leqslant \|w^k-w^*\|^2.$$

Hence $\{\|w^k-w^*\|^2\}$ is a descent sequence with lower bound. Thus it is convergent and therefore bounded, and so the sequence $\{w^k\}$ is also bounded. Thus it has at least one cluster point, say $\hat{w}$, and there is a subsequence $\{w^{k_j}\}$ such that

$$\lim_{j\to\infty} w^{k_j} = \hat{w}. \tag{20}$$

On the other hand, from (18) again, we obtain

$$\sum_{k=0}^{\infty} \|w^k-\tilde{w}^k\|^2_{(Q+Q^{\mathrm{T}})/2} \leqslant \frac{1}{c_0}\|w^0-w^*\|^2,$$

which together with the positive definiteness of the matrix $\frac{1}{2}(Q+Q^{\mathrm{T}})$ implies that

$$\lim_{k\to\infty} \|w^k-\tilde{w}^k\| = 0. \tag{21}$$

This and (20) show that

$$\lim_{j\to\infty} \tilde{w}^{k_j} = \hat{w}. \tag{22}$$

Then, taking the limit along the subsequence $\{w^{k_j}\}$ in (14), and by (21) and (22), we have

$$(w-\hat{w})^{\mathrm{T}} d_2(\hat{w}) \geqslant 0, \quad \forall w \in \mathscr{W}.$$

This together with $\hat{w} \in \mathscr{W}$ ($\mathscr{W}$ is a closed convex set and $w^k \in \mathscr{W}$, $\forall k$) shows that $\hat{w} \in \mathscr{W}^*$. That is, $\hat{w}$ is a solution of VIP($F, \mathscr{W}$). Then the global convergence of $\{w^k\}$ is obvious from (18). □

NUMERICAL RESULTS

In this section, we apply Algorithm 1 to some applications of (1) and report the numerical results. All the codes were written using MATLAB R2010a and were conducted on a ThinkPad notebook with 2 GB of memory.

Example 1 The following simple problem is the counterexample to show the divergence of the PDHG method[9]: $\min x$ s.t. $x = 1$, $x \geqslant 0$. Clearly, the single feasible solution $x = 1$ is also its optimal solution. Furthermore, $(x^*, \lambda^*) = (1, 1)$ is the unique saddle point of its Lagrangian function.

We now use Algorithm 1 to solve Example 1, and we set $r = s = 1$, $\gamma = 1.5$. The initial point is $w^0 = (0, 0)$. We see from Fig. 1 that Algorithm 1 is convergent in this case.

Example 2 Consider (2) with a nonsmooth objective function $\theta(x) = \|x\|_1$, whose Lagrangian function is $\mathscr{L}(x, \lambda) = \|x\|_1 - \lambda^{\mathrm{T}}(Ax - b)$. Now, let us elaborate on how to derive the closed-form solutions for the subproblems resulting from Algorithm 1.

For given $w^k = (x^k, \lambda^k)$, produce the predictor $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ by using (11) and obtain the following iterative scheme:

$$\tilde{x}^k = \arg\min_{x \in \mathbb{R}^n} \left\{ \mathscr{L}(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \right\},$$
$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}\,(A\tilde{x}^k - b),$$

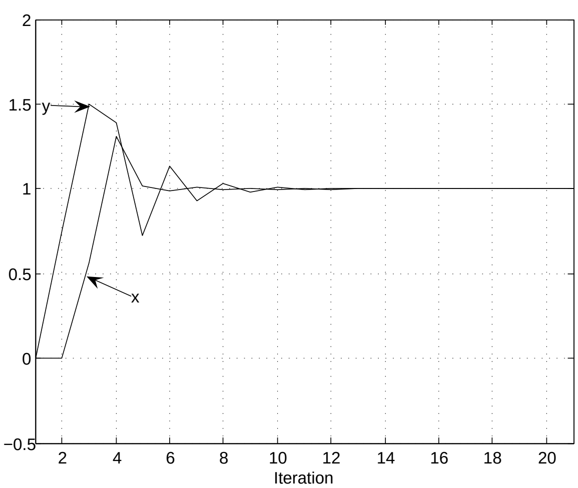

Fig. 1 Iterates generated by Algorithm 1 for Example 1.

and the first subproblem is equivalent to

$$\tilde{x}^k = \arg\min_{x \in \mathbb{R}^n} \left\{ \|x\|_1 + \frac{r}{2}\|x - \frac{1}{r}(A^{\mathrm{T}}\lambda^k + r x^k)\|^2 \right\},$$

which has a closed-from solution,

$$\tilde{x}^k = \mathrm{shrink}_{1,2}(a^k, 1/r) \doteq \mathrm{sgn}(a^k)\cdot\max\{0, |a^k| - 1/r\},$$

where $a^k = (A^{\mathrm{T}}\lambda^k + r x^k)/r$, and all the computation is componentwise. Then, with the updated $\tilde{w}^k$, the new iterate $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ is generated by

$$x^{k+1} = x^k - \gamma\alpha_k(\tilde{\xi}^k - A^{\mathrm{T}}\tilde{\lambda}^k),$$
$$\lambda^{k+1} = \lambda^k - \gamma\alpha_k(A\tilde{x}^k - b),$$

where

$$\tilde{\xi}^k \in \partial(\|\tilde{x}^k\|_1) = \begin{cases} \mathrm{sgn}(\tilde{x}^k), & \tilde{x}^k \neq 0, \\ \{z : -1 \leqslant z \leqslant 1\}, & \tilde{x}^k = 0. \end{cases}$$

In this experiment, we set the matrix $A \in \mathbb{R}^{m\times n}$ as a standard Gaussian matrix. The nonzero entries of the true signal x^* are selected at random from the standard Gaussian: $x^* = (0, 0, \dots, 0) \in \mathbb{R}^n$; p is generated by randomly permuting all the elements of the vector $(1, 2, \dots, n)$; then the elements of x^* whose subscripts are in the set contained by the first k elements p are reset to satisfy the standard Gaussian. The observed signal b is generated by $b = Ax^*$. We set $\mu = 0.01$, $\gamma = 1.5$, $r = 400$, $s = 2.01/r$, and the stopping criterion is RelErr $= \|x^k - x^*\|/\|x^*\| \leqslant 4\%$, or the number of iterations exceeds 10^4. The initial point, $w^0 = (A^{\mathrm{T}}b, 0)$. We set $m = \lfloor \alpha n \rfloor$ and $k = \lfloor \beta m \rfloor$ with $n = 500$, 1000, where k is the number of random nonzero elements contained in the original signal. For different combinations of α and β, the numerical results of Algorithm 1, the ALM method[3], and the MPDHG method of Ref. 9 are listed in Table 1. All the results are the average of 10 runs. For the ALM method, we use the fixed point method[16] to solve its x-related subproblem at each iteration, and the inner iteration is stopped when the number of inner iterations exceeds 100. The step length α of the MPDHG method should satisfy $\alpha \in (0, \alpha_{\max})$, where

$$\alpha_{\max} = \arg\max\{\alpha \mid Q^{\mathrm{T}} + Q - \alpha M^{\mathrm{T}}HM \succeq 0\},$$

where H and M are two matrices defined in Ref. 9. In this experiment, we use the Armijo line search to determine $\alpha_{\max}$. That is, we firstly set $\alpha = 1$, if the relationship $Q^{\mathrm{T}} + Q - \alpha M^{\mathrm{T}}HM \succeq 0$ does not hold, we reduce halve α until the relationship $Q^{\mathrm{T}} + Q -$

Table 1 Comparison of Algorithm 1 with the ALM and MPDHG methods.

n	α	β	ALM			MPDHG			Algorithm 1		
			Time	Iter	RelErr	Time	Iter	RelErr	Time	Iter	RelErr
500	0.3	0.2	2.3237	235.2	0.0397	6.3648	646.6	0.0396	1.7108	633.3	0.0395
	0.2	0.2	3.8273	482.3	0.0396	9.3549	1136.0	0.0397	2.2776	1010.0	0.0398
	0.2	0.1	2.1722	233.6	0.0393	2.5012	295.0	0.0391	1.0816	405.1	0.0390
1000	0.3	0.2	12.3329	210.8	0.0394	23.5978	609.3	0.0397	6.6872	580.1	0.0393
	0.2	0.2	15.8923	273.6	0.0397	32.2090	960.8	0.0399	8.6373	951.1	0.0396
	0.2	0.1	7.8271	198.7	0.0392	10.6081	321.4	0.0390	3.3800	364.7	0.0392

$\alpha M^{\mathrm{T}} H M \succeq 0$ holds. Clearly the procedure is quite time consuming if m and n are large. However, since these two terms are invariant at each iteration, we only need to compute it once before all iterations.

The numerical results given in Table 1 show that (i) all the tests completed successfully; (ii) Algorithm 1 is faster than the ALM and MPDHG methods since it always consumes less CPU time to achieve the prescribed accuracy.

CONCLUSIONS

In this paper, we have proposed a prediction-correction primal-dual hybrid gradient method for convex programming with linear constraints, whose global convergence can be guaranteed under mild conditions. Two sets of numerical results are given, which illustrate that the new method performs better than some state-of-the-art solvers. In the future, we will investigate the worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense of the new method.

Acknowledgements: The authors gratefully acknowledge the valuable comments of the anonymous referees. This work was supported by the National Natural Science Foundation of China (Grant No. 61601183) and the Major Science and Technology Project Education Department of Henan Province (Grant No. 17A510010).

REFERENCES

1. Qiao LB, Zhang BF, Sun JS (2017) A systematic review of structured sparse learning. *Front Inform Tech Electron Eng* **18**, 445–63.
2. He BS, Yuan XM, Zhang WX (2013) A customized proximal point algorithm for convex minimization with linear constraints. *Comput Optim Appl* **56**, 559–72.
3. Hestenes MR (1969) Multiplier and gradient methods. *J Optim Theor Appl* **4**, 303–20.
4. Zhu M, Chan TF (2008) An efficient primal-dual hybrid gradient algorithm for total variation image restoration, CAM Report 08-34, UCLA, USA.
5. Ma F, Ni MF (2018) A class of customized proximal point algorithms for linearly constrained convex optimization. *Comput Appl Math* **37**, 896–911.
6. Yang JF, Zhang Y (2011) Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM J Sci Comput* **33**, 250–78.
7. Yang JF, Yuan XM (2013) Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math Comput* **82**, 301–29.
8. Sun M, Liu J (2017) The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization. *J Inequal Appl* **2017**, 19.
9. He BS (2015) PPA-like contraction methods for convex optimization: A framework using variation inequality approach. *J Oper Res Soc China* **3**, 391–420.
10. He BS (2016) From the projection and contraction methods for variational inequalities to the splitting contraction methods for convex optimization. *Numer Math J Chin Univ* **38**, 74–96.
11. He BS, Liao LZ, Wang X (2012) Proximal-like contraction methods for monotone variational inequalities in a unifed framework I: Effective quadruplet and primary methods. *Comput Optim Appl* **51**, 681–708.
12. Yan XH, Han DR, Sun WY (2008) A self-adaptive projection method with improved step-size for solving variational inequalities. *Comput Math Appl* **55**, 819–32.
13. Rockafellar RT (1970) *Convex Analysis*, Princeton Univ Press, pp 233–53.
14. Bauschke HH, Borwein JM, Li W (1999) Strong conical hull intersection property, bounded linear regularity, Jamesoni's property (G), and error bounds in convex optimization. *Math Program* **86**, 135–60.
15. Yang HW, Han DR (2016) Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems. *SIAM J Numer Anal* **54**, 625–40.
16. Hale E, Yin WT, Zhang Y (2010) Fixed-point continuation applied to compressed sensing: Implementation and numerical experiments. *J Comput Math* **28**, 170–94.